

Clasificación lineal mediante algoritmo de perceptrón difuso

Valeria Paola González Duéñez

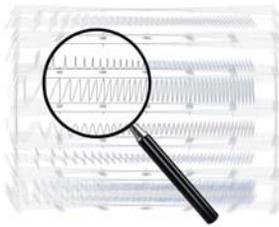
División de Admón. y Sistemas

vpaola@fime.uanl.mx

Óscar Leonel Chacón Mondragón

Programa de Postgrado en Ingeniería de Sistemas FIME-UANL.

ochacon@mail.uanl.mx



RESUMEN

La clasificación es una de las herramientas necesarias para llevar a cabo un buen reconocimiento de patrones. Las Redes de Neuronas Artificiales (RNA), como una sección del área de Inteligencia Artificial (IA), dispone del perceptrón que es un método simple y eficiente para aprender a través de ejemplos, a realizar clasificaciones lineales. Sin embargo, éste no es tan flexible cuando existen datos con gran incertidumbre. El Perceptrón Difuso (APD) resuelve este problema de convergencia utilizando la teoría de conjuntos difusos. En estas condiciones el APD presenta sus ventajas sobre los otros métodos.

PALABRAS CLAVE

Inteligencia artificial, clasificación lineal, reconocimiento de patrones, redes neuronales artificiales, perceptrón difuso.

ABSTRACT

The classification is a necessary tool to carry out good Pattern Recognition. The Artificial Neural Networks, as an Artificial Intelligence section, has the Perceptron, a simple and efficient method for learning a linear classifier from training examples. Although it is simple, it is non flexible when there are data of high uncertainty. The Fuzzy Perceptron solves convergence problem above using Fuzzy Theory. Under such conditions, the fuzzy perceptron overcomes other methods.

KEYWORDS

Artificial Intelligence, Linear Classification, Pattern Recognition, Artificial Neural Networks, Fuzzy Perceptron.

INTRODUCCIÓN

Actualmente el área de inteligencia artificial estudia el comportamiento inteligente del ser humano representando y simulando este comportamiento mediante la creación de modelos y algoritmos computacionales a fin de aplicarlo en la solución de problemas y en el caso de la industria^{1,2} con buen éxito; en sí la industria actualmente se ha apoyado principalmente, para el

área de toma de decisiones, con sistemas expertos, robótica, redes neuronales artificiales, lógica difusa, reconocimiento de patrones, etc.

El presente estudio está relacionado con el área de reconocimiento de patrones. Es importante mencionar que el reconocimiento de un objeto como miembro de una clase o grupo se denomina identificación. A su vez, la clasificación es el proceso de aprendizaje de agrupamiento de objetos en clases de acuerdo a su similitud. El área de reconocimiento de patrones incluye ambas: clasificación e identificación.³

Para su identificación el proceso de aprendizaje se lleva de manera implícita; éste se utiliza en aquellos sistemas en los cuales el registro de sus experiencias, dentro de un sistema interno, cambia de acuerdo a su comportamiento. La clasificación es una forma de aprendizaje que se inicia a partir de los antecedentes que existen de las clases. El razonamiento es un proceso en el que se aplican reglas generales, ecuaciones, relaciones y también una colección inicial de datos o hechos para deducir resultados o tomar decisiones. La lógica difusa utiliza el concepto de incertidumbre y mapeo de rasgos distintivos para la formación de agrupamientos denominándolos difusos.⁴⁻⁷ En el proceso de reconocimiento de patrones, al objeto se le asigna un grado característico o grado de membresía, identificando -en cada agrupamiento y de forma única- la similitud entre los atributos del objeto y los atributos de los agrupamientos o clases.⁸

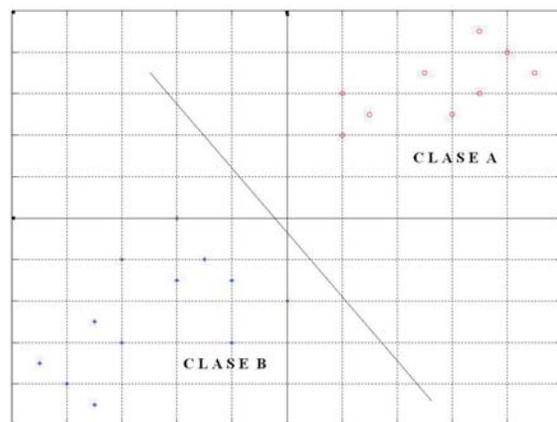
REPRESENTACIÓN DE PATRONES

El concepto de patrón puede establecerse como un grupo de datos que contienen rasgos distintivos asociados a una clase, que los hacen únicos e irrepetibles. Un objeto contiene atributos que son medidos para producir un patrón (vector), para asignarle un grado de membresía respecto a las clases en consideración o para definir un espacio característico que se utilizará para efectuar el proceso de reconocimiento. En general, los atributos de los objetos producen un grupo de características a las cuales se les aplican técnicas de similitud para lograr su reconocimiento.⁶

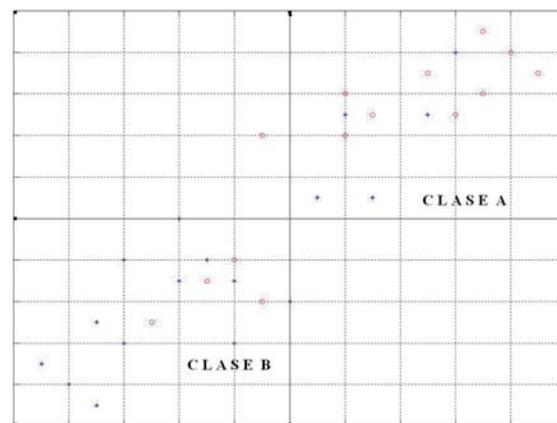
Un mecanismo para el reconocimiento de objetos o patrones es un sistema en el cual, el vector

que contiene las características dadas del patrón actúa como entrada al sistema, que el sistema opera sobre dicha entrada para producir una salida, la cual es una identificación única asociada con la clase del objeto a la cual pertenece. El reconocimiento está basado en la medida de los atributos físicos o mapeo de los mismos. Las N características de una población de objetos pueden considerarse tomadas como dimensiones dentro de un espacio vectorial, según se ilustra en la figura 1, para el caso de $N = 2$, con dos clases.

Existen diferentes técnicas que pueden utilizarse en el sistema de reconocimiento de patrones, siempre y cuando se tenga definido un espacio de rasgos distintivos, o características asociadas a una población de objetos. El modo operacional del reconocimiento es un sistema de mapeo, donde cada vector de entrada a través de un proceso,



a) Lineal



b) No lineal

Fig. 1. Separación de clases.

distingue o asocia los rasgos distintivos en un vector, para cada grupo de objetos. La acción que determina la región de aceptación se conoce como proceso de reconocimiento. En la clasificación se involucra el proceso de entrenamiento conocido como mecanismo de aprendizaje.

Las categorías de un mecanismo de aprendizaje se conocen como: *entrenamiento supervisado* y *entrenamiento no supervisado*. En el entrenamiento supervisado cada vector de entrada del sistema se identifica con un vector asociado de salida, el sistema aprende con cada vector presentado al sistema asociándolo a la salida correspondiente. En el aprendizaje no supervisado el sistema incluye ambos procesos entrenamiento y clasificación para cualquier entrada, asociando un prototipo para cada clase de objetos.³ Este aprendizaje realiza un proceso de agrupamiento donde se define un vector de rasgos distintivos a partir de las clases llamado "clustering". Cuando un vector de rasgos distintivos es una entrada en el sistema, se determinan las distancias a los centros de todos los agrupamientos generando una identificación del grupo al que pertenece, basado en la distancia mínima.⁸

CLASES LINEALMENTE SEPARABLES

Teniendo en cuenta que dentro de un espacio existen características o rasgos distintivos, un sistema clasificador puede describirse por una función discriminante, la cual identifica de manera única cada grupo.

El caso más simple de un mecanismo de entrenamiento es el que utiliza datos linealmente separables. Por ejemplo consideremos a X como un grupo de datos de entrenamiento $X = \{x_1, x_2, x_3, \dots, x_p\}$ $x_j \in R^s$. A partir de este conjunto se forman dos grupos (conjunto de vectores) A_1 y A_2 , que contienen los datos que serán identificados de acuerdo a su grado de membresía. Cuando ocurre el proceso de identificación se deberá obtener un vector que separe linealmente a estos grupos. El resultado de la etapa de aprendizaje está representado en el siguiente planteamiento: encontrar un vector w y un escalar b , tal que, todos los datos x_j , satisfagan las siguientes condiciones:

$$w^T x_j + b > 0, \quad \text{si } x_j \in A_1 \quad (1)$$

$$w^T x_j + b < 0, \quad \text{si } x_j \in A_2 \quad (2)$$

donde w^T representa transposición del vector w . Por lo tanto, si w y b existen, se puede decir que A_1 y A_2 son grupos o regiones linealmente separables; w es llamado también vector de separación y es el encargado de separar (identificar) los datos o elementos pertenecientes a los grupos A_1 y A_2 .

SEPARACIÓN LINEAL DE CLASES DIFUSAS

La condición para llevar a cabo la separación de clases difusas es considerar a los grupos A_1, A_2 como una partición difusa de $Y = \{y_1, y_2, \dots, y_p\}$, el cual es el grupo que contiene la totalidad de los datos $X = \{x_1, x_2, \dots, x_p\}$ $x_j \in R^s$ como vectores aumentados $y_j = [x_j \ 1]^T$. Consideremos

así mismo el vector $v = \begin{bmatrix} w \\ b \end{bmatrix}$ en relación a las ecuaciones (1) y (2).

Las clases difusas A_1 y A_2 serán linealmente separables si existe un vector $v \in R^{s+1}$, válido para cada $y \in Y$, tal que,

$$\mu_2(y) > 0.5 \Leftrightarrow v^T y > 0 \quad (3)$$

$$\mu_2(y) > 0.5 \Leftrightarrow v^T y < 0 \quad (4)$$

donde $\mu_1(y)$ y $\mu_2(y)$ representan una función equivalente a $v^T y = 0$ y se localizan en las ecuaciones (8) y (10). Si no se cumplen las condiciones anteriores, A_1 y A_2 serán considerados grupos difusos no lineales y por lo tanto es necesario aplicar otras herramientas de clasificación.

ALGORITMO DE PERCEPTRÓN DIFUSO

Es posible mencionar el modelo más simple de redes neuronales difusas; en este caso, el algoritmo de perceptrón difuso, el cual puede ser entrenado con los agrupamientos difusos A_1 y A_2 .

En la figura 2 se presenta el peso v_k asociado a la k -ésima conexión de entrada así como v_{s+1} , el umbral del perceptrón difuso. Para un elemento

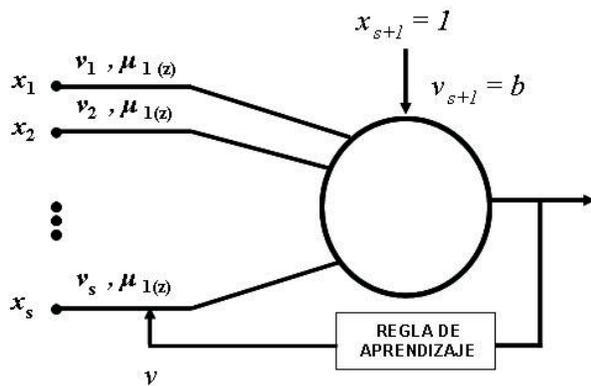


Fig. 2. Diagrama de perceptrón difuso.

$\mathbf{x} \in \mathbf{X}$, cada una de las conexiones de entrada x_j se conectan en el perceptrón difuso, etiquetadas con el grado de membresía correspondiente, según el agrupamiento A_1 y A_2 .

Considerando la normalización del signo, sea entonces

$$\mathbf{z}_j = \begin{cases} -\mathbf{y}_j, & \text{si } \mu_2(\mathbf{y}_j) > 0.5 \\ \mathbf{y}_j, & \text{por el contrario} \end{cases} \quad (5)$$

y sea $\mathbf{Z} = \{\mathbf{z}_1, \mathbf{z}_2, \mathbf{z}_3, \dots, \mathbf{z}_p\}$ el grupo de datos normalizados. Se puede establecer que los agrupamientos difusos A_1, A_2 , son linealmente separables, si existe un vector de separación $\mathbf{v} \in \mathbf{R}^{s+1}$, tal que

$$\mathbf{v}^T \mathbf{z}_j > 0, \quad j = 1, \dots, p \quad (6)$$

datos A_1, A_2 como grupos difusos linealmente separables en un espacio Z .

DETERMINACIÓN DE LOS GRADOS DE MEMBRESÍA

El APD establece que para un grupo de vectores simples, cada clase difusa se encuentra etiquetada como clase 1 y clase 2. Los vectores \mathbf{y}_j de la clase 2, tal como se muestra en la relación (5), están multiplicados por -1 .

La regla de aprendizaje generaliza el modelo de perceptrón clásico, de la siguiente manera:

$$\mathbf{v}^{k+1} = \mathbf{v}^k + \left| \mu_1(\mathbf{z}_k) - \mu_2(\mathbf{z}_k) \right|^m \mathbf{c} \mathbf{z}_k \quad (7)$$

considerado éste como un entrenamiento clásico

$$\mu_1(\mathbf{z}_k) = 0.5 + \frac{\exp\left(\frac{f(d_{k2} - d_{k1})}{d(\mathbf{u}_1, \mathbf{u}_2)}\right) - \exp(-f)}{2[\exp(f) - \exp(-f)]} \quad (8)$$

$$\mu_2(\mathbf{z}_k) = 1 - \mu_1(\mathbf{z}_k) \quad (9)$$

donde $d(\mathbf{u}_1, \mathbf{u}_2)$ es la distancia entre dos vectores y f es una constante positiva. Si \mathbf{z}_x es la clase 2 tenemos que

$$\mu_2(\mathbf{z}_k) = 0.5 + \frac{\exp\left(\frac{f(d_{k1} - d_{k2})}{d(\mathbf{u}_1, \mathbf{u}_2)}\right) - \exp(-f)}{2[\exp(f) - \exp(-f)]} \quad (10)$$

y

$$\mu_1(\mathbf{z}_k) = 1 - \mu_2(\mathbf{z}_k) \quad (11)$$

donde la constante f controla el rango en el cual el grado de membresía decrece a 0.5.

Si el vector \mathbf{z}_k es considerado de la clase 1 y es equidistante de los 2 prototipos, $d_{k1} = d_{k2}$, entonces el grado de membresía para A_1 es;

$$\mu_1(\mathbf{z}_k) = 0.5 + B \quad (12)$$

y si \mathbf{z}_k es considerado de la clase 2, el grado de membresía para A_2 es

$$\mu_2(\mathbf{z}_k) = 0.5 + B \quad (13)$$

en ambos casos

$$B = \frac{1 - e^{-f}}{2(e^f - e^{-f})} \quad (14)$$

La clasificación incorrecta de un punto depende de la clase que se desee considerar teniendo en cuenta la siguiente condición:

$$\mathbf{v}^{kT} \mathbf{z}_k \leq 0 \quad \wedge \quad (\mu_1(\mathbf{z}_k) \geq 0.5 + B \quad \vee \quad \mu_1(\mathbf{z}_k) \leq 0.5 - B) \quad (15)$$

Es necesario considerar situaciones en las cuales se requiere eliminar algunos puntos atípicos; en tal caso, la constante B debe ser reemplazada por $\beta = B + e$, donde e es un margen positivo. Bajo esta consideraciones, el APD utiliza la siguiente regla de entrenamiento:

$$v^{k+l} = \begin{cases} v^k + c|\mu_1(z_k) - \mu_2(z_k)|^m z_k, \\ \text{si } \left[(v^k)^T z_k \leq 0 \right] \\ \wedge (\mu_1(z_k) \geq 0.5 + \beta) \vee \\ (\mu_1(z_k) \geq 0.5 + \beta) \\ v^k \end{cases} \quad (16)$$

Si las clases son linealmente separables, este procedimiento converge a la separación de los vectores.⁸

PRUEBA Y RESULTADOS

Los datos usados para prueba se obtuvieron por simulación digital de un estudio de estabilidad para una falla trifásica de secuencia positiva realizado con anterioridad. En ellos se representa el comportamiento de las diferentes variables eléctricas, como la resistencia (*R*), reactancia (*X*), corriente (*I*), voltaje (*V*), impedancia (*Z*) y ángulo de desfase, según lo muestra la figura 3. Los datos de la figura 4 no contienen un pre-procesamiento para el ACP. Para esta muestra, se realizó un pre-procesamiento de los datos determinándose el subespacio de los dos mayores componentes principales, en donde se encontró el estado de falla y oscilación.⁹

Después del pre-procesamiento, los datos se encuentran distribuidos en cantidad no proporcional para los 3 estados: pre-falla, falla y oscilación de potencia; de las 2746 observaciones, 80 señalan el estado de pre-falla, 174 el estado de falla y 2491 el estado de oscilación. Después de haber identificado

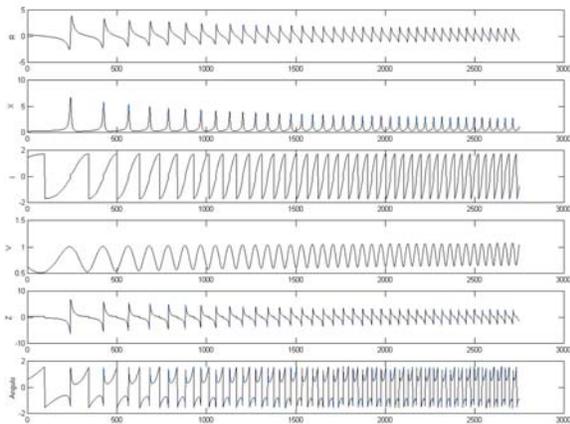


Fig. 3. Comportamiento de las variables eléctricas.

los datos según el comportamiento de las variables (*I*, *V* y *Z*), se tomó una muestra proporcional para el estado de falla y de oscilación, descartando la situación de pre-falla. Se seleccionó una muestra de 350 datos, donde 170 pertenecen al estado de falla y 180 al estado de oscilación. Quedando para esta situación el espacio representado por la figura 5.

Para cada caso, ya sea falla u oscilación se obtuvo el centro medio de cada agrupamiento, llamándolos u_1 y u_2 respectivamente.

$$u_1 = \frac{1}{170} \sum_{k=1}^{170} y_k \quad (17)$$

$$u_2 = \frac{1}{180} \sum_{k=171}^{350} y_k \quad (18)$$

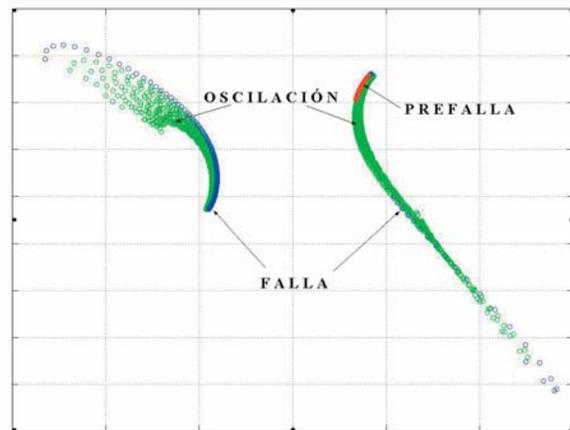


Fig. 4. Situación de prefalla, falla y oscilación.

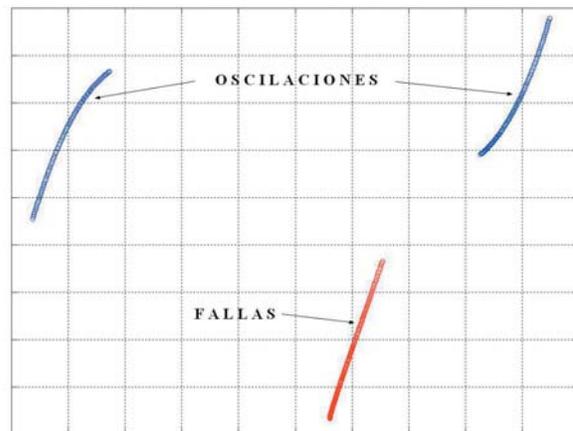


Fig. 5. Representación gráfica de los estados de falla y oscilación.

calculándose las distancias promedio entre los centros medios de cada agrupamiento,

$$d_{u_1, u_2} = \sqrt{(u_1 - u_2)^T (u_1 - u_2)} \quad (19)$$

Es importante mencionar que la simulación del algoritmo se realizó en Matlab. Se inicializaron los siguientes parámetros: $f = 3$ de la ecuación (14), $e = 0.001$ para β en (16), $c = 0.1$ y $m = 3$, éstas últimas mencionadas en la relaciones (7) y (16).

Para efectuar la clasificación, se requieren los grados de membresía de cada agrupamiento (falla y oscilación), por lo que es necesario calcular las distancias de los datos con respecto a cada uno de los centros, mediante las siguientes expresiones:

$$dk_1 = \sqrt{(y_k - u_1)^T (y_k - u_1)} \quad (20)$$

$$dk_2 = \sqrt{(y_k - u_2)^T (y_k - u_2)} \quad (21)$$

Para realizar la clasificación, se utilizaron los datos de oscilación y falla, identificándolos previamente según su grado de membresía. Para realizar la clasificación se tomaron los datos (los cuales son rasgos distintivos para cada situación diferente) y el valor asociado, +1 o -1, oscilación o falla, respectivamente. Se realiza la normalización del signo para obtener el dato z_k de acuerdo con la ecuación (5). Con estos datos se calculan los grados de membresía para las fallas mediante las ecuaciones (7) y (8), y para las oscilaciones mediante las ecuaciones (9) y (10).

El proceso iterativo de la regla de aprendizaje del

APD (16) requiere de un vector inicial $v^0 = \begin{bmatrix} w^0 \\ b^0 \end{bmatrix}$, sugiriéndose la siguiente estimación:

$$w^0 = \frac{c_d}{\|c_d\|} \quad \text{donde } c_d = u_1 - u_2 \quad (22)$$

$$b^0 = \frac{c_d^T}{\|c_d\|} c_m \quad \text{donde } c_m = \frac{u_1 + u_2}{2} \quad (23)$$

Cuando se aplicó el APD al total de los datos (ecuación(6)), éstos fueron plenamente identificados obteniendo una función lineal de separación. Para

los datos de entrenamiento de falla y oscilación la función lineal obtenida es $y = 0.067x - 0.1$,

donde $v = \begin{bmatrix} 0.067 \\ -0.1 \end{bmatrix}$ la cual se representa gráficamente en un período de -2.5 a 2.5 en las ordenadas, en la figura 6.

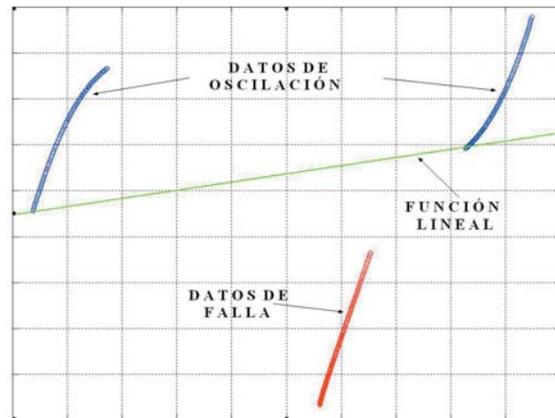


Fig. 6. Separación lineal de falla y oscilación.

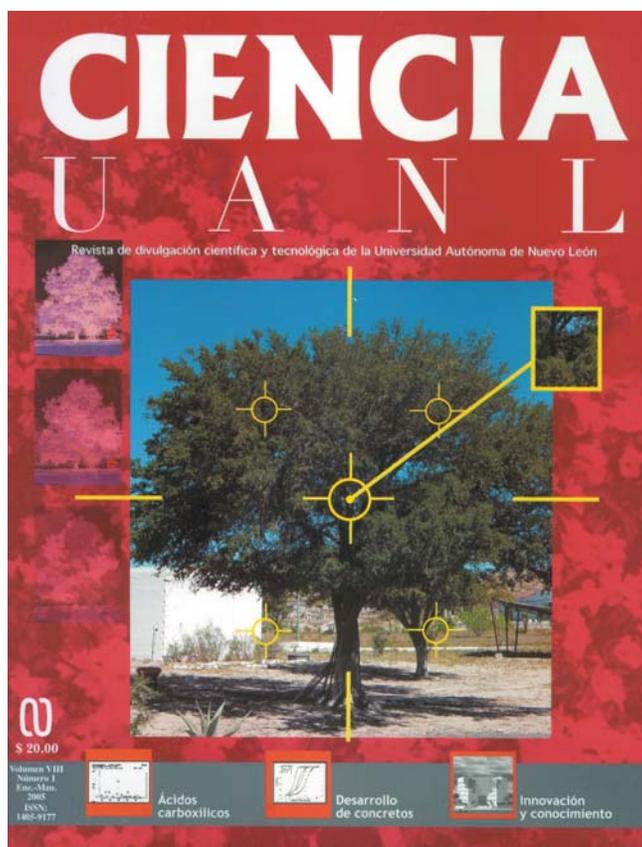
CONCLUSIONES

La finalidad de la aplicación del algoritmo es lograr una clasificación correcta, o determinar el hiperplano que logre la discriminación de los datos pertenecientes a cada grupo. Cuando se haya encontrado una separación lineal dentro de su fase de aprendizaje, el algoritmo termina generando los parámetros w (gradiente) y b (umbral) del hiperplano de separación logrando así el proceso de clasificación. Es necesario continuar con el estudio de los grados de membresía para los patrones de entrenamiento, principalmente cuando los datos no tienen muy bien definido su grado de pertenencia.

BIBLIOGRAFIA

1. Elaine Rich, "Artificial Intelligence," McGraw-Hill, 1983.
2. Stuart Russell, Peter Norving, "Inteligencia Artificial. Un enfoque moderno," Prentice-Hall 1996.
3. Carl G. Looney, "Pattern Recognition Using

- Neural Networks. Theory and Algorithms for Engineers and Scientists,” Oxford University Press 1997.
4. R. R. Yager, R.M. Tong, H. T. Nguyen, S. Ouchinnikov, “Fuzzy Sets and Applications,” John Willey & Sons 1987.
 5. Kevin M. Passino, Stephen Yurkovich, “Fuzzy Control,” Addison-Wesley 1998.
 6. Tina A. Folger, “Fuzzy Sets, Uncertainty and Information,” Prentice-Hall 1988.
 7. Timothy J. Ross, “Fuzzy Logic with Engineering Applications”. McGraw Hill, New York, New York 1995
 8. D. Dumitrescu, B. Lazzerini, L. C. Jain, “Fuzzy Sets and Their Application to Clustering and Training,” CRC 2000.
 9. V. González. Tesis de maestría “Clasificación de Fallas y Oscilaciones en un Sistema Eléctrico utilizando el Algoritmo de Perceptrón Difuso”. División de Posgrado, Facultad de Ingeniería Mecánica y Eléctrica, UANL 2004.



<http://www.uanl.mx/publicaciones/ciencia-uanl/>
rciencia@mail.uanl.mx

Tel. 01 (81) 8329 4000 Ext. 6622, Fax 6623