

FUNDAMENTO PSICOFISICO DEL RECONOCIMIENTO DE PATRONES

José M. SALINAS y Carmen PORRAS

Dpto. de Sociología y Psicología Social
Universidad de Granada

Este estudio presenta un clasificador de patrones que utiliza como criterio de asignación a una clase la mínima distancia logarítmica entre el patrón de entrada y el patrón de referencia de la clase. Se estudian los argumentos que basados en las leyes de la Psicofísica, recomiendan utilizar este algoritmo de clasificación. Mediante simulación en ordenador, se demuestra la superioridad, en múltiples tareas, de este clasificador sobre el clasificador gaussiano. Finalmente, un experimento de clasificación de patrones en dos clases, definidas de tal forma que si se utiliza un clasificador gaussiano la asignación corresponde a una clase y en caso de utilizar un clasificador basado en la distancia logarítmica corresponde a la otra clase, puso de manifiesto que los 33 sujetos experimentales realizaron una clasificación semejante a la que se obtendría con el algoritmo de mínima distancia logarítmica.

Palabras clave: Redes Neuronales; Reconocimiento de Patrones; Distancia logarítmica; Leyes Psicofísicas; Distancia entre sensaciones.

The Psychophysical basis of pattern recognition. This work proposes a pattern classifier that uses the minimum logarithmic distance between the input and the class reference pattern, as an assignation criterion. We show the reasonings based on the Psychophysical laws, which recommend to use this classification algorithm. By Montecarlo simulation, we have proved the superiority of this classifier over the gaussian classifier in many tasks. Finally, a two class pattern classification experiment was designed so that the assignation into a class or other depended on the use of our classifier or the gaussian one. The thirty three experimental subjects performed a classification like the classification obtained with the minimum logarithmic distance algorithm.

Key words: Neural Networks; Pattern Recognition; Logarithmic distance; Psychophysical Laws; Sensations distance.

El actual auge del paradigma conexionista y el intento de diseñar e implementar dispositivos artificiales semejantes a las redes neuronales que procesan la información en los sistemas biológicos, procede en parte de las dificultades encontradas por la Inteligencia Artificial al intentar resolver determinados problemas con la metodolo-

gía del procesamiento simbólico realizado secuencialmente. En este sentido se ha producido un cambio intentando resolver estos problemas con algoritmos y dispositivos inspirados en los elementos biológicos, no obstante este proceso se encuentra en su fase inicial y aún no ha llegado a profundizar suficientemente en esta lógica, así la mayoría de los dispositivos diseñados, aun-

que adoptan el procesamiento en paralelo, propio de los procesadores biológicos, no han apurado esta semejanza y operan y toman sus decisiones sobre la base de utilizar como inputs el sistema de magnitudes físicas de los fenómenos de que se trata; por el contrario los organismos vivos no operan ni computan estas magnitudes, sino que manejan las sensaciones que perciben al recibir un estímulo físico como input.

Para ilustrar cómo esta perspectiva "Psicofísica" que utiliza la sensación, como dato básico a partir del cual iniciar las distintas etapas de cálculo de un algoritmo, puede suponer un avance en la descripción de las reglas de operación de las redes neurales artificiales, presentaremos un clasificador de patrones basado en estos principios, demostraremos, mediante simulación, como este clasificador puede realizar determinadas tareas con mayor grado de exactitud que el clasificador gaussiano, arquetipo de los clasificadores utilizados tradicionalmente. Y finalmente describiremos un experimento donde se pone de manifiesto como el clasificador propuesto por nosotros, es mucho más relevante para predecir y describir la conducta de los sujetos humanos que otros clasificadores.

El reconocimiento de patrones es un proceso en el que los patrones de un determinado espacio son clasificados en un conjunto discreto de clases de patrones (muchos de los mecanismos fundamentales de la memoria pueden ser descritos formalmente en términos de reconocimiento o clasificación de patrones); en una de sus formas más habituales, cada una de estas clases es definida por un patrón que actúa como elemento representativo de la clase y que viene dado previamente al inicio del proceso de clasificación. En este caso la tarea que debe resolverse es asignar un patrón U , que representa un input percibido por el sistema, a una de las K clases que vienen definidas por los patrones de referencia U_i , siendo $i = 1, 2, \dots, K$.

Uno de los métodos más directo e intuitivo de abordar esta tarea consiste en calcular la distancia entre el patrón incógnita U y los patrones representantes U_i , y asignar el input a aquella clase cuyo patrón presente la mínima distancia. O dicho en otros términos, como quiera que distancia y similitud son conceptos contrapuestos, se trata de asignar cada patrón a la clase cuyo representante sea más semejante. Llegados a este punto surge la cuestión fundamental, ¿Qué distancia utilizar? La respuesta no es única, y la solución idónea depende de la estructura probabilística que posea el espacio de patrones. Diferentes tipos de distancias se han propuesto, véase por ejemplo Duda y Hart (1973) o Kohonen (1989), aunque la mayoría de ellas se encuentran fuertemente relacionadas al basarse en el producto interno de los patrones. Dentro de esta categoría de distancias cabe destacar la euclídea, ampliamente utilizada en aplicaciones diversas, e implementada en redes de componentes electrónicos para tareas de clasificación de patrones Winters y Rose (1989).

La distancia euclídea entre dos patrones x e y se define como:

$$d_E(x, y) = \sqrt{\sum_j (x_j - y_j)^2}$$

El clasificador de patrones basado en esta distancia euclídea se denomina habitualmente clasificador gaussiano, y es óptimo cuando las diferencias entre los componentes de los patrones de la clase correspondiente siguen una distribución Normal de media cero y una determinada varianza. No obstante, basta que aparezcan desviaciones con respecto a la hipótesis de normalidad, relativamente pequeñas, o que la distribuciones de las diferencias presenten heterocedasticidad para que este clasificador deje de ser óptimo. Por otra parte, no todas las situaciones en que debe realizarse una clasificación de patrones responden al esquema en que las fluctuaciones con respecto al patrón de referencia de la clase, vienen determinadas por la adic-

ción de un término estocástico, con distribución Normal; existen, como veremos, situaciones que pueden responder a esquemas multiplicativos o de otro tipo.

Debido a las razones anteriormente expuestas, es por lo que el clasificador que nosotros proponemos, basado en una distancia logarítmica, puede ser mas adecuado en la resolución de algunas tareas, como justificaremos mas adelante.

La distancia logarítmica entre dos vectores x e y se define como:

$$d_L(x, y) = \sum_j \left| \log \frac{x_j}{y_j} \right|$$

Es trivial comprobar que dicha función es una distancia.

$$d_L(x, y) \geq 0$$

$$y$$

$$d_L(x, y) = 0 \Leftrightarrow x = y$$

En efecto, esta función está definida como la suma de términos no negativos, luego evidentemente tiene que ser no negativa. Al mismo tiempo como no pueden existir términos negativos, para que el sumatorio sea igual a cero, deben ser iguales a cero todos los sumandos, o lo que es equivalente que todo los cocientes x_j/y_j sean iguales a la unidad, lo que supone que todos los componentes de ambos patrones sean iguales.

$$d_L(x, y) = d_L(y, x)$$

$$d_L(x, y) = \sum_j \left| \log \frac{x_j}{y_j} \right| = \sum_j \left| \log x_j - \log y_j \right| = \sum_j \left| \log y_j - \log x_j \right| = \sum_j \left| \log \frac{y_j}{x_j} \right| = d_L(y, x)$$

$$(d_L(x, z) \leq d_L(x, y) + d_L(y, z))$$

$$d_L(x, z) = \sum_j \left| \log \frac{x_j}{z_j} \right| = \sum_j \left| \log x_j - \log z_j \right| = \sum_j \left| \log x_j - \log z_j + \log y_j - \log y_j \right| = \sum_j \left| \log \frac{x_j}{y_j} + \log \frac{y_j}{z_j} \right| \leq \sum_j \left(\left| \log \frac{x_j}{y_j} \right| + \left| \log \frac{y_j}{z_j} \right| \right) = d_L(x, y) + d_L(y, z)$$

Veamos ahora los argumentos psicofísicos que apoyan y fundamentan la utilización del tipo de distancia aquí propuesto.

Parece obvio que si lo que perciben los organismos son las sensaciones que les producen los inputs estimulares, a la hora de clasificar un input como perteneciente a una clase u otra deben de operar computando las distancias entre la sensación de ese input y las que producen los patrones de referencia.

Sea γ , la sensación que produce un input percibido u , y sean γ_i las sensaciones que producen los distintos patrones u_i , entonces si cada una es un vector con j componentes, podemos definir la distancia entre las sensaciones como:

$$d(\gamma, \gamma_i) = \sum_j \left| \gamma_j - \gamma_{ij} \right|$$

Se ha preferido esta distancia, suma de los valores absolutos de las diferencias entre los componentes, por ser de forma general menos sensible a los valores extremos, que la distancia euclídea.

Si aceptamos como válida la formula de medición establecida por Pechner (1860) tendríamos que

$$\gamma = k (\log \beta - \log b)$$

y por consiguiente

$$\begin{aligned} d(\gamma, \gamma_i) &= \sum_j |k(\log u_j - \log b_j) - \\ &- k(\log u_{ij} - \log b_j)| = \sum_j k |\log u_j - \\ &- \log u_{ij}| = k \sum_j \left| \log \frac{u_j}{u_{ij}} \right| \end{aligned}$$

lo que nos indica que clasificar un input haciendo mínima la distancia logarítmica a los patrones, es equivalente a clasificarlo óptimamente de acuerdo con las sensaciones que producen los inputs y patrones.

Para los lectores que no acepten como válida la formula de medición, especialmente la suposición de que el cambio en la sensación es igual para todas las mínimas diferencias perceptibles, pueden encontrarse argumentos que avalan la utilización de esta distancia logarítmica en los trabajos de Stevens.

En efecto si asumimos que sobre continuos perceptuales tales como brillo, volumen, peso, longitud, duración, etc. estímulos de igual proporción producen sensaciones de igual proporción (Stevens 1957) y por consiguiente la sensación es una función potencial de la magnitud física; y además utilizamos una escala logarítmica de intervalos, basándonos en la posibilidad de que la dispersión discriminial pueda incrementarse proporcionalmente a la magnitud psicológica y esto haga posible escalar el continuo en intervalos que sean iguales en términos de logaritmos (Stevens 1958). En

tal escala la distancia que se obtendría entre dos magnitudes psicológicas ψ_1 y ψ_2 sería

$$\begin{aligned} d(\psi_1, \psi_2) &= |\log \psi_1 - \log \psi_2| = \\ &= \log \left| \frac{\psi_1}{\psi_2} \right| \end{aligned}$$

si ahora consideramos que la magnitud psicológica es una función potencial de las magnitudes físicas ϕ_1 y ϕ_2 tendríamos que

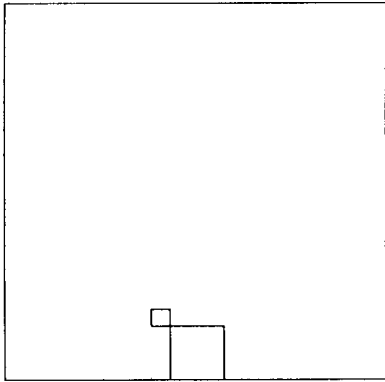
$$\begin{aligned} d(\psi_1, \psi_2) &= \log \left| \frac{k \Phi_1^n}{k \Phi_2^n} \right| = \\ &= n \log \left| \frac{\Phi_1}{\Phi_2} \right| \end{aligned}$$

Por consiguiente nuevamente llegamos a la conclusión de que clasificar el input teniendo en cuenta la distancia logarítmica es equivalente a clasificarlo de acuerdo con las sensaciones que produce.

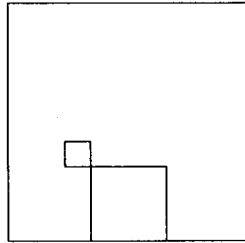
SIMULACION

Mediante simulación en ordenador hemos comparado las características de ejecución de una tarea por dos clasificadores de patrones, uno el tradicional clasificador gaussiano, basado en la distancia euclídea y otro basado en una distancia logarítmica.

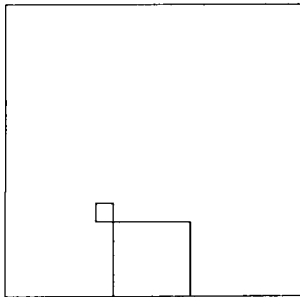
La tarea a resolver consistía en clasificar una entrada, generada aleatoriamente, en uno de los cuatro patrones definidos previamente. Los patrones fijados inicialmente eran cuatro gráficos, véase figura 1, constituidos cada uno de ellos por tres cuadrados de diferentes áreas. Los valores de los componentes del vector que define cada patrón son las áreas de los cuadrados que lo forman.



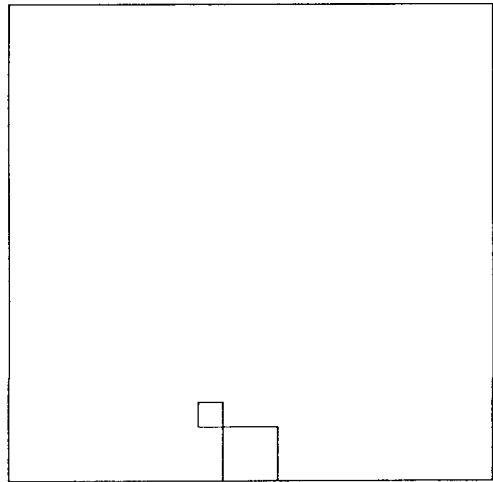
(0,1;1;50)



(0,2;2;20)



(0,1;2;30)



(0,2;1;80)

Figura 1. Gráficos correspondientes a los patrones

Las entradas a ambos clasificadores en cada prueba eran generadas seleccionando aleatoriamente un patrón y modificando sus componentes aleatoria e independientemente de acuerdo con una distribución de probabilidad.

Parece bastante obvio que el esquema tradicional de generar las entradas sumando a cada componente del patrón seleccionado

un término estocástico que sigue una distribución Normal de media cero y de varianza σ^2 , igual para todos los componentes, no es el más adecuado para esta tarea. Si la varianza fuese lo suficientemente importante como para producir cambios perceptibles en el mayor de los tres cuadrados que componen la figura, entonces ocurriría que los cambios en el menor de los cuadrados se-

rían enormes si el término de error fuese positivo y si fuese negativo haría que el área a representar fuese negativa lo cual es imposible. Por el contrario, si la varianza fuera de un orden de magnitud adecuada al menor de los componentes las alteraciones que experimentarían los otros dos serían inapreciables.

Todo lo anterior, parece indicar que si queremos, de algún modo, reflejar en nuestra tarea situaciones simplificadas, pero que reflejen los elementos esenciales de tareas que se presentan en el mundo real, debemos de considerar modificaciones de los componentes que sean proporcionales a sus valores.

En concreto, en nuestra simulación, cada componente era incrementado o disminuido proporcionalmente a su valor de acuerdo con el siguiente esquema:

En primer lugar se determinaba aleatoriamente y con probabilidades iguales a 0,5, si la componente sería incrementada o disminuida, el nuevo valor era obtenido multiplicando o dividiendo por $1+z$ el valor original. La cantidad z que determina la proporción de incremento o decremento era a su vez generada aleatoriamente de acuerdo con una distribución de probabilidad que hace que cuanto mayor sea la proporción de modificación menos probable sea esta, concretamente la función de densidad de esta distribución es:

$$f(z) = a \exp(-az)$$

En nuestra simulación se realizaron 1.000 pruebas con cada conjunto de patrones, siendo el parámetro de la distribución, a igual a 3. Los resultados obtenidos con los cuatro patrones representados en la figura 1 y otros patrones análogos, aparecen en la tabla 1.

Como puede verse, en todos los casos el clasificador basado en una distancia lo-

garítmica ha demostrado ser mas adecuado para resolver este tipo de tareas. Los porcentajes en que se mejora la ejecución de la tarea varían de acuerdo con las diferencias en el orden de magnitud de los distintos componentes y la dificultad de diferenciar los patrones, pero en cualquier caso hay una mejora sustancial en la ejecución de la tarea.

METODO

Sujetos.

Los Ss fueron 33 voluntarios de ambos sexos, estudiantes de segundo de Psicología de la Universidad de Granada.

Procedimiento

Se presentaron 10 láminas, en cada una de las cuales había tres patrones estímulares formados por tres cuadrados que variaban en su área, pero no en la disposición de éstos. Dos de los patrones constituían los patrones de clasificación y se situaron en la parte superior de la lámina, en la parte inferior se situaba el otro que denominamos patrón de prueba (como ejemplo veáse la primera lámina, figura 2).

La tarea de los Ss consistía en indicar a cual de los dos patrones de clasificación se asemejaba mas el patrón de prueba.

Las áreas de los cuadrados que constituían los patrones se manipularon de forma que la respuesta dada fuera diferente en función de si los Ss utilizaban para la clasificación la distancia euclídea o por el contrario utilizaban la logarítmica, exceptuando dos casos (los que corresponden a las láminas 5 y 6) en los que el resultado de la clasificación era el mismo independientemente de la distancia utilizada. Los valores de las áreas de los cuadrados en centímetros cuadrados se indican en la tabla 2.

La posición de los patrones de clasificación se aleatorizó, de manera que aquél cuya distancia logarítmica al de prueba era

Patrones	Clasificador gaussiano	Clasificador "psicofísico"
(1;5)	724	783
(2;2)		
(2;8)		
(1;3)		
(0,1;5)	636	764
(0,2;2)		
(0,2;8)		
(0,1;3)		
(0,01;5)	631	786
(0,02;2)		
(0,02;8)		
(0,01;3)		
(0,1;1;50)	656	839
(0,2;2;20)		
(0,2;1;80)		
(0,1;2;30)		
(0,01;0,1;5)	647	827
(0,02;0,2;2)		
(0,02;0,2;8)		
(0,01;0,1;3)		
(0,01;0,1;5)	651	858
(0,02;0,2;2)		
(0,02;0,1;8)		
(0,01;0,2;3)		

Tabla 1. Resultados de la simulación

menor, se situaba a veces en el lado derecho y otras veces en el izquierdo; también se aleatorizó el orden de presentación de las láminas. Otra variable que se tuvo en cuenta fue la dificultad de la elección, que venía dada por la diferencia entre las distancias del patrón de prueba a cada uno de los patrones de clasificación, de manera que

cuanto menor fuese esa diferencia mas difícil sería la elección.

Una vez realizada la prueba y con objeto de facilitar la exposición posterior de los resultados, consideramos como respuesta correcta aquella elección que coincide con la que esperamos en función del cálculo de la distancia logarítmica.

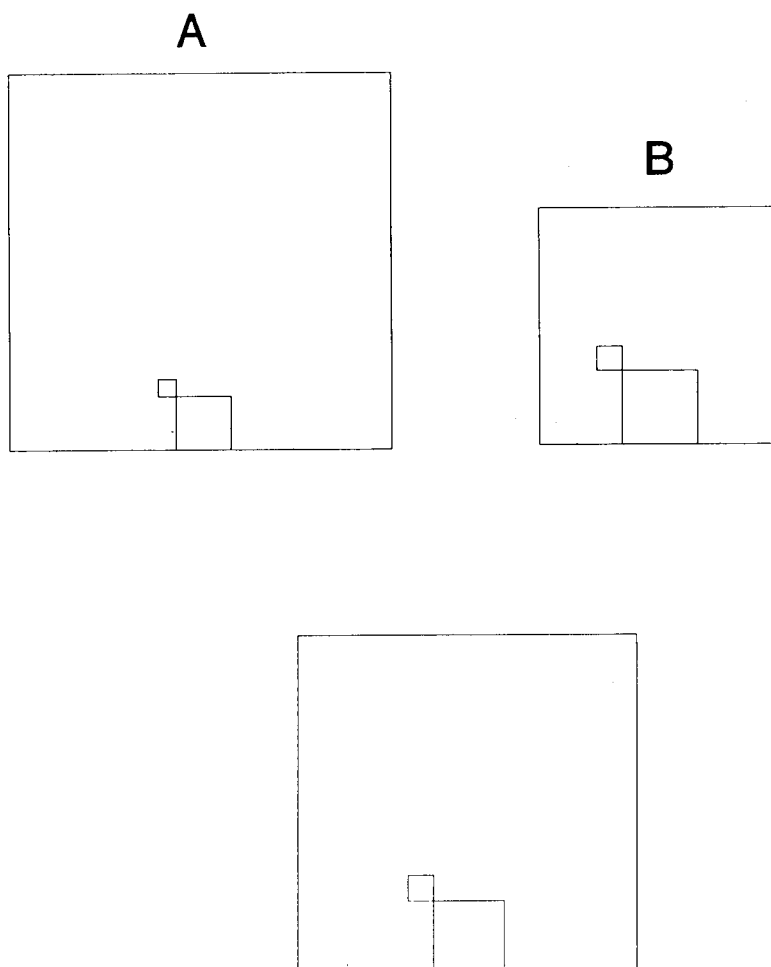


Figura 2. Lámina nº 1

RESULTADOS

Los resultados obtenidos se presentan en la tabla 3, en la que podemos observar que todos los Ss excepto dos han conseguido puntuaciones superiores a 5, siendo la media

de estas de 7,88, lo cual es indicativo de la utilización de la distancia logarítmica como criterio para la clasificación de patrones.

Los datos también se han organizado en función de la dificultad de cada una de las elecciones, contabilizándose el número

Lámina	Patrón izda.	Patrón Dcha.	Patrón Prueba
1	(0,1; 1; 50)	(0,4; 2; 20)	(0,3; 1,7; 40)
2	(0,1; 1; 50)	(0,4; 2; 20)	(0,16; 1,2; 30)
3	(0,4; 2; 20)	(0,1; 1; 50)	(0,09; 0,0; 30)
4	(0,4; 2; 20)	(0,1; 1; 50)	(0,5; 2,2; 40)
5	(0,1; 1; 50)	(0,4; 2; 20)	(0,16; 1; 6,40)
6	(0,4; 2; 20)	(0,1; 1; 50)	(0,50; 1,4; 30)
7	(0,1; 1; 30)	(0,5; 2; 50)	(0, 2; 1,2; 45)
8	(0,1; 1; 30)	(0,5;2;20)	(0,4; 1,8; 35)
9	(0,5; 2; 50)	(0,1; 1; 30)	(0,09; 1,6; 45)
10	(0,5; 2; 50)	(0,1; 1; 30)	(0,6; 1,2; 36)

Tabla 2. Patrones correspondientes a las figuras de las láminas.

de errores para cada elección; como queda indicado en la tabla 4 y sobre todo en la figura 3, la relación entre ambas variables tiene una tendencia lineal, de manera que a mayor dificultad, mayor es el número de Ss que cometen errores.

DISCUSION

En cuanto a los resultados de la simulación, parece fuera de toda duda la superioridad del clasificador basado en la distancia logarítmica sobre el clasificador gaussiano, en la condiciones consideradas;

entradas a los clasificadores generadas a partir de un patrón que sufre alteraciones aleatorias en sus componentes, proporcionales a la magnitud de los mismos.

Por consiguiente, la discusión debe centrarse en la capacidad del esquema probabilístico de contaminación para reflejar adecuadamente el entorno natural. Como ya se indicó, al exponer la tarea concreta utilizada en nuestra simulación, un esquema aditivo de contaminación con términos que se distribuyesen normalmente sería inadecuado, ya que conduciría a la aparición de áreas negativas lo cual carece de sentido.

Puntuaciones	10	9	8	7	6	5	4	3	2	1	0
Nº sujetos	7	7	7	7	3	—	—	—	2	—	—

Tabla 3. Distribución de las puntuaciones de los sujetos

Lámina	Diferencias entre distancias	Número de errores
3	1,98	4
4	1,61	2
6	1,48	5
8	1,44	3
9	1,06	4
10	0,95	8
1	0,71	12
5	0,67	6
2	0,67	10
7	0,25	16

Tabla 4. Número de fallos en cada lámina ordenadas según su dificultad

La cuestión trascendente es si de forma general puede considerarse que el esquema multiplicativo de contaminación refleja con mayor fidelidad la realidad circundante a los sistemas biológicos que el esquema aditivo con residuos distribuidos normalmente.

Parece probado que de forma general el ruido que se produce en un sistema es proporcional a la magnitud de la señal que se procesa en el mismo, lo cual avalaría nuestras propuestas en favor de un clasificador basado en la distancia logarítmica. Por otra parte, si tomamos como árbitro de la mejor adaptabilidad a la naturaleza, la evolución de los sistemas biológicos, tenemos como prueba irrefutable de la superioridad del clasificador "psicofísico" el resultado de nuestro experimento, que pone de manifiesto el hecho de que los sujetos humanos utilizan un criterio de clasificación que conduce a resultados análogos a los que se obtienen utilizando una distancia logarítmica.

En efecto, creemos que la afirmación anterior queda suficientemente probada por los resultados de la tabla 3, donde se observa que el 85% de los sujetos tienen puntuaciones superiores a 7, lo que indica una clara tendencia a utilizar la mínima distancia logarítmica como criterio de clasificac-

ción. Mas aún, cuando se analizan los casos en que los sujetos se han apartado del criterio anteriormente enunciado, puede observarse que corresponden a aquellos elementos que presentan una mayor dificultad, en términos de distancia logarítmica entre los patrones de referencia, véase figura 3, lo cual permite atribuir estas desviaciones del criterio a simples errores de los sujetos.

En base a los argumentos anteriormente expuestos, creemos poder afirmar que el clasificador basado en la distancia logarítmica será mas adecuado que los tradicionales clasificadores basados en la distancia euclídea, u otras relacionadas con ella, para resolver aquellas tareas complejas de clasificación de patrones, como reconocimiento de caracteres manuscritos, reconocimiento de rostros, etc. en las cuales los sistemas biológicos se han mostrado, hasta el momento, claramente superiores.

Por otra parte, la relativa facilidad con que este clasificador puede ser implementado en una red de resistores, con tecnología VLSI, (Salinas y Dobson 1991) hace que se le pueda considerar como una alternativa prometedora para la construcción de autómatas capaces de realizar una clasificación óptima de las señales que reciben.

La razón básica que justifica el mejor comportamiento del clasificador “psicofísico” en múltiples situaciones radica en el hecho de que este clasificador opera haciendo mínima una función de los errores relativos, mientras que los restantes clasificadores tratan de minimizar una función de los errores absolutos. De esta forma, en todas aquellas tareas en que el error relativo sea un mejor índice de la similitud entre patrones, el clasificador basado en la distancia logarítmica se mostrará superior a las alternativas tradicionales. Esta característica se pondrá especialmente de manifiesto cuando los componentes de los patrones presenten órdenes de magnitud muy diferentes.

En aquellos casos en que los componentes de los patrones presentan magnitudes semejantes minimizar los errores relativos o los errores absolutos es prácticamente

equivalente y por consiguiente ambos tipos de clasificadores conducen a resultados análogos. Por el contrario, cuando los componentes de los patrones difieren considerablemente en sus magnitudes, los criterios de error relativo y de error absoluto pueden diferir radicalmente. Esta es la razón de que en nuestra simulación los mayores porcentajes de mejora de las características de ejecución del clasificador “psicofísico” sobre el clasificador gaussiano se encuentren en aquellos casos en que las diferencias en el orden de magnitud de los componentes de los patrones son mayores.

Una ventaja adicional que presenta el trabajar con un clasificador basado en una distancia logarítmica es la propiedad que posee esta distancia de ser invariable frente a cambios de escala, lo cual nos hace pensar que con una adecuada definición de los

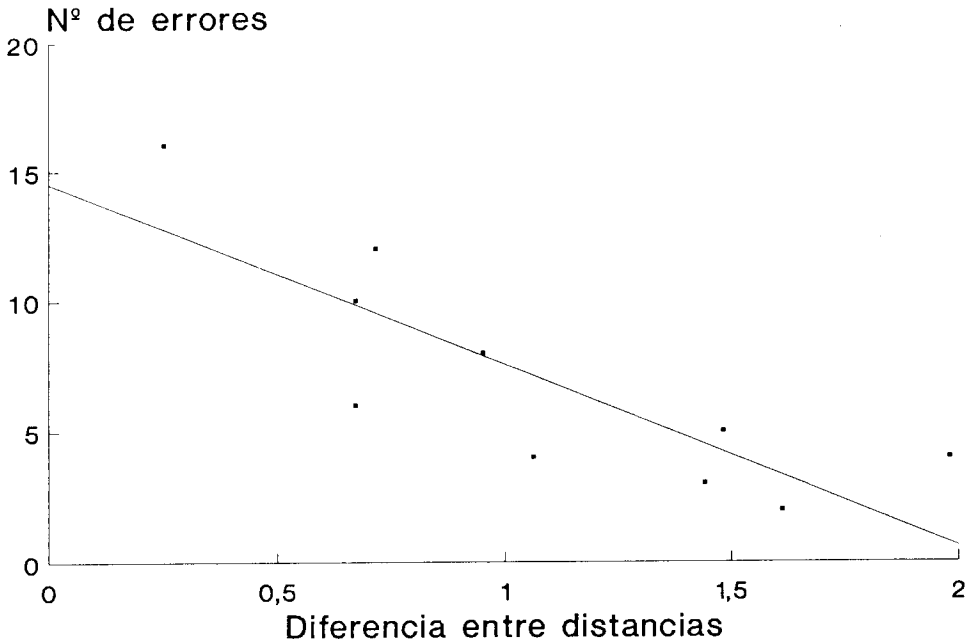


Figura 3 Tendencia del nº de errores en función de la dificultad.

patrones, un clasificador basado en esta distancia podría poseer las propiedades de invarianza que presentan los sistemas biológi-

cos y reproducir efectos como la constancia del color o la invariabilidad frente a traslaciones homotéticas.

REFERENCIAS

- Duda, R.O. y Hart, P.E. (1973). *Pattern Classification and Scene Analysis*. Londres: John Wiley.
- Fechner, G. T. (1964). *Elements of Psychophysics*. En: G.A. Miller Mathematics and Psychology. New York: John Wiley (Orig. 1860).
- Kohonen, T. (1989) *Self-Organization and Associative Memory*. Berlín: Springer-Verlag.
- Salinas, J.M. and Dobson, V.G. (1991) Implementing a "psychophysical" Pattern Classifier in a Decrementing Network. En: A. Prieto (Ed.), *Artificial Neural Networks* (pp.116-123). Berlin: Springer-Verlag.
- Stevens, S.S. (1957) On the Psychophysical law. *Psychological Review*, 64, 153-181.
- Stevens, S.S. (1958) Problems and Methods of Psychophysics. *Psychological Bulletin*, 55, 177-196.
- Winters, J.H. and Rose, C. (1989) Minimum Distance Autmata in Parallel Networks for Optimum Classification. *Neural Networks*, 2, 127-132.