EFECTOS DE LA ESTRUCTURA ESTADÍSTICA DE LOS DATOS EN LA IMPLEMENTACIÓN DE LA RED NEURONAL AUTOSUPERVISADA

EFFECTS OF STATISTICAL DATA STRUCTURE ON THE IMPLEMENTATION OF A SELF-SUPERVISED NEURAL NETWORK

Luis E. Huamanchumo de la Cuba[1], Luis A. Sánchez Alvarado[2]

RESUMEN

La presente investigación plantea como objetivo estudiar aspectos técnicos relacionados con la implementación de la red neuronal de Análisis de Componentes Principales (ACP) en términos de su capacidad predictiva, generalización y precisión con el fin de establecer criterios óptimos para su validación, evaluación del desempeño e implementación. Para ello, se plantea la hipótesis de que la estructura estadística de los datos influye significativamente en el óptimo desempeño de la red neuronal de ACP en el contexto no supervisado. Se demostró que el algoritmo Hebbiano de la fase de aprendizaje garantiza la calidad de representación de la red debido a que capitaliza eficientemente la información en escenarios con varianza generalizada grande.

Palabras clave.- Análisis de componentes principales, Algoritmo hebbiano, Reducción de dimensionalidad.

ABSTRACT

The purpose of this research is to study technical aspects involved in the implementation of a Principal Component Analysis (PCA) neural network in terms of predictive capacity, generalization and accuracy in order to establish optimal criteria for the validation and implementation thereof. Our hypothesis is that the statistical structure of the data affects the optimal performance of a PCA neural network in the unsupervised context. It was demonstrated that the Hebbian algorithm at the learning phase ensures enhanced quality of network representation as it makes efficient use of information where generalized variance is large.

Key words.- Principal component analysis, Hebbian algorithm, Dimensionality reduction.

INTRODUCCIÓN

La presente investigación plantea como objetivo estudiar los aspectos técnicos relacionados con la implementación de la red neuronal de Análisis de Componentes Principales (ACP) en términos de su capacidad de predicción, generalización y precisión con el fin de establecer criterios óptimos para su implementación, validación y evaluación del desempeño.

[1]Doctor, Maestro y Lic. Catedrático de la Escuela Profesional de Ingeniería Estadística (EPIES) de la Universidad Nacional de Ingeniería, [2]Ing. Catedrático de la de la Universidad Nacional de Ingeniería.

En el campo de la implementación de las redes neuronales ACP, no se ha desarrollado indicadores en un marco metodológico que permita implementar un modelo neuronal de acuerdo a sus características de aprendizaje y generalización, como sí sucede en el caso de las redes neuronales supervisadas en donde existen inclusive líneas de investigación muy bien consolidadas.

Debido a ello, la presente investigación centra su atención en la estructura multivariada de los datos de entrenamiento y su repercusión en el desempeño de la red neuronal ACP respecto a su capacidad predictiva, generalización y tiempo de entrenamiento [1].

Se plantea la hipótesis de que la estructura estadística de los datos influye significativamente en el óptimo desempeño de la red neuronal ACP en el contexto no supervisado. El algoritmo Hebbiano de la fase de aprendizaje, garantiza la calidad de representación de la red, más no la generalización cuando los datos presentan patrones asimétricos y diferentes grados de correlación y variabilidad multivariada.

Se acoge el modelo del 'cuello de botella' de la información para formular un modelo 'auto-supervisado' de la red ACP que en adelante se tratará según el enfoque tradicional de modelo no supervisado y se hará referencia a su cualidad de auto-supervisado para analizar sus propiedades de generalización. Seguidamente, desarrolla el marco teórico referencial tanto del modelo estadístico como del neuronal ACP. Se formula las hipótesis de investigación, las definiciones operativas y la correspondiente operacionalización de las variables.

Luego, se diseña el experimento correspondiente que permitirá alcanzar los objetivos establecidos en la investigación con rigurosidad científica. Finalmente, analiza los resultados experimentales obtenidos.

REDUCCIÓN DE DIMENSIONALIDAD EN LA MÁQUINA DE APRENDIZAJE

Las dos estrategias de modelamiento, reducción de datos y reducción de dimensionalidad, se clasifican en dos tipos de métodos: cuantización vectorial y reducción de dimensionalidad.

La presente investigación se centra en la reducción de dimensionalidad, en consecuencia, se desarrollará el enfoque neuronal y estadístico relativo a este problema.

La reducción de dimensionalidad consiste en encontrar el mapeo, desde un espacio *p*-dimensional, de los datos de entrenamiento hacia un espacio *m*-dimensional donde $m<n$,

$$G(x): \Re^p \rightarrow \Re^m \qquad (1)$$

de modo que resulte una codificación z=G(x) para cada observación *x*. Un 'buen' mapeo *G* funcionará como un codificador de dimensión reducida de la distribución original.

En particular, debería existir otro mapeo inverso tal que:

$$F(z): \Re^m \rightarrow \Re^p \qquad (2)$$

el cual producirá la decodificación $x'=F(z)$ de las observaciones *x* originales.

Así, un mapeo completo que incluya un proceso de codificación-decodificación sería:

$$x'=F(G(x)) \qquad (3)$$

Para encontrar el mejor mapeo es necesario especificar una clase de funciones aproximadas (mapeos):

$$f(x,\mu)=F(G(x)) \qquad (4)$$

parametrizadas por μ para encontrar una función que minimice el riesgo:

$$R(\omega)=\int L(x,x')p(x)dx=\int L(x,f(x,\omega))p(x)dx \qquad (5)$$

El ACP es una técnica de reducción de dimensionalidad, la cual, implementa una proyección (mapeo) z=G(x) que representa una transformación lineal de un vector x.

TRANSFORMACIÓN COMPONENTES PRINCIPALES

En un proceso de reducción de dimensionalidad el codificador está dado por la función G que produce el mapeo a partir de los datos de entrenamiento del espacio $\Re^p$ a un espacio de características de menor dimensión $\Re^m$. El decodificador está representado por la función F que mapea desde el espacio $\Re^m$ al espacio original $\Re^p$. Este proceso de codificación-decodificación se puede representar en términos del modelo 'cuello de botella' de la información, como se muestra en el Figura 1.

Dado un dato multivariado $x \in \Re^p$, se busca el mapeo:

$$f(x,\omega)=F(G(x)) \qquad (6)$$

tal que se minimice el riesgo. Cuando el riesgo se minimiza, la variable aleatoria z=G(x) provee una representación de los datos originales x en el espacio de menor dimensión $\Re^m$.

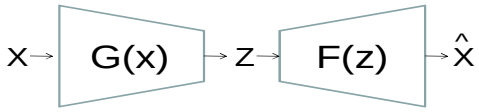

Fig. 1 *Proceso de reducción de dimensionalidad de la información como 'cuello de botella'*

La transformación componentes principales presenta propiedades óptimas en la clase de funciones $f(x,V)$:

Las componentes principales Z constituyen un conjunto de combinaciones lineales en el espacio proyectado de máxima variabilidad de los datos originales.

Provee la mejor representación lineal de menor dimensión en el sentido de que la suma total de las distancias al cuadrado de todos los puntos a sus proyecciones están minimizadas.

Si las funciones de mapeo F y G se restringen a la clase de funciones lineales, la composición $F(G(x))$ provee la mejor aproximación de los datos.

En el punto de vista estadístico, el ACP sólo depende de la matriz de covarianzas o correlación de $X=\{x_1, x_2, \ldots, x_n\}$. Su desarrollo no requiere que los datos se distribuyan como una normal multivariada, sin embargo, asumir esta distribución permitiría hacer inferencias a partir de las componentes obtenidas de una muestra [2].

APRENDIZAJE HEBBIANO NO SUPERVISADO

El aprendizaje Hebbiano no supervisado construye características cuantitativas en las cuales la variable se predice generalmente por regresión lineal. El aprendizaje Hebbiano se basa en la hipótesis de ajuste sináptico de Hebb, el cual sugiere que los pesos sinápticos w se ajustan proporcionalmente a la correlación entre el flujo de comunicación pre y post sináptica neuronal, x e y respectivamente.

La inestabilidad de w se trata con aplicaciones repetidas convergen a la unidad, $|w|=1$ donde $|w|^2=w'w$. Si el ratio de aprendizaje es pequeño se puede combinar los dos pasos anteriores en uno conocido como la regla de Oja [3]:

$$w < -w + \varepsilon y(x - yw) \qquad (7)$$

Se puede formular una generalización de este proceso para seleccionar, de igual forma, los M primeros vectores característicos con M neuronas lineales. Este proceso es conocido como la regla de Sanger y se formula de la siguiente manera:

$$w^k < -w^k + £y^k\left(x - \sum_{i=1}^{k} y^k w^k\right) \qquad (8)$$

Se puede notar, a partir de (8), que si k=1 se obtiene la regla de Oja. Para k=2,

$$w^2 < -w^2 + £y^2\left(x - y^1 w^1 - y^2 w^2\right) \qquad (9)$$

Cuando se alcanza la convergencia, los vectores w son ortogonales dos a dos y normalizados a 1. Así, w^k es un vector característico de C con valor característico asociado a^k.

HIPÓTESIS

La estructura estadística de los datos, es decir, patrón de deformación de los datos respecto a la normal, grados de correlación y variabilidad multivariada influyen significativamente en el óptimo desempeño de la red neuronal de análisis de componentes principales en el contexto no supervisado, es decir, en su calidad de representación de datos y generalización.

Así mismo, el algoritmo Hebbiano de la fase de aprendizaje garantiza la calidad de representación de los datos y no garantiza la generalización cuando los datos presentan patrones asimétricos al 10% de significación y variabilidad multivariada.

METODOLOGÍA

Descripción de la metodología [1] op cit. 2010):

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 de caracterización de datos de entrenamiento y validación, indicadores de aprendizaje y generalización. (Fase metodológica, capítulo IV).
3. Bloqueo.- Establecer los escenarios definidos por la cantidad de datos, varianza generalizada y distribución de los datos manteniendo las mismas condiciones experimentales en cada escenario. (Fase experimental, capítulo V).
4. Aleatorización.- Según los parámetros definidos en cada escenario se generan los números aleatorios p-dimensionales. (Fase experimental, capítulo V).
5. Réplica.- En cada escenario se realizan diez corridas con el fin de obtener resultados consistentes. (Fase experimental, capítulo V).
6. Control.- En el análisis de resultados el modelo estadístico ACP será utilizado como control dado su mejor rendimiento al conocerse los parámetros poblacionales.

Operacionalización de variables

Aprendizaje.- Proceso iterativo mediante el cual se determina el conjunto de pesos sinápticos a partir de valores iniciales nulos y con el acceso total de los datos (en 'batch') siguiendo la regla de aprendizaje 'hebbiano'.

Calidad de representación.- Mide la capacidad de la red para acumular la máxima variabilidad o máxima inercia contenida en los datos hasta las primeras tres componentes principales. Se mide como el ratio entre los tres primeros valores característicos y la varianza o inercia total de los datos (Indicador B.3: calidad de representación-fase entrenamiento).

Generalización.- Capacidad inherente de una red neuronal mediante la cual ésta logra capturar, en la etapa de aprendizaje, la información (variabilidad) subyacente en los datos de entrenamiento la cual le permite responder con precisión ante patrones o datos no utilizados en la fase de entrenamiento. Se miden mediante el Indicador C.1: calidad de representación-fase validación y C.2: error validación.

El error de generalización.- El error de generalización se mide en la fase de ejecución por la distancia euclidea media generada con los datos de validación. Para efectos del trabajo de investigación, se medirá el error de generalización (Indicador C.1) promedio obtenido en cada réplica.

EL EXPERIMENTO

Espacio parametral generador

Es necesario establecer la matriz de correlación como matriz definida positiva para que el cálculo mediante la descomposición de Cholesky sea posible. Teniendo en cuenta que el modelo generador es:

$$Y = LX \qquad (10)$$

Donde,

***Y*:** vector de números aleatorios con distribución normal multivariada,
***X*:** vector de números aleatorios con distribución normal univariada,
***L*:** la descomposición de Cholesky de la matriz de correlación.

Los 'outliers' definen los casos con presencia de datos asimétricos en una proporción del 3%.

Estos, están incluidos en la cantidad de datos de entrenamiento 'ne' para los casos III y IV en los tres tamaños de muestra, así como también, en los datos de validación 'nv' para los casos II y III.

Con 16 variables se tienen que definir 152 parámetros poblacionales; el añadir una variable más implicaría definir 48 parámetros poblacionales adicionales y así sucesivamente.

Tabla 1. *Cantidad de datos de entrenamiento y validación según casos.*

CASO I						
	Entrenamiento			Validación		
n	n_e	n_e(out)	Total	n_v	n_v(out)	Total
450	360	-	**360**	90	-	**90**
900	720	-	**720**	180	-	**180**
1800	1440	-	**1440**	360	-	**360**
CASO II						
	Entrenamiento			Validación		
n	n_e	n_e(out)	Total	n_v	n_v(out)	Total
450	360	-	**360**	87	3	**90**
900	720	-	**720**	174	6	**180**
1800	1440	-	**1440**	149	11	**360**
CASO III						
	Entrenamiento			Validación		
n	n_e	n_e(out)	Total	n_v	n_v(out)	Total
450	349	11	**360**	87	3	**90**
900	698	22	**720**	174	6	**180**
1800	1397	43	**1440**	149	11	**360**

Nota.-
'n' cantidad de datos de la muestra
'ne' cantidad de datos de entrenamiento
'np' cantidad de datos de prueba
'nV' cantidad de datos de validación
'nout' cantidad de datos 'outliers'

El conjunto de datos asimétricos se genera cambiando las observaciones limpias por observaciones 'outlier' en la proporción mostrada en la Tabla 1 para cada uno de los casos considerados y de acuerdo a la definición operativa correspondiente.

En este punto es necesario establecer los parámetros de diseño correspondiente al perceptrón que se ajustará a la aplicación final y que se someterá al análisis de sus propiedades según las condiciones experimentales ya establecidas.

En tal sentido, de manera arbitraria [4 y 5], se optará por la estructura neuronal explicitada en la ficha técnica mostrada en la Tabla 2.

El número de nodos considerado para la capa sensora responde a las 16 variables estudiadas en una investigación previa relativa a la validación por constructo de la escala de actitudes hacia la investigación [6].

Del mismo modo, los tres nodos considerados en la capa de respuesta responden a la estructura de los datos generados, los cuales provienen de una matriz de correlación de tres bloques en la diagonal principal (como puede verse en los anexos estadísticos IV.1 y IV.2 en [1] op cit). Esto definirá tres componentes principales que resumen una proporción significativamente grande de la variabilidad de los datos.

Tabla 2. *Ficha Técnica para la Red Neuronal ACP experimental.*

Número de nodos en la capa sensora	16
Número de nodos en la capa respuesta	03
Capas ocultas	01
Número de nodos en la capa oculta	03
Tipo de entrenamiento	Incremental
Tipo de Red Neuronal	Estática
Accesibilidad de datos	Batch
Ingreso de Datos	Concurrente
Sentido	Unidireccional
Ratio de Aprendizaje	0.1
Algoritmo de Aprendizaje	No Supervisado

Bases experimentales

Se utilizan técnicas estadísticas cuantitativas, gráficas y tabulares para el análisis de datos. Es importante recalcar, que este análisis no sólo se centra en la fase final de resultados sino que se enfoca también al análisis de los datos de entrenamiento en su fase de diseño, generación y algoritmo.

a) Aleatorización.- El principio de aleatorización ha sido respetado en la simulación de los datos de entrenamiento y validación puesto que han respondido a un proceso de generación pseudoaleatoria con posterior prueba de aleatoriedad. La distribución normal multivariada ha permitido también mantener la aleatoriedad e independencia en las particiones muestrales utilizadas durante el proceso de análisis. Así mismo, la asignación de los casos experimentales ha sido aleatoria dentro de bloques, ver Bloqueo (b).

b) Bloqueo.-Los escenarios sobre los que se desarrollan los casos analizados dependen de la varianza generalizada de los datos. Es decir, los tres casos definidos para el proceso experimental se generan bajo los mismos patrones de aleatoriedad en cada uno de los escenarios definido por la varianza generalizada: pequeña y grande respectivamente. El bloqueo permitirá identificar la existencia de patrones sistemáticos en la capacidad predictiva de la red que está influenciada por la variabilidad generalizada de los datos.

c) Réplica.- El principio de replicación está presente en cada uno de los tres casos definidos en cada bloque. De tal modo, que se podrá evaluar la consistencia científica de los resultados.

d) Control.- La técnica estadística de análisis de componentes principales se utiliza como control debido a que la formulación del modelo se desarrolla bajo un contexto de datos generados concordante con los supuestos estadísticos de partida. En consecuencia, se puede esperar que el modelo sea óptimo en su capacidad predictiva y captura de inercia de los datos.

RESULTADOS Y ANÁLISIS

Caracterización de los datos

La Tabla 3 resume los resultados obtenidos en las diez replicaciones para cada escenario según cantidad de datos. Los coeficientes de asimetría (A4) promedio obtenidos en el conjunto de datos de entrenamiento de los casos I y II son iguales debido a que el conjunto de datos es el mismo. Por las mismas razones de diseño, se observa resultados

similares en el conjunto de datos de validación de los casos II y III.

Sin embargo, en los seis escenarios la estructura de datos de entrenamiento y validación, en conjunto, es distinta para los efectos experimentales.

De la Tabla 4, se observa que la intensidad de multicolinealidad (indicador A.6) es severa cuando los datos se distribuyen normalmente y presentan una varianza generalizada pequeña. Para el caso que presenta deformaciones respecto a la distribución normal, el indicador se reduce significativamente. La presencia de 'outliers' afecta el grado de intensidad de multicolinealidad en un esquema parecido al observado en el caso I.

Asimismo, cuando los datos presentan un grado significativo de deformación, en un contexto de varianza generalizada grande, el indicador es inestable conforme aumenta el tamaño.

De esto último, no hay evidencia en los otros casos: en el caso I especialmente en donde las diferencias en la reducción de la intensidad de multicolinealidad son significativamente más notorias.

Tabla 3. *Coeficiente de Asimetría Promedio por Tipo y Cantidad de Datos según Caso*

Indicador A.4 Coeficiente de asimetría multivariado

Cantidad de datos	CASO I		CASO II		CASO III	
	VP	VG	VP	VG	VP	VG
Entrenamiento						
360	13,7	13,5	13,7	13,5	569,5	467,4
720	6,6	7,4	6,6	7,4	390,6	548,8
1440	3,4	3,3	3,4	3,3	293,3	272,7
Validación						
90	51,0	52,0	194,7	155,9	194,7	155,9
180	26,9	28,6	299,0	315,8	299,0	315,8
360	13,4	14,0	507,6	519,1	507,6	519,1

Nota.-
VP: Varianza generalizada pequeña (Indicador A.2)
VG: Varianza generalizada grande (Indicador A.2)
FUENTE: Anexo IV.3 y IV.5 en [1] op cit.

Tabla 4. *Intensidad de Multicolinealidad por Caso según Tamaño de Muestra*

Indicador A.6 Intensidad de multicolinealidad.

Tamaño de Muestra	CASO I		CASO III	
	VP	VG	VP	VG
360	69,27	23,17	26,78	22,13
720	69,25	23,03	26,35	16,48
1440	68,74	22,80	26,86	17,24

Nota.-
Modelo Estadístico ACP
VP: Varianza generalizada pequeña (Indicador A.2)
VG: Varianza generalizada grande (Indicador A.2)
FUENTE: Anexo IV.7 en [1] op cit.

Aprendizaje

El estudio de la capacidad de aprendizaje de la red neuronal ACP se centra en el estudio de la calidad de representación (indicador B.3) y el error de entrenamiento para fines predictivos (indicador B.1). De acuerdo a lo especificado en la hipótesis operativa, se espera que los efectos en la calidad de representación no estén influenciados por la estructura de correlación de los datos. Con base teórica suficiente se espera que el modelo estadístico sea más eficiente en capturar toda la información de los datos y, en consecuencia, se utiliza como control.

Calidad de representación en el modelo estadístico ACP

En la Tabla 5 se observa que la absorción de la inercia es limpia en los casos en que el conjunto de datos no presenta patrones con deformación respecto a la distribución normal. Por ejemplo, para el caso de varianza generalizada pequeña y tamaño de muestra es 360, las tres primeras componentes explican en promedio el 70.22% de la variabilidad de los datos y, en el caso de varianza generalizada grande, las tres primeras componentes explican el 79.35%.

Esta diferencia se mantiene para tamaños de muestra de 720 y 1,440 observaciones respectivamente.

Tabla 5. *Calidad de representación promedio para las tres primeras componentes por caso según tamaño de muestra*

Indicador B.3 Calidad de representación.

Tamaño de Muestra	CASO I		CASO III	
	VP	VG	VP	VG
360	70,22	79,35	76,19	75,01
720	70,25	77,87	74,85	70,83
1440	69,97	75,84	73,57	67,44

Nota.-
Modelo estadístico ACP
VP: Varianza generalizada pequeña (Indicador A.2)
VG: Varianza generalizada grande (Indicador A.2)
FUENTE: Anexos del IV.8 al IV.19 en [1] op cit.

Cuando los datos presentan deformaciones en distribución se eleva significativamente el ratio de inercia si se compara con el caso I. En efecto, la definición operativa de los 'outliers' considera observaciones proporcionalmente mayores (10 veces la desviación estándar) en las cinco primeras variables originales, consecuentemente se espera que la primera componente asuma la mayor proporción de la variabilidad dado su alta correlación con las ocho primeras variables originales.

Por otro lado, si se observa el efecto de la deformación en distribución, en un escenario de varianza generalizada grande, la absorción de inercia es significativamente menor conforme el tamaño de la muestra aumenta.

Si se compara los casos I y III, cuando la muestra es de 720 observaciones la caída en la inercia es de 7.04% y cuando la muestra es de 1,440 la caída es de 8.50%.

Estas cifras son significativas sobre todo si tenemos en cuenta que cada componente principal aporta en promedio 6.25% de inercia. Así, para un tamaño de muestra de 1,440 el ratio de inercia o calidad de representación para el caso normal es mayor al 75% en promedio mientras que para el caso con deformaciones es menor al 68%.

Calidad de representación en la red neuronal ACP

La capacidad de absorción de la red neuronal ACP en la fase de entrenamiento mantiene las mismas características que el modelo estadístico ACP. Este resultado se observa tanto para el caso I como para el caso III como puede verse en la Tabla 6. En el caso III, por ejemplo, la deformación de los datos respecto a la distribución normal producen los mismos efectos de absorción que el observado en la técnica estadística ACP – Tabla 5.

Tabla 6. *Calidad de representación en la fase de entrenamiento para las tres primeras componentes por caso según tamaño de muestra.*

Indicador B.3 Calidad de representación.

Tamaño de Muestra	CASO I		CASO III	
	VP	VG	VP	VG
360	69,44	79,25	76,73	76,02
720	68,78	77,86	75,36	71,66
1440	69,31	75,96	74,45	69,60

Nota.-
Modelo Neuronal ACP
VP: Varianza generalizada pequeña (Indicador A.2)
VG: Varianza generalizada grande (Indicador A.2)
FUENTE: Anexo [1]

Es importante notar, que de acuerdo al indicador de calidad de representación (indicador B.3) la red neuronal presenta un mal desempeño en capturar la variabilidad de los datos en escenarios con datos normales y varianza generalizada pequeña. Se observa así, que para el caso I con varianza pequeña el ratio de inercia promedio observado en el modelo estadístico ACP – Tabla 5 – fluctúa entre 69.97% y 70.22% para muestras de tamaño 1,440 y 360 respectivamente mientras que en el modelo neuronal la fluctuación es de 69.31% a 69.44%. El ratio de inercia obtenido es menor al obtenido por la técnica estadística ACP y este resultado es robusto al tamaño de muestra. Sin embargo, en el escenario con varianza generalizada grande, se observa una notable mejora en la calidad de representación de los datos

conforme aumenta la cantidad de datos de entrenamiento o tamaño de muestra para el caso de la técnica estadística.

La incorporación de observaciones 'outliers' aumenta la dispersión de los datos – caso III. Ésta es capturada por las primeras componentes principales; de allí que, se observa un aumento en el ratio de inercia (indicador B.3) registrado por la red neuronal.

Sin embargo, se reduce la capacidad de absorción de inercia en contextos de varianza generalizada grande cuando se presentan casos con deformación en distribución. Específicamente, en el caso de 1,440 datos de entrenamiento y varianza generalizada grande la incorporación de deformaciones de distribución – caso III comparado con el caso I - impacta al ratio de inercia de manera que cae 6.36% - Tabla 6 - en el caso neuronal comparado al 8.40% del caso estadístico ACP – Tabla 5.

Por otro lado, la Tabla 7 mide el grado de eficiencia de la red neuronal ACP en términos predictivos.

Como se indicó en la metodología, el EEN (Error de Entrenamiento) está definido como el cuadrado de la distancia euclidea entre el valor observado y el valor predicho por la red neuronal.

En el escenario correspondiente al caso I se observa que la capacidad predictiva de la red mejora notablemente cuando la cantidad de datos aumenta sobretodo en el caso de varianza generalizada grande.

En el caso en que existen deformaciones en la distribución de los datos respecto a la normal se observa un aumento significativo del EEN. En efecto, tanto para el caso con varianza generalizada pequeña como grande el aumento es notable, significativamente mayor al 300% con respecto al caso normal. En ambos casos se nota una reducción consistente en el EEN conforme aumenta la cantidad de datos de entrenamiento.

Según el indicador B.2, el EEN será menor cuando los datos presentan distribución normal y la varianza generalizada es pequeña.

Tabla 7. *Error de entrenamiento promedio por caso según tamaño de muestra.*

Indicador B.2 Error de entrenamiento (EEN).

Tamaño de Muestra	CASO I		CASO III	
	VP	VG	VP	VG
360	3,94	11,00	16,22	27,40
720	5,26	8,52	15,99	26,44
1440	4,12	5,34	13,99	23,07

Nota.-
Modelo Neuronal ACP
VP: Varianza generalizada pequeña (Indicador A.2)
VG: Varianza generalizada grande (Indicador A.2)
FUENTE: Anexo [1]

La evidencia confirma la superioridad del modelo estadístico ACP cuando los datos de entrenamiento tienen distribución normal multivariada inclusive con muestras pequeñas. Se espera que la calidad de representación se mantenga en los datos reconstruidos debido a que el modelo estadístico ACP al constituir una transformación lineal las propiedades primigenias de los datos se mantienen. La red neuronal ACP, en la fase de entrenamiento, supera al modelo estadístico para grandes cantidades de datos y varianza generalizada grande en un contexto de datos distribuidos normalmente. Al ser un modelo de distribución libre o no paramétrico, la mayor cantidad de datos e información redundarán en un mejor rendimiento de la red neuronal. El resultado se mantiene favorable para casos en que la red neuronal se entrena con datos que presentan deformaciones. Esto significa que la red neuronal no pierde la capacidad de aprendizaje o el algoritmo Hebbiano capitaliza eficientemente la información en contextos de varianza generalizada grande independientemente de la distribución de los datos.

Generalización

La generalización de la calidad de representación de los datos es un aspecto fundamental y de interés en la presente investigación. En el campo de las aplicaciones de ingeniería y/o investigación científica este resultado permitirá saber si la red incorpora características inherentes a la estructura de los datos o bajo qué condiciones esta capacidad de generalización de la red es óptima.

La Tabla 8, muestra el ratio de inercia o calidad de representación promedio en la fase de validación entre las réplicas para cada uno de los escenarios y cantidad de datos considerados. El caso I, por ejemplo, preserva su calidad de representación. El indicador C.1 (calidad de representación), muestra un ratio inercial que mantiene los niveles alrededor del 69% y no es afectado por el tamaño de la varianza generalizada ni por la cantidad de datos. Los pesos sinápticos calculados en la fase de entrenamiento con datos normales y diferentes niveles de varianza generalizada son utilizados en la fase de validación para calcular las proyecciones de los nuevos puntos sobre los ejes rotados. Al presentar la misma estructura en distribución, estos datos de validación mantienen la misma cantidad de información contenida en los datos. Es decir, no hay pérdida en la calidad de representación lo cual puede considerarse como un desempeño óptimo de la red neuronal. El caso II, se caracteriza por utilizar datos normales en la fase de entrenamiento y datos con deformaciones en su distribución en la fase de validación. Como se observa en la Tabla 8, la calidad de representación de los datos ha caído significativamente si se compara con el ratio de inercia obtenido en la fase de entrenamiento en el caso de varianza generalizada pequeña – caso I de la Tabla 6. Es más, en el caso de varianza generalizada grande la caída es del 10% en promedio independientemente de la cantidad de datos de validación, notándose una significativa y progresiva mejora cuando la cantidad de datos aumenta.

Tabla 8. *Calidad de representación en la fase de validación para las tres primeras componentes por caso según tamaño de muestra*

Indicador C.1 Calidad de representación.

Tamaño de Muestra	CASO I		CASO II		CASO III	
	VP	VG	VP	VG	VP	VG
90	68,63	79,48	59,02	67,62	72,42	73,95
180	68,38	77,57	64,40	66,06	72,15	65,95
360	69,45	75,96	62,57	66,07	72,27	68,11

Nota.-
Modelo Neuronal ACP
VP: Varianza generalizada pequeña (Indicador A.2)
VG: Varianza generalizada grande (Indicador A.2)
FUENTE: Anexo

En el caso III se utilizó datos con deformaciones en distribución tanto en la fase de entrenamiento como en la de validación. Se observa una caída significativa en la calidad de representación de los datos en la fase de validación – Tabla 5 - si se compara con el alcanzado por la técnica estadística ACP en el mismo escenario. Sólo en el caso de varianza generalizada grande con 360 datos de validación la red neuronal supera en calidad de representación (68.11%) al modelo estadístico ACP (67.44%).

Como se observa en la Tabla 9, el error de validación (EVA) mantiene los niveles observados en la fase de entrenamiento –Tabla7, al contrario de lo observado en los casos en que se presentan deformaciones en la distribución de los datos (caso II y III).

Por otro lado, si se analiza el caso en que la deformación de los datos está presente tanto en la fase de entrenamiento como de validación se notará que la mejora es independiente de la magnitud de la varianza generalizada. En el caso de varianza generalizada pequeña y grande la reducción es de 1.83 y 3.46 respectivamente.

Tabla 9. *Error de validación promedio por caso según tamaño de muestra.*

Indicador C.2 Error de validación (EVA).

Tamaño de Muestra	CASO I		CASO II		CASO III	
	VP	VG	VP	VG	VP	VG
360	2,94	11,30	15,66	24,77	15,75	29,23
720	5,29	8,45	13,77	23,02	16,50	27,45
1440	4,18	5,32	16,20	26,90	14,37	23,44

Nota.-
Modelo Neuronal ACP
VP: Varianza generalizada pequeña (Indicador A.2)
VG: Varianza generalizada grande (Indicador A.2)
FUENTE: Anexo [1]

CONCLUSIONES

Para contrastar la hipótesis se ha diseñado un experimento aleatorio en diferentes escenarios del espacio parametral, diferentes estructuras de datos y cantidad de datos de entrenamiento lo cual le otorga

la calidad de resultados generalizables por su naturaleza científico-experimental. De esta forma, se logra demostrar que las tareas de reducción de dimensionalidad desde el punto de vista de su calidad de representación y aprendizaje cuando los datos provienen de una población con distribución normal dan óptimos resultados. En tal contexto, la red ACP logra mejor rendimiento cuando se dispone de gran cantidad de datos y varianza generalizada grande. Así mismo, se comprueba que la red neuronal ACP no pierde su capacidad de aprendizaje en términos de calidad de representación de los datos frente a deformaciones en su distribución debido a que el algoritmo Hebbiano capitaliza eficientemente la información de los datos. Al contrario, ante la presencia de *outliers*, tanto en la fase de entrenamiento como validación, el algoritmo Hebbiano no logra construir eficientemente la función (i.e. mapeo) de codificación-decodificación que minimice la pérdida (EVA), lo cual afecta su propiedad de generalización.

Para estudiar la capacidad de generalización se utilizó la misma metodología utilizada por las redes supervisadas asumiendo que los datos de llegada eran los de entrada valiéndose así de la capacidad de representación (o reconstrucción) de los datos adquirido durante la fase de entrenamiento.

En escenarios con distribución normal - caso I - la red neuronal ACP demostró poseer la capacidad de almacenar en sus pesos sinápticos las características observadas en todos los patrones de entrenamiento que fueron usados durante la fase de entrenamiento. Los escenarios caracterizados por el caso III reafirmaron la capacidad de generalización de la red al ajustar los pesos sinápticos con la nueva información contenida en los datos asimétricos o con deformaciones respecto a la normal.

REFERENCIAS

1. **Huamanchumo, L.,** "Efectos de la Estructura Estadística de Datos en la Implementación de la Red Neuronal de Análisis de Componentes Principales". Tesis de Maestría. Facultad de Ingeniería Industrial y de Sistemas. Universidad Nacional de Ingeniería. 2010.
2. **Johnson, R., Wichern, D.,** "Applied Multivariate Statistical Analysis". 5th Edition. Prentice Hall. 2002.
3. **Spencer, R., Sánchez-Sinencio, E**., "A Fully-Differential CMOS Implementation of Oja´s Learning Rule in a Dual-Synapse Neuron for Extracting Principal Components for Face Recognition". Analog and Mixed Signal Center. Texas A&M University. USA.
4. **Aiken, Milan.,** "Artificial Neural Systems as a Research Paradigm for the Study of Group Decision Support Systems". Group Decision and Negotiation Kluwer Academic Publishers. Department of Management and Marketing, School of Business Administration, University of Mississippi. 1997. pp. 379.
5. **Kwak, N, King, C.,** "Dimensionality reduction based on ICA for regression analysis." Lecture notes on computer science. ICANN 2006, Part I, LNCS 4131. Springer-Verlag. Berlin Heidelberg. 2006. pp.7-8.
6. **Huamanchumo, L.** "Escala de Actitud hacia la Investigación, Estudiantes y Carreras Profesionales de Ingeniería y Ciencias de la UNI" **TECNIA.** Vol. 16 N° 2. Lima-Perú. 2006. pp.43-50.8

Correspondencia: lhdlc40@gmail.com

Recepción de originales: enero 2013
Aceptación de originales: junio 2013