

**Efecto de los datos atípicos discordantes y contaminantes en la
implementación de las redes neuronales supervisadas Back Propagation**

Tesis

Para obtener el Título de Ingeniero Estadístico en la Escuela Profesional de
Ingeniería Estadística, bajo la dirección del profesor MS EM Rolando Bazán
Gonzales, sustentado en público el 30 de Junio del 2011

por

Bach. Luis Alberto Sánchez Alvarado

Composición de la Comisión de Sustentación:

Prof. Lic. Rolando Bazán Gonzales, MS EM.
Presidente de la Comisión de Titulación

Prof. Lic. Luis Huamanchumo de la Cuba, MC.
Escuela Profesional de Ingeniería Estadística, Asesor

Prof. Lic. Rita Guzmán López
Escuela Profesional de Ingeniería Estadística, Revisora

Agradezco a mis padres Milciades y Angélica, y a todos mis seres queridos por todo el apoyo y la confianza puesta en mí.
A mis profesores por todos sus consejos y enseñanzas.

RESUMEN

La presente investigación plantea como objetivo estudiar aspectos técnicos relacionados con la implementación de la red neuronal *Back Propagation* en términos de su calidad de aprendizaje y generalización con el fin de establecer criterios óptimos para su validación, evaluación del desempeño e implementación. Para ello, se plantea la hipótesis de que la presencia de datos atípicos, ya sean del tipo discordante o contaminante tienen un efecto en el desempeño de la red neuronal supervisada y entrenada con el método de *Back Propagation*. Para verificar este supuesto se ha llevado a cabo un diseño experimental donde se tienen en cuenta diferentes tamaños de muestra, y la incorporación de dos tipos de datos atípicos: los discordantes y contaminantes. Se llega a la conclusión que, bajo las características descritas, el algoritmo *Back Propagation* da resultados óptimos cuando no se cuenta con datos atípicos y en especial en muestras grandes. A la vez, se aprecia que el efecto de los datos discordantes es mucho mayor que el provocado por los datos contaminantes.

Palabras claves.- *Back Propagation*, Análisis de regresión lineal, calidad de aprendizaje, generalización de datos.

INDICE

	Pag.
INTRODUCCIÓN	1
OBJETIVOS	3
JUSTIFICACIÓN	4
CAPITULO I ANTECEDENTES	5
1.1 Desarrollo histórico de las redes neuronales.....	5
1.2 Desarrollo histórico del algoritmo <i>Back Propagation</i>	6
1.3 Estudios comparativos de Redes neuronales con Análisis Estadísticos.....	7
1.4 Datos atípicos.....	9
CAPITULO II MARCO TEÓRICO.....	10
2.1 Redes Neuronales.....	10
2.1.1 Definición	10
2.1.2 Arquitectura de Redes Neuronales	11
2.1.3 Funciones de Transferencia	13
2.1.4 Pesos y bias	14
2.1.5 Tipos de Aprendizaje	15
2.1.6 Generalización de los datos	16
2.1.7 Ratio de aprendizaje	17
2.1.8 Perceptrón	18
2.1.9 Algoritmo <i>Back Propagation</i>	23
2.2 Análisis de Regresión Lineal Múltiple	29
2.2.1 Definición	29
2.2.2 Estimación de los coeficientes de regresión por mínimos cuadrados.....	30
2.2.3 Interpretación geométrica de mínimos cuadrados	33

2.2.4	Supuestos del Análisis de Regresión Lineal	35
2.2.4.1	Linealidad	35
2.2.4.2	Independencia	36
2.2.4.3	Homocedasticidad	37
2.2.4.4	Normalidad	40
2.2.4.5	No colinealidad	41
2.2.5	Validación del modelo	43
2.3	Datos atípicos.....	44
CAPÍTULO III HIPÓTESIS Y METODOLOGÍA		46
3.1	Hipótesis general	46
3.1.1	Hipótesis específica I.....	46
3.1.2	Hipótesis específica II	46
3.2	Descripción de la metodología	47
3.3	Definiciones y su operacionalización	48
3.3.1	Definiciones operativas	48
3.3.2	Indicadores	50
3.3.2.1	Caracterización de datos	51
3.3.2.2	Indicadores de aprendizaje	54
3.3.2.3	Indicadores de generalización	56
3.2.3	Operaciones Intermedias	56
3.2.3.1	Técnica estadística de Regresión Lineal Múltiple....	56
3.2.3.2	Método de la Red Neuronal Back Propagation	57
CAPITULO IV DISEÑO DEL EXPERIMENTO		58
4.1	Datos de entrenamiento validación y prueba	59
4.1.1	Espacio parametral generador	60
4.1.2	Casos de estudio.....	61
4.1.3	Generador datos normales mutivariados.....	65
4.2	Arquitectura de la red ACP	69

4.3	Bases experimentales	69
	4.3.1 Aleatorización.....	69
	4.3.2 Bloqueo.....	70
	4.3.3 Réplica.....	70
	4.3.4 Control	70
CAPITULO V RESULTADOS Y ANÁLISIS		71
5.1	Caracterización de los datos	71
5.2	Aprendizaje	75
	5.2.1 Resultados del análisis de Regresión Lineal Múltiple....	76
	5.2.2 Calidad del proceso de aprendizaje en redes neuronales	77
5.3	Generalización de los datos.....	81
CONCLUSIONES		85
RECOMENDACIONES		88
REFERENCIAS BIBLIOGRÁFICAS		89
ANEXO		92
	Anexo I.- Equivalencia de términos.....	93
	Anexo II.- Funciones de Transferencia.....	95
	Anexo III.- Programas en MATLAB.....	97
	Anexo IV.- Programas usados en el diseño de experimento.....	109
	Anexo V.- Datos Estadísticos.....	116

INDICE DE CUADROS

Cuadro	Título	Pág.
Cuadro N° 4.1	Grados de Correlación	60
Cuadro N° 4.2	Escenarios por tipo de dato según casos	64
Cuadro N° 4.3	Pesos dados a cada Variable Independiente.....	66
Cuadro N° 5.1	Intensidad de Multicolinealidad – Caso I (Sin valores atípicos).....	72
Cuadro N° 5.2	Estadístico Durbin Watson – Caso I (Sin valores atípicos).....	73
Cuadro N° 5.3	Prueba de Levene – Caso I (Sin valores atípicos)	74
Cuadro N° 5.4	Prueba de Kolmogorov– Caso I (Sin valores atípicos)	75
Cuadro N° 5.5	Error Cuadrado medio del Análisis de Regresión – Datos atípicos discordantes.....	76
Cuadro N° 5.6	Error Cuadrado medio del Análisis de Regresión –Datos atípicos contaminantes	77
Cuadro N° 5.7	Error cuadrado Medio del Análisis de Redes Neuronales en la etapa de aprendizaje – Datos atípicos discordantes.....	78
Cuadro N° 5.8	Error cuadrado Medio del Análisis de Redes Neuronales en la etapa de aprendizaje– Datos atípicos contaminantes.....	79
Cuadro N° 5.9	Número de épocas realizadas en la etapa de aprendizaje	80
Cuadro N° 5.10	Tiempos de procesamiento en la etapa de aprendizaje (Segundos)	80
Cuadro N° 5.11	Error cuadrado Medio del Análisis de Redes Neuronales en la etapa de generalización– Datos atípicos discordantes	82
Cuadro N° 5.12	Error cuadrado Medio del Análisis de Regresión Lineal en la etapa de generalización	83
Cuadro N° 5.13	Error cuadrado Medio del Análisis de Regresión Lineal en la etapa de generalización	84

INDICE DE GRÁFICOS

Gráfico	Título	Pág.
Gráfico N° 1.1	Aplicaciones de las Redes Neuronales ligados a Predicción.....	7
Gráfico N° 1.2	Terminología equivalente entre Redes Neuronales y Modelos Estadísticos.....	8
Gráfico N° 1.3	Comparación de metodologías.....	8
Gráfico N° 2.1	Principales Funciones de Transferencias.....	14
Gráfico N° 3.1	Indicador A.1 Normalidad Multivariada de Mardia	51
Gráfico N° 3.2	Indicador A.2 Coeficiente de Asimetría Multivariado	51
Gráfico N° 3.3	Indicador A.3 Coeficiente de Curtosis Multivariado.....	52
Gráfico N° 3.4	Indicador A.4 Intensidad de Multicolinealidad	52
Gráfico N° 3.5	Indicador A.5 Grado de Independencia de Residuos	53
Gráfico N° 3.6	Indicador A.6 Homocedasticidad	54
Gráfico N° 3.7	Indicador B.1 Valor adaptivo.....	54
Gráfico N° 3.8	Indicador B.2 Épocas	55
Gráfico N° 3.9	Indicador B.3 Error Cuadrado Medio (ECM)	55
Gráfico N° 3.10	Indicador C.1 Error Cuadrado Medio (ECM)	56
Gráfico N° 4.1	Características de la Red Neuronal	69

INDICE DE FIGURAS

Figura	Título	Pág.
Figura N° 1.1	Historia cronológica de las Redes Neuronales.....	6
Figura N° 2.1	Comparación entre una neurona y un modelo neuronal	11
Figura N° 2.2	Representación de Regresión lineal simple en un esquema de Redes.....	13
Figura N° 2.3	Ajuste de los datos y la plasticidad.....	16
Figura N° 2.4	El gradiente de la función error.....	17
Figura N° 2.5	La influencia de la variación de los pesos	18
Figura N° 2.6	Aprendizaje de la red neuronal Perceptrón.....	19
Figura N° 2.7	Estructura de una Red Neuronal con una sola capa.....	20
Figura N° 2.8	Notación abreviada de una Red Neuronal con una sola capa.....	21
Figura N° 2.9	Arquitectura de una red Multicapa.....	22
Figura N° 2.10	Arquitectura abreviada de una red Multicapa.....	23
Figura N° 2.11	Fase de propagación en la primera capa.....	26
Figura N° 2.12	Fase de propagación en la segunda capa.....	27
Figura N° 2.13	Fase de propagación en la tercera capa.....	27
Figura N° 2.14	Cálculo de las sensibilidades de la capa 3 a la 1.....	28
Figura N° 2.15	Actualización de pesos y bias en la primera iteración.....	29
Figura N° 2.16	Interpretación geométrica de los mínimos cuadrados.....	34
Figura N° 2.17	Homocedasticidad	38
Figura N° 2.18	Patrones en la gráfica de residuales	39
Figura N° 2.19	Gráficos de Probabilidad Normal	41
Figura N° 2.20	Efectos de la multicolinealidad: Regresores Inestables.....	43
Figura N° 2.21	Datos atípicos contaminantes y discordantes	45
Figura N° 4.1	Algoritmo generador de datos para el estudio	68

Introducción

El presente trabajo de investigación tiene como objetivo principal estudiar el efecto de los datos atípicos discordantes y contaminantes en redes entrenadas con la técnica de Back Propagation, con la finalidad de establecer en qué grado afectaría la calidad del proceso de aprendizaje y la posterior generalización de los datos. La revisión bibliográfica señala que existen gran cantidad de trabajos comparativos entre metodologías estadísticas y las redes neuronales y el efecto de datos atípicos discordantes en su desempeño, no obstante el efecto de los datos atípicos contaminantes ha sido dejado de lado. De esta forma este estudio se encarga, bajo la fundamentación estadística, revisar los efectos de ambos tipos de datos atípicos: el discordante y el contaminante, y ver sus efectos en la calidad del proceso de aprendizaje y generalización de datos. Esto ayudará a los investigadores a llevar su trabajo siguiendo criterios plenamente estudiados.

Como hipótesis se plantea que la presencia de datos atípicos, ya sean del tipo discordante o contaminante afectan el desempeño de la red neuronal supervisada y entrenada con el método de *Back Propagation*. Para contrastar la hipótesis señalada se implementa un experimento el cual maneja diferentes escenarios determinados por la cantidad de muestra (40, 160, 1000 y 3000 datos) y la presencia o no de datos atípicos, así mismo se trata que las muestras cumplan con los supuestos del análisis de Regresión lineal múltiple. Los resultados obtenidos reflejan que la presencia de datos atípicos del tipo discordante o contaminante influyen en el desempeño de la red neuronal entrenada con el algoritmo de *Back Propagation*, y en todos los casos la técnica estadística de regresión lineal da mejores resultados en especial en la

fase de formación del modelo o la parte aprendizaje en el caso de las redes neuronales. Así mismo se tiene que el efecto ocasionado por los datos atípicos contaminantes es menor que el producido por los datos atípicos discordantes.

El estudio comienza con una breve reseña histórica de la redes neuronales, su evolución y las primeras apariciones de la técnica Back Propagation así como también se menciona los estudios comparativos que se han hecho hasta la fecha, esto permitirá tener una mejor visión del problema planteado y la forma como podrá contribuir al campo de la investigación. En el capítulo II se presenta el marco teórico que tiene como fin dar las pautas teóricas tanto de las redes neuronales como de la parte estadística para así comprender más claramente los términos tratados en la investigación. En el capítulo III, se describe toda la metodología empleada, esto incluye las definiciones operacionales y los indicadores que son de gran ayuda en la parte de análisis. En el capítulo IV explica y se fundamenta todo el diseño experimental con el fin de garantizar que el estudio se lleva de una manera adecuada. En el capítulo V se explican los resultados obtenidos y finalmente en el capítulo VI se dan algunas recomendaciones para estudios posteriores.

Objetivos de la investigación

Objetivo general

Estudiar el efecto de los datos atípicos discordantes y contaminantes en redes entrenadas con la técnica de *Back Propagation*, con la finalidad de establecer en qué grado afectaría la calidad del proceso de aprendizaje y la posterior generalización de los datos.

Objetivos específicos

- i. Analizar que tipo de datos atípicos, discordantes o contaminantes, son los que dan una mejor calidad del proceso de aprendizaje y generalización de los datos

- ii. Establecer criterios óptimos para la implementación de redes neuronales entrenadas con *Back Propagation*, con el fin de aminorar los posibles efectos de los datos atípicos.

Justificación

Tras hacer una profunda revisión bibliográfica se ha encontrado que existen muchos estudios comparativos entre las redes neuronales y las metodologías estadísticas y uno de esos puntos analizados es el comportamiento de la red ante la presencia de datos atípicos, específicamente del tipo **discordante** ya que este tipo cuentan con metodologías para su identificación. No obstante es importante ver también los efectos que podrían tener otros tipos de atípicos, como es el caso de los denominados **contaminantes**, lo cuales tiene como principal problema su identificación dentro de la muestra.

Bajo esta premisa el presente trabajo busca estudiar bajo la fundamentación estadística, los efectos de ambos tipos de datos atípicos: el discordante y el contaminante, y ver sus efectos en la calidad del proceso de aprendizaje y generalización de datos. Esto conllevará a mejorar y ampliar los conocimientos que se tiene en esta área de estudio y así ayudar a los investigadores a llevar su trabajo siguiendo criterios plenamente estudiados.

Capítulo I

Antecedentes

Una explicación breve y clara respecto a los orígenes y estudios previos que se han hecho dentro del campo de las Redes Neuronales, así como su relación con la Estadística contribuirán a tener un mayor entendimiento sobre el tema que se trata en el presente trabajo de Investigación.

1.1 Desarrollo Histórico de las Redes Neuronales

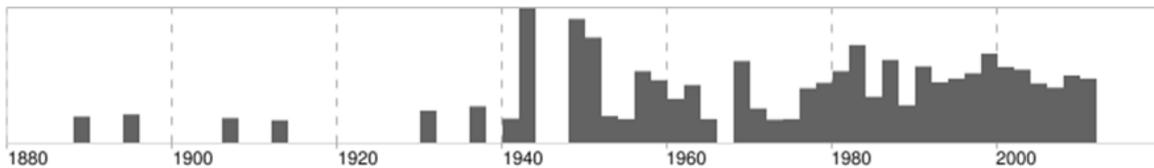
Algunos investigadores indican que el inicio de este campo de investigación se da en 1888 año en que Santiago Ramón y Cajal describió la estructura de la neurona y su forma de interconexión estableciendo así las bases para el estudio de las redes neuronales. A partir de esa fecha se fueron dando algunos avances sobre el estudio del cerebro humano y sobretodo de las neuronas, y es hasta el año 1936 donde aparece el primer precedente propiamente dicho de las redes neuronales, este año el científico Alan Turing realizó estudios de como se relacionaba el cerebro con la computación.

Finalmente en 1943 McCulloch and Walter Pitts publican los primeros estudios sobre las redes neuronales dando así nacimiento a este nuevo campo de investigación, no obstante al no contar con los avances computacionales suficientes los estudios referentes a las redes neuronales recién tuvieron su apogeo a partir de los años 60 y desde ahí los estudios no han finalizado hasta la actualidad. El más usado buscador de Internet **Google** nos brinda una figura

que nos ayuda a visualizar mejor la historia de las redes neuronales desde sus inicios hasta la actualidad (Ver Figura N°1.1).

Figura N° 1.1

Historia Cronológica de las Redes Neuronales



Fuente: Buscador www.Google.com

1.2 Desarrollo Histórico del algoritmo *Back Propagation*

Los inicios de las redes neuronales se da con el estudio del perceptrón, la regla de aprendizaje fue dada por Frank Rosenblatt, conforme los años fueron pasando y los estudios en este campo también lo hacían los investigadores hicieron notar la gran desventaja de este tipo de redes, el principal problema era que sólo eran utilizadas para problemas linealmente separables. Este mismo problema lo tenía el método de "Least Mean Square" dado por Bernard Widrow y Marcian Hoff.

Ante la presencia de este problema, en 1960 Kelly and Bryson desarrollaron métodos numéricos que permiten dar una primera solución al problema con un cálculo recursivo de la gradiente de una función costo. En 1962, Dreyfus estableció que el problema se solucionaría con un sistema multinivel, este sistema luego llevaría el nombre de algoritmo *Back Propagation*. En 1974 aparentemente aparece la primera definición del algoritmo para entrenar redes multicapa en la tesis de Paul Werbos (1974).

En los años ochenta el algoritmo *Back Propagation* se da a conocer como tal y es ampliamente difundido en forma independiente por David Rumelhart (1986) , Geoffrey Hinton y Ronald Williams (1986), David Parker(1985) y Yann Le Cun (1986). De todos estos autores Rumelhart resalta más por su obra *Parallel Distributed Processing* que tuvo gran acogida por los investigadores de redes neuronales. Actualmente el perceptrón multicapa entrenada con el algoritmo de *Back Propagation* es uno de los más usados.

Revisando la bibliografía existente se encuentra que en el transcurso de estos años han sido usados en diversos campos para fines predictivos, entre los que se puede citar tenemos los siguientes:

Gráfico N° 1.1

Aplicaciones de las Redes Neuronales ligados a Predicción

Área	Referencia
Medicina	Aplicación de las Redes Neuronales artificiales a la predicción del resultado a corto, mediano, largo plazo del trasplante Renal Pediátrico (Serrano , 2004)
Finanzas	Estudio comparativo de Redes Neuronales y Análisis de Regresión para la predicción de nuevos billetes emitidos (Rurkhamet B. et al, 1998)
Robótica	Aprendizaje de movimiento fino de robots: Diseño y experimentos (Versino C et al, 2000)
Ingeniería Civil	Predicción de la consistencia de concreto por medias y el uso de Redes Neuronales Artificiales (Gonzales <i>et al</i> , 2006)
Industria y Manufactura	Optimización de termoeléctricas con Redes Neuronales (Capuma M. <i>et al</i> , 2009)
Turismo	Pronóstico de la demanda del sector turismo (Law <i>et al</i> , 2003)
Investigación de Mercados	Uso de las Redes Neuronales para predecir la respuesta del mercado. (Pearson <i>et al</i> , 2003)

Fuente: Elaboración Propia

1.3 Estudios Comparativos de Redes Neuronales con Análisis Estadísticos

En todos estos campos de acción, las redes neuronales han sido usadas como modelos de predicción y de clasificación, entrando a competir con los modelos clásicos estadísticos, es así como varios autores comenzaron a desarrollar trabajos comparando ambas metodologías, siendo considerado el

trabajo de Ripley (1993) como el mejor desarrollado hasta la fecha. Tras estos estudios se comienza a encontrar cierta relación entre ambas metodologías llegando a conceptualizar a las redes neuronales como modelos de regresión no lineales (Warren. S. Sarle-1994).

Sarle (1994) como fruto de sus estudios comparativos, logra presentar una serie de equivalencias entre los términos usadas por ambas metodologías, (Ver Gráfico N° 1.2.), esto ayuda a tener una mejor comprensión de las redes neuronales basado en el conocimiento estadístico.

Gráfico N° 1.2

Terminología equivalente entre Redes Neuronales y Modelos Estadísticos

Modelos Estadístico	Redes Neuronales
Variables	Características
Variables independientes	Input
Valores predichos	Outputs
Variables dependientes	Target o valores de entrenamiento
Residuales	Errores
Estimación	Entrenamiento, aprendizaje adaptabilidad
Criterio de estimación	Función error o costo
Observación	Patrones o pares de entranamiento
Parámetros estimados	Pesos
Interpolación o extrapolación	Generalización

Fuente: Sarle (1994)

Gráfico 1.3

Comparación de metodologías

Área	Referencia	Nº Variables	Muestra	Mejor Desempeño
Medicina	Gaudart <i>et al.</i> (2004)	5	3000	Ambas
Manufactura	Smith y Mason(1997)	2 y 3	5,10,40,80	Regresión Lineal
Manufactura	Finnie <i>et al.</i> (1997)	4	299	Redes Neuronales
Manufactura	Shuhui <i>et al.</i> (2001)	2	1500	Redes Neuronales
Manufactura	Feng y Wang (2002)	5	48	Ambas
Marketing	Dutta <i>et al.</i> (1994)	7	138	Regresión Lineal
Marketing	Thieme <i>et al.</i> (2000)	43	612	Redes Neuronales
Finanzas	Duliba(1991)	5-10	600	Redes Neuronales
Finanzas	Spear an Leis(1997)	4	328	Ambas

Fuente: Paliwal y Kumar (2009)

1.4 Datos atípicos

Los datos atípicos son observaciones que se presentan en la gran mayoría de estudios y por lo general son identificables pudiendo tener un carácter positivo o negativo. Tendrá un carácter positivo cuando puede mostrar nuevos segmentos que no se habían tomado en cuenta en la investigación, por el contrario tendrán un efecto negativo cuando son productos de errores de recojo de información provocando así malas estimaciones. Por tal motivo a través de los años diversos investigadores se han preocupado por buscar técnicas que los ayuden a identificar y estudiar, llegando a convertirse en un análisis requerido previamente a cualquier análisis estadístico.

Actualmente se dispone una gran variedad de test para identificar datos atípicos tanto univariados como multivariados, dentro de este último grupo se tiene la D^2 de Mahalanobis que es el test que se usará en el presente estudio.

Beckman y Cook (1983) define dos tipos de datos atípicos: los **discordantes** y los **contaminantes**, el primer tipo pueden ser detectados con los test tanto univariados como multivariados, mientras los del grupo restante son muy difíciles de detectar en especial cuando sus valores se traslapan con los de la población en estudio.

En los estudios comparativos entre las redes neuronales y los modelos estadísticos también se ha visto el efecto de los datos atípicos pero sólo del tipo **discordante**, por ello el aporte que se hace con esta investigación es analizar ambos tipos de datos atípicos.

Capítulo II

Marco Teórico

2.1 Redes Neuronales

2.1.1 Definición

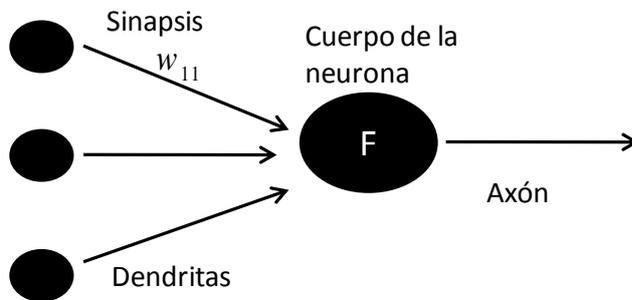
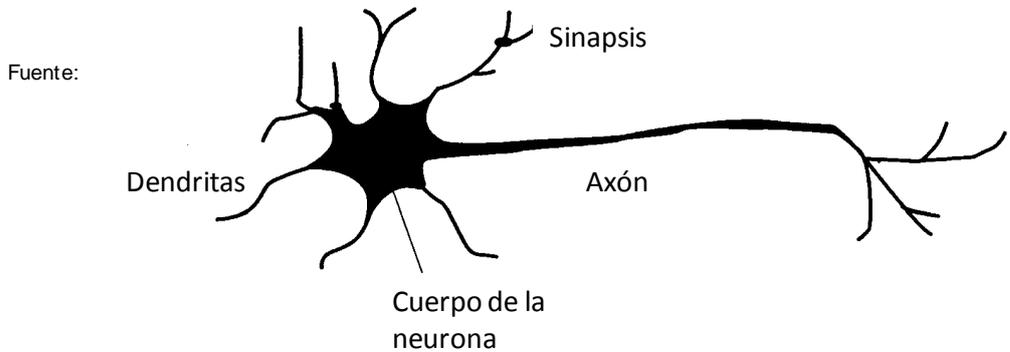
La mayoría de investigadores definen a las redes neuronales como modelos artificiales y simplificados del cerebro humano, capaces de aprender a través de sus sistemas interconectados, y que tienen como unidades elementales a los nodos que vendrían a simbolizar las unidades básicas del cerebro humano, **las neuronas**.

En la figura 2.1 se puede visualizar una analogía entre los componentes de una neurona y un modelo de red neuronal. Las **dendritas** en una neurona son las encargadas de recibir la información proveniente de otras neuronas, luego esta información es procesada en el **cuerpo de la neurona** y la respuesta resultante es enviada hacia otras neuronas a través del **axón**, este traspaso de información se hace a través de un impulso eléctrico que determina el grado de excitación de la neurona, denominado sinapsis. En la red neuronal el proceso es similar, para explicar esto se ha mostrado una red simple, donde las dendritas son las conexiones que se dan desde los valores de entrada (*inputs*) hacia el nodo de la capa oculta, este nodo viene hacer el cuerpo de la neurona que por medio de una función matemática procesa la información la cual finalmente da una salida (*output*) , esto a través de la conexión del nodo hacia afuera, esto sería equivalente al axón, finalmente la

fuerza con que se traspassa la información (sinapsis) es representada por los pesos de la red neuronal.

Figura N° 2.1

Comparación entre una neurona y un modelo neuronal



Veelenturf (1995)

2.1.2 Arquitectura de Redes Neuronales

Giudici (2003) define al término arquitectura como la organización de la red neuronal: el número de capas, el número de unidades (neuronas) que siguen en cada capa, y la manera en que éstas son conectadas.

Respecto al tipo de capas que puede tener una red neuronal, existen tres tipos:

- Input: Son las encargadas de recibir sólo la información del ambiente externo, cada neurona en este caso corresponde a una variable explicatoria, en esta capa no se realiza ningún tipo de cálculo.
- Output: Es la capa que produce los resultados finales, los cuales son enviados al ambiente exterior.
- Oculta: Son las capas que se encuentran entre las capas *input* y *output*, y reciben este nombre porque no tienen contacto con el ambiente externo, son capas utilizadas exclusivamente para el análisis.

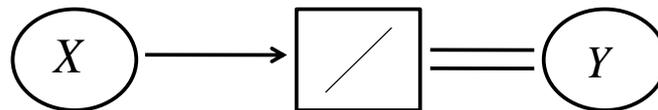
La arquitectura de la red neuronal por lo general es presentada gráficamente, por esta razón en ocasiones recibe el nombre de “topología de la red”. Sarle (1994) señala como poder graficar redes neuronales, las indicaciones que señala son:

- Círculos representan variables observadas: Con el nombre mostrado dentro del círculo.
- Cuadrados representan valores calculados con una función de uno o más argumentos (Ver 2.1.3). El símbolo dentro de la caja indica el tipo de función. La mayoría de cajas también disponen del parámetro bias (Ver 2.1.4).
- -Las flechas indican de donde provienen los recursos que servirán de argumentos para la función, por lo general están asociados a pesos que deben ir ajustándose conforme la red vaya aprendiendo (Ver 2.1.4).
- Dos líneas paralelas indican que el resultado final ha sido calculado por algún método de estimación.

Un ejemplo de topología, se muestra en la figura 2.2, esta red neuronal simboliza o equivale a una regresión lineal simple:

Figura N° 2.2

Representación de Regresión lineal simple en un esquema de Redes



Fuente: Sarle (1994)

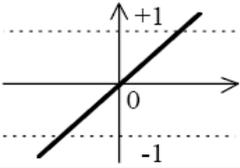
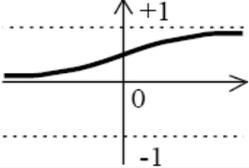
Al trabajar con redes neuronales los términos de función de transferencia, pesos, bias, tipos de aprendizaje, ratio de aprendizaje son comunes, por ello a continuación se explicará cada uno con el fin de tener una mejor comprensión.

2.1.3 Función de transferencia

Son funciones matemáticas que permiten obtener el valor del potencial post sináptico a partir de los pesos y los valores de entrada. Una completa lista de estas funciones son presentadas en el Anexo 2.1, en esta parte presentaremos sólo las dos más usadas en el algoritmo de *Back Propagation*, la lineal y la sigmoideal, en el gráfico 2.1 se aprecia su representación matemática, su gráfica matemática y la forma como representarlas dentro de una topología de redes neuronales.

Gráfico N° 2.1

Principales Funciones de Transferencias

Función de Transferencia		
Nombre	Lineal	Log Sigmoidal
Notación Matemática	$purelin \quad (x) = x$	$log \ sig \ (x) = \frac{1}{1 + e^{-x}}$
Gráfica Matemática		
Representación en Topología de Redes		

Fuente: Demuth *et al.* (2005)

De estas dos la función lineal será empleada en los experimentos que usaremos en el presente trabajo.

2.1.4 Pesos y bias

Giudici (2003) define a los **pesos** como coeficientes adaptivos que en analogía con los modelos biológicos, son modificados en respuesta a varias señales que viajan a través de la red acorde a un algoritmo de aprendizaje. Este mismo autor señala que las **bias** son un valor umbral, que es usualmente introducido, siendo similares a un intercepto dentro de un modelo de regresión.

El establecer los valores iniciales de los pesos y bias es un paso importante, por ello existen varios criterios para este punto, Leondes (1998) explica dos de ellos: “El método de mínimos cuadrados ortogonales” y “El método de máxima covarianza”; también Freeman y Skapura (1991) señalan que los valores iniciales óptimos deben ser números aleatorios que estén entre

-0.5 y +0.5, Fernandez (2001) da también una serie de metodologías al respecto. En este caso, se usará una inicialización de 0.05 tanto para las bias como para los pesos.

Un alcance final referente a los pesos y bias es que desde los inicios de las redes neuronales su cálculo interno ha sido considerado como una **caja negra**, porque lo único que el usuario visualiza son los resultados finales y no se puede tener una idea de cuál de las variables es la más importante o cual no lo es, pregunta que sí se puede contestar en los modelos de regresión; al respecto Montañó (2002) explica que ante este problema están surgiendo metodologías conocidos con el nombre de análisis de sensibilidad cuyo objetivo es ayudar a obtener un carácter explicativo de cada una de las variables que ingresan a la red.

2.1.5 Tipos de Aprendizaje

Pueden ser divididos en dos, el supervisado y el no supervisado. El **supervisado** se da cuando un conjunto de vectores de ingreso son recolectados y presentados a la red, las salidas son calculadas y comparadas con los valores esperados, con el error calculado los pesos son corregidos de acuerdo al algoritmo de aprendizaje, Rojas (1996) los denomina también como **aprendizaje con profesor** porque el proceso sabe las respuestas correctas para los vectores de ingreso. Dada la naturaleza del estudio se usará un entrenamiento supervisado.

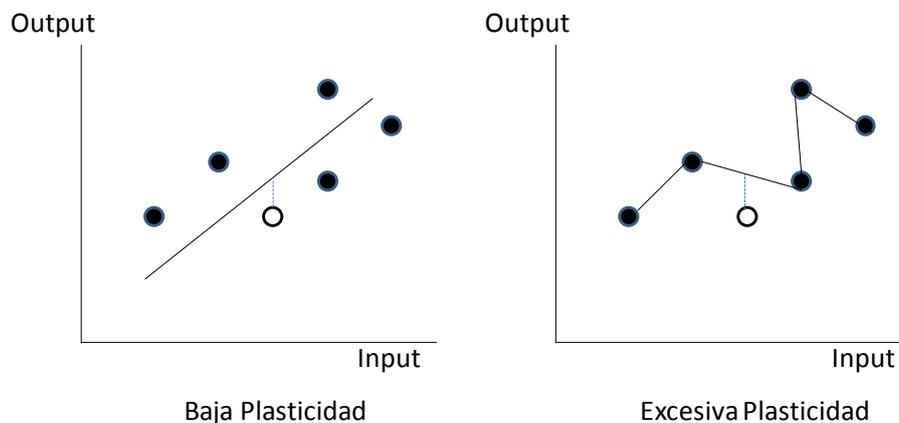
El **no supervisado o auto supervisado es** usado cuando, para un conjunto de vectores de ingreso el valor exacto que debería producir la red es desconocida. Estas redes deben encontrar las características, correlaciones o regularidades que se pueda establecer entre los datos de entrada; dentro de estos tipos de entrenamiento resalta el aprendizaje hebbiano.

2.1.6 Generalización de los datos

Uno de los objetivos de la red neuronal es ser capaz de **generalizar**, es decir tener un buen desempeño ante valores desconocidos, lo que se conoce en estadística como interpolación y extrapolación de los datos. Aquí ocurre un problema contradictorio porque por un lado debe dar una buena aproximación con los valores presentados y a la vez debe ser capaz de tener un buen desempeño ante valores nuevos. Esto se muestra claramente con el siguiente ejemplo planteado por Rojas (1996), aquí habla de la plasticidad del sistema, si tenemos una baja plasticidad podemos tener un ajuste aceptable pero con un error mayor a cero, sin embargo al tener una excesiva plasticidad vemos que se logra ajustar de forma exacta los valores dados en el entrenamiento con un error igual a cero, el problema surge al presentar nuevos valores en donde la mejor alternativa es la primera, ya que como se observa en el segundo gráfico la red sólo ha aprendido a pronosticar los valores ingresados, esto se conoce con el nombre de **sobre ajuste**, a mayor plasticidad mayor error en los datos de prueba.

Figura N° 2.3

Ajuste de los datos y la plasticidad



Fuente: Rojas (1996)

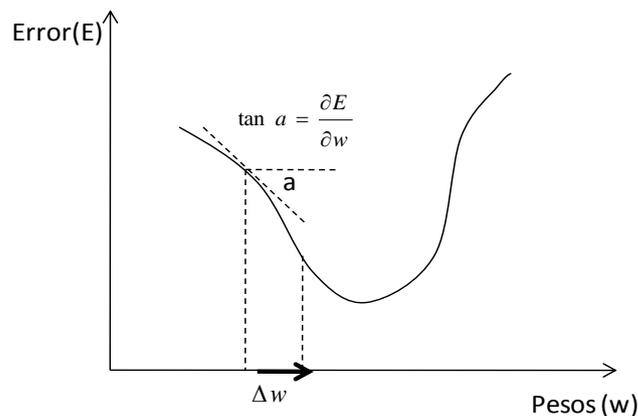
Una solución para evitar estos problemas es usar **datos de validación**, de esta manera la muestra es dividida en tres partes: los **datos de entrenamiento** que son los usados para el cálculo de gradientes y así actualizar pesos y bias, los **datos de validación** son los encargados de monitorear el proceso durante el entrenamiento, el error de validación durante las primeras iteraciones tiende a bajar pero llega un momento cuando la red empieza a tener sobreajuste y el error de los datos de validación se incrementan, es aquí donde se detiene el entrenamiento evitando el sobreajuste y una buena generalización que será corroborada con el tercer segmento de la data, **los datos de test o prueba**.

2.1.7 Ratio de aprendizaje

Es el encargado de determinar la velocidad de aprendizaje de la red, así su función es determinar la proporción con que irán variando los pesos. En el algoritmo de *Back Propagation* tiene su equivalencia en la gradiente de la función error (Ver figura 2.4), el valor seleccionado también es de gran importancia por el motivo que podemos caer en mínimos locales y no en un mínimo global (Ver figura 2.5).

Figura 2.4

El gradiente de la función error

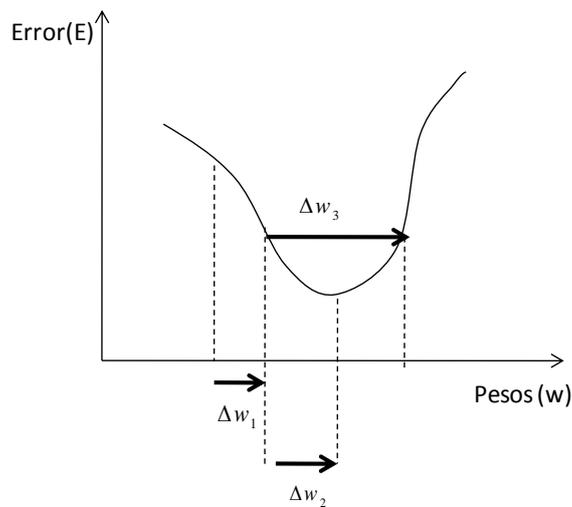


Fuente: Veelenturf(1995)

En la figura N° 2.5 se ve el efecto de tomar un ratio de aprendizaje equivocado, un valor pequeño ayuda a tomar mínimos globales tal como lo hace la variación 2, en cambio al tomar un valor grande podemos saltar los mínimos globales y tomar sólo mínimos locales esto ocurre en la variación de pesos 3.

Figura 2.5

La influencia de la variación de los pesos



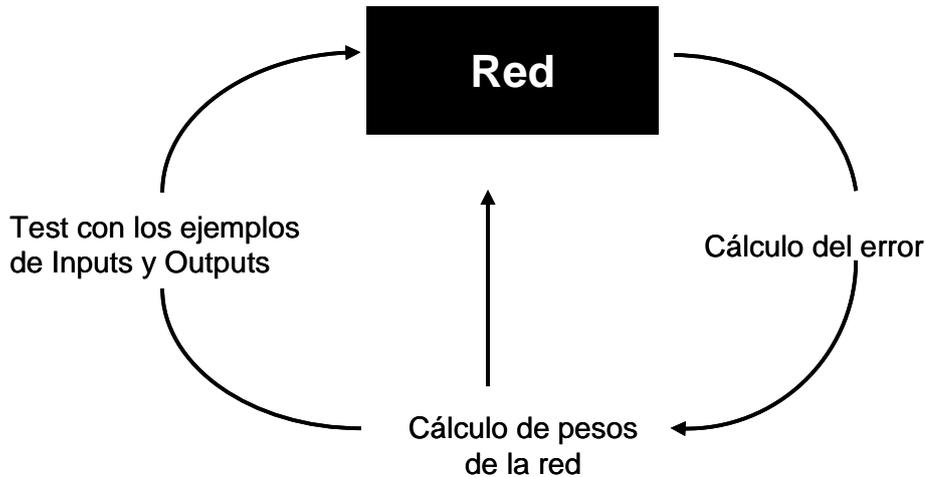
Fuente: Veelenurf(1995)

2.1.8 Perceptrón

En 1958 Frank Rosenblatt propone el perceptrón, que fue básicamente usado para el reconocimiento de patrones, y trabaja con un aprendizaje del tipo supervisado, la corrección de pesos se da en forma iterativa hasta que la red aprenda y llegue al valor deseado, tal como se muestra en la figura 2.6.

Figura N° 2.6

Aprendizaje de la red neuronal Perceptrón



Fuente: Rojas(1996)

La estructura de un perceptrón se muestra en la figura 2.7, en este caso cada elemento del vector de *input* **p** es conectado para cada neurona a través de la matriz de pesos **W**. La neurona *i* ésima procesa la información usando los elementos inputs ponderados y agregando el valor de la bia generando de esta forma el input $n(i)$ que será tomado por la función de transferencia, dando como resultado el output $a(i)$; matemáticamente este procesamiento es mostrado en 2.1 y 2.2.

$$n(i) = w_{i1}p_1 + w_{i2}p_2 + w_{i3}p_3 + \dots w_{iR}p_R + b(i) \quad (2.1)$$

$$a(i) = f(n(i)) \quad (2.2)$$

En la ecuación 2.1, la notación usada por lo pesos es la siguiente w_{ij} , haciendo referencia al peso que va hacia la neurona *i* desde el elemento *j* ésimo del vector input. Por ejemplo w_{12} es la fuerza con que viaja desde el

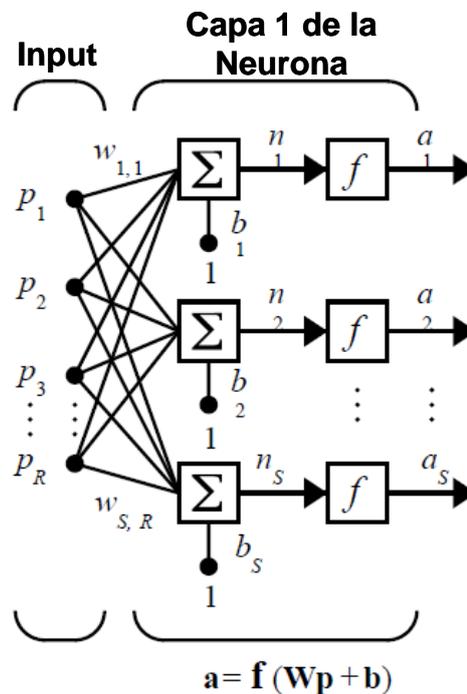
elemento 2 del vector input hacia la primera neurona. En forma matricial los pesos los podemos representar como sigue:

$$W = \begin{bmatrix} w_{1,1} & w_{1,2} & \dots & w_{1,R} \\ w_{2,1} & w_{2,2} & \dots & w_{2,R} \\ \vdots & \vdots & \ddots & \vdots \\ w_{S,1} & w_{S,2} & \dots & w_{S,R} \end{bmatrix} \quad (2.3)$$

Es importante de resaltar que el número de elementos inputs (**R**) no necesariamente tiene que ser igual al número de neuronas (**S**).

Figura N° 2.7

Estructura de una Red Neuronal con una sola capa

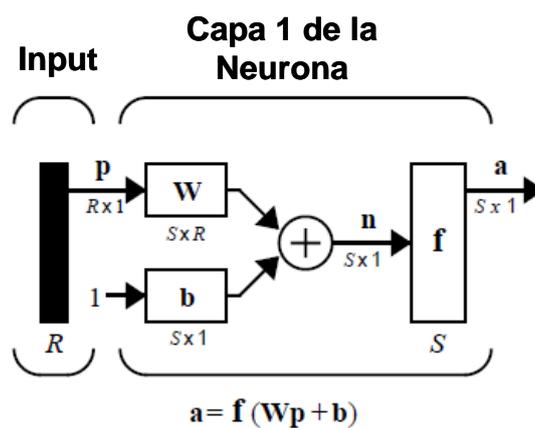


Fuente: Demuth *et al.* (2005)

Demuth et al. (2005) también muestra la forma como poder abreviar la representación de este tipo de neuronas, usando vectores y matrices , de acuerdo a ello se tendrá un vector \mathbf{p} de tamaño \mathbf{R} , una matriz de pesos $\mathbf{W}_{S \times R}$ (ver 2.3) , un vector de bias de tamaño \mathbf{S} , y un vector de resultados también de tamaño \mathbf{S} .

Figura N° 2.8

Notación abreviada de una Red Neuronal con una sola capa



Fuente: Demuth et al. (2005)

El perceptrón tiene varias limitaciones, muchas de las cuales fueron estudiadas por Minsky and Papert, la primera es que suele trabajar sólo con valores binarios, luego y lo más importante es que sólo es usado para clasificar problemas linealmente separables; ante esto las redes multicapas surgieron pudiendo ahora aproximar cualquier tipo de función.

En las redes multicapas cada capa tiene una matriz de pesos \mathbf{W} , un vector de bias \mathbf{b} , y un vector de outputs \mathbf{a} . Para distinguir entre la matriz de pesos, los vectores de outputs, etc. en cada una de las capas (ver figura N° 2.9) se usa un superíndice que señala la capa de pertenencia.

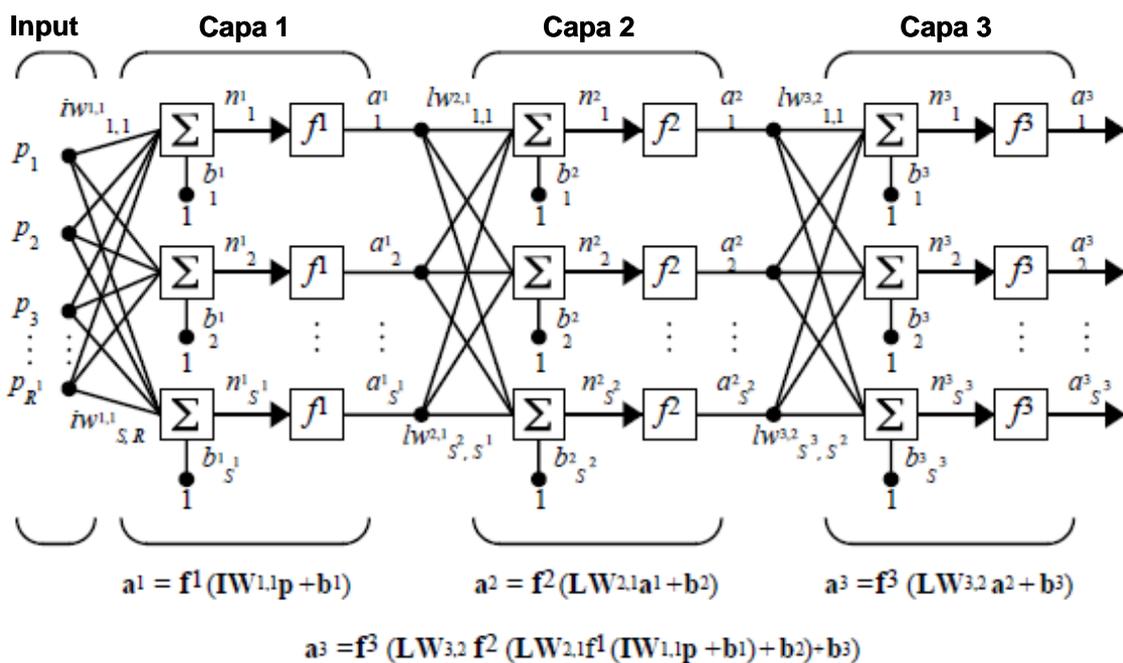
Al igual que en el caso de una capa, no es necesario que los \mathbf{R}^1 ingresos sean igual en número a las \mathbf{S}^1 neuronas de la primera capa, a las \mathbf{S}^2 neuronas

de la segunda capa, etc., al contrario es muy común que diferentes capas tengan diferentes números de neuronas.

Las salidas u *outputs* de cada capa intermedia son los *inputs* para las siguientes capas, de esta manera la capa 2 puede ser analizada como una red de una capa con inputs \mathbf{S}^1 y \mathbf{S}^2 neuronas, con matriz de pesos \mathbf{W}^2 de orden $\mathbf{S}^2 \times \mathbf{S}^1$, el vector de inputs es \mathbf{a}^1 y el de outputs \mathbf{a}^2 .

Cada una de las capas de las redes multicapa cumplen diferentes roles, la capa que da los resultados es la capa denominada **output** (en la figura N° 2.9 ver capa 3), todo el resto de capas son denominadas capas ocultas (en la figura N° 2.9 son las capas 1 y 2).

Figura N° 2.9
Arquitectura de una red Multicapa

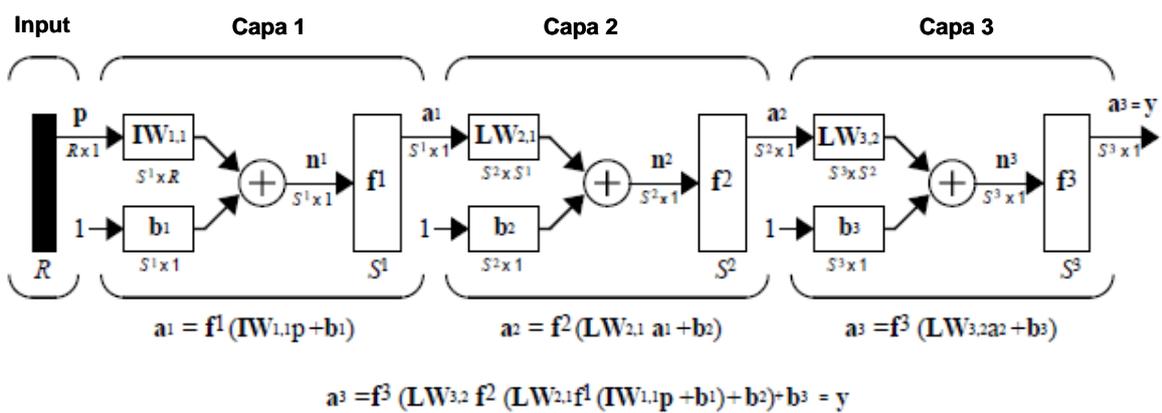


Fuente: Demuth et al. (2005)

La representación gráfica de las redes multicapas también pueden ser reducidas, siguiendo la misma lógica de la representación de las redes de una sola capa, por ejemplo la representación abreviada de la red mostrada en el figura N° 2.9 es la siguiente:

Figura N° 2.10

Arquitectura abreviada de una red Multicapa



Fuente: Demuth *et al.* (2005)

2.1.9 Algoritmo Back Propagation

Es uno de los más importantes algoritmos dentro de las redes multicapas, y su potencia radica en su capacidad de entrenar capas ocultas superando así las limitaciones de las redes de una sola capa, esto gracias al uso de la regla del gradiente descendente que busca ir ajustando los pesos de tal forma que minimice el error global.

$$w_{ij}^m(k+1) = w_{ij}^m(k) - \alpha \frac{\partial \hat{F}}{\partial w_{ij}^m} \tag{2.4}$$

$$b_i^m(k+1) = b_i^m(k) - \alpha \frac{\partial \hat{F}}{\partial b_i^m} \quad (2.5)$$

En las redes multicapas el error no es una función explícita de los pesos en las capas ocultas, por consiguiente sus derivadas no son fácilmente calculadas, por lo que se tendrá que hacer el cálculo en forma indirecta, para ello se asume que la función f del error es una función explícita de n .

$$\frac{\partial f(n(w))}{\partial w} = \frac{\partial f(n)}{\partial n} x \frac{\partial n(w)}{\partial w} \quad (2.6)$$

Basándonos en las definiciones planteadas en 2.4 y 2.5 se tiene:

$$\frac{\partial \hat{F}}{\partial w_{ij}^m} = \frac{\partial \hat{F}}{\partial n_i^m} x \frac{\partial n_i^m}{\partial w_{ij}^m} \quad (2.7)$$

$$\frac{\partial \hat{F}}{\partial b_i^m} = \frac{\partial \hat{F}}{\partial n_i^m} x \frac{\partial n_i^m}{\partial b_i^m} \quad (2.8)$$

El segundo término de las ecuaciones 2.7 y 2.8 puede ser reemplazado por:

$$n_i^m = \sum_{j=1}^{S^{m-1}} w_{ij}^m a_j^{m-1} + b_i^m \quad (2.9)$$

Por consiguiente:

$$\frac{\partial n_i^m}{\partial w_{ij}^m} = a_j^{m-1}, \frac{\partial n_i^m}{\partial b_i^m} = 1 \quad (2.10)$$

Ahora definimos la sensibilidad de \hat{F} ante cambios en el elemento i ésimo de la capa m :

$$S_i^m = \frac{\partial \hat{F}}{\partial n_i^m} \quad (2.11)$$

La expresión 2.11 para la última capa (M) y el resto de capas se puede simplificar como:

$$s^M = -2F^M (n^M)(t - a) \quad (2.12)$$

$$s^m = F^m (n^m)(W^{m+1})^T s^{m+1}, m = M - 1, \dots, 2, 1 \quad (2.13)$$

Simplificando 2.7 y 2.8 resulta:

$$\frac{\partial \hat{F}}{\partial w_{ij}^m} = s_i^m a_j^{m-1} \quad (2.14)$$

$$\frac{\partial \hat{F}}{\partial b_i^m} = s_i^m \quad (2.15)$$

Estos resultados los reemplazamos en 2.4 y 2.5:

$$w_{ij}^m(k+1) = w_{ij}^m(k) - \alpha s_i^m a_j^{m-1} \quad (2.16)$$

$$b_i^m(k+1) = b_i^m(k) - \alpha s_i^m \quad (2.17)$$

En forma matricial, se tiene:

$$W^m(k+1) = W^m(k) - \alpha s^m (a^{m-1})^T \quad (2.18)$$

$$b^m(k+1) = b^m(k) - \alpha s^m \quad (2.19)$$

Con ello el proceso iterativo se realizará hasta conseguir un error aceptable.

Dado la explicación matemática, se tiene que el algoritmo de *Back Propagation* se da en dos partes, en la primera que es la **propagación hacia adelante**, en cada una de las capas se van efectuando sus salidas respectivas a_j^i desde la primera hasta la última, esto se proceso se puede apreciar desde la figura 2.11 hasta la 2.13 donde se presenta un caso de una red con tres capas. Como se puede notar las salidas de una capa forman parte de los inputs de la capa siguiente.

Figura N° 2.11

Fase de propagación en la primera capa

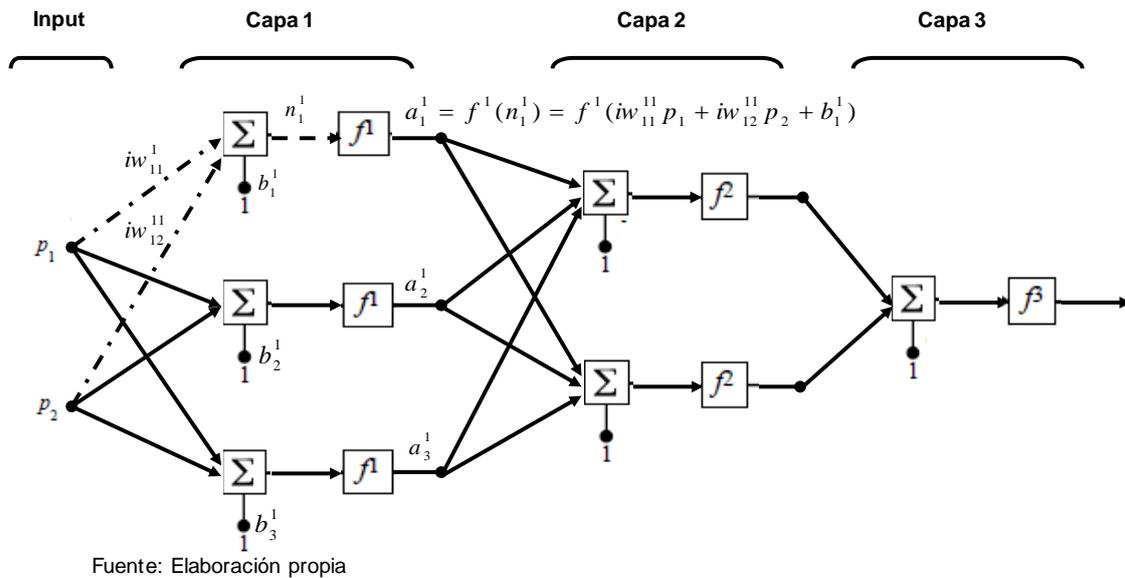
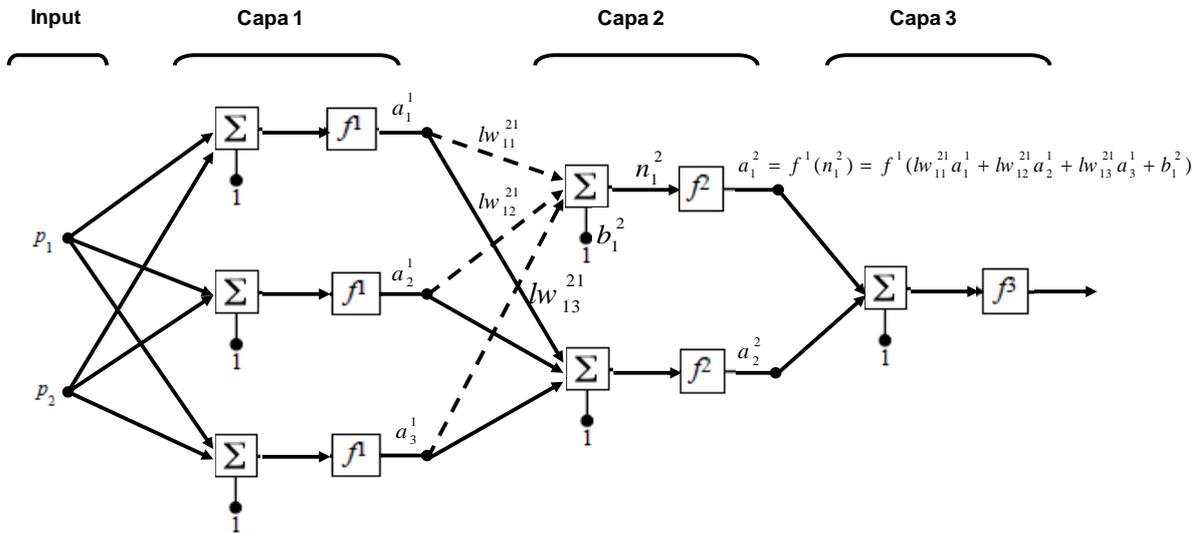


Figura N° 2.12

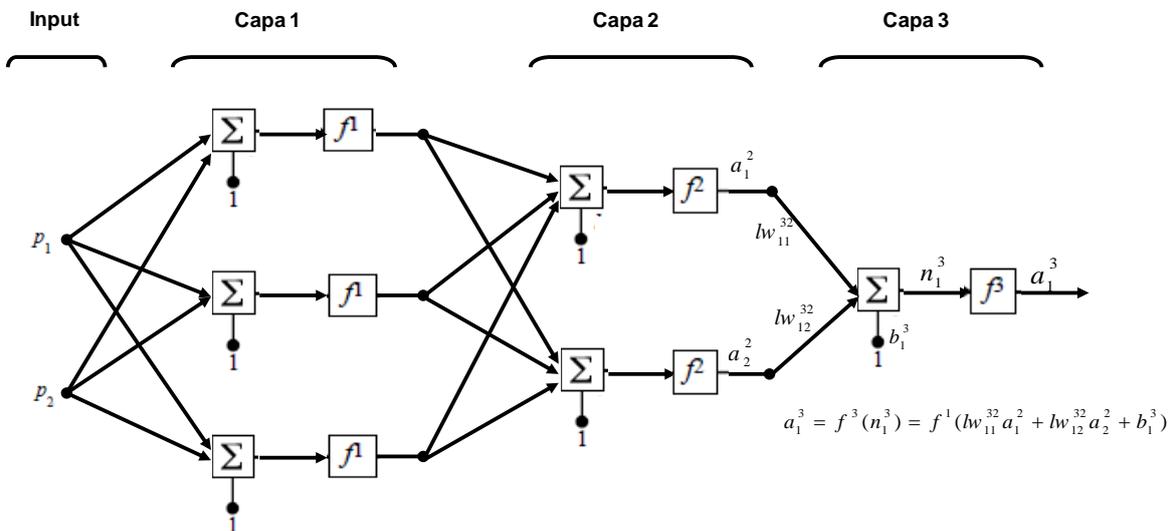
Fase de propagación en la segunda capa



Fuente: Elaboración propia

Figura N° 2.13

Fase de propagación en la tercera capa

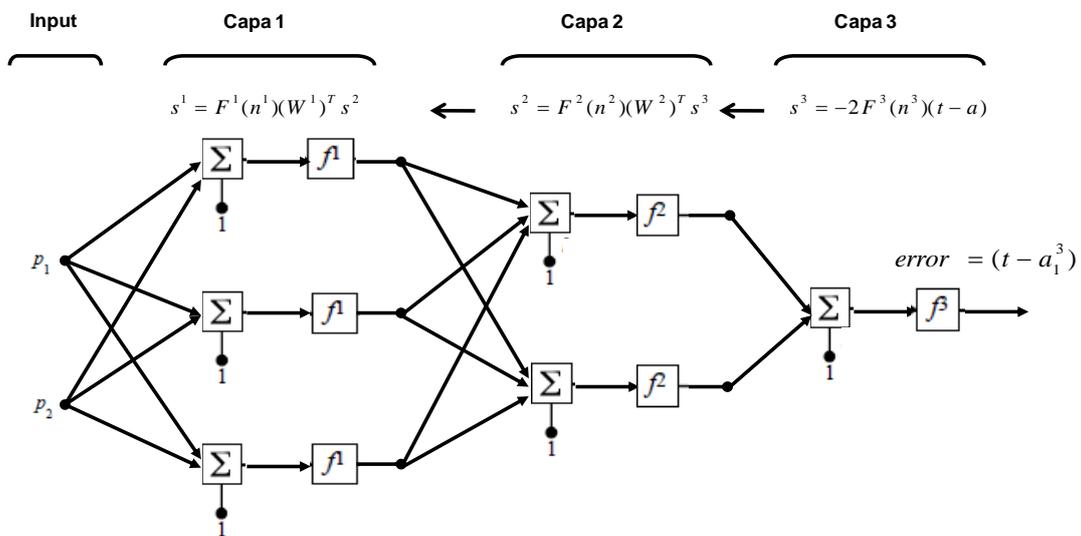


Fuente: Elaboración propia

Una vez que se tiene la salida de la última capa, en nuestro ejemplo la salida de la capa 3, comienza la segunda fase denominada **propagación hacia atrás**, en la figura 2.14 vemos como se inicia este proceso con el cálculo del error y luego con el cálculo de las sensibilidades desde la última capa hacia la primera, en esta parte se usa las ecuaciones 2.12 y 2.13.

Figura N° 2.14

Cálculo de las sensibilidades de la capa 3 a la 1

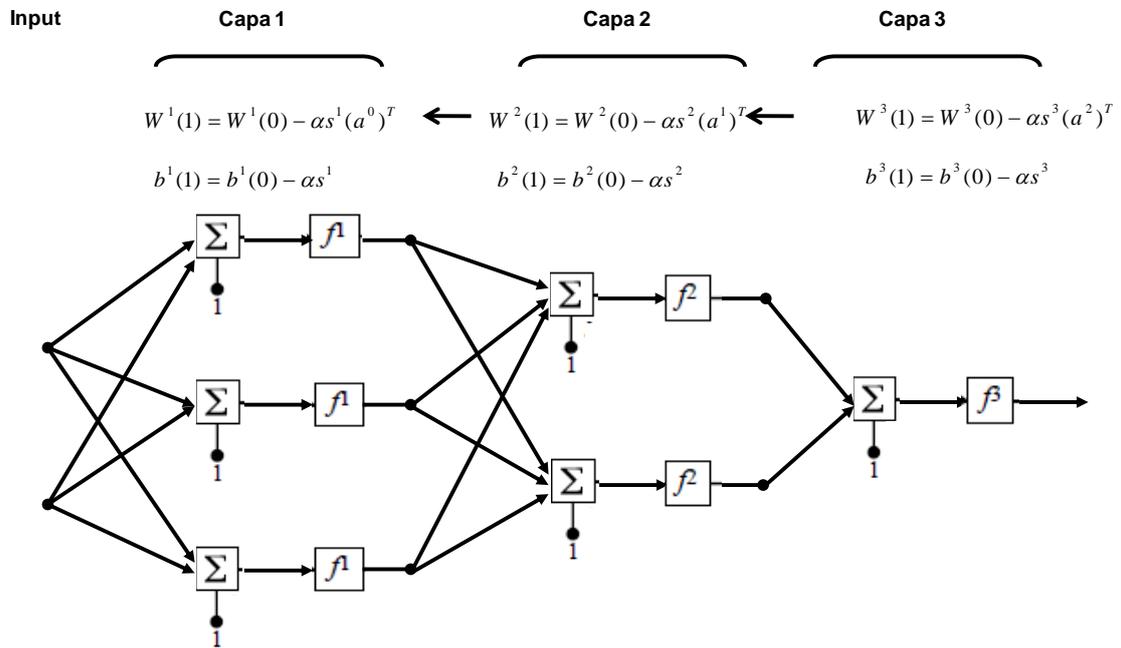


Fuente: Elaboración propia

Finalmente la actualización de los pesos y bias se dan también desde la última capa hasta la primera capa, en la figura 2.15 se aprecia como se realiza el cálculo para la primera iteración usando las ecuaciones indicadas en 2.18 y 2.19.

Figura N° 2.15

Actualización de pesos y bias en la primera iteración



Fuente: Elaboración propia

2.2 Análisis de Regresión Lineal Múltiple

2.2.1 Definición

En análisis de regresión es una técnica de dependencia en la cual se busca analizar la relación entre una variable criterio (variable dependiente) y varias variables independientes (predictores), Anderson et al (2001).

En general el modelo para relacionar la variable respuesta y con k regresores o variables predictoras sería:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_k x_k + \varepsilon \quad (2.20)$$

2.2.2 Estimación de los Coeficientes de Regresión por Mínimos Cuadrados

En el análisis de regresión lineal la estimación de los coeficientes es un paso muy importante, de esto dependerá que tan precisas sean las estimaciones, con este fin se hace uso de los mínimos cuadrados.

El modelo de regresión lineal planteado en la ecuación 2.20 se puede escribir como:

$$\begin{aligned} y_i &= \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_k x_{ik} + \varepsilon_i \\ &= \beta_0 + \sum_{j=1}^k \beta_j x_{ij} + \varepsilon_i, \quad i = 1, 2, \dots, n \end{aligned} \quad (2.21)$$

La función de mínimos cuadrados es:

$$\begin{aligned} S(\beta_0, \beta_1, \dots, \beta_k) &= \sum_{i=1}^n \varepsilon_i^2 \\ &= \sum_{i=1}^n (y_i - \beta_0 - \sum_{j=1}^k \beta_j x_{ij})^2 \end{aligned} \quad (2.22)$$

Para ello se debe minimizar la función S respecto a $\beta_0, \beta_1, \dots, \beta_k$, estos deben satisfacer:

$$\left. \frac{\partial S}{\partial \beta_0} \right|_{\hat{\beta}_0, \hat{\beta}_1, \dots, \hat{\beta}_k} = -2 \sum_{i=1}^n (y_i - \hat{\beta}_0 - \sum_{j=1}^k \hat{\beta}_j x_{ij}) = 0$$

$$\left. \frac{\partial S}{\partial \beta_j} \right|_{\hat{\beta}_0, \hat{\beta}_1, \dots, \hat{\beta}_k} = -2 \sum_{i=1}^n (y_i - \hat{\beta}_0 - \sum_{j=1}^k \hat{\beta}_j x_{ij}) x_{ij} = 0 \quad (2.23)$$

Al simplificar se obtiene las ecuaciones normales de mínimos cuadrados:

$$\begin{aligned} n\hat{\beta}_0 + \hat{\beta}_1 \sum_{i=1}^n x_{i1} + \hat{\beta}_2 \sum_{i=1}^n x_{i2} + \dots + \hat{\beta}_k \sum_{i=1}^n x_{ik} &= \sum_{i=1}^n y_i \\ \hat{\beta}_0 \sum_{i=1}^n x_{i1} + \hat{\beta}_1 \sum_{i=1}^n x_{i1}^2 + \hat{\beta}_2 \sum_{i=1}^n x_{i1} x_{i2} + \dots + \hat{\beta}_k \sum_{i=1}^n x_{i1} x_{ik} &= \sum_{i=1}^n x_{i1} y_i \\ \cdot & \cdot \cdot \cdot \cdot \\ \cdot & \cdot \cdot \cdot \cdot \cdot \\ \cdot & \cdot \cdot \cdot \cdot \cdot \\ \hat{\beta}_0 \sum_{i=1}^n x_{ik} + \hat{\beta}_1 \sum_{i=1}^n x_{ik} x_{i1} + \hat{\beta}_2 \sum_{i=1}^n x_{ik} x_{i2} + \dots + \hat{\beta}_k \sum_{i=1}^n x_{ik}^2 &= \sum_{i=1}^n x_{ik} y_i \end{aligned} \quad (2.24)$$

En total tenemos $p=k+1$ ecuaciones normales, si lo colocamos en forma matricial nos quedaría:

$$y = X\beta + \varepsilon \quad (4.12)$$

Donde:

$$y = \begin{bmatrix} y_1 \\ y_2 \\ \cdot \\ \cdot \\ y_n \end{bmatrix}$$

$$X = \begin{bmatrix} 1 & x_{11} & x_{12} & \dots & x_{1k} \\ 1 & x_{21} & x_{22} & \dots & x_{2k} \\ \cdot & \cdot & \cdot & & \cdot \\ \cdot & \cdot & \cdot & & \cdot \\ \cdot & \cdot & \cdot & & \cdot \\ 1 & x_{n1} & x_{n2} & \dots & x_{nk} \end{bmatrix}$$

$$\beta = \begin{bmatrix} \beta_1 \\ \beta_2 \\ \cdot \\ \cdot \\ \beta_n \end{bmatrix}$$

$$\varepsilon = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \cdot \\ \cdot \\ \varepsilon_n \end{bmatrix}$$

(2.25)

Minimizando la función de mínimos cuadrados se tiene:

$$S(\beta) = \sum_{i=1}^n \varepsilon_i^2 = \varepsilon' \varepsilon = (y - X\beta)'(y - X\beta)$$

$$S(\beta) = y'y - \beta'Xy - y'X\beta + \beta'X'X\beta$$

$$S(\beta) = y'y - 2\beta'X'y + \beta'X'X\beta$$

$$\left. \frac{\partial S}{\partial \beta} \right|_{\hat{\beta}} = -2X'y + 2X'X\hat{\beta} = 0$$

$$X'X\hat{\beta} = X'y$$

$$\hat{\beta} = (X'X)^{-1}X'y \quad (2.26)$$

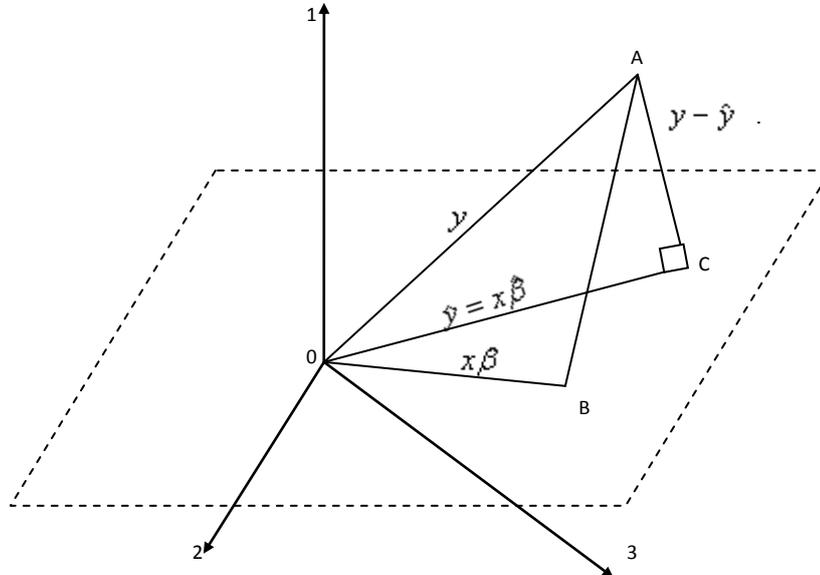
2.2.3 Interpretación geométrica de mínimos cuadrados

Montómery (2004) presenta una interpretación geométrica de los mínimos cuadrados, dado el vector de las observaciones $y' = [y_1, y_2, \dots, y_n]$ define un vector del origen al punto A de la figura 2.16, y_1, y_2, \dots, y_n formando las coordenadas de un espacio muestral n dimensional. La matriz consiste en p vectores columna de orden $n \times 1$, estos p vectores forman un espacio p dimensional llamado espacio de estimación. Se puede representar cualquier punto en este espacio de estimación mediante una combinación lineal de los vectores $1, x_1, \dots, x_k$, de esta forma cualquier punto del espacio de estimación tiene la forma $X\beta$, por ejemplo el punto de B al punto A elevado al cuadrado se define como:

$$S(\beta) = (y - X\beta)'(y - X\beta) \quad (2.27)$$

Figura N° 2.16

Interpretación geométrica de los mínimos cuadrados



Fuente: Montgomery *et.al* (2004)

Para minimizar la distancia al cuadrado del punto A definido por el vector de observaciones y el espacio de estimación es necesario determinar el punto del espacio de estimación que esté más cercano al punto A, este punto es el punto C y se ubica al pie de la recta que llega desde el punto A y es perpendicular o normal al espacio de estimación, el punto C se define como :

$\hat{y} = X\hat{\beta}$, con ello podemos establecer:

$$y - y' = y - X\hat{\beta} \tag{2.28}$$

$$X'X\hat{\beta} = X'y \tag{2.29}$$

Esta última ecuación es conocida con el nombre de ecuaciones normales de mínimos cuadrados.

2.2.4 Supuestos del análisis de Regresión Lineal

Los supuestos son usados para comprobar la adecuación del modelo, las violaciones a estos supuestos pueden provocar un modelo inestable, produciendo resultados totalmente diferentes y en muchos casos conclusiones opuestas al trabajar con distintas muestras. Montgomery (2004) señala que no basta con observar los estadísticos de resumen como el t , R^2 , F , porque son sólo propiedades **globales** y como tales no aseguran la adecuación del modelo. Los supuestos que se deben verificar son:

- Linealidad
- Independencia
- Homocedasticidad
- Normalidad
- No - Colinealidad

A continuación se explicará brevemente cada uno de ellos:

2.2.4.1 Linealidad

Señala que la relación entre la variable dependiente es una combinación lineal de las variables predictoras y los residuos, El incumplimiento de este supuesto suele denominarse **error de especificación**. Si se tiene variables que no cumplen con este supuesto se puede aplicar transformaciones. Una forma práctica de verificar el cumplimiento o no de la linealidad es visualmente, el programa estadístico SPSS propone los gráficos de regresión parcial para la variable dependiente con cada una de las variables independientes.

Estos diagramas son similares a los diagramas de dispersión pero no están basados en las puntuaciones originales, sino en los residuos obtenidos al efectuar el análisis de regresión lineal con el resto de variables independientes.

2.2.4.2 Independencia

Xin and Xiao (2009) señalan que el supuesto de independencia está asociado con la selección aleatoria de las observaciones de la muestra, porque esto conlleva a que los errores sean independientes entre sí, es decir los residuos forman una variable aleatoria.

Uno de los test más usados para comprobar este supuesto es el estadístico Durbin – Watson (1951) :

$$DW = \frac{\sum_{i=2}^n (e_i - e_{i-1})^2}{\sum_{i=1}^n e_i^2} \quad (2.30)$$

Donde $e_i = Y_i - \hat{Y}_i$ se refiere a los residuos, este estadístico suele oscilar entre 0 y 4, pudiendo asumirse independencia cuando se tiene valores entre 1.5 y 2.5.

Entre las causas de la correlación de los errores se tiene la omisión de variables relevantes, en tal caso las variables omitidas pasan a formar parte del término del error, así si hay correlación entre distintas observaciones de las variables omitidas, también lo habrá entre distintos valores de los términos de perturbación (Escalante (2010)).

Como posible consecuencias de no cumplir este supuesto, Kikut (2003) menciona:

- Los estimadores β son lineales, insesgados y consistentes pero no eficientes (varianza mínima).
- Al existir la posibilidad de que la varianza estimada subestime la verdadera, podrían sobreestimarse el R^2 y las pruebas t y F dejar de ser válidas, si se aplican es probable que conduzcan a conclusiones erróneas sobre la significancia estadística de los estimadores.

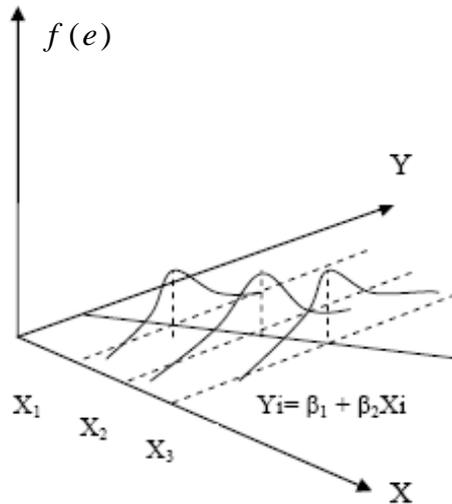
2.2.4.3 Homocedasticidad

Se cumplirá la Homocedasticidad cuando la varianza de los errores es constante, o cuando las varianzas de los errores de las estimaciones condicionales a los valores de las variables explicativa son idénticas; si no se cumpliera este supuesto se denomina heterocedasticidad. Las causas que pueden provocar la heterocedasticidad son:

- La variable respuesta cambia su magnitud para valores de uno o más variables independientes.
- Se tiene en la data valores atípicos, la inclusión o exclusión de tales observaciones, especialmente si la muestra es pequeña, puede afectar los resultados de la regresión.
- Mala especificación del modelo debido a omisión de variables importantes para el modelo.

Figura N° 2.17

Homocedasticidad



Fuente: Kikut *et al.*(2003)

Entre las consecuencias de no cumplir con este supuesto se tiene:

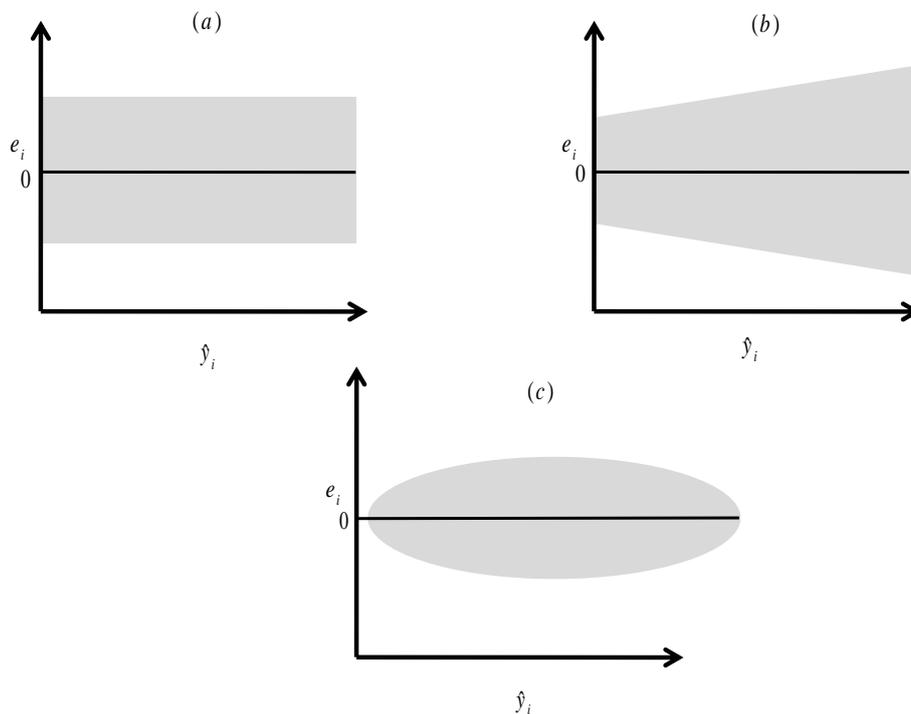
- Los coeficientes de regresión estimados son lineales e insesgados.
- Disminución de la eficiencia del estimador mínimo cuadrático. Éste deja de ser el de mínima varianza entre todos los estimadores lineales e insesgados.

Para la comprobar si se cumple o no este supuesto se pueda hacer uso de gráficos de los residuales e_i con los valores ajustados correspondientes \hat{y}_i , por ejemplo si se tiene una gráfica como la figura N° 2.18 (a) se puede concluir que se tiene errores con una varianza constante, por lo contrario la figura N° 2.18 (b) y (c) señalan que la varianza de los errores no es contante.

Figura N° 2.18

Patrones en la gráfica de residuales

Fuente:



Montgomery (2004)

Un estadístico empleado también para ver la homogeneidad de la varianza es Levene, y se define como:

$$W = \frac{(N - k) \sum_{i=1}^k n_i (\bar{Z}_i - \bar{Z}_{..})^2}{(k - 1) \sum_{i=1}^k \sum_{j=1}^{n_i} (Z_{ij} - \bar{Z}_i)^2} \quad (2.31)$$

Donde $z_{ij} = |X_{ij} - \bar{X}_i|$, con \bar{X}_i , es la media del i - ésimo subgrupo,

2.2.4.4 Normalidad

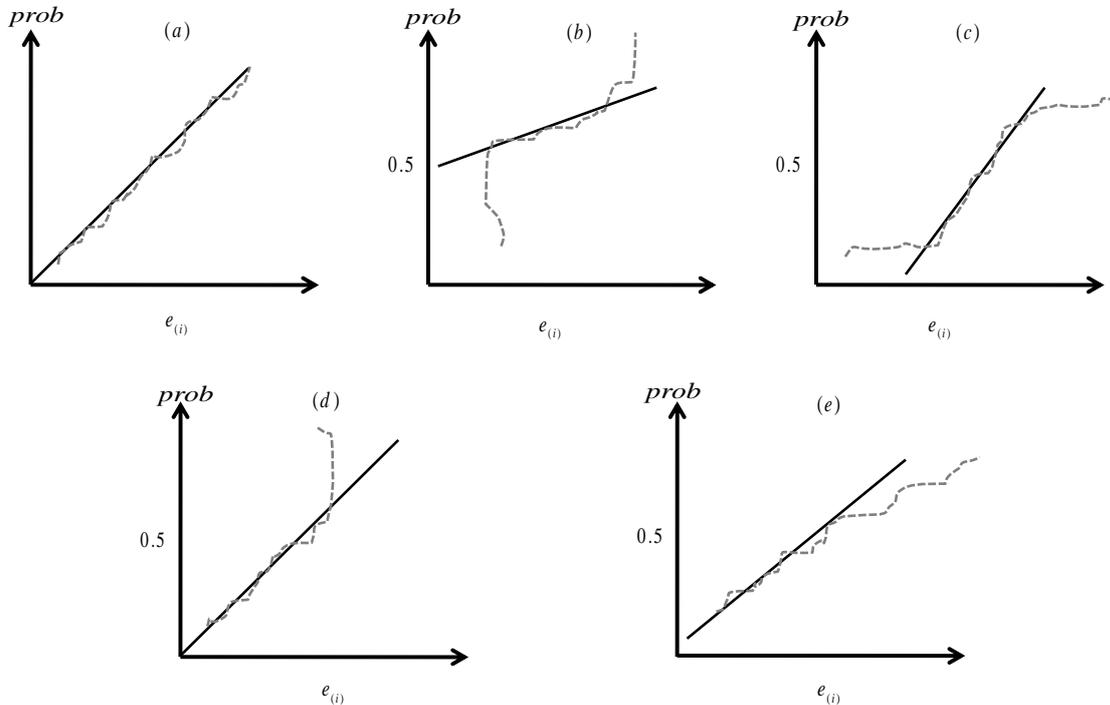
Las pequeñas desviaciones respecto a la hipótesis de normalidad no afectan mucho al modelo, pero una no normalidad grande es más seria, porque los estadísticos t o F y los intervalos de confianza y predicción dependen de la suposición de la normalidad, más aun si los errores provienen de una distribución con colas más gruesas que la normal, el ajuste de mínimos cuadrados será sensible a un subconjunto menor de datos.

Al igual que en la comprobación de los supuestos anteriores un análisis gráfico es de gran ayuda; en este caso se hace uso de gráficos de probabilidad normal, la figura N° 2.19 (a) muestra una distribución normal idealizada, los puntos caen aproximadamente cerca a la recta. Las figuras desde la (b) hasta la (e), muestran problemas característicos, en la (b) se tiene colas gruesas en ambos extremos; la (c) es una distribución con colas más delgadas que una distribución normal, finalmente las figuras (d) y (e) muestran patrones de asimetría positiva y negativa respectivamente.

Hay autores como Andrews (1979) y Gnanadesikan (1977) quienes indican que el uso de gráficos de probabilidad normal no son del todo confiable, por ello es necesario también el uso de test estadísticos entre ellos tenemos el test de Shapiro – Willks y el de Kolmogorov – Smirnov.

Figura 2.19

Gráficos de Probabilidad Normal



Fuente: Montgomery (2004)

2.2.4.5 No colinealidad

Anderson et al. (2001) define la colinealidad como la asociación o correlación entre dos variables independientes y la multicolinealidad indican que es la asociación entre tres o más variables independientes.

Montgomery (2004) señala que existen cuatro principales fuentes de multicolinealidad:

- El método de recolección de datos que se empleó
- Restricciones en el modelo o en la población
- Especificación del modelo

- Un modelo sobre definido

También nos señala que la multicolinealidad tiene diversos efectos en la regresión lineal, siendo el más grave la inestabilidad de los coeficientes de regresión. Por ejemplo suponiendo que se escalan x_1, x_2 y y a la longitud unitaria, tendremos un modelo:

$$y = \beta_1 x_1 + \beta_2 x_2 + \varepsilon \quad (2.32)$$

Sus ecuaciones normales son definidas como:

$$(X'X)\hat{\beta} = X'y$$

$$\begin{bmatrix} 1 & r_{12} \\ r_{12} & 1 \end{bmatrix} \begin{bmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \end{bmatrix} = \begin{bmatrix} r_{1y} \\ r_{2y} \end{bmatrix} \quad (2.33)$$

Donde r_{12} es la correlación simple entre x_1 y x_2 y r_{jy} es la correlación entre x_j y y , $j = 1, 2$. La inversa de $(X'X)$ es:

$$C = (X'X)^{-1} = \begin{bmatrix} \frac{1}{(1-r_{12}^2)} & \frac{-r_{12}}{(1-r_{12}^2)} \\ \frac{-r_{12}}{(1-r_{12}^2)} & \frac{1}{(1-r_{12}^2)} \end{bmatrix} \quad (2.34)$$

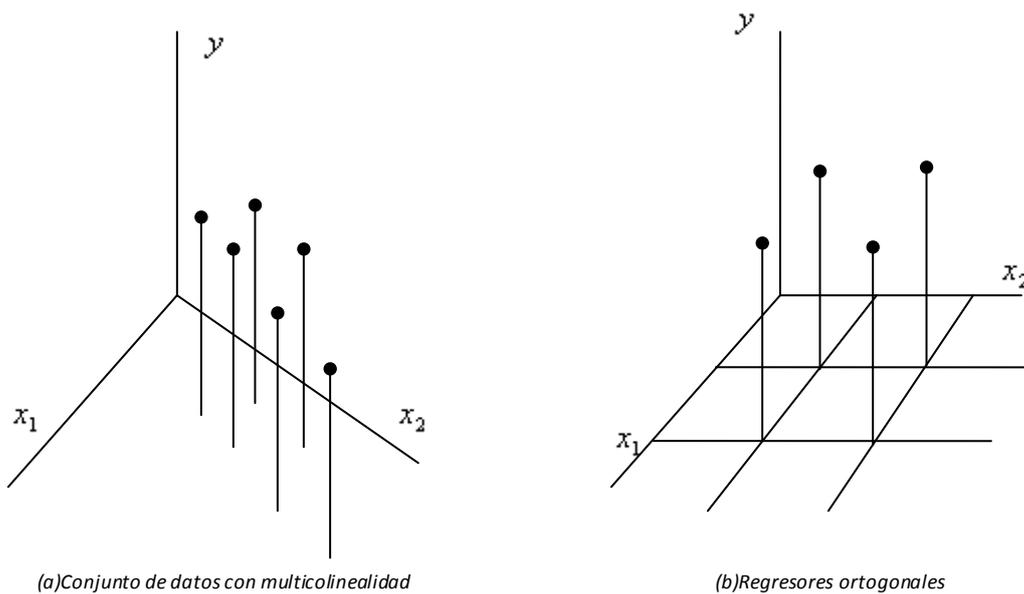
Con ello los estimados de los coeficientes de regresión son:

$$\hat{\beta}_1 = \frac{r_{1y} - r_{12}r_{2y}}{(1-r_{12}^2)} \quad \hat{\beta}_2 = \frac{r_{2y} - r_{12}r_{1y}}{(1-r_{12}^2)} \quad (2.35)$$

Si hay fuerte multicolinealidad se cumplirá que $|r_{12}| \rightarrow 1$, $Var(\hat{\beta}_j) = c_{jj}\sigma^2 \rightarrow \infty$ y además $Cov(\hat{\beta}_1, \hat{\beta}_2) = c_{12}\sigma^2 \rightarrow \pm\infty$, esto provocaría grandes varianzas y covarianzas de los estimadores de regresión por mínimos cuadrados, esto implicaría que distintas muestras tomadas con los mismos valores de x podrían ocasionar estimaciones muy diferentes de los parámetros del modelo. Geométricamente la multicolinealidad formaría planos inestables, caso contrario a los regresores sin multicolinealidad (regresores ortogonales) cuyo plano sería mucho más estable, tal como se puede apreciar en la siguiente figura:

Figura N° 2.20

Efectos de la multicolinealidad: Regresores inestables



Fuente: Montgomery et.al (2004)

2.2.5 Validación del Modelo

La validación de un modelo de regresión lineal es demostrar o confirmar si el modelo es efectivo para los propósitos que han sido creados, esto no sólo consiste en ver si el modelo se ajusta bien a los datos de la muestra con que se

trabaja si no también a datos independientes a la muestra. Así para hacer una buena validación se debe revisar los siguientes criterios:

- La ecuación de regresión provee predicciones no sesgadas de las cantidades de interés.
- La precisión de la predicción es lo suficientemente buena para los objetivos del estudio.

Ambos criterios, el sesgo y la varianza, son incorporados en la medida llamada error cuadrado medio de predicción (MSE), y se define como el promedio de la diferencia al cuadrado del valor observado con el valor predicho, ver ecuación N° 2.32.

$$MSE = \frac{(Y_i - \hat{Y}_i)^2}{n} \quad (2.36)$$

2.3 Datos atípicos

Un *outlier* o dato atípico es una observación que es numéricamente distante del resto de datos. Beckman y Cook (1983) tras una amplia revisión de documentos sobre este tema, llegan a definir a los datos atípicos como “un colectivo que puede ser catalogado como una observación contaminante o discordante”. Una observación discordante es una observación que aparece en forma sorpresiva o discrepante ante el investigador, mientras que una observación contaminante es una observación que no proviene de la población objetiva.

Los datos atípicos del tipo discordante tienen diversos métodos tanto univariados como multivariados para ser detectados. Entre los métodos multivariados tenemos el basado en la distancia de Mahalanobis, la forma y el

tamaño de una data multivariada son cuantificadas por la matriz de varianza y covarianza (C) , una distancia asociada a esta matriz es la distancia de Mahalanobis, en una muestra p dimensional se define como:

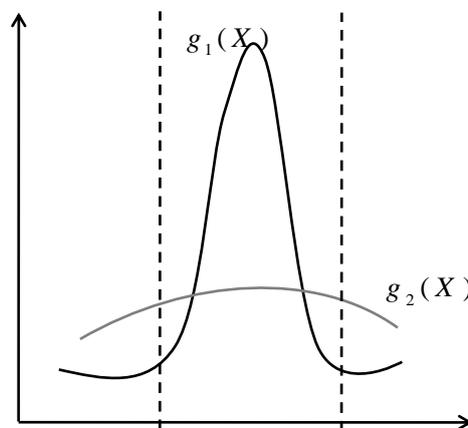
$$DM_i = ((x_i - t)^T C^{-1} (x_i - t))^{1/2} \quad (2.37)$$

Donde t es la media, para una data multivariada estos valores siguen una distribución Chi cuadrada con p grados de libertad (χ_p^2), los datos atípicos son las observaciones que tienen una mayor distancia.

A diferencia de los datos atípicos discordantes los del tipo contaminante son difíciles de detectar, en la figura 2.21 tenemos dos distribuciones, la $g_1(x)$ pertenece a la población de estudio en cambio $g_2(x)$ son datos diferentes a la población objetivo y serían catalogados como contaminantes, pero si se aplica un método de detección de *outliers* solo se puede detectar los que se encuentran en las colas dejando en la muestra una gran proporción de atípicos por ello se dice que sólo con la experiencia del investigador podría detectarlos.

Figura 2.21

Datos atípicos contaminantes y discordantes



Fuente: Iglewicz y Hoaglin(1993)

Capítulo III

Hipótesis y metodología

3.1 Hipótesis General

La presencia de datos atípicos, ya sean del tipo discordante o contaminante afectan en forma significativa (a un 10%) el desempeño de la red neuronal supervisada y entrenada con el método de *Back Propagation*.

3.1.1 Hipótesis específica I

El método de *Back Propagation*, no garantiza la calidad del proceso de aprendizaje (obtención de un buen ajuste) y la generalización de los datos, al tener datos atípicos ya sea en la parte de entrenamiento o validación.

3.1.2 Hipótesis específica II

El efecto de los datos atípicos del tipo contaminante sobre la calidad del proceso de aprendizaje y la generalización de los datos, es menor que el efecto provocado por los datos atípicos discordantes.

3.2 Descripción de la metodología

Para contrastar las hipótesis planteadas ha sido necesario desarrollar un experimento aleatorio para lo cual se tiene que seguir los siguientes pasos:

- (1) Establecer definiciones operativas que permitan cuantificar los conceptos sistematizados en el marco teórico (capítulo II) -Fase metodológica, capítulo IV.
- (2) Operacionalizar dichas definiciones mediante indicadores (acápites 4.3.2) de caracterización de datos de entrenamiento y validación, indicadores de aprendizaje y generalización. (Fase metodológica, capítulo III).
- (3) Bloqueo.- Establecer los escenarios definidos por la cantidad de datos, la presencia o no de datos atípicos (ubicación dentro de la data y el tipo de dato atípico) manteniendo las mismas condiciones experimentales en cada escenario. (Fase experimental, capítulo IV).
- (4) Aleatorización.- Según los parámetros definidos en cada escenario se generan los números aleatorios p-dimensionales. (Fase experimental, capítulo IV).
- (5) Réplica.- En cada escenario se realizan diez corridas con el fin de obtener resultados consistentes. (Fase experimental, capítulo IV).
- (6) Control.- En el análisis de resultados el modelo estadístico de Análisis de Regresión Múltiple será utilizado como control dado su buen rendimiento al cumplirse todos los supuestos que exige este modelo.

3.3 Definiciones y su Operacionalización

3.3.1 Definiciones Operativas

i. Estimador Estadístico

Un estimador es un estadístico (que se obtiene a partir de una muestra) usado para estimar un parámetro desconocido de la población. Un buen estimador debe poseer las propiedades de insesgadez, eficiencia, convergencia y robustez.

ii. Estructura de datos

En general se hace referencia a la estructura de los datos para especificar las características poblacionales de un conjunto de datos, en nuestro caso se garantizará que cumplan los supuesto del análisis multivariado de Regresión Lineal: distribución normal multivariada (Indicador A.1, A.2 y A.3), la no colinealidad (Indicador A.4), independencia (Indicador A.5), la homocedasticidad (Indicador A.6) y la linealidad la cual se ha trabajado en forma gráfica.

iii. Dato atípico

En el estudio como ya se explicó contaremos con dos tipos: los datos atípicos discordantes que son observaciones extremas que se presentan en cualquiera de las 8 variables independientes y pueden tener una desviación de 6 a 25 veces la de la variable univariada. En el caso de los datos atípicos funcionales todas las variables provienen de una población distinta, por ende tienen media y desviación diferente.

iv. Datos de Entrenamiento

Conjunto de datos utilizados por la red neuronal para modificar los pesos y las bias y el gradiente durante la etapa de entrenamiento, de tal forma que las salidas arrojadas por la red neuronal sean lo más cercanas a los datos de salida o *target*. Los datos de entrenamiento constituyen el 70% de los datos de la muestra. En este conjunto de datos se incorpora la mayor cantidad de datos atípicos.

v. Datos de Validación

Conjunto de datos utilizados para monitorear el error durante la etapa de entrenamiento. Representan el 15% de los datos de la muestra, y en su estructura también serán incorporados los datos atípicos.

vi. Datos de Prueba

Es el conjunto de datos usados después del entrenamiento y función básica es ver la eficiencia de la red neuronal y compararla con diferentes modelos. Representan el 15% de los datos de la muestra y no presentan casos atípicos.

vii. Criterio de Parada

El algoritmo se detendrá si: se excede las 100 épocas, se obtiene un gradiente menor a $1e-10$, o se logra tener un error de 0.

viii. Aprendizaje

Es un proceso que hace que el algoritmo calcule el gradiente y actualice los pesos y bias de la red neuronal, cuyos valores iniciales son aleatorios.

ix. Calidad del Proceso de Aprendizaje

Mide la capacidad de la red neural para aproximarse a los valores objetivos o *target* presentados durante el entrenamiento, en este experimento usaremos el error cuadrado medio como su indicador (Indicador4.9). A menor ECM mejor la calidad del proceso de aprendizaje.

x. Generalización

Es una característica de las redes neuronales, si se entrena a una red neuronal con un número suficiente de patrones, y luego se le presenta un nuevo conjunto de datos y dan salidas correctas entonces se dirá que la red generaliza en forma correcta. La generalización se medirá con el indicador C.1.

xi. Tiempo de entrenamiento

Tiempo empleado por la red para aprender o calcular los pesos sinápticos. Se mide en segundos o también se puede medir en forma indirecta por el número épocas que pasan por el algoritmo hasta lograr la convergencia. El número de épocas se usará presenta como indicador en el cuadro N° 3.8.

3.3.2 Indicadores

Los indicadores se clasifican en: a) indicadores de caracterización de datos de entrenamiento y validación, b) indicadores de aprendizaje e c) indicadores de generalización.

3.3.2.1 Caracterización de Datos

Gráfico N° 3.1

Indicador A.1 Normalidad Multivariada de Mardia

NOMBRE	Test de Normalidad Multivariada de Mardia
FUNDAMENTO	Permite contrastar la hipótesis de normalidad de un conjunto de n – observaciones p- variadas.
FORMA DE CÁLCULO	Se basa en la prueba individual del test de simetría y curtosis multivariada.
FUENTE	Datos generados por simulación de una población normal con parámetros especificados en los Anexos IV.1
PUNTOS CRÍTICOS	Ver puntos críticos del Cuadro N° 4.2 Y 4.3

Gráfico N° 3.2

Indicador A.2 Coeficiente de Asimetría Multivariado

NOMBRE	Coeficiente de Asimetría Multivariado (γ_1)
FUNDAMENTO	Medida de asimetría multivariada
FORMA DE CÁLCULO	$\hat{\gamma}_1 = \frac{1}{n^2} \sum_{i=1}^n \sum_{j=1}^n m_{ij}^3$ <p>El estimador del coeficiente de asimetría poblacional (γ_1) se definió en (3.29).</p>
FUENTE	Datos generados por simulación de una población normal con parámetros especificados en los Anexos IV.1
PUNTOS CRÍTICOS	<p>$\gamma_1 > 0$ Asimetría esférica</p> <p>$\gamma_1 = 0$ Distribución esférica uniforme</p> <p>$(1/6)n\gamma_1 \sim \chi^2(f)$ al $\alpha=10\%$ de significación; $f=p(p+1)(p+2)/6$ si la población es normal multivariada y $n \rightarrow \infty$; p: 8 variables univariadas</p>

Gráfico N° 3.3

Indicador A.3 Coeficiente de Curtosis Multivariado

NOMBRE	Coeficiente de Curtosis Multivariado (γ_2)
FUNDAMENTO	Medida de grado de curtosis multivariada.
FORMA DE CÁLCULO	$\hat{\gamma}_2 = \frac{1}{n} \sum_{i=1}^n m_i^4$ <p>El estimador del coeficiente de curtosis poblacional está definido en (3.27).</p>
FUENTE	Datos generados por simulación de una población normal con parámetros especificados en los Anexos IV.1
PUNTOS CRÍTICOS	<p>$(\gamma_2) = 80$ p=8 variables</p> <p>$(\gamma_2) > 80$ Alto grado de curtosis multivariada comparado a la distribución normal multivariada para p=8</p> <p>$(\gamma_2) < 80$ Bajo grado de curtosis multivariada comparado a la distribución normal multivariada para p=8</p> <p>$(\gamma_2 - p(p+2))/(8p(p+2)/n) ^{1/2} \sim N(0,1)$ con $\alpha=10\%$ de significación; si la población es normal multivariada y $n \rightarrow \infty$</p>

Gráfico N° 3.4

Indicador A.4 Intensidad de Multicolinealidad

NOMBRE	Intensidad de Multicolinealidad (k)
FUNDAMENTO	Medida de grado de colinealidad de las variables

FORMA DE CÁLCULO	$k = \sqrt{\frac{\lambda_{MAX}}{\lambda_{MIN}}}$ <p>Donde λ_{MIN} y λ_{MAX} son los valores característicos mínimo y máximo respectivamente de la matriz varianza-covarianza 'S'.</p>
FUENTE	Datos generados por simulación de una población normal con parámetros especificados en los Anexos IV.1
PUNTOS CRÍTICOS	<p>k=1 No hay presencia de multicolinealidad</p> <p>K > 20 Multicolinealidad Elevada</p> <p>K > 30 Multicolinealidad Severa</p> <p>K= +∞ Multicolinealidad perfecta</p>

Gráfico N° 3.5

Indicador A.5 Grado de Independencia de Residuos

NOMBRE	Grado de Independencia de Residuos – Estadístico Durbin - Watson (DW)
FUNDAMENTO	Proporciona información sobre el grado de independencia existente de los residuos, para verificar el supuesto de Independencia.
FORMA DE CÁLCULO	$DW = \frac{\sum_{i=2}^n (e_i - e_{i-1})^2}{\sum_{i=1}^n e_i^2}$ <p>Donde $e_i = Y_i - \hat{Y}_i$ son los residuos de un modelo de regresión.</p>
FUENTE	Datos generados por simulación de una población normal con parámetros especificados en los Anexos IV.1
PUNTOS CRÍTICOS	<p>Se puede asumir independencia entre residuos cuando:</p> $1.5 \leq DW \leq 2.5$

Gráfico N° 3.6

Indicador A.6 Homocedasticidad

NOMBRE	Homocedasticidad – Estadístico Levene
FUNDAMENTO	Proporciona información para evaluar la varianza de los residuos
FORMA DE CALCULO	$W = \frac{(N - k) \sum_{i=1}^k n_i (\bar{Z}_i - \bar{Z}_{..})^2}{(k - 1) \sum_{i=1}^k \sum_{j=1}^{n_i} (Z_{ij} - \bar{Z}_{i.})^2}$ <p>Donde $z_{ij} = X_{ij} - \bar{X}_{i.}$, con \bar{X}_{ij}, es la media del i - ésimo subgrupo,</p>
FUENTE	Datos generados por simulación de una población normal con parámetros especificados en los Anexos IV.1
PUNTOS CRÍTICOS	<p>La hipótesis nula planteada es:</p> <p>Ho: Las varianzas son iguales</p> <p>Se rechaza la hipótesis con un nivel de significancia $\alpha=10\%$ si $W > F_{\alpha, k-1, N-k}$ es el valor crítico superior de la distribución F con $k - 1$ grados de libertad en el numerador y $N - k$ grados en el denominador de significancia α.</p>

3.3.2.2 Indicadores de Aprendizaje

Gráfico N° 3.7

Indicador B.1 Valor adaptivo

NOMBRE	Valor adaptivo (μ)
FUNDAMENTO	Parámetro usado para el ajuste de los pesos y bias, que varía controlando en incremento o decrecimiento del error.
FORMA DE CÁLCULO	Parámetro que caracteriza el proceso de aprendizaje de la red neuronal entrenados con el algoritmo de <i>Levenberg - Mardquardt</i> que pertenece a uno de los tipos de

	entrenamiento <i>Back Propagation</i> .
FUENTE	Datos generados por simulación de una población normal con parámetros especificados en los Anexos IV.1 y IV.2
PUNTOS CRÍTICOS	μ (valor inicial)=0.001 Incremento de $\mu=10$ Decrecimiento de $\mu=0.1$

Gráfico N° 3.8

Indicador B.2 Épocas

NOMBRE	Épocas
FUNDAMENTO	Mide en forma indirecta el tiempo de procesamiento de la red neuronal.
FORMA DE CÁLCULO	Una época es el paso de todas los datos de entrenamiento por el algoritmo de aprendizaje.
FUENTE	Datos de entrenamiento (x)
PUNTOS CRÍTICOS	Máximo número de épocas permitido durante el entrenamiento es igual a 100.

Gráfico N° 3.9

Indicador B.3 Error Cuadrado Medio (ECM)

NOMBRE	Error Cuadrado Medio (ECM)
FUNDAMENTO	Mide la precisión o el ajuste obtenido durante el entrenamiento de la red neuronal.
FORMA DE CÁLCULO	$ECM = (Y_i - \hat{Y}_i)^2 / n$ <p>Donde Y_i es la variable dependiente o valor objetivo en el caso de las redes neuronales, \hat{Y}_i es el valor dado por la red neuronal después del entrenamiento del mismo.</p>

FUENTE	Datos de entrenamiento y validación (x).
PUNTOS CRÍTICOS	ECM=0 Nivel óptimo de precisión o ajuste. ECM>0 Nivel de precisión o ajuste alcanzado.

3.3.2.3 Indicador de Generalización

Gráfico N° 3.10

Indicador C.1 Error Cuadrado Medio (ECM)

NOMBRE	Error Cuadrado Medio (ECM)
FUNDAMENTO	Mide la precisión obtenida al usar los parámetros de la red neuronal ya entrenada.
FORMA DE CÁLCULO	$ECM = (Y_i - \hat{Y}_i)^2 / n$ <p>Donde Y_i es la variable dependiente o valor objetivo en el caso de las redes neuronales, \hat{Y}_i es el valor dado por la red neuronal después del entrenamiento del mismo.</p>
FUENTE	Datos de prueba (x).
PUNTOS CRÍTICOS	ECM=0 Nivel óptimo de precisión. ECM>0 Nivel de precisión alcanzado.

3.2.3 Operaciones Intermedias

3.2.3.1 Técnica Estadística de Regresión Lineal Múltiple

El análisis de Regresión Lineal Múltiple cuantifica la relación entre la variable dependiente dado por un vector $Y(n \times 1)$ y las variables independientes X dado en una matriz $(n \times p)$, a partir de ello se desarrolla una ecuación lineal con fines predictivos. Con el fin de usar como control y para efectos comparativos la técnica de Regresión Lineal en cada uno de los casos tomará los datos de entrenamiento y validación, los mismos que usa las redes

neuronales, para formar el modelo estadístico. Luego de obtener el modelo se procederá a calcular los residuales y el indicador de precisión (Indicador B.3)

Una vez obtenido el modelo es importante realizar una validación y comprobar su desempeño con fines predictivos. En nuestro caso se tomará los datos destinados para la prueba o test en las redes neuronales, y así realizar la validación calculando el ECM (Indicador C.1) y hacer las comparaciones respectivas con las redes neuronales.

3.2.3.2 Método de la Red Neuronal *Back Propagation*

Las redes neuronales tras su procesamiento permitirán obtener los indicadores ya definidos y hacer las comparaciones determinadas por los casos presentados. En un primer instante se analizará la parte de representación de los datos, el tiempo que demora el procesamiento tanto en segundos como en épocas (Indicador B.2) y el error obtenido durante la parte de entrenamiento (Indicador B.3) con ello se podrá analizar el efecto de la presencia de los datos atípicos en cada caso trabajado.

Terminada la etapa de entrenamiento de la red, se tomarán los valores de prueba y se obtendrán el indicador definido para la etapa de generalización (Indicador C.1) teniendo la comparación final y tal vez la importante, ver en que caso la red neuronal es mejor para predecir nuevos valores y si es mejor que un modelo estadístico bajo el cumplimiento de todos los supuestos del análisis multivariado.

Capítulo IV

Diseño del experimento

Montgomery (2002) señala que todo diseño estadístico de experimentos se debe plantear de tal forma que se recaben datos adecuados y puedan analizarse dando conclusiones válidas y objetivas. Para ello se debe cumplir con los tres principios básicos del diseño experimental: la realización de réplicas, aleatorización y formación de bloques.

Para garantizar el principio de aleatorización se usa técnicas de simulación en cada uno de los pasos, así se genera datos con distribución normal multivariada. De acuerdo a los objetivos del estudio se desea analizar la precisión de las redes neuronales ante la presencia de datos atípicos en diferentes escenarios, esto permite tener un diseño de bloques determinado por la cantidad de muestra, tipo de dato atípico, y el porcentaje de los mismos. La realización de réplicas da la seguridad de tener resultados consistentes, por ello en cada escenario se emplea un total de 10 réplicas. Finalmente como comparación se hace uso del análisis estadístico de Regresión Lineal Múltiple.

Las etapas del diseño experimental son:

- i. Generación de los datos de entrenamiento, validación y prueba.
- ii. Establecimiento de los parámetros que se tendrá en cada escenario de análisis.
- iii. Revisión de las bases experimentales.

Para la generación de datos y el análisis de redes neuronales se escoge el software MATLAB debido a la libertad que da para la implementación y modificación de programas que junto a los módulos de redes neuronales que contiene facilita y permite obtener salidas óptimas con gran precisión. Además como herramienta para la validación y análisis estadístico se usa el software estadístico SPSS 15.0. Además como herramienta secundaria se usa macros trabajadas en Microsoft Excel, esto sobretodo en la formación de las bases experimentales.

4.1 Datos de entrenamiento, validación y prueba.

Como ya se explicó se usa como metodología comparativa al análisis de Regresión Lineal Múltiple, esto conlleva a cumplir con la mayoría de supuestos que exige la técnica estadística de tal forma de no dar ninguna ventaja a las redes neuronales y ver así netamente el efecto de los datos atípicos en la precisión de los datos. El principal supuesto a cumplir en la muestra base (muestra sin valores atípicos) es la normalidad de los datos, lo que lleva a trabajar con datos provenientes de poblaciones normal p – variada (incluye tanto a la variable dependiente como a las independientes). Una vez garantizado este supuesto, se procede a la deformación de la distribución de los datos usando cierto porcentaje de valores atípicos. En el experimento se usa una proporción de 0% y 3%, tal como lo plantea Chen – Chia , C. et al. (2002), en su estudio de aproximación de funciones con outliers, en donde propone trabajar con un porcentaje no mayor a 5%.

La cantidad de datos de entrenamiento, validación y prueba varía según el escenario, manteniendo sí la proporción entre ellos. Los porcentajes se han establecido en base a lo señalado en el manual de redes neuronales del software Matlab, 75% de entrenamiento, 15% validación y 15% para datos de prueba.

Finalmente es importante resaltar que los datos de validación y entrenamiento son usados para la formación del modelo estadístico, y los datos de prueba se emplean para analizar su generalización datos. Esto se fundamenta ya que la forma como trabaja las redes neuronales es similar, para establecer los pesos emplea datos de entrenamiento y con los de validación los va corrigiendo para evitar el sobre ajuste, usando el modelo final en los datos de prueba.

4.1.1 Espacio Parametral Generador

En primer lugar se genera las variables independientes, el algoritmo de generación de datos necesita la matriz de correlaciones; esta matriz debe ser definida positiva y los valores dentro de ella son generados en forma aleatoria en Microsoft Excel (Ver Anexo IV.1). Algo importante de resaltar es el uso de valores por debajo de 0.4 para las correlaciones ya que de acuerdo a lo presentado por Bisquerra, R. (1987), estos valores son catalogados como correlaciones bajas o muy bajas, evitando de antemano la multicolinealidad de los datos, tema tratado en los supuestos del análisis de Regresión Lineal.

Cuadro N ° 4.1
Grados de Correlación

Correlación	Clasificación
[0.8,1]	Muy Alta
[0.6,0.8>	Alta
[0.4,0.6>	Moderada
[0.2,0.4>	Baja
<0,0.2>	Muy Baja

Fuente: Bisquerra, R. (1987),

Teniendo presente todo lo mencionado en el párrafo anterior se creó dos matrices de correlación, la primera mostrada en el Anexo N° IV.2 es empleada para los datos de todos los escenarios (es como la población objetivo del estudio) y la segunda matriz del Anexo N° IV.3 empleada para la generación de datos atípicos contaminantes (datos que provienen de una población diferente a la objetivo).

4.1.2 Casos de estudio

Para la determinación de los escenarios se toma en cuenta:

a. Presencia o no de datos atípicos

El objetivo de esta investigación es ver el efecto de los datos atípicos en las redes *Back Propagation* y su uso para fines predictivos. Así se tiene:

- Datos limpios (sin valores atípicos).
- Datos con outliers del tipo discordante.
- Datos con outliers del tipo contaminante.

Los outliers del tipo discordante son incorporados en forma aleatoria dándoles una distancia respecto a la media de 6 a 25 desviaciones estándar, para comprobar que son realmente atípicos se hace uso de las distancias de Mahalanobis y sus respectivas pruebas de hipótesis, este análisis se trabaja en el SPSS y la sintaxis puede verse en el Anexo IV.5. Para los datos atípicos del tipo contaminante se ha incorporado en forma aleatoria registros provenientes de otra población diferente a la población objetivo.

b. Ubicación de los datos atípicos

Otro punto que se toma en cuenta es la ubicación de los datos atípicos, es decir si están en la parte de entrenamiento o en el de validación; no se considera la incorporación de atípicos en los datos de prueba por la razón que estamos más interesados en el efecto de los datos atípicos en la formación de la red neuronal. De este modo se plantea 4 casos:

El Caso I, no se tiene datos atípicos, ni en los datos de entrenamiento ni de validación, esto permite en condiciones normales, es decir bajo el cumplimiento de los supuestos estadísticos, analizar cuál de las técnicas tiene una mejor performance.

En el Caso II, se comienza con la incorporación de datos atípicos en los datos de entrenamiento, esto se hace con el fin de ver cómo afecta en la formación de los pesos de las redes neuronales.

En el Caso III, se incorpora datos atípicos en los datos de validación, con esto se busca ver si los datos de validación cumplen o no con su labor, es decir evitar o no el sobre ajuste de la red.

Como último caso tenemos el IV, en donde se encontrará datos atípicos tanto en la parte de entrenamiento como en la de validación, es un caso extremo en donde se espera tener problemas en todo el proceso.

c. Tamaño de Muestra

El tamaño muestral es un factor determinante en problemas de predicción, Anderson *et al.* (2001) indica que tienen un impacto directo en la potencia estadística de la regresión múltiple, a partir de ello muestras pequeñas (menor a 20 observaciones) sólo son apropiadas para análisis de regresión lineal simple y sólo pueden detectar relaciones muy fuertes con cierto grado de incertidumbre. Contrariamente muestras muy grandes, superiores a 1000

observaciones, hacen los test de significancia demasiados sensibles indicando que cualquier relación es estadísticamente sensible. Finalmente ellos recomiendan tener entre 15 a 20 observaciones por variable independiente como lo ideal y lo mínimo aceptable 5 observaciones.

Basado en este criterio y teniendo en cuenta las 8 variables independientes, se plantea plantea los siguientes tamaños muestrales:

- Tamaño muestral de 40: Tomando el valor mínimo de observaciones de 5.
- Tamaño muestral de 160: Tomando máximo ideal de 20 observaciones.
- Tamaño muestral de 1000: Tamaño muestral mínimo considerado como muy grande para el análisis de Regresión Lineal Múltiple.
- Tamaño muestral de 3000: Tamaño considerablemente muy grande y óptimos para redes neuronales.

Considerando todos los puntos tratados los escenarios a trabajar son:

Cuadro Nº 4.2

Escenarios por tipo de dato según casos

CASO I					
Entrenamiento		Validación		Prueba	
Tamaño de muestra	ne	ne(out)	nv	nv(out)	np
40	28	-	6	-	6
160	112	-	24	-	24
1000	700	-	150	-	150
3000	2100	-	450	-	450

CASO II					
Entrenamiento		Validación		Prueba	
Tamaño de muestra	ne	ne(out)	nv	nv(out)	np
40	27	1	6	-	6
160	109	3	24	-	24
1000	679	21	150	-	150
3000	2037	63	450	-	450

CASO III					
Entrenamiento		Validación		Prueba	
Tamaño de muestra	ne	ne(out)	nv	nv(out)	np
40	28	-	5	1	6
160	112	-	23	1	24
1000	700	-	145	5	150
3000	2100	-	436	14	450

CASO IV					
Entrenamiento		Validación		Prueba	
Tamaño de muestra	ne	ne(out)	nv	nv(out)	np
40	27	1	5	1	6
160	109	3	23	1	24
1000	679	21	145	5	150
3000	2037	63	436	14	450

ne: número de datos de entrenamiento sin valores atípicos

ne(out) : número de datos de entrenamiento con valores atípicos

nv: número de datos de validación sin valores atípicos

nv(out) : número de datos de validación con valores atípicos

np: número de datos de prueba.

Caso I: Muestra sin valores atípicos

Caso II: Muestra con valores atípicos en la parte de entrenamiento

Caso III: Muestra con valores atípicos en la parte de validación

Caso IV: Muestra con valores atípicos en la parte de entrenamiento y validación.

4.1.3 Generador de datos normales multivariados

Como punto inicial se procede a generar las variables independientes, usando como base generadora la matriz de correlaciones mostrada en el punto 4.1.1. Para la generación se usa el algoritmo del software MVN, este es el más apropiado debido a que permite generar datos normales multivariados y a la vez usa la matriz de correlaciones, siendo de gran ayuda para el control sobre la multicolinealidad de las variables independientes.

Este algoritmo comienza generando datos normales univariados aleatorios, para tal fin MVN usa el algoritmo de Joseph L. Leva (1992), quien mejora el algoritmo de ratio de desviación uniforme propuesto por A.J. Kinderman and J. F. Monohan en 1977.

El algoritmo de Joseph L. Leva es:

- i. Generar los números aleatorios uniformes $u = \xi_1$ y $v = 2\sqrt{2/e(\xi_2 - 1/2)}$.
- ii. Evaluar la forma cuadrática $Q = x^2 + y(ay - bx)$ con $x = u - s$ y $y = |v| - t$.
- iii. Si $Q < r_1$ entonces ir al paso vi.
- iv. Si $Q > r_2$ entonces ir al paso i.
- v. Si $v^2 > -4v^2 \ln u$ entonces ir al paso i
- vi. Retornar el ratio v/u como un número aleatorio proveniente de una distribución normal estándar.

Luego los datos normales multivariados son producidos al multiplicar las variables normales univariadas con la descomposición de Cholesky de la matriz de correlaciones de las variables a generar, conforme a la siguiente fórmula:

$$Y = LX \tag{4.1}$$

Donde:

Y: Vector de números aleatorios con distribución normal multivariada.

X: Vector de números aleatorios con distribución normal univariada.

L: Descomposición de Cholesky de la matriz de correlaciones.

Una vez que se tiene las variables independientes, se genera la variable dependiente como sigue:

- En forma aleatoria se da pesos a cada variable independiente, con estos pesos se forma un valor previo de la variable dependiente, los pesos dados a cada variable independiente son

Cuadro N° 4.3
Pesos dados a cada Variable Independiente

V. Independiente	Peso
X1	0.37
X2	0.74
X3	0.40
X4	0.56
X5	0.91
X6	0.20
X7	0.25
X8	0.60

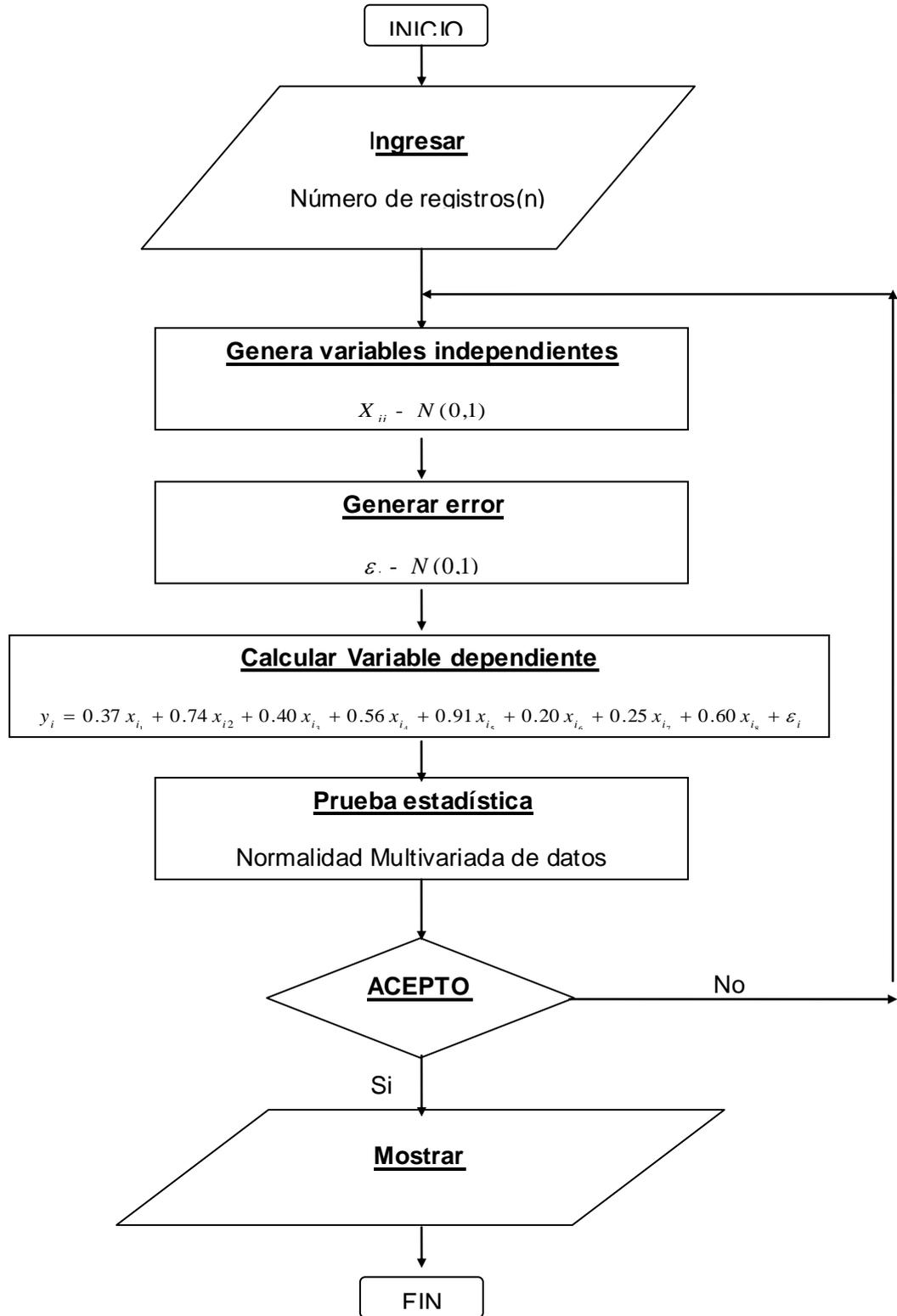
- Como se puede ver en la ecuación 4.3, la variable dependiente se formará tras multiplicar los pesos de cada variable independiente por su respectivo valor y al final de esto se le sumará un error, denotado en la ecuación con la letra ξ . Este error tiene una distribución normal estándar y también se genera dentro del programa de simulación.

Luego de tener toda la matriz generada se aplica el test de Mardia de tal forma que se garantice que los datos provienen de una población normal multivariada.

Todo este algoritmo ha sido implementado en Matlab 7.0 (Ver Anexo III.1) , la función que hemos creado se llama **generación** y necesita como parámetros de ingreso el número de registros y el número de variables que se desea generar, además se deberá ingresar la matriz de correlaciones, con todo estos datos hacemos uso del algoritmo ya descrito.

Figura N°4.1

Algoritmo generador de datos para el estudio



4.2 Arquitectura de la red

La red *Back Propagation* tiene como especificaciones las características mostradas en el gráfico 4.1.

Gráfico N° 4.1

Características de la Red Neuronal

Neuronas de entrada	08
Neuronas de salida	01
Capas Ocultas	01
Función de Transferencia	Lineal
Tipo de Entrenamiento	Incremental
Algoritmo de aprendizaje	Supervisado, <i>Back Propagation</i>

4.3 Bases Experimentales

Para el análisis de datos comparativo se usa indicadores comunes a las técnicas estadísticas y a las redes neuronales, tales como ECM (Error Cuadrado Medio), entre otros. Adicionalmente se hará uso de gráficos que nos ayudarán a visualizar mejor el desempeño de las redes, por ejemplo el gráfico evolución del error.

4.3.1 Aleatorización

La aleatorización es garantizada en cada uno de los pasos de la generación de datos del experimento, desde la estructura de la matriz de

correlación hasta los pesos y el error que se usan para la generación de la variable dependiente.

4.3.2 Bloqueo

Para el bloqueo de los datos se tiene en cuenta el tipo de dato atípico, la cantidad de muestra y la ubicación donde se encuentra, en datos de entrenamiento o validación.

4.3.3 Réplica

Para cada caso o escenario se ha decidido tomar un total de 10 réplicas, así los resultados serán más precisos.

4.3.4. Control

El análisis de Regresión Lineal Múltiple se usará como metodología de control, debido que esta técnica se usa frecuentemente para problemas de predicción.

Capítulo V

Resultados y análisis

5.1 Caracterización de los datos:

Como se estableció en el capítulo de diseño del experimento se cuenta con varios escenarios determinados por el tipo de dato atípico. Los valores atípicos se incorporan a una base que cumple con todos los supuestos de un análisis de regresión lineal múltiple.

Los datos deben provenir de una población con distribución normal, para comprobar ello se usó el test de Normalidad multivariada de Mardia (Indicador N° A.1, A.2 y A.3) , los resultados pueden revisarse el Anexo N° V.1, en todos ellos usando un nivel de significancia de 0.01 se aprueban la hipótesis nula tanto de simetría como de curtosis indicando así de una distribución normal multivariada para cada uno de los casos.

Otro supuesto importante a probar es ver si existe o no multicolinealidad, con este fin se ha calculado la Intensidad de multicolinealidad (Indicador A.4), como se observa en el cuadro N° 5.1 (caso sin valores atípicos que forma el resto de casos) en todas las réplicas y tamaños de muestra se tiene un valor alrededor de 3 lo que señala que no existe signos de correlación elevada entre las variables independientes. Para el resto de casos también se ha calculado el indicador (Ver Anexo N° V.2 y V.3) y se puede afirmar que para todos ellos

no se tendrá problemas de multicolinealidad, a pesar de la existencia de los diversos datos atípicos.

Cuadro N° 5.1

Intensidad de Multicolinealidad – Caso I (Sin valores atípicos)

CASO I				
Réplicas	Muestra 40	Muestra 160	Muestra 1000	Muestra 3000
R1	5.035	2.987	2.559	2.617
R2	3.999	2.613	2.587	2.612
R3	3.829	3.281	2.707	2.637
R4	3.557	2.711	2.665	2.636
R5	3.076	3.037	2.798	2.648
R6	4.445	3.112	2.740	2.701
R7	3.389	2.572	2.634	2.647
R8	4.670	2.854	2.528	2.681
R9	2.974	3.057	2.683	2.712
R10	3.948	3.026	2.468	2.683

Fuente: Anexo N° V.2 y V.3

El análisis de independencia de residuos también ha sido tomado en cuenta; el estadístico de Durbin Watson (Indicador A.5) indica en el cuadro N° 5.2 para la muestra inicial la independencia de los residuos, como se aprecia el estadístico toma valores alrededor de 2 estando así dentro de los límites aceptables. En los Anexos N° V.4 y V.5 se cuenta con el valor del estadístico Durbin Watson para los casos con valores atípicos discordantes y contaminantes respectivamente, sus números reflejan el cumplimiento del supuesto de independencia.

Cuadro N° 5.2

Estadístico Durbin Watson – Caso I (Sin valores atípicos)

CASO I				
Réplicas	Muestra 40	Muestra 160	Muestra 1000	Muestra 3000
R1	1.493	2.282	1.916	2.053
R2	1.668	1.960	2.066	1.931
R3	1.800	1.744	2.029	1.934
R4	1.223	2.288	1.960	2.082
R5	2.071	1.692	1.888	2.040
R6	1.508	1.867	2.037	2.095
R7	1.959	2.039	2.162	2.043
R8	1.883	2.032	1.993	2.023
R9	2.025	1.883	2.087	1.984
R10	1.703	1.451	1.994	1.986

Fuente: Anexo N° V.4 y V.5

El siguiente supuesto a probar fue la homogeneidad de varianzas de los errores (Homocedasticidad), el cuadro N° 5.3 se muestra los valores del estadístico de Levene (Indicador A.6) para las 10 réplicas del caso I en donde se ve que si se toma un nivel de significancia de 0.01 aceptaríamos la hipótesis nula de homogeneidad de varianzas dando por cumplido este supuesto. Los Anexos N° V.5 y V.7 muestran a detalle el análisis para los casos con valores atípicos, todos cumplen también la homogeneidad de varianzas.

Cuadro Nº 5.3

Prueba de Levene – Caso I (Sin valores atípicos)

CASO I								
Réplicas	Muestra 40		Muestra 160		Muestra 1000		Muestra 3000	
	Estadístico- Levene	Sig.	Estadístico- Levene	Sig.	Estadístico- Levene	Sig.	Estadístico- Levene	Sig.
R1	0.059	0.809	0.981	0.324	0.260	0.610	0.921	0.337
R2	0.114	0.738	2.851	0.094	2.203	0.138	0.324	0.569
R3	0.088	0.769	0.000	0.985	0.297	0.586	1.355	0.245
R4	0.043	0.837	0.028	0.867	0.328	0.567	0.750	0.386
R5	6.316	0.017	0.388	0.534	1.327	0.250	0.121	0.728
R6	0.319	0.576	0.290	0.591	0.006	0.936	0.027	0.869
R7	1.925	0.175	6.065	0.015	0.427	0.514	0.857	0.355
R8	3.991	0.054	3.968	0.048	0.039	0.843	1.798	0.180
R9	2.083	0.159	0.728	0.395	0.142	0.706	0.072	0.789
R10	0.017	0.898	0.019	0.891	2.274	0.132	0.036	0.849

Fuente: Anexo Nº V.6 y V.7

El estadístico de Kolmogorov Smirnov nos ayuda a determinar la normalidad de los residuos, usando un nivel de significancia de 0.01 y observando el cuadro Nº 5.4 se ve que el caso sin valores atípicos cumple con el supuesto, no obstante esto no se cumple en los casos de tamaño 1000 y 3000 cuando se incorporan los datos atípicos discordantes y contaminantes (Ver Anexo Nº V.8 y V.9).

Luego de analizar los diversos supuestos del análisis de regresión lineal múltiple se puede afirmar que los diferentes escenarios tomados en cuenta para este estudio cumplen con los principales supuestos: la no multicolinealidad, la independencia de residuos, la homocedasticidad y la linealidad (que se ha garantizado a la hora de generar los datos, cuidando que exista una combinación lineal entre la variable dependiente y las independientes); el único supuesto que no se cumple en algunos escenarios es la normalidad de los residuos y se da por la presencia de los valores atípicos ya sea del tipo discordante o contaminante.

Cuadro Nº 5.4

Prueba de Kolmogorov– Caso I (Sin valores atípicos)

CASO I								
Réplicas	Muestra 40		Muestra 160		Muestra 1000		Muestra 3000	
	Kolmogorov-Smirnov(a)	Sig.	Kolmogorov-Smirnov(a)	Sig.	Kolmogorov-Smirnov(a)	Sig.	Kolmogorov-Smirnov(a)	Sig.
R1	0.114	.200(*)	0.055	.200(*)	0.020	.200(*)	0.014	.200(*)
R2	0.070	.200(*)	0.066	.200(*)	0.027	.151(*)	0.016	.111(*)
R3	0.095	.200(*)	0.043	.200(*)	0.017	.200(*)	0.013	.200(*)
R4	0.131	.142(*)	0.064	.200(*)	0.023	.200(*)	0.012	.200(*)
R5	0.122	.200(*)	0.044	.200(*)	0.027	.175(*)	0.009	.200(*)
R6	0.109	.200(*)	0.056	.200(*)	0.015	.200(*)	0.013	.200(*)
R7	0.114	.200(*)	0.077	.051(*)	0.018	.200(*)	0.014	.200(*)
R8	0.110	.200(*)	0.071	.085(*)	0.020	.200(*)	0.011	.200(*)
R9	0.153	.043(*)	0.045	.200(*)	0.019	.200(*)	0.009	.200(*)
R10	0.100	.200(*)	0.043	.200(*)	0.030	0.077	0.007	.200(*)

Fuente: Anexo Nº V.8 y V.9

5.2 Aprendizaje

En esta parte se analizará la calidad del proceso de aprendizaje teniendo como indicadores al número de épocas (Indicador B2) y el error cuadrado medio (Indicador B3). Este análisis será útil para probar las hipótesis planteadas en el estudio, esperando así el efecto de los datos atípicos en el proceso de aprendizaje de la red neuronal entrenada con el algoritmo de Back – Propagation. A su vez se analizará los resultados obtenidos con el modelo estadístico de análisis de regresión lineal múltiple que sirve como control en el presente estudio, dado que se ha cumplido con la gran mayoría de supuestos se espera que el análisis de regresión múltiple tenga un mejor desempeño en esta parte del análisis. Los casos que se analizarán son los ya señalados en los capítulos anteriores, el caso I que servirá como guía para ver en condiciones normales cual es el óptimo desempeño de la red; luego se verá el caso II en el que se tiene atípicos en la muestra de entrenamiento, el caso III que presenta atípicos en la parte de validación y el IV que tiene atípicos en la parte de validación y prueba, esto a su vez para los datos atípicos discordantes y contaminantes.

5.2.1 Resultados del Análisis de Regresión Lineal Múltiple

Para realizar el análisis de regresión lineal múltiple se ha tomado los datos de entrenamiento y validación destinados al análisis de redes neuronales, el análisis consiste en ver los resultados tomando en cuenta los dos casos de datos atípicos en cada escenario establecido. Los primeros resultados son dados en el cuadro 5.5, haciendo notar el incremento del error cuadrado medio al aumentar la cantidad de datos atípicos discordantes.

Cuadro N° 5.5
Error Cuadrado medio del Análisis de Regresión –
Datos atípicos discordantes

Muestra	Caso I	Caso II	Caso III	Caso IV
40	0.652904	0.862137	0.743219	1.007371
160	0.911423	1.452628	1.062701	1.683008
1000	0.963760	3.386975	1.507299	3.643545
3000	0.975191	3.306835	1.775214	3.570709

Caso I: Muestra sin valores atípicos

Caso II : 3% de atípicos en los datos de entrenamiento

Caso III: 3% de atípicos en los datos de validación

Caso IV: 3% de atípicos en los datos de entrenamiento y validación.

Fuente: Anexo N° V.6

El cuadro siguiente (Cuadros N° 5.6) muestran los resultados tras incorporar datos atípicos del tipo contaminante. Como se aprecia el efecto de este tipo de datos es menor en comparación a los discordantes, esto se debe al traslape de los datos contaminantes con los datos de la muestra, debemos recordar tal como se explicó en el marco teórico los datos contaminantes al sólo diferenciarse por la población de origen se confunden con la muestra no afectando mucho a los resultados obtenidos en la regresión lineal y por ende sin formar una tendencia definida tal como sucede con los datos discordantes. Todos los resultados explicados en esta parte son usados como control en el análisis de redes neuronales.

Cuadro N° 5.6

Error Cuadrado medio del Análisis de Regresión –Datos atípicos contaminantes

Muestra	Caso I	Caso II	Caso III	Caso IV
40	0.652904	0.673820	0.639504	0.702534
160	0.911423	0.930919	0.909645	0.940619
1000	0.963760	0.986034	0.972927	0.990339
3000	0.975191	1.004830	0.981956	1.014138

Caso I: Muestra sin valores atípicos

Caso II : 3% de atípicos en los datos de entrenamiento

Caso III: 3% de atípicos en los datos de validación

Caso IV: 3% de atípicos en los datos de entrenamiento y validación.

Fuente: Anexo N° V.7

5.2.2 Calidad del Proceso de Aprendizaje en Redes Neuronales

Al igual que ocurre en el análisis de regresión lineal, la incorporación de datos atípicos tiene un efecto en el error cuadrado medio. Para el análisis de las redes neuronales se trabaja de dos formas la primera dividiendo la muestra en entrenamiento y prueba y la otra dividiéndola en la forma que ya se explicó: entrenamiento, prueba y validación. Los resultados del cuadro N° 5.7 muestra algo que ya se anticipó en la parte del marco teórico, específicamente el uso o no de muestras de validación, al hacer uso de ello el error cuadrado medio tiende a ser un poco superior disminuyendo esta diferencia conforme aumenta el tamaño de muestra, llegando a ser esta diferencia menor a 1% en muestras por encima de los 1000 casos a excepción del caso III, donde se trabaja con datos atípicos en la parte de validación y se llega a tener incrementos de 3%.

Se hace un análisis sin tomar datos de validación por cada tamaño de muestra, uniéndose así los datos destinados al entrenamiento y validación en uno solo quedando los casos como sigue: el caso I no tiene datos atípicos, el caso III es el que tiene la menor cantidad de atípicos, seguido del caso II y finalmente se tiene el caso IV con la mayor cantidad de atípicos, por consecuente el incremento del error cuadrado medio se debe dar en este mismo orden y esto es lo que ocurre en cada tamaño de muestra, no obstante el incremento va disminuyendo conforme se aumenta el tamaño de muestra, por lo que una conclusión sería que el efecto de datos atípicos en la parte de

aprendizaje se aminora conforme aumenta el tamaño de muestra. Ahora si se analiza el entrenamiento de la red con datos de validación, un mayor efecto tiene la presencia de datos atípicos discordantes en la parte de entrenamiento que si se ubican en la parte de validación con excepción de la muestra pequeña de 40, esto se debe que al tener una muestra pequeña la red trata de asegurarse con la poca información disponible en llegar a un punto de convergencia que logre un error cuadrado medio de generalización dentro de un margen razonable; finalmente como se mencionó al definir los casos el peor escenario es el IV que muestra el peor desempeño de la red en la parte de aprendizaje debido a la existencia de datos atípicos discordantes tanto en la parte de validación como de entrenamiento.

Cuadro Nº 5.7

Error cuadrado Medio del Análisis de Redes Neuronales en la etapa de aprendizaje – Datos atípicos discordantes

Muestra	Muestra Entrenamiento, Prueba				Muestra Entrenamiento, Validación, Prueba			
	Caso I	Caso II	Caso III	Caso IV	Caso I	Caso II	Caso III	Caso IV
40	0.652904	0.862137	0.743219	1.007371	0.699562	0.942967	1.245566	1.634966
160	0.911423	1.452628	1.062701	1.683008	0.919350	1.464477	1.364951	1.975833
1000	0.963760	3.386975	1.507299	3.643545	0.965102	3.395952	1.676829	3.676510
3000	0.975192	3.306835	1.775214	3.570709	0.975712	3.313272	1.832976	3.577830

Caso I: Muestra sin valores atípicos

Caso II : 3% de atípicos en los datos de entrenamiento

Caso III: 3% de atípicos en los datos de validación

Caso IV: 3% de atípicos en los datos de entrenamiento y validación.

Fuente: Anexos Nº V.14 Y V.15

Los resultados del caso de datos atípicos contaminantes figuran en el cuadro Nº 5.8, el análisis muestra que los efectos ocasionados son similares a los producidos por los atípicos discordantes pero el efecto se da en menor grado; por ejemplo el uso de validación incrementa el error cuadrado medio del periodo de aprendizaje de la red sin embargo por ejemplo en el caso IV al usar tipos discordante y usar datos de validación representaba un incremento del 62% mientras para este mismo caso , pero teniendo datos atípicos contaminantes , este incremento se reduce a 15% aproximadamente. Respecto a las variaciones respecto a la incorporación de datos contaminantes por caso

y dentro de cada tamaño de muestra, también se reducen manteniendo en promedio variaciones por debajo del 8%; el cuadro N° 5.7 muestra que para el caso de datos discordante estas variaciones son mucho mayor llegando en algunos casos duplicar el error cuadrado medio.

Cuadro N° 5.8

Error cuadrado Medio del Análisis de Redes Neuronales en la etapa de aprendizaje– Datos atípicos contaminantes

Muestra	Muestra Entrenamiento,Prueba				Muestra Entrenamiento,Validacion,Prueba			
	Caso I	Caso II	Caso III	Caso IV	Caso I	Caso II	Caso III	Caso IV
40	0.652904	0.673820	0.639504	0.702535	0.699562	0.735933	0.700718	0.808014
160	0.911423	0.930919	0.909645	0.940619	0.919350	0.939072	0.918970	0.948389
1000	0.963760	0.986034	0.972927	0.990339	0.965102	0.987329	0.974685	0.991674
3000	0.975192	1.004830	0.981957	1.014138	0.975712	1.005413	0.982466	1.014605

Caso I: Muestra sin valores atípicos

Caso II : 3% de atípicos en los datos de entrenamiento

Caso III: 3% de atípicos en los datos de validación

Caso IV: 3% de atípicos en los datos de entrenamiento y validación.

Fuente: Anexo N° V.18 Y V.22

El siguiente indicador a analizar es el número de épocas (Indicador B2) y el tiempo de procesamiento, los valores obtenidos figuran en los cuadros 5.9 y 5.10, si bien es cierto se ve un incremento ante la presencia de datos atípicos de cualquiera de los dos tipos estos no son muy considerables, teniendo épocas y tiempos de procesamiento similares en ambos casos. Tal vez lo más resaltante es el incremento en el tiempo de procesamiento al aumentar el tamaño de muestra, como se ve en el caso de pasar de una muestra de 1000 a 3000 se incrementa el tiempo en alrededor de 80%, esto es importante resaltar debido a que el consumo de recursos computacionales es importante en proyectos a gran escala.

Cuadro Nº 5.9

Número de épocas realizadas en la etapa de aprendizaje

Datos atípicos discordantes								
Muestra	Muestra Entrenamiento,Prueba				Muestra Entrenamiento,Validacion,Prueba			
	Caso I	Caso II	Caso III	Caso IV	Caso I	Caso II	Caso III	Caso IV
40	7.80	6.70	6.50	7.80	7.00	7.60	6.80	7.40
160	8.40	8.40	7.70	8.50	6.40	6.00	7.40	6.60
1000	9.50	8.40	9.40	9.80	6.60	7.10	7.60	7.80
3000	10.30	9.50	9.00	9.50	7.50	6.80	7.40	7.10

Datos atípicos contaminantes								
Muestra	Muestra Entrenamiento,Prueba				Muestra Entrenamiento,Validacion,Prueba			
	Caso I	Caso II	Caso III	Caso IV	Caso I	Caso II	Caso III	Caso IV
40	7.80	6.20	7.50	7.30	7.00	6.70	7.20	6.20
160	8.40	8.00	8.90	7.30	6.40	7.10	6.60	6.60
1000	9.50	8.00	9.70	10.10	6.60	6.90	7.70	6.20
3000	10.30	10.50	10.00	10.00	7.50	6.80	7.60	7.60

Caso I: Muestra sin valores atípicos

Caso II : 3% de atípicos en los datos de entrenamiento

Caso III: 3% de atípicos en los datos de validación

Caso IV: 3% de atípicos en los datos de entrenamiento y validación.

Fuente: Anexos Nº V.11, V.15, V19 y V.23

Cuadro Nº 5.10

Tiempos de procesamiento en la etapa de aprendizaje (Segundos)

Datos atípicos discordantes								
Muestra	Muestra Entrenamiento,Prueba				Muestra Entrenamiento,Validacion,Prueba			
	Caso I	Caso II	Caso III	Caso IV	Caso I	Caso II	Caso III	Caso IV
40	5.93	6.34	5.91	6.00	7.04	6.78	6.40	6.53
160	6.90	6.60	6.61	6.86	7.87	7.00	7.33	7.50
1000	9.67	10.67	10.20	10.83	12.24	11.74	10.46	11.18
3000	17.05	19.17	18.72	18.93	16.88	16.77	15.24	17.16

Datos atípicos contaminantes								
Muestra	Muestra Entrenamiento,Prueba				Muestra Entrenamiento,Validacion,Prueba			
	Caso I	Caso II	Caso III	Caso IV	Caso I	Caso II	Caso III	Caso IV
40	5.93	5.67	6.19	6.01	7.04	6.78	6.40	6.53
160	6.90	6.34	6.85	6.49	7.87	7.38	7.20	7.15
1000	9.67	10.17	10.79	11.02	12.24	11.41	10.68	10.50
3000	17.05	18.55	20.56	17.95	16.88	16.46	16.76	15.99

Caso I: Muestra sin valores atípicos

Caso II : 3% de atípicos en los datos de entrenamiento

Caso III: 3% de atípicos en los datos de validación

Caso IV: 3% de atípicos en los datos de entrenamiento y validación.

Fuente: Anexos Nº V.12, V.16, V20 y V.24

Para terminar con este punto es necesario comparar los resultados de las redes neuronales con los que se obtiene en las regresiones lineales, lo primero que resalta es la similitud entre los valores obtenidos en el análisis de regresión lineal múltiple y las redes neuronales que trabajan sin datos de validación, esta similitud se debe a que en este caso ambos procedimientos se

encargan de minimizar el error entre los datos observados y los pronosticados, existiendo sólo una mínima diferencia (aproximadamente de 0.000005% en el caso de datos discordantes y de 0.00001% en el caso de contaminante) como respuesta a su forma de trabajo, esto es una constante en cada uno de los casos.

Al trabajar con datos de validación, se comienzan a notar las diferencias, situando en todos los casos al análisis de regresión lineal múltiple como una mejor técnica de ajuste, siendo los datos contaminantes los que se acercan más a los resultados dados por el análisis estadístico. Esta superioridad del análisis estadístico se da por el cumplimiento de supuestos y también debido que al trabajar con datos de validación, la red neuronal tiende a sacrificar un poco el ajuste en la parte de entrenamiento para así poder tener una mejor generalización de los datos. Con esto podemos concluir que en cualquiera de los casos, si deseamos sólo lograr un buen ajuste de datos y dejar de lado la generalización, la técnica estadística es la mejor opción sobre todo cuando los supuestos son cumplidos.

5.3. Generalización de los datos

Lograr una buena generalización de datos es importante debido a que nos garantiza que el modelo no sólo servirá para los datos de muestra sino que también tendrá un buen desempeño al presentarse nuevos conjuntos de datos, en esta parte se analizará el error cuadrado medio de generalización (Indicador C.1). Los primeros resultados a analizar son los obtenidos por los datos atípicos discordante, presentados en el cuadro 5.11, para ambas formas de trabajo cuando la data no presenta datos atípicos (Caso I) el error cuadrado medio disminuye conforme se aumenta el tamaño de muestra, lo que nos indica que en condiciones normales la red neuronal tendrá un mejor desempeño en tamaños de muestras grandes. El uso de datos de validación no ha logrado tener un menor error cuadrado medio salvo en las muestras pequeñas (Muestra de 40), esto señala que cuando se tiene datos que cumplen

con los principales supuestos del análisis de regresión lineal una mejor alternativa es el no uso del datos de validación para así tener una mayor cantidad de datos y corregir el efecto de datos atípicos.

Cuadro N° 5.11

Error cuadrado Medio del Análisis de Redes Neuronales en la etapa de generalización– Datos atípicos discordantes

Muestra	Muestra Entrenamiento,Prueba				Muestra Entrenamiento,Validacion,Prueba			
	Caso I	Caso II	Caso III	Caso IV	Caso I	Caso II	Caso III	Caso IV
40	1.026084	1.604527	1.362435	1.802459	1.017850	1.595951	1.587754	1.788490
160	1.074303	1.508204	1.259437	1.838651	1.080549	1.538829	1.306123	1.928936
1000	0.981813	2.401464	1.201091	2.665020	0.984035	2.459814	1.138112	2.651789
3000	0.957248	2.097518	1.133842	2.337825	0.958701	2.129925	1.150673	2.337251

Caso I: Muestra sin valores atípicos

Caso II : 3% de atípicos en los datos de entrenamiento

Caso III: 3% de atípicos en los datos de validación

Caso IV: 3% de atípicos en los datos de entrenamiento y validación.

Fuente: Anexo N° V.13 y N° V.17

Los datos atípicos contaminantes tienen un comportamiento diferente en la parte de generalización en ambas formas de trabajo (con o sin datos de validación), como se hace notar en el cuadro 5.12 el indicador de generalización disminuye conforme aumenta el tamaño de muestra en todos los casos esto a partir de la muestra de 160 hacia adelante. También algo importante de señalar es que el efecto de los datos atípicos discordantes en la parte de generalización no es determinante ya que en todos los casos los valores el error cuadrado medio obtenido con muy cercanos con una diferencia de alrededor del 1%. Si comparamos el efecto en la generalización ocasionados por los dos tipos de casos atípicos, se concluye que los datos atípicos discordante tienen un mayor efecto en la generalización debido a que forman un mayor error cuadrado medio.

Cuadro N° 5.12

Error cuadrado Medio del Análisis de Redes Neuronales en la etapa de generalización– Datos atípicos contaminantes

Muestra	Muestra Entrenamiento,Prueba				Muestra Entrenamiento,Validacion,Prueba			
	Caso I	Caso II	Caso III	Caso IV	Caso I	Caso II	Caso III	Caso IV
40	1.026084	1.061248	0.953771	1.118420	1.017850	1.041328	0.934650	0.992062
160	1.074303	1.060749	1.071133	1.058774	1.080549	1.056370	1.088245	1.075800
1000	0.981813	0.985777	0.982932	0.986694	0.984035	0.988233	0.983788	0.989375
3000	0.957248	0.957144	0.957420	0.957500	0.958701	0.957715	0.958589	0.958149

Caso I: Muestra sin valores atípicos

Caso II : 3% de atípicos en los datos de entrenamiento

Caso III: 3% de atípicos en los datos de validación

Caso IV: 3% de atípicos en los datos de entrenamiento y validación.

Fuente: Anexo N° V.21 y V.25

Lo último por analizar es ver si la técnica estadística de regresión lineal tiene un mejor desempeño que las redes neuronales cuando se cumplen la mayoría de sus supuestos y se cuentan con datos atípicos, los resultados se resumen en el cuadro 5.13. Si se compara los resultados se va a tener que no se puede determinar con exactitud cuál de las dos metodologías es mejor en la parte de generalización por el motivo que en algunos casos las redes neuronales dan un menor error cuadrado medio que la regresión lineal múltiple y viceversa. Al trabajar sin datos de validación los resultados son muy parecidos para todos los casos y para ambos tipos de datos atípicos siendo la diferencia por alrededor de 0.005%, mientras que al trabajar con datos de validación esta diferencia está en torno a 2.7% para los datos atípicos discordantes y 1.33% para los datos atípicos contaminantes.

Cuadro N° 5.13

Error cuadrado Medio del Análisis de Regresión Lineal en la etapa de generalización

Muestra	Datos atípicos discordantes				Datos atípicos contaminantes			
	Caso I	Caso II	Caso III	Caso IV	Caso I	Caso II	Caso III	Caso IV
40	1.025947	1.604447	1.362305	1.802485	1.025947	1.061230	0.953962	1.118094
160	1.074364	1.508226	1.259419	1.838652	1.074364	1.060703	1.071147	1.058753
1000	0.981814	2.401430	1.201132	2.664976	0.981814	0.985778	0.982939	0.986690
3000	0.957243	2.097426	1.133854	2.337748	0.957243	0.957141	0.957425	0.957494

Caso I: Muestra sin valores atípicos

Caso II : 3% de atípicos en los datos de entrenamiento

Caso III: 3% de atípicos en los datos de validación

Caso IV: 3% de atípicos en los datos de entrenamiento y validación.

Conclusiones

1. Respecto a la Hipótesis General.- Se ha demostrado que la presencia de datos atípicos del tipo discordante o contaminante influyen en el desempeño de la red neuronal entrenada con el algoritmo de *Back Propagation*.
 - 1.1. Es conocido que las redes neuronales son consideradas como modelos de distribución libre, sin embargo en este estudio se ha comprobado que bajo el cumplimiento de los supuestos estadísticos del modelo de regresión lineal y con la incorporación de los dos tipos de datos atípicos estudiados, las redes neuronales presentan un desempeño por debajo que el obtenido del modelo estadístico, tanto en la parte de calidad del proceso de aprendizaje como en la parte de generalización de los datos.
 - 1.2. Tras analizar los resultados obtenidos en los escenarios planteados, se ha visto conveniente trabajar las redes neuronales sin datos de validación, esto permite incrementar el número de casos en la etapa de entrenamiento llegando a acercarse a lo obtenido por los modelos estadísticos. Esta diferencia puede llegar a sólo 0.000005% en el caso de datos discordantes y de 0.00001% en el caso de contaminante en el caso de la etapa de aprendizaje y 0.005% en la etapa de generalización.

2. Respecto a la hipótesis Específica I.- El uso del algoritmo *Back Propagation*, no garantiza un óptimo desempeño de las redes neuronales ni en la etapa de aprendizaje ni en la etapa de generalización. Su mejor desempeño dentro de los escenarios planteados se da cuando no se dispone de datos atípicos y se cuenta con tamaños de muestras grandes.

3. Respecto a la hipótesis Específica II.- Tras comparar los resultados obtenidos por los datos discordante y contaminantes, se concluye que estos últimos tienen menor efecto en todas las partes del proceso de redes neuronales, manteniendo resultados muy cercanos en los diferentes casos de estudio y no teniendo grandes variaciones tal como si ocurre con los del tipo discordante.

4. Como pautas a seguir en estos casos para los investigadores de redes neuronales se pueden establecer los siguientes puntos:

Paso I: Ingreso de los datos, tamaño de muestra y número de variables a trabajar.

Paso II: Comprobación de supuestos:

2.1. Análisis de la normalidad Multivariada: Test de Mardia (Indicador A.1, A2, A3).

2.2. Análisis de la Multicolinealidad: Intensidad de multicolinealidad (Indicador A.4)

2.3. Análisis de Independencia de residuos: Estadístico Durbin Watson (Indicador A.5).

2.4 Análisis de Homocedasticidad: Estadístico de Levene (Indicador A.6).

2.5 Análisis de Linealidad: Gráficos parciales de residuales.

Paso III: Detección de valores atípicos por medio de la distancia de Mahalanobis.

Paso IV: Separación de datos atípicos de la muestra.

Paso V: División de la muestra en forma aleatoria, en dos partes una de entrenamiento (85%) y otra de prueba (15%).

Paso VI: Entrenamiento de la red con un valor adaptivo específico (Indicador B.1.), medida de las épocas y el tiempo de la etapa de aprendizaje (Indicador B.2) y cálculo del error cuadrado medio (Indicador B.3.)

Paso VII: Aplicación del modelo a los datos de prueba, y cálculo del error cuadrado medio del error de generalización (Indicador C.1).

Paso VIII: Si el modelo tiene un error cuadrado medio aceptable tanto en la etapa de aprendizaje como de generalización ir al siguiente paso , caso contrario volver al paso VI.

Paso IX: Implementación del modelo neuronal.

Recomendaciones

1. Tal como ya se explicó los datos atípicos del tipo contaminantes son muy difíciles de ubicar y por tanto de tratar, ante ello se recomienda hacer más estudios sobre el efecto que podrían tener en los diferentes tipos de análisis de redes neuronales, permitiendo así no sólo tener planes operacionales en casos de predicción sino también en otros tipos de tareas realizadas por las redes neuronales, tales como los algoritmos de clasificación.
2. Se ha demostrado que tener un conocimiento estadístico de los datos con que se trabaja es muy importante, porque permite saber la manera de proceder correctamente en un análisis de redes neuronales, tal como se vio en esta investigación que bajo el cumplimiento de los supuestos del análisis de regresión lineal es mejor trabajar sin datos de validación y así lograremos una mejor calidad de aprendizaje y generalización de los datos. Esto es una prueba más que el análisis estadístico y el de redes neuronales pueden trabajar muy bien de la mano, llegando a convertirse en modelos híbridos con un buen desempeño; por eso se recomienda realizar investigaciones que creen criterios sobre cómo y cuándo combinar ambas metodologías en distintos campos de la ciencia.

REFERENCIAS BIBLIOGRÁFICAS

Basogarin, X. (2009). Redes Neuronales Artificiales y sus Aplicaciones. Universidad del País Vasco.

Bisquerra, R. (1987). Introducción a la Estadística Aplicada a la Investigación Educativa. Barcelona : PPV.

Chatterjee, S. (2006). Regression Analysis by Example. Cuarta Edición. Wiley Interscience A John Wiley & Sons, INC. Publication.

Cheng, B. y Titterington, m. (1994). Neural Network: A review from a Statistical Perspective. Statistical Science, Volumen 8, Nº1, 2-54.

Chen, C. Shun, F. Jin, T., Ching Ching, H. (2002). Robust Support Regression Network for Function Aproximation with Outliers. IEE Transactions on Neural Network.

Demuth, H. Beale, M. y Hagan, M (2005). Neural Network Toolbox User Guide Version 4. The Math Works, Inc.

Escalante, D. Mayorga, W. Vergara, C. (2010). Manual de Aplicación del Modelo de Regresión Lineal Múltiple. www.eumed.net/libros/2010c/720/

Fernandez, M. y Hernández, E. (Abril 2001). Weight Initialization Methods for Multilayer Feedforward. European Symposium on Artificial Neural Network Bruges, pp 119-124.

Freeman, J. y Skapura, D. (1991). Neural Networks Algorithms, Applications, and Programming Techniques. Addison Wesley Publishing Company.

Filzmoser, p. (2004). A Multivariate Outlier Detection Method. Belarusian State University, Minsk.

Galushkin, A. (2007). Neural Networks Theory. Springer Berlin Heidelberg New York.

Giudici, P. (2003). Applied Data Mining: Statistical Method for Business and Industry. John Wiley & Ltd.

Hair. J. Anderson, R. Tatham, R y Black, W.(2001). Análisis Multivariante. Quinta Edición. Prentice Hall.

Hagan, M. Demuth, H. y Beale, M.(1996). Neural Network Design. China Machine Press y Citic Publishing House.

Iglewicz, B. y Hoaglin, D (1985). How to Detect and Handle Outliers. Volumen16. The ASQ Basic References in Quality Control: Statistical Techniques. USA.

Jain, L. y Fanelli, A. (2000). Recent Advances in Artificial Neural Design and Applications. CRC Press LLC.

Johnson, R. y Wichern D. (2007). Applied Multivariate Statistical Analysis. Sexta Edición. Prentice Hall.

Kasabov, N (1998). Foundations of Neural Network, Fuzzy Systems, and Knowledge Engineering. Massachuset Institute of Technology.

Kikut, C. *et. al.* (2003). Principales Indicadores para el Diagnóstico del Análisis de Regresión Lineal. Banco Central de Costa Rica.

Kröse, B. y Van Der Smagt, P. (1996). An Introduction to Neural Network. Octava Edición. Universidad de Amsterdam.

Kumar, U. y Paliwal, M. (2009). Neural Network and Statistical Techniques : A Review of Applications. Expert Systems with Applications, Volumen 36.pp.2-17.

Leondes, C (1998) .Algorithms and Architectures: Neural Network Systems Techniques and Applications. Volumen I. Academic Press.

Leva, J. (1992). Algorithm 712: A Normal Random Number Generator. ACM Transactions on Mathematical Software. Dec 01 1992 v 18 n 4. p 454.

Montaño, J. (2002). Redes Neuronales Artificiales Aplicadas al Análisis de Datos. Tesis Doctoral. Universidad Les Illes Lalears.

Montgomery, D. Peck E. y Vining G. (2004). Introducción al Análisis de Regresión Lineal. Mexico: Compañía Editorial Continental

Montgomery, D. (2002), Diseño y Análisis de Experimentos.Universidad Estatal de Arizona.

Pitarque, A. Roy, F. y Ruiz, C. (1998). Redes Neuronales vs MODELOS Estadísticos: Simulaciones sobre Tareas de Predicción y Clasificación. Universidad de Valencia.

Rabunal, J. y Dorado, J. (2006). Artificial Neural Networks in Real Life Applications. Idea Group Publishing.

Rawlings, J. Pantula, S. y Dickey, D (1998). Applied Regression Analysis :A Research Tool. Segunda edición. Springer.

Raykov, T. y Marcoulides, G. (2008). An Introduction to Applied Multivariate. New York: Routledge.

Rencher, A. (2002). Methods of Multivariate Analysis. Segunda Edición. Wiley Interscience A John Wiley & Sons, INC. Publication.

Rojas, R.(1996). Neural Network: A Systematic Introduction. Springer

Rurkhamet, B. Chutima, P. Reodecha, M. (1998). Comparative Study of Artificial Neural Network and Regression Analysis for Forecasting New Issued Banknotes. Thammasat, Int. Volumen 3.

Sarle, W. (Abril 1994). Neural Network and Statistical Models. Proceedings of the Nineteenth Annual SAS Users Group International Conference.

Serrano, A. (2004). Aplicación de las Redes Neuronales Artificiales a la Predicción del Resultado a Corto, Medio y Largo Plazo del Trasplante Renal Pediátrico. Universidad de Valencia.

Simar, L. (2007). Applied Multivariate Statistical Analysis. Segunda Edición. Springer.

Timm, N. (2002). Applied Multivariate Analysis. Springer

Uebersax JS. MVN program for random multivariate normal numbers. (2006). Available at the *Statistical Methods for Rater Agreement* website: <http://www.john-uebersax.com/stat/agree.htm>.

Valdez, A. (Mayo 2009). Redes Neuronales. Revista Informática Tecnología y Sociedad. Bogota.

Veelenturf, L. (1995). Analysis Application of Artificial Neural Networks. Prentice Hall International.

Walck, C. (1998). Hand – book on Statistical Distributions for Experimentalists. University of Stockholm.

Yan, X. y Gang, X. (2009). Linear Regression Analysis: Theory and Computing. World Scientific.

Zhang, P (2004). Neural Network in Business Forecasting. Idea Group Publishing.

ANEXOS

Anexos I
Términos y Equivalencias

Anexo 1.1

Equivalencias entre modelos estadísticos y modelos de redes neuronales

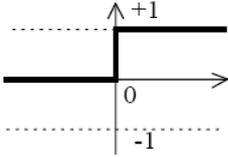
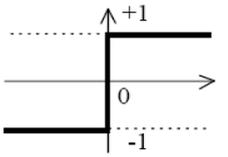
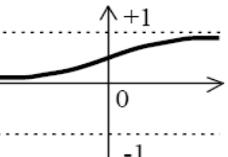
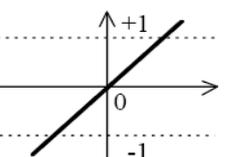
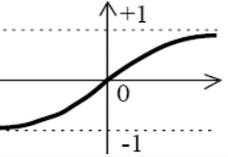
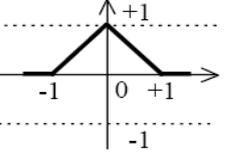
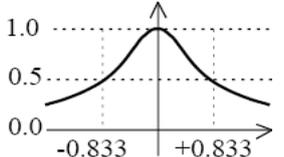
Modelo Estadístico	Modelo de Red Neuronal
Regresión lineal múltiple	Perceptrón simple con función lineal
Regresión logística	Perceptrón simple con función logística
Función discriminante lineal	Perceptrón simple con función umbral
Regresión no lineal múltiple	Perceptrón multicapa con función lineal en la salida
Función discriminante no lineal	Perceptrón multicapa con función logística en la salida
Análisis de componente principales	Regla de la Oja
Análisis de clusters	Mapas autoorganizados de Kohonen
Regresión Kernel	Función de base radial (RBF)

Fuente:Montaño (2002)

Anexos II
Funciones de Transferencias

Anexo II.1

Principales funciones de transferencia

Función de transferencia	Gráfica Matemática	Representación en topología de redes
Hard Limit		
Hard Limit Simétrica		
Log Sigmoidal		
Lineal		
Tangencial Sigmoidal		
Función Triangular		
Función Radial		

Fuente: Demuth *et al.* (2005)

Anexos III

Programas en Matlab

Anexo III.1.

Generación de los datos

```
function generacion(n,m)
%*****
%INDICACIONES PREVIAS
%*****
%Este programa genera datos normales en base a la matriz de correlaciones
%En nuestro caso generaremos primero las variables dependientes xi,i=1:8
%El programa funciona de la siguiente manera:
%Paso 1; Se indica la cantidad de registros(n) y el número de variables(m)
%Paso 2: Se ingresa la matriz de correlaciones
%Paso 3: Se genera el bloque de los datos degenerados por el algoritmo de
%dato por Joseph Leva
%Paso 4: Se genera de acuerdo a la regla la variable Y
%Paso 5: Se genera una variable aleatoria normal con media cero y
%desviación estándar 1, estos valores serán los errores y serán adjuntados
%a la variable Y.
%Paso 6: Se aplica el test de Mardia y se muestra los valores del test
%Paso 7: Esto se hace para cada una de las 10 Réplicas del estudio
%*****
%*****
%PASO 2
%*****
%Se Ingresa la matriz de correlación (debe ser de orden m*m)
matriz_corr=[
1      0.387  0.269  0.379  0.211  0.297  0.241  0.243;
0.387  1      0.386  0.254  0.257  0.31  0.372  0.384;
0.269  0.386  1      0.205  0.215  0.369  0.293  0.272;
0.379  0.254  0.205  1      0.392  0.237  0.374  0.367;
0.211  0.257  0.215  0.392  1      0.282  0.234  0.346;
0.297  0.31  0.369  0.237  0.282  1      0.232  0.202;
0.241  0.372  0.293  0.374  0.234  0.232  1      0.297;
0.243  0.384  0.272  0.367  0.346  0.202  0.297  1
];
for replica=1:10
    control=0;
    while control==0
```

```

%*****
%PASO 3
%*****
matriz=ones(n,m);
%Se comienza a generar variables normales univariadas
%usando el algoritmo de Joseph Leva
%Los parámetros usados se obtuvieron de HandBook on Statistical -141
%e: número de euler
e=2.71828;
a=0.196;
b=0.25472;
r1=0.27597;
r2=0.27876;
s=0.449871;
t=0.386595;
for i=1:n
    for j=1:m
        object=0;
        while object==0
            e1=rand();
            e2=rand();
            u=e1;
            v=2*sqrt(2/e)*(e2-0.5);
            x=u-s;
            y=abs(v)-t;
            Q=x*x+y*(a*y-b*x);
            if Q<r1
                object=1;
                matriz(i,j)=v/u;
            end
            if Q>r2
                object=0;
            end
            if v*v>-4*u*u*log(u)
                object=0;
            end
        end
    end
end
matriz;
l=chol(matriz_corr);
gx=matriz*l;

```

```

%*****
%PASO 4
%*****

%Para generar se tendrá los siguientes pesos generado en forma aleatoria
%Y  x1      x2      x3  x4      x5      x6      x7  x8
%  0.37  0.74  0.40  0.56  0.91  0.20  0.25  0.60

matriz_generada=ones(n,m+1);

for i=1:n
    for j=2:m+1
        matriz_generada(i,j)=gx(i,j-1);
    end
end

for i=1:n
matriz_generada(i,1)=0.37*gx(i,1)+0.74*gx(i,2)+0.40*gx(i,3)+0.56*gx(i,4)+0.91*gx(i,5)+0.2*gx(i,6
)+0.25*gx(i,7)+0.60*gx(i,8);
end

```

```

%*****
%PASO 5
%*****
%Se genera una variable univariada con N(0,1)esto representará a los
%errores
errores=ones(n,1);
for i=1:n
    object=0;
    while object==0
        e1=rand();
        e2=rand();
        u=e1;
        v=2*sqrt(2/e)*(e2-0.5);
        x=u-s;
        y=abs(v)-t;
        Q=x*x+y*(a*y-b*x);
        if Q<r1
            object=1;
            errores(i,1)=v/u;
        end
        if Q>r2
            object=0;
        end
        if v*v>-4*u*u*log(u)
            object=0;
        end
    end

end
end
%Se aplica los errores a la variable Y
for i=1:n
    matriz_generada(i,1)=matriz_generada(i,1)+errores(i,1);
end

```

```

%*****
%PASO 6
%*****

%Aplicamos el test de Mardia para ver si los datos son normales o no
%Los datos se generarán hasta que se tenga datos con distribución normal
%H = [stats.Hs stats.Ps stats.Ms stats.CVs stats.Hk stats.Pk stats.Mk
%stats.CVk]
Prueba_Mardia=mardiatest(matriz_generada,0.01);
N_Hs=Prueba_Mardia(1);
N_Hk=Prueba_Mardia(5);

if N_Hs==0 & N_Hk==0
    disp('Las matriz generada')
    replica
    matriz_generada
    Prueba_Mardia
    control=1;
end
end
end
end

```

Anexo III.2.

Implementación de la red neuronal sin datos de validación

```
function redesf2()
%*****
%*****
%Se importa todo el Excel de la carpeta work de MATLAB
data=xlsread('CASO4B','M3000');
%Indicar la cantidad de datos de entrenamiento(ne),validacion(nv)y
%prueba(np)
ne=2100;
nv=450;
np=450;
n=ne+nv+np;
s_entre=0;
s_entre_val=0;
s_prueba=0;
%Vector de tiempos
tiempom=zeros(1,10);
%Vector de pesos
pesosm=zeros(10,8);
%Vector de peso de la capa media a la final
pesosmlw=zeros(1,10);
%Vector de bias de la capa intermedia
biasm1=zeros(1,10);
%Vector de bias de la capa externa
biasm2=zeros(1,10);
%vector de epocas;
epocasm=zeros(10,1);
%El proceso se repite para cada réplica
for r=1:10
    paramos=6
    clc
    disp('***ESTAMOS EN LA REPLICA:');
    r
    %Se define los datos input y target de los datos de entrenamiento
    %Los datos son ingresados en forma transpuesta es decir cada columna es
    %un registro y cda fila una variable
    %Se define los datos input y target de los datos de entrenam+validacion
    p=(data((r-1)*n+1:(r-1)*n+ne+nv,2:9))';
    t=(data((r-1)*n+1:(r-1)*n+ne+nv,1:1))'
    %Se define los datos input y target de los datos de prurba
    dpp=(data((r-1)*n+ne+nv+1:(r-1)*n+ne+nv+np,2:9))';
    dpt=(data((r-1)*n+ne+nv+1:(r-1)*n+ne+nv+np,1:1))'
```

```

%SE COMIENZA A TRABAJAR LA RED
net=newff(minmax(p),[1,1],{'purelin','purelin'},'trainlm');
%Buscamos los pesos adecuados
pv=(data((r-1)*n+1:(r-1)*n+ne+nv,2:9));
tv=(data((r-1)*n+1:(r-1)*n+ne+nv,1:1) );
pesos=[0.05 0.05 0.05 0.05 0.05 0.05 0.05 0.05];
pesoslw=0.05;
bias1=0.05;
bias2=0.05;
pesos_sel=[0.05 0.05 0.05 0.05 0.05 0.05 0.05 0.05];
pesoslw_sel=0.05;
bias1_sel=0.05;
bias2_sel=0.05;
MSE_inicial=100;
t1=clock;
for j=1:100
    net=init(net);
    pesos=net.IW{1,1};
    pesoslw=net.LW{2,1};
    bias1=net.b{1,1};
    bias2=net.b{2,1};
    [net,tr]=train(net,p,t);
    net.trainParam.show=NaN;
    a_prueba=sim(net,pv);
    MSE_prueba=sum((tv-a_prueba).*(tv-a_prueba))/(ne+nv);
    if MSE_prueba<MSE_inicial
        MSE_inicial=MSE_prueba;
        pesos_sel=pesos;
        pesoslw_sel=pesoslw;
        bias1_sel=bias1;
        bias2_sel=bias2;
    end
end
net.IW{1,1}=pesos_sel;
net.LW{2,1}=pesoslw_sel;
net.b{1,1}=bias1_sel;
net.b{2,1}=bias2_sel;
[net,tr]=train(net,p,t);
net.trainParam.show=NaN;
pesosm(r,:)=net.IW{1,1};
biasm1(r)=net.b{1,1};
biasm2(r)=net.b{2,1};
pesosmlw(r)=net.LW{2,1};
epocasm(r)=max(tr.epoch);
t2=clock;
tiempom(r)=etime(t2,t1);
errores_entrena=tr.perf

```

```

%Viendo el MSE de los datos de entrenamiento
a=sim(net,p);
prediccion_entre=a
MSE_entre=sum((t-a).*(t-a))/(ne+nv);
%Viendo el MSE de los datos de prueba
a=sim(net,dpp);
prediccion_prueba=a
MSE_prueba=sum((dpt-a).*(dpt-a))/(np);
s_entre=s_entre+MSE_entre;
s_prueba=s_prueba+MSE_prueba;
end
%Arrojando Resultados deseados

MSE_Prom_entre=s_entre/10
MSE_Prom_prueba=s_prueba/10
epocas_trabajadas=epocasm'
tiempo_trabajados=tiempom
pesosm
pesosmlw
biasm1
biasm2

```

Anexo III.3.

Implementación de la red neuronal con datos de validación

```
function redesf4()

%Se importa todo el Excel de la carpeta work de MATLAB
data=xlsread('CASO4B','M3000');
%Indicar la cantidad de datos de entrenamiento(ne),validacion(nv)y
%prueba(np)
ne=2100;
nv=450;
np=450;
n=ne+nv+np;
s_entre=0;
s_entre_val=0;
s_prueba=0;
%Vector de tiempos
tiempom=zeros(1,10);
%Vector de pesos
pesosm=zeros(10,8);
%Vector de peso de la capa media a la final
pesosmlw=zeros(1,10);
%Vector de bias de la capa intermedia
biasm1=zeros(1,10);
%Vector de bias de la capa externa
biasm2=zeros(1,10);
%vector de epocas;
epocasm=zeros(10,1);
%El proceso se repite para cada réplica
for r=1:10

    paramos=6
    clc

    disp('***ESTAMOS EN LA REPLICA:');
    r
    %Se define los datos input y target de los datos de entrenamiento
    %Los datos son ingresados en forma transpuesta es decir cada columna es
    %un registro y cda fila una variable
    %Se define los datos input y target de los datos de entrenam+validacion
    p=(data((r-1)*n+1:(r-1)*n+ne,2:9))';
    t=(data((r-1)*n+1:(r-1)*n+ne,1:1))'
```

```

%Se define los datos input y target de los datos de validacion
VV.P=(data((r-1)*n+ne+1:(r-1)*n+ne+nv,2:9));
VV.T=(data((r-1)*n+ne+1:(r-1)*n+ne+nv,1:1))'
%Se define los datos input y target de los datos de entrenam+validacion
pev=(data((r-1)*n+1:(r-1)*n+ne+nv,2:9));
tev=(data((r-1)*n+1:(r-1)*n+ne+nv,1:1));
%Se define los datos input y target de los datos de prueba
dpp=(data((r-1)*n+ne+nv+1:(r-1)*n+ne+nv+np,2:9));
dpt=(data((r-1)*n+ne+nv+1:(r-1)*n+ne+nv+np,1:1))'
%SE COMIENZA A TRABAJAR LA RED
net=newff(minmax(p),[1,1],{'purelin','purelin'},'trainlm');
%Buscamos los pesos adecuados
pv=(data((r-1)*n+1:(r-1)*n+ne+nv,2:9));
tv=(data((r-1)*n+1:(r-1)*n+ne+nv,1:1))';
pesos=[0.05 0.05 0.05 0.05 0.05 0.05 0.05 0.05];
pesoslw=0.05;
bias1=0.05;
bias2=0.05;
pesos_sel=[0.05 0.05 0.05 0.05 0.05 0.05 0.05 0.05];
pesoslw_sel=0.05;
bias1_sel=0.05;
bias2_sel=0.05;
MSE_inicial=100;
t1=clock;
for j=1:100
    net=init(net);
    pesos=net.IW{1,1};
    pesoslw=net.LW{2,1};
    bias1=net.b{1,1};
    bias2=net.b{2,1};
    [net,tr]=train(net,p,t,[],[],VV);
    net.trainParam.show=NaN;
    a_prueba=sim(net,pv);
    MSE_prueba=sum((tv-a_prueba).*(tv-a_prueba))/(ne+nv);
    if MSE_prueba<MSE_inicial
        MSE_inicial=MSE_prueba;
        pesos_sel=pesos;
        pesoslw_sel=pesoslw;
        bias1_sel=bias1;
        bias2_sel=bias2;
    end
end
net.IW{1,1}=pesos_sel;
net.LW{2,1}=pesoslw_sel;
net.b{1,1}=bias1_sel;
net.b{2,1}=bias2_sel;
[net,tr]=train(net,p,t,[],[],VV);
net.trainParam.show=NaN;
pesosm(r,:)=net.IW{1,1};
biasm1(r)=net.b{1,1};
biasm2(r)=net.b{2,1};
pesosmlw(r)=net.LW{2,1};
epocasm(r)=max(tr.epoch);
t2=clock;

```

```

tiempom(r)=etime(t2,t1);
errores_entrena=tr.perf
errores_validac=tr.vperf
%Viendo el MSE de los datos de entrenamiento
a=sim(net,pev);
prediccion_entre=a
MSE_entre=sum((tev-a).*(tev-a))/(ne+nv);
%Viendo el MSE de los datos de prueba
a=sim(net,dpp);
prediccion_prueba=a
MSE_prueba=sum((dpt-a).*(dpt-a))/(np);

s_entre=s_entre+MSE_entre;
s_prueba=s_prueba+MSE_prueba;
end
%Arrojando Resultados deseados

MSE_Prom_entre=s_entre/10
MSE_Prom_prueba=s_prueba/10
epocas_trabajadas=epocasm'
tiempo_trabajados=tiempom
pesosm
pesosmlw
biasm1
biasm2

```

Anexos IV
Programas usados en el diseño de experimento

Anexo IV.1
Generación de la Matriz de Correlaciones

```
Public Sub genera_cor()
Randomize
For i = 5 To 12
  For j = 3 To 10
    fila = i - 4
    columna = j - 2
    If fila = columna Then
      Cells(i, j) = 1
    Else
      If fila > columna Then
        valor = (Int((200 * Rnd) + 1) + 199)
        Cells(i, j) = valor / 1000
      End If
    End If
  Next j
Next i
End Sub
```

Anexo IV.2

Matriz de Correlación usada para la Población objetivo

	x1	x2	x3	x4	x5	x6	x7	x8
x1	1	0.387	0.269	0.379	0.211	0.297	0.241	0.243
x2	0.387	1	0.386	0.254	0.257	0.31	0.372	0.384
x3	0.269	0.386	1	0.205	0.215	0.369	0.293	0.272
x4	0.379	0.254	0.205	1	0.392	0.237	0.374	0.367
x5	0.211	0.257	0.215	0.392	1	0.282	0.234	0.346
x6	0.297	0.31	0.369	0.237	0.282	1	0.232	0.202
x7	0.241	0.372	0.293	0.374	0.234	0.232	1	0.297
x8	0.243	0.384	0.272	0.367	0.346	0.202	0.297	1

Anexo IV.3

Matriz de Correlación usada para la Población Ajena (Outliers contaminante)

	x1	x2	x3	x4	x5	x6	x7	x8
x1	1.000	0.400	0.350	0.290	0.390	0.120	0.180	0.243
x2	0.400	1.000	0.120	0.350	0.290	0.240	0.290	0.102
x3	0.350	0.120	1.000	0.400	0.360	0.233	0.293	0.420
x4	0.290	0.350	0.400	1.000	0.230	0.410	0.250	0.360
x5	0.390	0.290	0.360	0.230	1.000	0.120	0.190	0.170
x6	0.120	0.240	0.233	0.410	0.120	1.000	0.232	0.202
x7	0.180	0.290	0.293	0.250	0.190	0.232	1.000	0.180
x8	0.243	0.102	0.420	0.360	0.170	0.202	0.180	1.000

Anexo IV.4

Incorporación de valores atípicos del Tipo Discordante

```
Public Sub atipicos_tipoa()  
'Esto ingresa atípicos del ti  
'DISCORDANTE, la cual es un dato que aparece en forma sorpresiva  
'Definimos las cantidades de entrenamiento(t), validación(v), prueba(p)  
ne = 2037  
ne_out = 63  
nv = 436  
nv_out = 14  
np = 450  
  
'Comenzamos con el llenado de atípicos  
fila = 1  
For i = 1 To 10  
  
    'Datos de entrenamiento  
    For j = 1 To ne  
        fila = fila + 1  
        Cells(fila, 13) = 0  
        Cells(fila, 14) = 1  
    Next j  
  
    If ne_out > 0 Then  
        For j = 1 To ne_out  
            fila = fila + 1  
            Cells(fila, 13) = 1  
            Cells(fila, 14) = 1  
            'Incorporando atípicos que pueden ir de 6 a 25  
            For k = 1 To 1  
                varx = Int((8 * Rnd) + 1) + 4  
                numero = Int((19000 * Rnd) + 1) + 6000  
                If Cells(fila, varx) < 0 Then  
                    signo = -1  
                Else  
                    signo = 1  
                End If  
                Cells(fila, varx) = signo * numero / 1000  
                Cells(fila, varx).Interior.ColorIndex = 45  
            Next k  
        Next j  
    End If  
  
    'Datos de validación  
    For j = 1 To nv  
        fila = fila + 1  
        Cells(fila, 13) = 0  
        Cells(fila, 14) = 2  
    Next j  
  
    If nv_out > 0 Then  
        For j = 1 To nv_out  
            fila = fila + 1
```

```

Cells(fila, 13) = 1
Cells(fila, 14) = 2
For k = 1 To 1
    varx = Int((8 * Rnd) + 1) + 4
    numero = Int((19000 * Rnd) + 1) + 6000
    If Cells(fila, varx) < 0 Then
        signo = -1
    Else
        signo = 1
    End If
    Cells(fila, varx) = signo * numero / 1000
    Cells(fila, varx).Interior.ColorIndex = 45
Next k
Next j
End If

'Datos de prueba
For j = 1 To np
    fila = fila + 1
    Cells(fila, 13) = 0
    Cells(fila, 14) = 3
Next j

Next i
End Sub

```

Anexo IV.5

Identificación de datos atípicos en SPSS usando Distancia de Mahalanobis

```
USE ALL.
COMPUTE filter_$=(Rep = 1).
VARIABLE LABEL filter_$ 'Rep = 1 (FILTER)'.
VALUE LABELS filter_$ 0 'No seleccionado' 1 'Seleccionado'.
FORMAT filter_$ (f1.0).
FILTER BY filter_$.
EXECUTE .

REGRESSION
/MISSING LISTWISE
/STATISTICS COEFF OUTS R ANOVA
/CRITERIA=PIN(.05) POUT(.10)
/NOORIGIN
/DEPENDENT Y
/METHOD=ENTER x1 x2 x3 x4 x5 x6 x7 x8
/SAVE MAHAL .

COMPUTE p_mah_1 = 1-CDF.CHISQ(MAH_1,8) .
EXECUTE .

FILTER OFF.
USE ALL.
EXECUTE .
```

Anexos V
Datos estadísticos

Anexo V.1

Prueba de Normalidad Multivariada – Test de Mardia

MUESTRA DE 40									
Réplica	Simetría				Curtosis				
	Hipótesis Rechazada	P valor	Estadístico	Valor crítico	Hipótesis Rechazada	P valor	Estadístico	Valor crítico	
1	No	0.9571	135.1316	210.1758	No	0.0403	-1.7475	2.3263	
2	No	0.76	151.8633	210.1758	No	0.0485	-1.6591	2.3263	
3	No	0.9547	135.5406	210.1758	No	0.011	-2.2903	2.3263	
4	No	0.8189	148.3714	210.1758	No	0.046	-1.6854	2.3263	
5	No	0.9913	124.9172	210.1758	No	0.0209	-2.0362	2.3263	
6	No	0.6999	155.0189	210.1758	No	0.0403	-1.7467	2.3263	
7	No	0.504	164.1527	210.1758	No	0.0172	-2.116	2.3263	
8	No	0.8859	143.4515	210.1758	No	0.0824	-1.3888	2.3263	
9	No	0.5767	160.8527	210.1758	No	0.0318	-1.8548	2.3263	
10	No	0.9832	128.7901	210.1758	No	0.0103	-2.3165	2.3263	

MUESTRA DE 160									
Réplica	Simetría				Curtosis				
	Hipótesis Rechazada	P valor	Estadístico	Valor crítico	Hipótesis Rechazada	P valor	Estadístico	Valor crítico	
1	No	0.9676	133.1074	210.1758	No	0.0262	-1.9395	2.3263	
2	No	0.5544	161.8694	210.1758	No	0.0311	-1.8653	2.3263	
3	No	0.853	146.0415	210.1758	No	0.2231	-0.7617	2.3263	
4	No	0.7243	153.7723	210.1758	No	0.2885	0.5577	2.3263	
5	No	0.4079	168.5928	210.1758	No	0.1433	-1.0656	2.3263	
6	No	0.4745	165.494	210.1758	No	0.1182	-1.1843	2.3263	
7	No	0.6326	158.2691	210.1758	No	0.1416	-1.0733	2.3263	
8	No	0.9825	129.0148	210.1758	No	0.0119	-2.2602	2.3263	
9	No	0.4743	165.5052	210.1758	No	0.1459	-1.054	2.3263	
10	No	0.6748	156.2548	210.1758	No	0.1339	-1.108	2.3263	

MUESTRA DE 1000									
Réplica	Simetría				Curtosis				
	Hipótesis Rechazada	P valor	Estadístico	Valor crítico	Hipótesis Rechazada	P valor	Estadístico	Valor crítico	
1	No	0.5054	164.0891	210.1758	No	0.062	-1.5384	2.3263	
2	No	0.927	139.3777	210.1758	No	0.3209	-0.4653	2.3263	
3	No	0.1752	181.8433	210.1758	No	0.16	-0.9946	2.3263	
4	No	0.6693	156.5258	210.1758	No	0.0341	-1.8237	2.3263	
5	No	0.571	161.1134	210.1758	No	0.0924	-1.3264	2.3263	
6	No	0.7473	152.5592	210.1758	No	0.4856	-0.036	2.3263	
7	No	0.5379	162.6166	210.1758	No	0.1444	-1.0607	2.3263	
8	No	0.4517	166.5431	210.1758	No	0.2838	-0.5716	2.3263	
9	No	0.0454	196.9171	210.1758	No	0.4267	0.1847	2.3263	
10	No	0.0906	189.77	210.1758	No	0.0439	-1.7068	2.3263	

MUESTRA DE 3000									
Réplica	Simetría				Curtosis				
	Hipótesis Rechazada	P valor	Estadístico	Valor crítico	Hipótesis Rechazada	P valor	Estadístico	Valor crítico	
1	No	0.2397	177.4852	210.1758	No	0.0222	-2.0097	2.3263	
2	No	0.8171	148.4844	210.1758	No	0.0743	-1.4444	2.3263	
3	No	0.9867	127.3757	210.1758	No	0.1607	-0.9915	2.3263	
4	No	0.9562	135.2846	210.1758	No	0.0145	-2.1842	2.3263	
5	No	0.8767	144.2206	210.1758	No	0.0172	-2.1162	2.3263	
6	No	0.6435	157.7561	210.1758	No	0.1041	-1.2588	2.3263	
7	No	0.6275	158.5079	210.1758	No	0.0352	-1.8091	2.3263	
8	No	0.7167	154.1662	210.1758	No	0.0134	-2.2131	2.3263	
9	No	0.6965	155.1862	210.1758	No	0.3959	0.264	2.3263	
10	No	0.8346	147.3338	210.1758	No	0.0121	-2.2529	2.3263	

Anexo V.2

Intensidad de Multicolinealidad (Atípicos Discordante)

CASO I				
Réplicas	Muestra 40	Muestra 160	Muestra 1000	Muestra 3000
R1	5.035	2.987	2.559	2.617
R2	3.999	2.613	2.587	2.612
R3	3.829	3.281	2.707	2.637
R4	3.557	2.711	2.665	2.636
R5	3.076	3.037	2.798	2.648
R6	4.445	3.112	2.740	2.701
R7	3.389	2.572	2.634	2.647
R8	4.670	2.854	2.528	2.681
R9	2.974	3.057	2.683	2.712
R10	3.948	3.026	2.468	2.683

CASO II				
Réplicas	Muestra 40	Muestra 160	Muestra 1000	Muestra 3000
R1	8.091	3.433	2.180	1.769
R2	7.851	3.945	2.157	1.862
R3	11.950	4.281	1.988	1.871
R4	6.213	3.642	1.938	1.679
R5	8.214	3.612	2.043	1.745
R6	8.422	4.298	2.127	1.894
R7	7.493	4.036	2.224	1.820
R8	8.035	3.461	2.120	1.903
R9	6.131	4.352	2.151	1.773
R10	6.303	3.209	2.140	1.762

CASO III				
Réplicas	Muestra 40	Muestra 160	Muestra 1000	Muestra 3000
R1	8.345	3.459	2.487	2.331
R2	9.229	3.914	2.275	2.286
R3	8.679	3.728	2.888	2.201
R4	9.945	3.603	2.557	2.303
R5	5.250	3.059	2.758	2.168
R6	6.151	3.779	2.455	2.099
R7	10.437	2.777	2.557	2.333
R8	7.849	3.609	2.455	2.298
R9	6.404	3.426	2.594	2.227
R10	8.186	3.210	2.548	2.476

CASO IV				
Réplicas	Muestra 40	Muestra 160	Muestra 1000	Muestra 3000
R1	9.142	3.587	2.035	1.693
R2	9.142	3.990	2.052	1.851
R3	11.279	4.357	1.828	1.815
R4	9.658	3.655	1.794	1.679
R5	7.806	3.326	1.960	1.585
R6	8.219	3.439	2.025	1.826
R7	9.925	3.527	2.072	1.726
R8	8.389	3.337	2.048	1.683
R9	7.804	4.556	2.023	1.727
R10	9.336	2.910	2.042	1.712

Caso I: Muestra sin valores atípicos

Caso II : 3% de atípicos en los datos de entrenamiento

Caso III: 3% de atípicos en los datos de validación

Caso IV: 3% de atípicos en los datos de entrenamiento y validación.

Anexo V.3

Intensidad de Multicolinealidad (Atípicos Contaminantes)

CASO I				
Réplicas	Muestra 40	Muestra 160	Muestra 1000	Muestra 3000
R1	5.035	2.987	2.559	2.617
R2	3.999	2.613	2.587	2.612
R3	3.829	3.281	2.707	2.637
R4	3.557	2.711	2.665	2.636
R5	3.076	3.037	2.798	2.648
R6	4.445	3.112	2.740	2.701
R7	3.389	2.572	2.634	2.647
R8	4.670	2.854	2.528	2.681
R9	2.974	3.057	2.683	2.712
R10	3.948	3.026	2.468	2.683

CASO II				
Réplicas	Muestra 40	Muestra 160	Muestra 1000	Muestra 3000
R1	4.992	2.972	2.552	2.608
R2	3.791	2.677	2.565	2.591
R3	3.356	3.271	2.736	2.611
R4	3.672	2.750	2.608	2.611
R5	3.051	3.054	2.787	2.608
R6	4.339	3.071	2.679	2.683
R7	3.339	2.577	2.644	2.634
R8	4.404	2.894	2.487	2.660
R9	3.406	3.006	2.661	2.689
R10	3.868	3.023	2.470	2.654

CASO III				
Réplicas	Muestra 40	Muestra 160	Muestra 1000	Muestra 3000
R1	4.849	2.961	2.555	2.611
R2	3.892	2.677	2.585	2.611
R3	3.416	3.268	2.721	2.638
R4	3.565	2.710	2.657	2.625
R5	2.919	3.032	2.784	2.651
R6	4.580	3.146	2.734	2.709
R7	3.444	2.641	2.636	2.649
R8	4.207	2.861	2.520	2.681
R9	3.220	3.042	2.666	2.709
R10	4.118	3.053	2.450	2.678

CASO IV				
Réplicas	Muestra 40	Muestra 160	Muestra 1000	Muestra 3000
R1	4.168	2.964	2.554	2.604
R2	3.757	2.660	2.581	2.592
R3	3.107	3.249	2.749	2.609
R4	3.876	2.762	2.599	2.601
R5	3.067	3.063	2.771	2.607
R6	4.419	3.106	2.661	2.683
R7	3.396	2.598	2.642	2.632
R8	3.916	2.900	2.483	2.653
R9	3.305	2.989	2.663	2.679
R10	3.707	3.053	2.478	2.648

Caso I: Muestra sin valores atípicos

Caso II : 3% de atípicos en los datos de entrenamiento

Caso III: 3% de atípicos en los datos de validación

Caso IV: 3% de atípicos en los datos de entrenamiento y validación.

Anexo V.4

Estadístico Durbin Watson – Independencia (Atípicos Discordante)

CASO I				
Réplicas	Muestra 40	Muestra 160	Muestra 1000	Muestra 3000
R1	1.493	2.282	1.916	2.053
R2	1.668	1.960	2.066	1.931
R3	1.800	1.744	2.029	1.934
R4	1.223	2.288	1.960	2.082
R5	2.071	1.692	1.888	2.040
R6	1.508	1.867	2.037	2.095
R7	1.959	2.039	2.162	2.043
R8	1.883	2.032	1.993	2.023
R9	2.025	1.883	2.087	1.984
R10	1.703	1.451	1.994	1.986

CASO II				
Réplicas	Muestra 40	Muestra 160	Muestra 1000	Muestra 3000
R1	1.601	2.233	2.120	1.769
R2	1.333	2.120	2.085	1.862
R3	1.834	1.821	2.254	1.871
R4	1.015	2.535	1.761	1.679
R5	2.228	1.866	2.218	1.745
R6	1.797	1.663	1.514	1.894
R7	1.953	2.143	2.044	1.820
R8	1.475	1.921	1.985	1.903
R9	1.862	1.832	2.310	1.773
R10	1.681	1.670	1.847	1.762

CASO III				
Réplicas	Muestra 40	Muestra 160	Muestra 1000	Muestra 3000
R1	1.569	2.190	2.487	2.104
R2	1.579	1.916	2.275	2.158
R3	1.851	1.828	2.888	2.097
R4	1.501	2.290	2.557	1.990
R5	1.962	1.533	2.758	2.107
R6	1.535	1.883	2.455	2.325
R7	1.648	2.050	2.557	1.922
R8	1.895	1.944	2.455	1.714
R9	1.668	2.029	2.594	2.184
R10	1.570	1.481	2.548	1.948

CASO IV				
Réplicas	Muestra 40	Muestra 160	Muestra 1000	Muestra 3000
R1	1.850	2.179	2.108	1.868
R2	1.273	2.086	2.088	1.897
R3	1.853	1.928	2.066	2.106
R4	1.452	2.551	1.840	2.102
R5	2.067	1.734	2.139	1.866
R6	1.865	1.682	1.594	2.050
R7	1.657	2.143	1.993	1.955
R8	1.508	1.794	1.994	1.952
R9	1.878	2.039	2.356	2.061
R10	1.514	1.677	1.811	2.037

Caso I: Muestra sin valores atípicos

Caso II : 3% de atípicos en los datos de entrenamiento

Caso III: 3% de atípicos en los datos de validación

Caso IV: 3% de atípicos en los datos de entrenamiento y validación.

Anexo V.5

Estadístico Durbin Watson – Independencia (Atípicos Contaminante)

CASO I				
Réplicas	Muestra 40	Muestra 160	Muestra 1000	Muestra 3000
R1	1.493	2.282	1.916	2.053
R2	1.668	1.960	2.066	1.931
R3	1.800	1.744	2.029	1.934
R4	1.223	2.288	1.960	2.082
R5	2.071	1.692	1.888	2.040
R6	1.508	1.867	2.037	2.095
R7	1.959	2.039	2.162	2.043
R8	1.883	2.032	1.993	2.023
R9	2.025	1.883	2.087	1.984
R10	1.703	1.451	1.994	1.986

CASO II				
Réplicas	Muestra 40	Muestra 160	Muestra 1000	Muestra 3000
R1	1.542	2.288	1.897	2.060
R2	1.585	2.069	2.082	1.962
R3	1.879	1.697	2.009	1.952
R4	1.369	2.249	1.957	2.121
R5	2.106	1.679	1.894	2.074
R6	1.499	1.756	2.019	2.100
R7	1.939	2.085	2.136	2.055
R8	1.812	1.992	2.023	2.028
R9	1.962	1.949	2.070	2.005
R10	1.713	1.501	1.979	1.971

CASO III				
Réplicas	Muestra 40	Muestra 160	Muestra 1000	Muestra 3000
R1	1.489	2.297	1.921	2.052
R2	1.661	1.962	2.068	1.928
R3	1.553	1.738	2.032	1.944
R4	1.195	2.290	1.967	2.092
R5	2.073	1.689	1.883	2.036
R6	1.506	1.860	2.025	2.083
R7	1.906	2.071	2.130	2.049
R8	1.937	2.016	2.017	2.024
R9	2.238	1.854	2.076	1.982
R10	1.685	1.423	1.998	1.986

CASO IV				
Réplicas	Muestra 40	Muestra 160	Muestra 1000	Muestra 3000
R1	1.720	2.302	1.905	2.064
R2	1.450	2.073	2.076	1.954
R3	1.874	1.702	2.008	1.965
R4	1.920	2.250	1.962	2.113
R5	2.038	1.682	1.882	2.079
R6	1.499	1.739	2.021	2.082
R7	1.773	2.117	2.145	2.061
R8	1.723	1.968	2.015	2.054
R9	2.076	1.945	2.072	1.996
R10	1.744	1.502	1.957	1.948

Caso I: Muestra sin valores atípicos

Caso II: 3% de atípicos en los datos de entrenamiento

Caso III: 3% de atípicos en los datos de validación

Caso IV: 3% de atípicos en los datos de entrenamiento y validación.

Anexo V.6

Prueba de Homocedasticidad – Test de Levene (Atípicos Discordante)

CASO I								
Réplicas	Muestra 40		Muestra 160		Muestra 1000		Muestra 3000	
	Estadístico- Levene	Sig.	Estadístico- Levene	Sig.	Estadístico- Levene	Sig.	Estadístico- Levene	Sig.
R1	0.059	0.809	0.981	0.324	0.260	0.610	0.921	0.337
R2	0.114	0.738	2.851	0.094	2.203	0.138	0.324	0.569
R3	0.088	0.769	0.000	0.985	0.297	0.586	1.355	0.245
R4	0.043	0.837	0.028	0.867	0.328	0.567	0.750	0.386
R5	6.316	0.017	0.388	0.534	1.327	0.250	0.121	0.728
R6	0.319	0.576	0.290	0.591	0.006	0.936	0.027	0.869
R7	1.925	0.175	6.065	0.015	0.427	0.514	0.857	0.355
R8	3.991	0.054	3.968	0.048	0.039	0.843	1.798	0.180
R9	2.083	0.159	0.728	0.395	0.142	0.706	0.072	0.789
R10	0.017	0.898	0.019	0.891	2.274	0.132	0.036	0.849

CASO II								
Réplicas	Muestra 40		Muestra 160		Muestra 1000		Muestra 3000	
	Estadístico- Levene	Sig.	Estadístico- Levene	Sig.	Estadístico- Levene	Sig.	Estadístico- Levene	Sig.
R1	0.093	0.762	1.973	0.162	1.048	0.306	0.393	0.531
R2	0.230	0.635	0.329	0.567	0.013	0.910	0.358	0.550
R3	0.124	0.728	0.003	0.954	0.174	0.677	0.311	0.577
R4	0.360	0.553	0.121	0.729	1.977	0.160	0.156	0.693
R5	0.043	0.837	3.038	0.084	0.002	0.964	1.263	0.261
R6	0.515	0.478	0.149	0.700	0.008	0.928	1.832	0.176
R7	1.308	0.261	0.752	0.388	0.314	0.575	1.498	0.221
R8	0.313	0.580	9.169	0.003	1.010	0.315	0.415	0.519
R9	1.741	0.196	0.031	0.861	0.041	0.839	0.011	0.917
R10	0.032	0.859	0.007	0.932	1.789	0.181	0.066	0.798

CASO III								
Réplicas	Muestra 40		Muestra 160		Muestra 1000		Muestra 3000	
	Estadístico- Levene	Sig.	Estadístico- Levene	Sig.	Estadístico- Levene	Sig.	Estadístico- Levene	Sig.
R1	1.699	0.202	2.541	0.113	1.007	0.316	0.072	0.789
R2	0.179	0.675	1.166	0.282	0.443	0.506	0.000	0.993
R3	0.249	0.621	0.265	0.608	0.001	0.977	1.346	0.246
R4	0.333	0.568	0.052	0.820	0.698	0.404	2.133	0.144
R5	1.806	0.188	0.514	0.475	2.440	0.119	0.275	0.600
R6	2.317	0.138	0.009	0.925	1.735	0.188	0.133	0.715
R7	2.245	0.144	5.664	0.019	0.269	0.604	0.014	0.906
R8	0.026	0.872	3.938	0.049	0.705	0.401	3.502	0.061
R9	0.458	0.503	0.527	0.469	0.681	0.409	0.027	0.869
R10	0.786	0.382	0.076	0.783	0.021	0.885	0.619	0.431

CASO IV								
Réplicas	Muestra 40		Muestra 160		Muestra 1000		Muestra 3000	
	Estadístico- Levene	Sig.	Estadístico- Levene	Sig.	Estadístico- Levene	Sig.	Estadístico- Levene	Sig.
R1	0.112	0.740	0.938	0.335	0.007	0.935	0.076	0.783
R2	0.024	0.877	0.236	0.628	0.085	0.770	0.325	0.569
R3	0.199	0.659	0.970	0.326	0.136	0.713	0.799	0.372
R4	0.495	0.487	0.504	0.479	2.409	0.121	0.002	0.967
R5	0.002	0.969	2.756	0.099	0.332	0.565	1.056	0.304
R6	0.107	0.746	0.360	0.549	0.061	0.805	1.656	0.198
R7	0.522	0.475	0.906	0.343	0.427	0.514	1.928	0.165
R8	0.231	0.634	3.220	0.075	1.601	0.206	0.053	0.817
R9	0.007	0.936	0.030	0.862	0.012	0.911	0.016	0.900
R10	1.556	0.221	0.182	0.671	1.465	0.226	0.401	0.527

Caso I: Muestra sin valores atípicos

Caso II: 3% de atípicos en los datos de entrenamiento

Caso III: 3% de atípicos en los datos de validación

Caso IV: 3% de atípicos en los datos de entrenamiento y validación.

Anexo V.7

Prueba de Homocedasticidad – Test de Levene (Atípicos Contaminante)

CASO I								
Réplicas	Muestra 40		Muestra 160		Muestra 1000		Muestra 3000	
	Estadístico- Levene	Sig.	Estadístico- Levene	Sig.	Estadístico- Levene	Sig.	Estadístico- Levene	Sig.
R1	0.059	0.809	0.981	0.324	0.260	0.610	0.921	0.337
R2	0.114	0.738	2.851	0.094	2.203	0.138	0.324	0.569
R3	0.088	0.769	0.000	0.985	0.297	0.586	1.355	0.245
R4	0.043	0.837	0.028	0.867	0.328	0.567	0.750	0.386
R5	6.316	0.017	0.388	0.534	1.327	0.250	0.121	0.728
R6	0.319	0.576	0.290	0.591	0.006	0.936	0.027	0.869
R7	1.925	0.175	6.065	0.015	0.427	0.514	0.857	0.355
R8	3.991	0.054	3.968	0.048	0.039	0.843	1.798	0.180
R9	2.083	0.159	0.728	0.395	0.142	0.706	0.072	0.789
R10	0.017	0.898	0.019	0.891	2.274	0.132	0.036	0.849

CASO II								
Réplicas	Muestra 40		Muestra 160		Muestra 1000		Muestra 3000	
	Estadístico- Levene	Sig.	Estadístico- Levene	Sig.	Estadístico- Levene	Sig.	Estadístico- Levene	Sig.
R1	0.440	0.512	0.788	0.376	0.093	0.761	0.713	0.398
R2	0.092	0.763	4.170	0.043	1.968	0.161	1.100	0.294
R3	0.004	0.949	0.107	0.745	1.781	0.182	1.896	0.169
R4	0.006	0.936	0.165	0.685	0.176	0.675	0.381	0.537
R5	2.163	0.151	0.477	0.491	2.368	0.124	0.000	0.988
R6	0.334	0.567	0.285	0.594	0.029	0.865	0.266	0.606
R7	2.175	0.150	5.263	0.023	0.552	0.458	0.333	0.564
R8	4.459	0.043	4.434	0.037	0.046	0.831	1.446	0.229
R9	1.710	0.200	1.148	0.286	0.045	0.833	0.003	0.956
R10	0.290	0.594	0.177	0.675	1.432	0.232	0.001	0.980

CASO III								
Réplicas	Muestra 40		Muestra 160		Muestra 1000		Muestra 3000	
	Estadístico- Levene	Sig.	Estadístico- Levene	Sig.	Estadístico- Levene	Sig.	Estadístico- Levene	Sig.
R1	0.990	0.327	1.074	0.302	0.344	0.558	0.655	0.418
R2	0.213	0.648	2.746	0.100	2.151	0.143	0.472	0.492
R3	1.052	0.313	0.006	0.940	0.354	0.552	1.053	0.305
R4	0.003	0.956	0.021	0.885	0.221	0.639	0.836	0.361
R5	5.899	0.021	0.634	0.427	1.214	0.271	0.001	0.980
R6	0.243	0.625	0.494	0.484	0.000	0.995	0.076	0.782
R7	2.414	0.130	4.432	0.037	0.225	0.635	1.163	0.281
R8	3.643	0.065	3.739	0.055	0.201	0.654	1.947	0.163
R9	1.475	0.233	0.414	0.521	0.018	0.894	0.082	0.775
R10	0.318	0.577	0.010	0.922	2.515	0.113	0.007	0.933

CASO IV								
Réplicas	Muestra 40		Muestra 160		Muestra 1000		Muestra 3000	
	Estadístico- Levene	Sig.	Estadístico- Levene	Sig.	Estadístico- Levene	Sig.	Estadístico- Levene	Sig.
R1	0.077	0.783	1.149	0.286	0.005	0.941	0.522	0.470
R2	0.000	0.985	3.567	0.061	1.637	0.201	1.236	0.266
R3	0.183	0.672	0.197	0.657	2.137	0.144	1.724	0.189
R4	0.106	0.747	0.177	0.674	0.376	0.540	0.529	0.467
R5	1.291	0.264	0.868	0.353	3.386	0.066	0.039	0.844
R6	1.209	0.280	0.534	0.466	0.034	0.854	0.598	0.439
R7	2.212	0.147	4.879	0.029	0.293	0.588	0.148	0.700
R8	1.379	0.249	4.188	0.043	0.020	0.887	1.180	0.278
R9	0.883	0.354	1.024	0.313	0.091	0.763	0.005	0.943
R10	0.558	0.461	0.057	0.812	1.110	0.292	0.071	0.790

Caso I: Muestra sin valores atípicos

Caso II: 3% de atípicos en los datos de entrenamiento

Caso III: 3% de atípicos en los datos de validación

Caso IV: 3% de atípicos en los datos de entrenamiento y validación.

Anexo V.4

Prueba de Normalidad de errores – Kolmogorov (Atípicos Discordante)

CASO I								
Réplicas	Muestra 40		Muestra 160		Muestra 1000		Muestra 3000	
	Kolmogorov-Smirnov(a)	Sig.	Kolmogorov-Smirnov(a)	Sig.	Kolmogorov-Smirnov(a)	Sig.	Kolmogorov-Smirnov(a)	Sig.
R1	0.114	.200(*)	0.055	.200(*)	0.020	.200(*)	0.014	.200(*)
R2	0.070	.200(*)	0.066	.200(*)	0.027	.151(*)	0.016	.111(*)
R3	0.095	.200(*)	0.043	.200(*)	0.017	.200(*)	0.013	.200(*)
R4	0.131	.142(*)	0.064	.200(*)	0.023	.200(*)	0.012	.200(*)
R5	0.122	.200(*)	0.044	.200(*)	0.027	.175(*)	0.009	.200(*)
R6	0.109	.200(*)	0.056	.200(*)	0.015	.200(*)	0.013	.200(*)
R7	0.114	.200(*)	0.077	.051(*)	0.018	.200(*)	0.014	.200(*)
R8	0.110	.200(*)	0.071	.085(*)	0.020	.200(*)	0.011	.200(*)
R9	0.153	.043(*)	0.045	.200(*)	0.019	.200(*)	0.009	.200(*)
R10	0.100	.200(*)	0.043	.200(*)	0.030	0.077	0.007	.200(*)
CASO II								
Réplicas	Muestra 40		Muestra 160		Muestra 1000		Muestra 3000	
	Kolmogorov-Smirnov(a)	Sig.	Kolmogorov-Smirnov(a)	Sig.	Kolmogorov-Smirnov(a)	Sig.	Kolmogorov-Smirnov(a)	Sig.
R1	0.101	.200(*)	0.067	.200(*)	0.049	0.000	0.051	0.000
R2	0.082	.200(*)	0.047	.200(*)	0.054	0.000	0.056	0.000
R3	0.083	.200(*)	0.028	.200(*)	0.065	0.000	0.052	0.000
R4	0.102	.200(*)	0.057	.200(*)	0.070	0.000	0.053	0.000
R5	0.095	.200(*)	0.045	.200(*)	0.052	0.000	0.052	0.000
R6	0.082	.200(*)	0.055	.200(*)	0.051	0.000	0.057	0.000
R7	0.103	.200(*)	0.052	.200(*)	0.062	0.000	0.059	0.000
R8	0.135	.120(*)	0.084	.019(*)	0.054	0.000	0.048	0.000
R9	0.096	.200(*)	0.049	.200(*)	0.067	0.000	0.053	0.000
R10	0.121	.200(*)	0.048	.200(*)	0.052	0.000	0.058	0.000
CASO III								
Réplicas	Muestra 40		Muestra 160		Muestra 1000		Muestra 3000	
	Kolmogorov-Smirnov(a)	Sig.	Kolmogorov-Smirnov(a)	Sig.	Kolmogorov-Smirnov(a)	Sig.	Kolmogorov-Smirnov(a)	Sig.
R1	0.117	.200(*)	0.068	.200(*)	0.037	0.009	0.060	0.000
R2	0.073	.200(*)	0.059	.200(*)	0.030	.060(*)	0.065	0.000
R3	0.099	.200(*)	0.051	.200(*)	0.037	0.008	0.064	0.000
R4	0.094	.200(*)	0.061	.200(*)	0.050	0.000	0.057	0.000
R5	0.087	.200(*)	0.047	.200(*)	0.042	0.001	0.055	0.000
R6	0.124	.200(*)	0.053	.200(*)	0.047	0.000	0.056	0.000
R7	0.166	.019(*)	0.084	.020(*)	0.053	0.000	0.048	0.000
R8	0.078	.200(*)	0.060	.200(*)	0.027	.149(*)	0.048	0.000
R9	0.093	.200(*)	0.053	.200(*)	0.070	0.000	0.064	0.000
R10	0.100	.200(*)	0.050	.200(*)	0.051	0.000	0.056	0.000
CASO IV								
Réplicas	Muestra 40		Muestra 160		Muestra 1000		Muestra 3000	
	Kolmogorov-Smirnov(a)	Sig.	Kolmogorov-Smirnov(a)	Sig.	Kolmogorov-Smirnov(a)	Sig.	Kolmogorov-Smirnov(a)	Sig.
R1	0.100	.200(*)	0.057	.200(*)	0.044	0.001	0.046	0.000
R2	0.115	.200(*)	0.036	.200(*)	0.052	0.000	0.052	0.000
R3	0.078	.200(*)	0.041	.200(*)	0.059	0.000	0.045	0.000
R4	0.115	.200(*)	0.049	.200(*)	0.062	0.000	0.048	0.000
R5	0.075	.200(*)	0.070	.010(*)	0.049	0.000	0.048	0.000
R6	0.102	.200(*)	0.041	.200(*)	0.047	0.000	0.050	0.000
R7	0.140	.088(*)	0.062	.200(*)	0.055	0.000	0.054	0.000
R8	0.153	.043(*)	0.063	.200(*)	0.058	0.000	0.042	0.000
R9	0.074	.200(*)	0.048	.200(*)	0.060	0.000	0.047	0.000
R10	0.099	.200(*)	0.064	.200(*)	0.053	0.000	0.050	0.000

Caso I: Muestra sin valores atípicos

Caso II : 3% de atípicos en los datos de entrenamiento

Caso III: 3% de atípicos en los datos de validación

Caso IV: 3% de atípicos en los datos de entrenamiento y validación.

Anexo V.5

Prueba de Normalidad de errores – Kolmogorov (Atípicos Contaminante)

CASO I								
Réplicas	Muestra 40		Muestra 160		Muestra 1000		Muestra 3000	
	Kolmogorov-Smirnov(a)	Sig.	Kolmogorov-Smirnov(a)	Sig.	Kolmogorov-Smirnov(a)	Sig.	Kolmogorov-Smirnov(a)	Sig.
R1	0.114	.200(*)	0.055	.200(*)	0.020	.200(*)	0.014	.200(*)
R2	0.070	.200(*)	0.066	.200(*)	0.027	.152(*)	0.016	.111(*)
R3	0.095	.200(*)	0.043	.200(*)	0.017	.200(*)	0.013	.200(*)
R4	0.131	.148(*)	0.064	.200(*)	0.023	.200(*)	0.012	.200(*)
R5	0.122	.200(*)	0.044	.200(*)	0.027	.175(*)	0.009	.200(*)
R6	0.109	.200(*)	0.056	.200(*)	0.015	.200(*)	0.013	.200(*)
R7	0.114	.200(*)	0.077	.045(*)	0.018	.200(*)	0.014	.200(*)
R8	0.110	.200(*)	0.071	.085(*)	0.020	.200(*)	0.011	.200(*)
R9	0.153	.043(*)	0.045	.200(*)	0.019	.200(*)	0.009	.200(*)
R10	0.100	.200(*)	0.043	.200(*)	0.030	.077(*)	0.007	.200(*)
CASO II								
Réplicas	Muestra 40		Muestra 160		Muestra 1000		Muestra 3000	
	Kolmogorov-Smirnov(a)	Sig.	Kolmogorov-Smirnov(a)	Sig.	Kolmogorov-Smirnov(a)	Sig.	Kolmogorov-Smirnov(a)	Sig.
R1	0.113	.200(*)	0.050	.200(*)	0.018	.200(*)	0.012	.200(*)
R2	0.068	.200(*)	0.067	.200(*)	0.024	.200(*)	0.015	.140(*)
R3	0.069	.200(*)	0.062	.200(*)	0.022	.200(*)	0.016	.107(*)
R4	0.141	.084(*)	0.054	.200(*)	0.026	.200(*)	0.011	.200(*)
R5	0.116	.200(*)	0.046	.200(*)	0.030	.077(*)	0.012	.200(*)
R6	0.090	.200(*)	0.034	.200(*)	0.017	.200(*)	0.016	.120(*)
R7	0.111	.200(*)	0.077	.050(*)	0.019	.200(*)	0.013	.200(*)
R8	0.124	.200(*)	0.067	.200(*)	0.023	.200(*)	0.011	.200(*)
R9	0.098	.200(*)	0.043	.200(*)	0.017	.200(*)	0.009	.200(*)
R10	0.097	.200(*)	0.044	.200(*)	0.027	.181(*)	0.012	.200(*)
CASO III								
Réplicas	Muestra 40		Muestra 160		Muestra 1000		Muestra 3000	
	Kolmogorov-Smirnov(a)	Sig.	Kolmogorov-Smirnov(a)	Sig.	Kolmogorov-Smirnov(a)	Sig.	Kolmogorov-Smirnov(a)	Sig.
R1	0.099	.200(*)	0.052	.200(*)	0.021	.200(*)	0.014	.200(*)
R2	0.075	.200(*)	0.057	.200(*)	0.028	.119(*)	0.016	.119(*)
R3	0.102	.200(*)	0.048	.200(*)	0.018	.200(*)	0.014	.200(*)
R4	0.094	.200(*)	0.062	.200(*)	0.026	.193(*)	0.009	.200(*)
R5	0.132	.1432(*)	0.045	.200(*)	0.027	.168(*)	0.013	.200(*)
R6	0.106	.200(*)	0.053	.200(*)	0.015	.200(*)	0.015	.200(*)
R7	0.121	.200(*)	0.071	.089(*)	0.016	.200(*)	0.012	.200(*)
R8	0.151	.048(*)	0.057	.200(*)	0.021	.200(*)	0.010	.200(*)
R9	0.135	.117(*)	0.047	.200(*)	0.016	.200(*)	0.009	.200(*)
R10	0.100	.200(*)	0.046	.200(*)	0.032	.035(*)	0.009	.200(*)
CASO IV								
Réplicas	Muestra 40		Muestra 160		Muestra 1000		Muestra 3000	
	Kolmogorov-Smirnov(a)	Sig.	Kolmogorov-Smirnov(a)	Sig.	Kolmogorov-Smirnov(a)	Sig.	Kolmogorov-Smirnov(a)	Sig.
R1	0.108	.200(*)	0.049	.200(*)	0.016	.200(*)	0.012	.200(*)
R2	0.066	.200(*)	0.070	.196(*)	0.023	.200(*)	0.015	.147(*)
R3	0.083	.200(*)	0.059	.200(*)	0.021	.200(*)	0.017	.099(*)
R4	0.101	.200(*)	0.055	.200(*)	0.029	.096(*)	0.012	.200(*)
R5	0.103	.200(*)	0.050	.200(*)	0.030	.059(*)	0.013	.200(*)
R6	0.082	.200(*)	0.033	.200(*)	0.018	.200(*)	0.017	.098(*)
R7	0.137	.104(*)	0.073	.072(*)	0.018	.200(*)	0.013	.200(*)
R8	0.122	.200(*)	0.062	.200(*)	0.021	.200(*)	0.013	.200(*)
R9	0.077	.200(*)	0.044	.200(*)	0.018	.200(*)	0.010	.200(*)
R10	0.075	.200(*)	0.046	.200(*)	0.027	.171(*)	0.011	.200(*)

Caso I: Muestra sin valores atípicos

Caso II : 3% de atípicos en los datos de entrenamiento

Caso III: 3% de atípicos en los datos de validación

Caso IV: 3% de atípicos en los datos de entrenamiento y validación.

Anexo V.6

Error Cuadrado del Análisis de Regresión Múltiple – Datos atípicos discordantes

Réplica	Muestra 40				Muestra 160				Muestra 1000				Muestra 3000			
	CASO				CASO				CASO				CASO			
	I	II	III	IV	I	II	III	IV	I	II	III	IV	I	II	III	IV
R1	0.435327	0.451369	0.558760	0.881707	0.874002	1.649750	1.295769	1.844707	0.954210	3.335091	1.466977	3.568112	0.983314	3.455395	1.808253	3.728327
R2	0.682184	1.263398	0.710629	1.275333	0.781346	1.293751	0.944827	1.529118	1.031938	3.646719	1.260721	3.689143	0.967536	3.322898	1.730072	3.611876
R3	0.748625	0.678051	0.739069	0.577076	0.861210	1.135803	0.985742	1.357833	0.930762	3.102127	1.426170	3.477356	0.925508	3.610115	1.900868	3.865443
R4	0.470419	0.483077	0.636589	0.614874	0.842234	1.514529	0.841517	1.520591	0.930035	3.313896	1.360340	3.505437	0.950448	3.179467	1.661704	3.399959
R5	0.826863	1.020653	0.931421	1.127133	0.986227	1.665239	1.267068	1.969059	1.054060	3.231441	1.630107	3.636813	0.961247	3.346484	1.662791	3.556728
R6	0.873401	1.503649	0.886430	1.713929	1.052675	1.368892	1.222125	1.374508	0.950476	3.790033	1.665847	4.012278	0.969532	3.283159	1.811097	3.541559
R7	0.627383	0.692626	0.686700	0.744436	0.925303	2.358056	0.925569	2.367836	0.879239	3.188337	1.360953	3.513012	1.034770	3.026884	1.609559	3.311108
R8	0.572206	1.226895	0.669926	1.317374	1.021549	1.402644	1.114975	1.755425	1.020199	3.473102	1.472954	3.596926	0.985834	3.317190	1.745498	3.520355
R9	0.812378	0.822147	0.983131	1.033173	0.806374	0.864759	1.040568	1.757674	0.899348	3.067664	1.777622	3.497022	0.997989	3.372384	2.072509	3.719922
R10	0.480258	0.479510	0.629536	0.788671	0.963308	1.272856	0.988850	1.353331	0.987328	3.721336	1.651299	3.939351	0.975737	3.154370	1.749788	3.451810
Promedio	0.652904	0.862137	0.743219	1.007371	0.911423	1.452628	1.062701	1.683008	0.963760	3.386975	1.507299	3.643545	0.975191	3.306835	1.775214	3.570709

Caso I: Muestra sin valores atípicos

Caso II : 3% de atípicos en los datos de entrenamiento

Caso III: 3% de atípicos en los datos de validación

Caso IV: 3% de atípicos en los datos de entrenamiento y validación.

Anexo V.7

Error Cuadrado del Análisis de Regresión Múltiple – Datos atípicos contaminantes

Réplica	Muestra 40				Muestra 160				Muestra 1000				Muestra 3000			
	CASO				CASO				CASO				CASO			
	I	II	III	IV	I	II	III	IV	I	II	III	IV	I	II	III	IV
R1	0.435327	0.421572	0.447604	0.508739	0.874002	0.879050	0.865737	0.875007	0.954210	0.956080	0.952149	0.953606	0.983314	0.995289	0.981678	0.990741
R2	0.682184	0.671156	0.691710	0.708651	0.781346	0.842300	0.776870	0.838869	1.031938	1.055237	1.034325	1.061154	0.967536	1.000224	0.977039	1.012577
R3	0.748625	0.636695	0.778742	0.619414	0.861210	0.857348	0.862444	0.909287	0.930762	0.967604	0.952144	0.967309	0.925508	0.973847	0.934769	0.984246
R4	0.470419	0.589307	0.518714	0.551418	0.842234	0.865813	0.841621	0.865577	0.930035	0.947232	0.940621	0.948868	0.950448	0.989169	0.960272	0.996937
R5	0.826863	0.947124	0.827782	1.000784	0.986227	1.012400	0.989183	1.066549	1.054060	1.088149	1.054698	1.101415	0.961247	1.009853	0.972593	1.017601
R6	0.873401	0.876755	0.882540	0.886317	1.052675	1.107690	1.050311	1.117032	0.950476	0.971918	0.952378	0.969659	0.969532	0.988286	0.971821	1.000831
R7	0.627383	0.629119	0.639729	0.684391	0.925303	0.929162	0.915949	0.917470	0.879239	0.916728	0.892360	0.932768	1.034770	1.061318	1.037679	1.076142
R8	0.572206	0.651962	0.469792	0.693973	1.021549	1.000768	1.018357	1.008742	1.020199	1.055807	1.039858	1.060330	0.985834	1.011183	0.991079	1.039170
R9	0.812378	0.824868	0.651842	0.721307	0.806374	0.846468	0.806296	0.840274	0.899348	0.886940	0.927714	0.890352	0.997989	1.012989	1.008751	1.012194
R10	0.480258	0.489637	0.486585	0.650350	0.963308	0.968189	0.969678	0.967381	0.987328	1.014644	0.983025	1.017931	0.975737	1.006147	0.983881	1.010939
Promedio	0.652904	0.673820	0.639504	0.702534	0.911423	0.930919	0.909645	0.940619	0.963760	0.986034	0.972927	0.990339	0.975191	1.004830	0.981956	1.014138

Caso I: Muestra sin valores atípicos

Caso II : 3% de atípicos en los datos de entrenamiento

Caso III: 3% de atípicos en los datos de validación

Caso IV: 3% de atípicos en los datos de entrenamiento y validación.

Anexo V.8

Error Cuadrado del Análisis de Regresión Múltiple (Generalización) – Datos atípicos discordantes

Réplica	Muestra 40				Muestra 160				Muestra 1000				Muestra 3000			
	CASO				CASO				CASO				CASO			
	I	II	III	IV	I	II	III	IV	I	II	III	IV	I	II	III	IV
R1	0.811358	0.843794	0.966334	1.162947	1.031965	1.475843	1.273126	1.705271	0.872769	2.568393	1.192556	2.930979	0.963637	2.440804	1.183324	2.717803
R2	2.206373	3.142463	2.541916	3.365798	0.788889	1.146735	0.965077	1.321667	0.996350	2.552934	1.053812	2.588971	0.995946	2.105483	1.171558	2.331624
R3	1.376689	1.896613	1.680778	2.245868	1.041030	1.303174	1.156237	1.137400	1.245200	2.327245	1.549839	2.641326	0.976360	2.423580	1.181529	2.718773
R4	2.013916	2.667655	2.279046	2.824559	1.126047	1.870881	1.125123	1.898134	1.098040	2.211569	1.147799	2.381368	0.931670	2.040346	1.059757	2.268874
R5	0.304337	0.530224	0.342292	0.390219	1.176566	1.439920	1.263907	1.968919	0.839382	1.947101	1.068536	2.305779	1.078676	2.299735	1.190371	2.502850
R6	0.501319	2.280516	0.394214	1.322377	1.023018	1.952200	1.701350	2.019464	1.058639	3.236829	1.331414	3.589988	0.941934	2.196465	1.235272	2.460910
R7	0.349635	0.527469	0.319956	0.464001	1.295223	2.291644	1.262016	2.355079	0.930146	2.255068	1.057830	2.563430	0.877081	1.661640	0.994821	1.852212
R8	0.659568	2.073521	1.114228	2.248335	1.418977	1.694583	1.663722	2.681531	0.749370	2.319348	1.031304	2.450222	0.947541	2.052912	1.141119	2.249217
R9	1.369333	1.425572	3.037656	3.501119	0.835341	0.948485	1.327400	2.032935	0.977098	2.190077	1.322406	2.609049	0.841343	1.908057	1.040955	2.193070
R10	0.666937	0.656641	0.946633	0.499630	1.006588	0.958791	0.856230	1.266121	1.051148	2.405740	1.255826	2.588653	1.018239	1.845237	1.139834	2.082146
Promedio	1.025947	1.604447	1.362305	1.802485	1.074364	1.508226	1.259419	1.838652	0.981814	2.401430	1.201132	2.664976	0.957243	2.097426	1.133854	2.337748

Caso I: Muestra sin valores atípicos

Caso II : 3% de atípicos en los datos de entrenamiento

Caso III: 3% de atípicos en los datos de validación

Caso IV: 3% de atípicos en los datos de entrenamiento y validación.

Anexo V.9

Error Cuadrado del Análisis de Regresión Múltiple (Generalización) – Datos atípicos contaminantes

Réplica	Muestra 40				Muestra 160				Muestra 1000				Muestra 3000		
	CASO				CASO				CASO				CASO		
	I	II	III	IV	I	II	III	IV	I	II	III	IV	I	II	III
R1	0.811358	0.809534	0.738947	0.907729	1.031965	1.043149	1.030974	1.041878	0.872769	0.870228	0.870197	0.869393	0.963637	0.963112	0.962764
R2	2.206373	2.090765	2.088886	2.328563	0.788889	0.773062	0.794737	0.773557	0.996350	0.995094	0.999919	0.999362	0.995946	0.994684	0.995428
R3	1.376689	1.900793	1.218727	2.199244	1.041030	0.959025	1.042639	0.919143	1.245200	1.239128	1.249918	1.242422	0.976360	0.980169	0.978112
R4	2.013916	1.499924	1.566783	1.558145	1.126047	1.121341	1.124135	1.122521	1.098040	1.104409	1.101933	1.108529	0.931670	0.934528	0.931894
R5	0.304337	0.561045	0.287991	0.551146	1.176566	1.176581	1.148089	1.186754	0.839382	0.840973	0.839571	0.839448	1.078676	1.072860	1.078431
R6	0.501319	0.534229	0.596658	0.565792	1.023018	0.994697	1.004994	0.998102	1.058639	1.060088	1.059039	1.059382	0.941934	0.945161	0.944010
R7	0.349635	0.339499	0.364293	0.369918	1.295223	1.291223	1.298210	1.294102	0.930146	0.934122	0.930396	0.935108	0.877081	0.877267	0.877112
R8	0.659568	0.740844	0.758337	0.736793	1.418977	1.418840	1.410376	1.423353	0.749370	0.771237	0.756068	0.770215	0.947541	0.946181	0.947383
R9	1.369333	1.469744	1.364404	1.595553	0.835341	0.837736	0.833561	0.840773	0.977098	0.985991	0.971585	0.984820	0.841343	0.840782	0.841259
R10	0.666937	0.665925	0.554597	0.368055	1.006588	0.991377	1.023752	0.987345	1.051148	1.056506	1.050765	1.058219	1.018239	1.016668	1.017858
Promedio	1.025947	1.061230	0.953962	1.118094	1.074364	1.060703	1.071147	1.058753	0.981814	0.985778	0.982939	0.986690	0.957243	0.957141	0.957425

Caso I: Muestra sin valores atípicos

Caso II : 3% de atípicos en los datos de entrenamiento

Caso III: 3% de atípicos en los datos de validación

Caso IV: 3% de atípicos en los datos de entrenamiento y validación.

Anexo V.10

Error Cuadrado del Análisis de Redes Neuronales (Proceso de Aprendizaje) –Datos atípicos discordantes (Usando datos de Entrenamiento y Validación)

Réplica	Muestra 40				Muestra 160				Muestra 1000				Muestra 3000			
	CASO				CASO				CASO				CASO			
	I	II	III	IV	I	II	III	IV	I	II	III	IV	I	II	III	IV
R1	0.447705	0.466495	0.783857	1.328307	0.883204	1.664618	1.921234	2.240099	0.955239	3.344191	1.684746	3.606667	0.983657	3.461423	1.906651	3.739671
R2	0.686246	1.305718	0.872380	1.471577	0.795835	1.300128	1.569571	1.828393	1.033185	3.655314	1.302757	3.695879	0.968190	3.326218	1.768272	3.616708
R3	0.772736	0.712646	0.907744	0.635216	0.869015	1.142542	1.265048	1.749017	0.931692	3.113065	1.722075	3.522852	0.925684	3.611776	1.972439	3.872913
R4	0.509731	0.538985	1.356630	1.208935	0.847378	1.522779	0.852575	1.556331	0.931565	3.325242	1.437425	3.540307	0.951242	3.192332	1.705340	3.403247
R5	0.852931	1.078556	1.090405	1.911634	0.993668	1.669214	1.469807	2.553271	1.056725	3.244927	1.824499	3.710973	0.961750	3.356232	1.708479	3.562221
R6	0.964150	1.837907	1.153980	2.381589	1.055826	1.373542	1.792533	1.388909	0.951592	3.794788	1.910440	4.036088	0.970245	3.285226	1.849916	3.551119
R7	0.763508	0.760704	0.901593	1.187839	0.930086	2.377127	0.931375	2.435287	0.881588	3.202356	1.460696	3.560186	1.035136	3.031355	1.652670	3.326269
R8	0.586629	1.281715	1.326709	2.505544	1.035645	1.420225	1.221886	2.003569	1.020616	3.480975	1.719050	3.617234	0.986224	3.325093	1.818507	3.528123
R9	0.851696	0.871584	2.766541	2.335566	0.808453	0.869370	1.595452	2.563138	0.899838	3.073595	1.872442	3.524136	0.998767	3.377169	2.134044	3.723107
R10	0.560290	0.575357	1.295820	1.383454	0.974390	1.305223	1.030033	1.440321	0.988977	3.725072	1.834159	3.950782	0.976229	3.165896	1.813446	3.454919
Promedio	0.699562	0.942967	1.245566	1.634966	0.919350	1.464477	1.364951	1.975833	0.965102	3.395952	1.676829	3.676510	0.975712	3.313272	1.832976	3.577830

Caso I: Muestra sin valores atípicos

Caso II : 3% de atípicos en los datos de entrenamiento

Caso III: 3% de atípicos en los datos de validación

Caso IV: 3% de atípicos en los datos de entrenamiento y validación.

Anexo V.11

Épocas del Análisis de Redes Neuronales (Proceso de Aprendizaje) – Datos atípicos discordantes (Usando datos de Entrenamiento y Validación)

Réplica	Muestra 40				Muestra 160				Muestra 1000				Muestra 3000			
	CASO				CASO				CASO				CASO			
	I	II	III	IV	I	II	III	IV	I	II	III	IV	I	II	III	IV
R1	6.00	6.00	7.00	7.00	7.00	5.00	7.00	5.00	8.00	8.00	8.00	7.00	6.00	8.00	7.00	4.00
R2	8.00	8.00	9.00	7.00	5.00	8.00	8.00	6.00	6.00	7.00	7.00	8.00	9.00	7.00	4.00	8.00
R3	6.00	6.00	6.00	8.00	9.00	8.00	6.00	7.00	6.00	6.00	8.00	9.00	8.00	8.00	9.00	10.00
R4	8.00	7.00	6.00	7.00	5.00	6.00	7.00	5.00	7.00	5.00	7.00	6.00	10.00	7.00	6.00	7.00
R5	7.00	7.00	7.00	7.00	6.00	6.00	8.00	6.00	6.00	10.00	9.00	8.00	7.00	6.00	8.00	7.00
R6	6.00	9.00	8.00	10.00	7.00	5.00	8.00	7.00	5.00	8.00	8.00	7.00	6.00	4.00	7.00	7.00
R7	10.00	11.00	5.00	9.00	7.00	4.00	8.00	8.00	8.00	8.00	9.00	9.00	6.00	7.00	13.00	7.00
R8	4.00	9.00	7.00	7.00	5.00	5.00	8.00	7.00	5.00	8.00	9.00	8.00	9.00	9.00	8.00	6.00
R9	6.00	7.00	6.00	6.00	6.00	7.00	8.00	7.00	8.00	5.00	7.00	9.00	7.00	5.00	7.00	7.00
R10	9.00	6.00	7.00	6.00	7.00	6.00	6.00	8.00	7.00	6.00	4.00	7.00	7.00	7.00	5.00	8.00
Promedio	7.00	7.60	6.80	7.40	6.40	6.00	7.40	6.60	6.60	7.10	7.60	7.80	7.50	6.80	7.40	7.10

Caso I: Muestra sin valores atípicos

Caso II : 3% de atípicos en los datos de entrenamiento

Caso III: 3% de atípicos en los datos de validación

Caso IV: 3% de atípicos en los datos de entrenamiento y validación.

Anexo V.12

Tiempo en segundos del Análisis de Redes Neuronales (Proceso de Aprendizaje) – Datos atípicos discordantes (Usando datos de Entrenamiento y Validación)

Réplica	Muestra 40				Muestra 160				Muestra 1000				Muestra 3000			
	CASO				CASO				CASO				CASO			
	I	II	III	IV	I	II	III	IV	I	II	III	IV	I	II	III	IV
R1	14.72	9.73	8.50	7.89	14.89	9.36	11.56	13.72	23.81	13.47	15.24	16.72	17.69	18.13	17.70	17.91
R2	6.00	6.33	6.14	6.36	9.56	6.70	9.27	6.58	11.86	10.58	10.56	10.02	15.84	17.67	14.55	18.34
R3	5.97	6.92	6.59	6.78	7.03	6.55	6.64	6.56	11.33	11.70	10.08	10.89	15.84	17.78	15.02	18.44
R4	6.47	5.97	5.78	6.08	6.44	7.34	6.38	7.75	11.11	10.22	10.24	9.81	16.41	15.36	15.39	16.11
R5	6.06	6.16	5.91	6.24	7.22	6.61	6.38	6.52	10.94	15.67	9.49	9.97	16.00	16.34	15.28	17.23
R6	6.08	6.61	6.05	7.70	7.06	6.75	6.56	6.83	11.17	11.02	9.09	11.09	14.58	15.63	14.44	16.42
R7	6.03	6.03	5.91	6.14	6.56	6.95	6.75	7.03	9.92	11.13	10.16	10.73	17.72	16.80	15.75	16.59
R8	6.22	6.16	6.05	6.13	6.56	6.47	6.42	6.47	11.17	10.64	10.59	10.67	23.22	16.08	15.05	15.80
R9	5.98	6.19	5.81	6.20	6.59	6.64	6.47	6.92	9.89	11.44	9.25	11.11	15.58	17.28	15.19	18.13
R10	6.86	6.31	6.08	6.55	6.75	6.64	6.89	6.61	11.19	11.53	9.92	10.80	15.91	16.66	14.08	16.63
Promedio	7.04	6.64	6.28	6.61	7.87	7.00	7.33	7.50	12.24	11.74	10.46	11.18	16.88	16.77	15.24	17.16

Caso I: Muestra sin valores atípicos

Caso II : 3% de atípicos en los datos de entrenamiento

Caso III: 3% de atípicos en los datos de validación

Caso IV: 3% de atípicos en los datos de entrenamiento y validación.

Anexo V.13

Error Cuadrado Medio de Redes Neuronales (Generalización) – Datos atípicos discordantes (Usando datos de Entrenamiento y Validación)

Réplica	Muestra 40				Muestra 160				Muestra 1000				Muestra 3000			
	CASO				CASO				CASO				CASO			
	I	II	III	IV	I	II	III	IV	I	II	III	IV	I	II	III	IV
R1	0.690008	0.809374	0.591228	0.780677	1.044187	1.442988	1.752256	1.856313	0.869172	2.716689	1.072551	2.810734	0.964604	2.417664	1.307527	2.767467
R2	2.257028	2.534055	1.719428	3.151891	0.782831	1.152945	0.912079	1.175825	1.006119	2.608887	1.027590	2.584305	0.998328	2.128443	1.098822	2.303078
R3	1.395378	1.558671	1.979259	2.080844	1.002583	1.371035	1.018500	1.394750	1.252370	2.400047	1.643301	2.536113	0.976771	2.416191	1.250941	2.716661
R4	2.217116	2.828814	3.728678	1.917352	1.089211	1.984865	1.118257	2.051077	1.093978	2.124997	1.146766	2.447615	0.933077	2.218158	1.068924	2.220269
R5	0.222840	0.498599	0.345709	1.133166	1.257443	1.469907	1.200566	1.639088	0.841530	2.119828	0.909669	2.226069	1.080090	2.221778	1.272452	2.543205
R6	0.503438	3.030835	0.289948	1.378543	1.029985	2.011593	1.316691	2.032712	1.067511	3.228592	1.277251	3.538068	0.945800	2.248290	1.231899	2.463074
R7	0.178049	0.641361	0.245131	0.417353	1.286929	2.325054	1.243560	2.417856	0.937220	2.308378	1.097422	2.534626	0.876775	1.740135	0.965433	1.816291
R8	0.624534	2.141778	1.288896	2.739964	1.480991	1.719816	1.727919	2.344222	0.743859	2.432265	0.850782	2.501300	0.950062	2.022865	1.127898	2.300080
R9	1.575351	1.452012	5.541357	3.817014	0.838049	0.936927	1.865576	3.100598	0.977733	2.226351	1.239967	2.744426	0.841842	1.958169	1.008550	2.156542
R10	0.514756	0.464009	0.147906	0.468096	0.993279	0.973159	0.905827	1.276914	1.050855	2.432102	1.115824	2.594634	1.019665	1.927558	1.174280	2.085845
Promedio	1.017850	1.595951	1.587754	1.788490	1.080549	1.538829	1.306123	1.928936	0.984035	2.459814	1.138112	2.651789	0.958701	2.129925	1.150673	2.337251

Caso I: Muestra sin valores atípicos

Caso II : 3% de atípicos en los datos de entrenamiento

Caso III: 3% de atípicos en los datos de validación

Caso IV: 3% de atípicos en los datos de entrenamiento y validación.

Anexo V.14

Error Cuadrado del Análisis de Redes Neuronales (Proceso de Aprendizaje) – Datos atípicos discordantes (Uniando datos de validación y entrenamiento)

Réplica	Muestra 40				Muestra 160				Muestra 1000				Muestra 3000			
	CASO				CASO				CASO				CASO			
	I	II	III	IV	I	II	III	IV	I	II	III	IV	I	II	III	IV
R1	0.435327	0.451369	0.558760	0.881707	0.874002	1.649750	1.295769	1.844707	0.954210	3.335091	1.466977	3.568112	0.983314	3.455395	1.808253	3.728327
R2	0.682184	1.263398	0.710629	1.275333	0.781346	1.293751	0.944827	1.529118	1.031938	3.646719	1.260721	3.689143	0.967536	3.322899	1.730072	3.611876
R3	0.748625	0.678051	0.739069	0.577076	0.861210	1.135803	0.985742	1.357833	0.930762	3.102127	1.426170	3.477356	0.925508	3.610115	1.900868	3.865444
R4	0.470419	0.483077	0.636589	0.614874	0.842234	1.514529	0.841517	1.520591	0.930035	3.313896	1.360340	3.505437	0.950448	3.179467	1.661704	3.399959
R5	0.826863	1.020653	0.931421	1.127133	0.986227	1.665239	1.267068	1.969059	1.054060	3.231441	1.630107	3.636813	0.961247	3.346484	1.662791	3.556728
R6	0.873401	1.503649	0.886430	1.713929	1.052675	1.368892	1.222125	1.374508	0.950476	3.790033	1.665848	4.012278	0.969532	3.283159	1.811097	3.541559
R7	0.627383	0.692626	0.686700	0.744436	0.925303	2.358056	0.925569	2.367836	0.879239	3.188337	1.360953	3.513012	1.034770	3.026884	1.609559	3.311108
R8	0.572206	1.226895	0.669926	1.317374	1.021551	1.402644	1.114975	1.755425	1.020199	3.473102	1.472954	3.596926	0.985834	3.317190	1.745498	3.520355
R9	0.812378	0.822147	0.983131	1.033173	0.806374	0.864760	1.040568	1.757674	0.899348	3.067664	1.777622	3.497022	0.997989	3.372384	2.072509	3.719922
R10	0.480259	0.479510	0.629536	0.788672	0.963308	1.272856	0.988851	1.353331	0.987328	3.721336	1.651299	3.939351	0.975737	3.154370	1.749789	3.451810
Promedio	0.652904	0.862137	0.743219	1.007371	0.911423	1.452628	1.062701	1.683008	0.963760	3.386975	1.507299	3.643545	0.975192	3.306835	1.775214	3.570709

Caso I: Muestra sin valores atípicos

Caso II : 3% de atípicos en los datos de entrenamiento

Caso III: 3% de atípicos en los datos de validación

Caso IV: 3% de atípicos en los datos de entrenamiento y validación.

Anexo V.15

Épocas del Análisis de Redes Neuronales (Proceso de Aprendizaje) – Datos atípicos discordantes (Uniando datos de validación y entrenamiento)

Réplica	Muestra 40				Muestra 160				Muestra 1000				Muestra 3000			
	CASO				CASO				CASO				CASO			
	I	II	III	IV	I	II	III	IV	I	II	III	IV	I	II	III	IV
R1	6.00	10.00	5.00	8.00	10.00	8.00	8.00	8.00	10.00	7.00	9.00	9.00	12.00	12.00	7.00	10.00
R2	8.00	5.00	6.00	10.00	7.00	10.00	10.00	8.00	7.00	10.00	10.00	12.00	9.00	12.00	10.00	13.00
R3	9.00	5.00	5.00	9.00	12.00	8.00	7.00	9.00	12.00	10.00	8.00	9.00	8.00	7.00	10.00	8.00
R4	6.00	6.00	8.00	7.00	7.00	5.00	8.00	11.00	8.00	10.00	10.00	12.00	9.00	8.00	6.00	8.00
R5	7.00	8.00	5.00	7.00	8.00	8.00	7.00	9.00	7.00	7.00	8.00	10.00	10.00	9.00	10.00	7.00
R6	10.00	5.00	5.00	8.00	4.00	8.00	10.00	8.00	8.00	9.00	8.00	9.00	13.00	9.00	6.00	11.00
R7	7.00	5.00	8.00	7.00	10.00	12.00	7.00	8.00	10.00	8.00	11.00	9.00	10.00	10.00	8.00	8.00
R8	9.00	7.00	8.00	10.00	5.00	10.00	8.00	9.00	9.00	7.00	10.00	5.00	10.00	10.00	11.00	8.00
R9	9.00	10.00	7.00	5.00	11.00	8.00	7.00	6.00	14.00	7.00	10.00	12.00	12.00	9.00	11.00	14.00
R10	7.00	6.00	8.00	7.00	10.00	7.00	5.00	9.00	10.00	9.00	10.00	11.00	10.00	9.00	11.00	8.00
Promedio	7.80	6.70	6.50	7.80	8.40	8.40	7.70	8.50	9.50	8.40	9.40	9.80	10.30	9.50	9.00	9.50

Caso I: Muestra sin valores atípicos

Caso II : 3% de atípicos en los datos de entrenamiento

Caso III: 3% de atípicos en los datos de validación

Caso IV: 3% de atípicos en los datos de entrenamiento y validación.

Anexo V.16

Tiempo del Análisis de Redes Neuronales (Proceso de Aprendizaje) – Datos atípicos discordantes (Uniando datos de validación y entrenamiento)

Réplica	Muestra 40				Muestra 160				Muestra 1000				Muestra 3000			
	CASO				CASO				CASO				CASO			
	I	II	III	IV	I	II	III	IV	I	II	III	IV	I	II	III	IV
R1	12.47	14.44	10.58	11.08	14.88	12.31	13.22	15.34	12.56	16.30	16.25	16.95	20.30	21.98	21.36	21.58
R2	5.17	5.25	5.86	5.45	6.34	6.31	7.49	5.58	10.20	10.41	9.78	10.05	15.59	17.61	18.48	20.44
R3	5.44	5.56	5.19	5.33	6.61	5.69	5.50	6.47	9.72	11.33	10.17	10.59	15.88	17.02	19.36	18.30
R4	5.13	5.55	5.45	5.31	5.56	5.92	5.64	5.97	9.06	10.22	9.69	10.24	15.47	20.53	17.13	19.89
R5	5.08	5.39	5.41	5.72	5.58	5.55	5.63	5.70	9.06	10.25	9.11	9.38	17.44	20.19	17.99	17.22
R6	5.39	5.56	5.28	5.27	5.69	5.52	5.88	5.83	9.00	9.14	8.91	10.53	18.25	19.30	19.42	19.74
R7	5.06	5.36	5.58	5.56	5.50	6.22	5.52	5.80	9.95	9.17	9.64	9.95	16.39	17.92	17.50	18.94
R8	5.05	5.28	5.38	5.45	6.41	6.23	6.02	6.09	9.58	9.81	8.89	10.33	17.59	18.11	20.09	16.55
R9	5.34	5.34	5.24	5.34	6.25	6.53	5.72	5.88	8.94	10.19	8.75	9.81	16.08	18.99	17.69	18.25
R10	5.14	5.66	5.11	5.53	6.17	5.70	5.48	5.95	8.67	9.91	10.84	10.44	17.56	20.02	18.22	18.41
Promedio	5.93	6.34	5.91	6.00	6.90	6.60	6.61	6.86	9.67	10.67	10.20	10.83	17.05	19.17	18.72	18.93

Caso I: Muestra sin valores atípicos

Caso II : 3% de atípicos en los datos de entrenamiento

Caso III: 3% de atípicos en los datos de validación

Caso IV: 3% de atípicos en los datos de entrenamiento y validación.

Anexo V.17

Error Cuadrado Medio de Redes Neuronales (Generalización) – Datos atípicos discordantes (Uniando datos de validación y entrenamiento)

Réplica	Muestra 40				Muestra 160				Muestra 1000				Muestra 3000			
	CASO				CASO				CASO				CASO			
	I	II	III	IV	I	II	III	IV	I	II	III	IV	I	II	III	IV
R1	0.811443	0.843822	0.966266	1.162894	1.031971	1.475896	1.273114	1.705075	0.872773	2.568744	1.192407	2.930752	0.963644	2.440779	1.183406	2.718439
R2	2.206998	3.143260	2.542906	3.365566	0.788957	1.146744	0.965117	1.321619	0.996349	2.553099	1.053773	2.588493	0.995945	2.105342	1.171598	2.331051
R3	1.377239	1.896578	1.681546	2.245753	1.041002	1.303214	1.156313	1.137325	1.245177	2.327327	1.549826	2.641462	0.976368	2.423618	1.181429	2.718459
R4	2.013838	2.667407	2.278827	2.824934	1.126092	1.870585	1.125050	1.898311	1.098041	2.211553	1.147684	2.381194	0.931674	2.040526	1.059886	2.269863
R5	0.304363	0.530437	0.342282	0.390233	1.176506	1.439868	1.263899	1.968930	0.839387	1.947128	1.068595	2.305036	1.078664	2.299610	1.190382	2.502797
R6	0.501302	2.280458	0.394186	1.322340	1.022962	1.952226	1.701500	2.019324	1.058583	3.236799	1.331572	3.590577	0.941986	2.196743	1.234974	2.461337
R7	0.349640	0.527484	0.319960	0.463954	1.295175	2.291703	1.262008	2.355159	0.930147	2.254862	1.057742	2.564151	0.877085	1.662015	0.994710	1.852005
R8	0.659631	2.073680	1.114097	2.248503	1.418424	1.694661	1.663734	2.681648	0.749528	2.319060	1.031279	2.450174	0.947532	2.053038	1.141231	2.248962
R9	1.369370	1.425473	3.037522	3.501083	0.835319	0.948391	1.327367	2.032993	0.976965	2.190051	1.322178	2.609327	0.841342	1.907937	1.041086	2.193390
R10	0.667016	0.656671	0.946756	0.499334	1.006627	0.958750	0.856270	1.266130	1.051182	2.406015	1.255851	2.589033	1.018240	1.845573	1.139720	2.081945
Promedio	1.026084	1.604527	1.362435	1.802459	1.074303	1.508204	1.259437	1.838651	0.981813	2.401464	1.201091	2.665020	0.957248	2.097518	1.133842	2.337825

Caso I: Muestra sin valores atípicos

Caso II : 3% de atípicos en los datos de entrenamiento

Caso III: 3% de atípicos en los datos de validación

Caso IV: 3% de atípicos en los datos de entrenamiento y validación.

Anexo V.18

Error Cuadrado del Análisis de Redes Neuronales (Proceso de Aprendizaje) – Datos contaminantes (Usando datos de Entrenamiento y Validación)

Réplica	Muestra 40				Muestra 160				Muestra 1000				Muestra 3000			
	CASO				CASO				CASO				CASO			
	I	II	III	IV	I	II	III	IV	I	II	III	IV	I	II	III	IV
R1	0.447705	0.431574	0.465329	0.671696	0.883204	0.890828	0.875397	0.880256	0.955239	0.957267	0.953448	0.954552	0.983657	0.995734	0.982062	0.991011
R2	0.686246	0.676125	0.701210	0.723778	0.795835	0.850009	0.794475	0.855399	1.033185	1.056365	1.035830	1.063111	0.968190	1.000776	0.977700	1.013024
R3	0.772736	0.660206	0.953661	0.737338	0.869015	0.865354	0.875700	0.920032	0.931692	0.968305	0.953766	0.967818	0.925684	0.974036	0.935242	0.984564
R4	0.509731	0.663945	0.621587	0.613453	0.847378	0.873128	0.848406	0.871677	0.931565	0.949239	0.941564	0.950138	0.951242	0.989958	0.961026	0.997719
R5	0.852931	0.996543	0.858054	1.137141	0.993668	1.019582	0.996913	1.070302	1.056725	1.090981	1.056738	1.104011	0.961750	1.010477	0.973055	1.018147
R6	0.964150	1.014421	0.996012	0.977578	1.055826	1.110455	1.051849	1.119635	0.951592	0.972936	0.953853	0.970717	0.970245	0.988856	0.972209	1.001372
R7	0.763508	0.807917	0.694071	0.796710	0.930086	0.933085	0.920334	0.921307	0.881588	0.918993	0.894651	0.933682	1.035136	1.061653	1.037934	1.076413
R8	0.586629	0.677607	0.481673	0.804693	1.035645	1.014308	1.032627	1.023393	1.020616	1.056032	1.041232	1.061131	0.986224	1.011482	0.991445	1.039625
R9	0.851696	0.857862	0.708991	0.796454	0.808453	0.850334	0.809246	0.842483	0.899838	0.887415	0.929646	0.890807	0.998767	1.014279	1.009670	1.012711
R10	0.560290	0.573131	0.526591	0.821302	0.974390	0.983635	0.984753	0.979408	0.988977	1.015757	0.986124	1.020773	0.976229	1.006879	0.984312	1.011463
Promedio	0.699562	0.735933	0.700718	0.808014	0.919350	0.939072	0.918970	0.948389	0.965102	0.987329	0.974685	0.991674	0.975712	1.005413	0.982466	1.014605

Caso I: Muestra sin valores atípicos

Caso II : 3% de atípicos en los datos de entrenamiento

Caso III: 3% de atípicos en los datos de validación

Caso IV: 3% de atípicos en los datos de entrenamiento y validación.

Anexo V.19

**Épocas del Análisis de Redes Neuronales (Proceso de Aprendizaje) – Datos atípicos contaminantes
(Usando datos de Entrenamiento y Validación)**

Réplica	Muestra 40				Muestra 160				Muestra 1000				Muestra 3000			
	CASO				CASO				CASO				CASO			
	I	II	III	IV	I	II	III	IV	I	II	III	IV	I	II	III	IV
R1	6.00	7.00	7.00	5.00	7.00	8.00	7.00	7.00	8.00	4.00	10.00	5.00	6.00	6.00	6.00	6.00
R2	8.00	6.00	7.00	5.00	5.00	6.00	8.00	7.00	6.00	6.00	8.00	5.00	9.00	7.00	9.00	7.00
R3	6.00	7.00	7.00	6.00	9.00	7.00	8.00	6.00	6.00	9.00	7.00	7.00	8.00	7.00	8.00	9.00
R4	8.00	6.00	7.00	7.00	5.00	7.00	5.00	8.00	7.00	7.00	5.00	9.00	10.00	7.00	10.00	10.00
R5	7.00	5.00	10.00	6.00	6.00	6.00	6.00	6.00	6.00	9.00	10.00	7.00	7.00	8.00	7.00	5.00
R6	6.00	5.00	8.00	7.00	7.00	7.00	4.00	6.00	5.00	5.00	7.00	5.00	6.00	7.00	6.00	7.00
R7	10.00	7.00	6.00	9.00	7.00	4.00	7.00	6.00	8.00	7.00	7.00	5.00	6.00	6.00	6.00	9.00
R8	4.00	6.00	7.00	6.00	5.00	8.00	8.00	8.00	5.00	5.00	8.00	5.00	9.00	5.00	10.00	8.00
R9	6.00	10.00	6.00	6.00	6.00	11.00	6.00	7.00	8.00	10.00	6.00	6.00	7.00	8.00	7.00	7.00
R10	9.00	8.00	7.00	5.00	7.00	7.00	7.00	5.00	7.00	7.00	9.00	8.00	7.00	7.00	7.00	8.00
Promedio	7.00	6.70	7.20	6.20	6.40	7.10	6.60	6.60	6.60	6.90	7.70	6.20	7.50	6.80	7.60	7.60

Caso I: Muestra sin valores atípicos

Caso II : 3% de atípicos en los datos de entrenamiento

Caso III: 3% de atípicos en los datos de validación

Caso IV: 3% de atípicos en los datos de entrenamiento y validación.

Anexo V.20

Tiempo en segundos del Análisis de Redes Neuronales (Proceso de Aprendizaje) – Datos atípicos contaminantes (Usando datos de Entrenamiento y Validación)

Réplica	Muestra 40				Muestra 160				Muestra 1000				Muestra 3000		
	CASO				CASO				CASO				CASO		
	I	II	III	IV	I	II	III	IV	I	II	III	IV	I	II	III
R1	14.72	11.72	9.05	10.66	14.89	12.11	10.69	11.33	23.81	21.17	16.72	14.89	17.69	18.67	14.59
R2	6.00	7.59	5.95	6.42	9.56	8.47	8.83	7.63	11.86	10.55	9.97	10.77	15.84	14.92	15.44
R3	5.97	6.05	5.88	5.95	7.03	6.81	6.83	6.81	11.33	10.61	10.20	10.05	15.84	16.89	24.30
R4	6.47	6.20	7.16	6.02	6.44	6.56	6.39	6.50	11.11	10.67	10.16	10.28	16.41	17.67	16.31
R5	6.06	6.22	5.95	6.06	7.22	6.64	6.84	6.73	10.94	10.61	10.17	9.58	16.00	15.11	15.89
R6	6.08	6.08	5.81	6.13	7.06	6.42	6.59	6.94	11.17	10.31	9.89	10.23	14.58	17.70	14.75
R7	6.03	5.94	6.25	6.03	6.56	6.59	6.50	6.48	9.92	9.91	9.52	9.73	17.72	16.56	17.72
R8	6.22	6.03	5.88	6.16	6.56	6.53	6.41	6.38	11.17	9.56	9.95	9.61	23.22	16.33	17.45
R9	5.98	6.06	5.91	5.91	6.59	6.41	6.44	6.53	9.89	11.08	9.55	10.91	15.58	15.28	15.44
R10	6.86	5.94	6.22	5.97	6.75	7.25	6.44	6.20	11.19	9.61	10.72	9.00	15.91	15.42	15.70
Promedio	7.04	6.78	6.40	6.53	7.87	7.38	7.20	7.15	12.24	11.41	10.68	10.50	16.88	16.46	16.76

Caso I: Muestra sin valores atípicos

Caso II : 3% de atípicos en los datos de entrenamiento

Caso III: 3% de atípicos en los datos de validación

Caso IV: 3% de atípicos en los datos de entrenamiento y validación.

Anexo V.21

Error Cuadrado Medio de Redes Neuronales (Generalización) – Datos atípicos contaminantes (Usando datos de Entrenamiento y Validación)

Réplica	Muestra 40				Muestra 160				Muestra 1000				Muestra 3000			
	CASO				CASO				CASO				CASO			
	I	II	III	IV	I	II	III	IV	I	II	III	IV	I	II	III	IV
R1	0.690008	0.729092	0.697972	0.814530	1.044187	1.045955	0.994692	1.043024	0.869172	0.869767	0.876446	0.875802	0.964604	0.961299	0.963459	0.961406
R2	2.257028	2.231525	2.269137	2.238888	0.782831	0.829057	0.821647	0.801277	1.006119	1.003833	0.995916	1.003258	0.998328	0.995567	0.998328	0.996293
R3	1.395378	1.781411	0.758233	1.744318	1.002583	0.919972	1.006461	0.931470	1.252370	1.241159	1.253608	1.242694	0.976771	0.980268	0.976771	0.982141
R4	2.217116	1.584950	1.720780	1.281352	1.089211	1.093721	1.126558	1.096899	1.093978	1.100501	1.107524	1.101036	0.933077	0.936473	0.933077	0.934931
R5	0.222840	0.490887	0.194121	0.248976	1.257443	1.232266	1.244687	1.227853	0.841530	0.846087	0.845756	0.843769	1.080090	1.074116	1.080090	1.073818
R6	0.503438	0.531332	0.595024	0.488074	1.029985	1.019429	1.031369	1.010029	1.067511	1.067538	1.056977	1.065874	0.945800	0.947365	0.945800	0.949044
R7	0.178049	0.550031	0.389570	0.228157	1.286929	1.276379	1.332918	1.310938	0.937220	0.940593	0.937524	0.942907	0.876775	0.877575	0.876775	0.877579
R8	0.624534	0.737891	0.826358	0.809912	1.480991	1.390974	1.478971	1.495132	0.743859	0.769613	0.749633	0.768273	0.950062	0.948255	0.950079	0.947662
R9	1.575351	1.292298	1.495024	1.497231	0.838049	0.822239	0.864549	0.848565	0.977733	0.988150	0.966030	0.991980	0.841842	0.839391	0.841842	0.840552
R10	0.514756	0.483866	0.400283	0.569185	0.993279	0.933707	0.980599	0.992816	1.050855	1.055090	1.048465	1.058162	1.019665	1.016841	1.019665	1.018066
Promedio	1.017850	1.041328	0.934650	0.992062	1.080549	1.056370	1.088245	1.075800	0.984035	0.988233	0.983788	0.989375	0.958701	0.957715	0.958589	0.958149

Caso I: Muestra sin valores atípicos

Caso II : 3% de atípicos en los datos de entrenamiento

Caso III: 3% de atípicos en los datos de validación

Caso IV: 3% de atípicos en los datos de entrenamiento y validación.

Anexo V.22

Error Cuadrado del Análisis de Redes Neuronales (Proceso de Aprendizaje) – Datos atípicos contaminantes (Uniando datos de validación y entrenamiento)

Réplica	Muestra 40				Muestra 160				Muestra 1000				Muestra 3000			
	CASO				CASO				CASO				CASO			
	I	II	III	IV	I	II	III	IV	I	II	III	IV	I	II	III	IV
R1	0.435327	0.421572	0.447604	0.508739	0.874002	0.879050	0.865737	0.875007	0.954210	0.956080	0.952149	0.953606	0.983314	0.995289	0.981679	0.990741
R2	0.682184	0.671156	0.691711	0.708651	0.781346	0.842300	0.776870	0.838869	1.031938	1.055237	1.034325	1.061154	0.967536	1.000224	0.977039	1.012577
R3	0.748625	0.636696	0.778742	0.619416	0.861210	0.857348	0.862444	0.909287	0.930762	0.967604	0.952144	0.967309	0.925508	0.973847	0.934769	0.984246
R4	0.470419	0.589308	0.518714	0.551418	0.842234	0.865813	0.841621	0.865577	0.930035	0.947233	0.940621	0.948868	0.950448	0.989169	0.960273	0.996937
R5	0.826863	0.947124	0.827782	1.000785	0.986227	1.012400	0.989183	1.066549	1.054060	1.088149	1.054699	1.101415	0.961247	1.009853	0.972594	1.017601
R6	0.873401	0.876755	0.882540	0.886317	1.052675	1.107690	1.050311	1.117032	0.950476	0.971918	0.952379	0.969659	0.969532	0.988286	0.971821	1.000831
R7	0.627383	0.629119	0.639729	0.684391	0.925303	0.929168	0.915949	0.917470	0.879239	0.916729	0.892360	0.932768	1.034770	1.061318	1.037679	1.076142
R8	0.572206	0.651962	0.469792	0.693973	1.021551	1.000768	1.018357	1.008742	1.020199	1.055807	1.039858	1.060330	0.985834	1.011183	0.991079	1.039170
R9	0.812378	0.824868	0.651842	0.721307	0.806374	0.846468	0.806296	0.840274	0.899348	0.886940	0.927714	0.890352	0.997989	1.012989	1.008751	1.012194
R10	0.480259	0.489637	0.486585	0.650350	0.963308	0.968189	0.969678	0.967381	0.987328	1.014644	0.983025	1.017931	0.975737	1.006147	0.983881	1.010939
Promedio	0.652904	0.673820	0.639504	0.702535	0.911423	0.930919	0.909645	0.940619	0.963760	0.986034	0.972927	0.990339	0.975192	1.004830	0.981957	1.014138

Caso I: Muestra sin valores atípicos

Caso II : 3% de atípicos en los datos de entrenamiento

Caso III: 3% de atípicos en los datos de validación

Caso IV: 3% de atípicos en los datos de entrenamiento y validación.

Anexo V.23

**Épocas del Análisis de Redes Neuronales (Proceso de Aprendizaje) – Datos atípicos contaminantes
(Uniando datos de validación y entrenamiento)**

Réplica	Muestra 40				Muestra 160				Muestra 1000				Muestra 3000			
	CASO				CASO				CASO				CASO			
	I	II	III	IV	I	II	III	IV	I	II	III	IV	I	II	III	IV
R1	6.00	5.00	5.00	11.00	10.00	8.00	8.00	7.00	10.00	10.00	12.00	10.00	12.00	9.00	9.00	10.00
R2	8.00	5.00	7.00	6.00	7.00	7.00	7.00	10.00	7.00	6.00	10.00	10.00	9.00	8.00	8.00	9.00
R3	9.00	6.00	8.00	5.00	12.00	10.00	13.00	11.00	12.00	10.00	11.00	11.00	8.00	8.00	10.00	8.00
R4	6.00	6.00	7.00	8.00	7.00	6.00	10.00	7.00	8.00	7.00	8.00	11.00	9.00	10.00	10.00	10.00
R5	7.00	6.00	9.00	9.00	8.00	6.00	9.00	4.00	7.00	7.00	9.00	9.00	10.00	9.00	11.00	13.00
R6	10.00	5.00	8.00	6.00	4.00	11.00	12.00	8.00	8.00	9.00	7.00	14.00	13.00	11.00	10.00	9.00
R7	7.00	8.00	9.00	6.00	10.00	5.00	11.00	8.00	10.00	5.00	10.00	9.00	10.00	10.00	11.00	8.00
R8	9.00	7.00	10.00	8.00	5.00	9.00	7.00	7.00	9.00	7.00	10.00	10.00	10.00	15.00	10.00	10.00
R9	9.00	9.00	6.00	8.00	11.00	8.00	7.00	7.00	14.00	9.00	10.00	8.00	12.00	12.00	14.00	9.00
R10	7.00	5.00	6.00	6.00	10.00	10.00	5.00	4.00	10.00	10.00	10.00	9.00	10.00	13.00	7.00	14.00
Promedio	7.80	6.20	7.50	7.30	8.40	8.00	8.90	7.30	9.50	8.00	9.70	10.10	10.30	10.50	10.00	10.00

Caso I: Muestra sin valores atípicos

Caso II : 3% de atípicos en los datos de entrenamiento

Caso III: 3% de atípicos en los datos de validación

Caso IV: 3% de atípicos en los datos de entrenamiento y validación.

Anexo V.24

Tiempo del Análisis de Redes Neuronales (Proceso de Aprendizaje) – Datos atípicos contaminantes (Uniando datos de validación y entrenamiento)

Réplica	Muestra 40				Muestra 160				Muestra 1000				Muestra 3000			
	CASO				CASO				CASO				CASO			
	I	II	III	IV	I	II	III	IV	I	II	III	IV	I	II	III	IV
R1	12.47	10.50	13.47	12.58	14.88	9.59	13.59	12.27	12.56	11.38	18.17	17.33	20.30	25.45	20.52	16.70
R2	5.17	5.80	5.97	5.13	6.34	6.78	6.45	7.08	10.20	9.50	11.88	9.75	15.59	16.42	15.74	18.70
R3	5.44	4.94	5.22	4.97	6.61	5.77	6.23	5.66	9.72	10.11	10.61	11.11	15.88	17.67	20.05	18.63
R4	5.13	5.22	6.63	5.80	5.56	5.83	7.44	5.38	9.06	9.86	9.55	11.03	15.47	16.50	21.42	17.13
R5	5.08	5.17	5.13	5.28	5.58	6.06	5.67	5.52	9.06	10.36	9.91	10.25	17.44	16.33	20.45	16.13
R6	5.39	4.77	4.89	5.05	5.69	6.00	6.50	5.98	9.00	9.86	9.98	10.72	18.25	18.08	21.52	16.45
R7	5.06	5.14	5.55	5.19	5.50	5.63	5.78	5.52	9.95	9.22	8.41	10.69	16.39	18.91	20.22	19.86
R8	5.05	5.22	4.89	5.58	6.41	6.45	5.86	5.77	9.58	10.41	9.91	9.42	17.59	19.52	21.45	18.38
R9	5.34	5.05	4.99	5.34	6.25	5.81	5.50	6.05	8.94	10.38	9.67	10.20	16.08	18.20	22.42	18.55
R10	5.14	4.86	5.19	5.24	6.17	5.52	5.52	5.66	8.67	10.67	9.78	9.70	17.56	18.44	21.81	19.00
Promedio	5.93	5.67	6.19	6.01	6.90	6.34	6.85	6.49	9.67	10.17	10.79	11.02	17.05	18.55	20.56	17.95

Caso I: Muestra sin valores atípicos

Caso II : 3% de atípicos en los datos de entrenamiento

Caso III: 3% de atípicos en los datos de validación

Caso IV: 3% de atípicos en los datos de entrenamiento y validación.

Anexo V.25

**Error Cuadrado Medio de Redes Neuronales (Generalización) – Datos atípicos contaminantes
(Uniendo datos de validación y entrenamiento)**

Réplica	Muestra 40				Muestra 160				Muestra 1000				Muestra 3000			
	CASO				CASO				CASO				CASO			
	I	II	III	IV	I	II	III	IV	I	II	III	IV	I	II	III	IV
R1	0.811443	0.809648	0.739067	0.907601	1.031971	1.043203	1.030907	1.041793	0.872773	0.870236	0.870149	0.869412	0.963644	0.963137	0.962792	0.962178
R2	2.206998	2.090339	2.086491	2.328294	0.788957	0.772936	0.794689	0.773584	0.996349	0.995096	0.999914	0.999353	0.995945	0.994668	0.995445	0.995060
R3	1.377239	1.900909	1.219168	2.203209	1.041002	0.959045	1.042557	0.919146	1.245177	1.239173	1.249874	1.242470	0.976368	0.980175	0.978133	0.981515
R4	2.013838	1.500403	1.566737	1.557686	1.126092	1.121397	1.124058	1.122515	1.098041	1.104399	1.101931	1.108542	0.931674	0.934526	0.931914	0.934717
R5	0.304363	0.560968	0.288009	0.550829	1.176506	1.176564	1.148167	1.186910	0.839387	0.840957	0.839580	0.839444	1.078664	1.072875	1.078352	1.073361
R6	0.501302	0.534193	0.596677	0.565780	1.022962	0.994671	1.004953	0.998180	1.058583	1.060058	1.059109	1.059351	0.941986	0.945115	0.943963	0.947975
R7	0.349640	0.339519	0.364285	0.369914	1.295175	1.291985	1.298164	1.294072	0.930147	0.934140	0.930404	0.935118	0.877085	0.877275	0.877116	0.877290
R8	0.659631	0.740832	0.758409	0.737146	1.418424	1.418530	1.410401	1.423443	0.749528	0.771253	0.756063	0.770239	0.947532	0.946232	0.947366	0.946129
R9	1.369370	1.469658	1.364346	1.595653	0.835319	0.837783	0.833619	0.840768	0.976965	0.985985	0.971567	0.984790	0.841342	0.840778	0.841263	0.841185
R10	0.667016	0.666015	0.554524	0.368088	1.006627	0.991373	1.023810	0.987331	1.051182	1.056472	1.050730	1.058221	1.018240	1.016656	1.017859	1.015592
Promedio	1.026084	1.061248	0.953771	1.118420	1.074303	1.060749	1.071133	1.058774	0.981813	0.985777	0.982932	0.986694	0.957248	0.957144	0.957420	0.957500

Caso I: Muestra sin valores atípicos

Caso II : 3% de atípicos en los datos de entrenamiento

Caso III: 3% de atípicos en los datos de validación

Caso IV: 3% de atípicos en los datos de entrenamiento y validación.