



## 2.1 Redes Neuronales Artificiales.

### 2.1.1 La neurona Biológica.

Una neurona biológica es una célula especializada en procesar información. Está compuesta por el cuerpo de la célula (soma) y dos tipos de ramificaciones: el axón y las dendritas (ver figura 2.1). La neurona recibe las señales (impulsos) de otras neuronas a través de sus dendritas y transmite señales generadas por el cuerpo de la célula a través del axón.

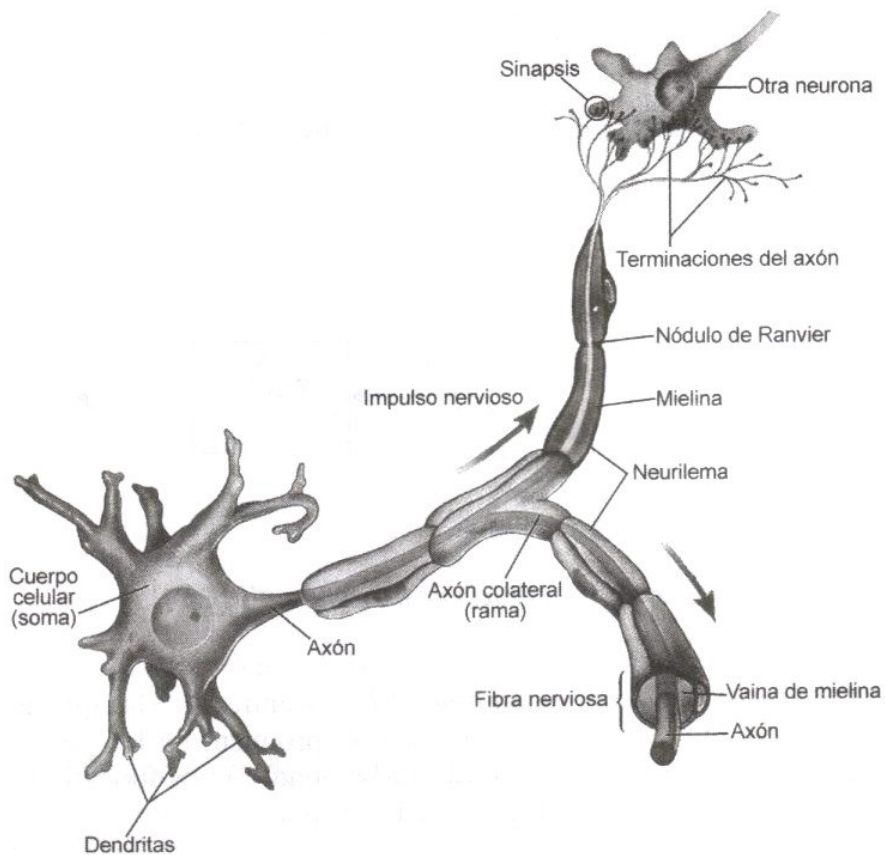


Figura 2.1 Neurona Biológica [Faúndez 00]

El córtex cerebral humano contiene alrededor de  $10^{11}$  neuronas. Cada neurona está conectada de forma masiva a un número variable de entre 1000 y 10,000 neuronas. Por tanto, el cerebro humano contiene de  $10^{14}$  a  $10^{15}$  interconexiones [Faúndez 00].

En el córtex cerebral se aprecia la existencia de una organización horizontal en capas (se suelen señalar unas seis capas), coexistiendo una organización vertical en forma de columnas. Hay grupos neuronales compuesto por millones de neuronas pertenecientes a una determinada región del cerebro, que constituyen unidades funcionales, especializadas en ciertas tareas (por ejemplo, existe un área visual, un área auditiva, un córtex sensorio-motor, etc.) [Martí & Sanz 01].

Las neuronas se comunican entre sí mediante trenes de pulsos de corta duración (del orden de milisegundos). El mensaje está modulado en la frecuencia de transmisión de los pulsos. Esta frecuencia varía sobre los 100 Hz. Por tanto, es un millón de veces inferior a la velocidad de conmutación de los circuitos electrónicos. Sin embargo, el ser humano es capaz de realizar tareas complejas en un tiempo inferior y con un porcentaje de aciertos superior al conseguido actualmente mediante computadoras. Por ejemplo tareas de reconocimiento del habla. Ello es debido a la paralelización masiva que realiza el cerebro, que está extremadamente lejos de poder ser implementada en un circuito electrónico [Faúndez 00].

### **2.1.2 Definición de una Red Neuronal.**

Existen diferentes formas de definir que son las redes neuronales algunas de ellas son:

- Un modelo matemático compuesto por un gran número de elementos procesales organizados en niveles [Hilera & Martínez 00].
- Las redes neuronales artificiales son modelos matemáticos inspirados en sistemas biológicos, adaptados y simulados en computadoras convencionales [Wasserman 89].

- “Redes neuronales artificiales son redes interconectadas masivamente en paralelo de elementos simples (usualmente adaptativos) y con organización jerárquica, las cuales intentan interactuar con los objetos del mundo real del mismo modo que lo hace el sistema nervioso biológico” [Kohonen 88].

Una red Neuronal Artificial es un sistema de procesamiento de información que tiene ciertas características de comportamiento en común con las Redes Neuronales Biológicas:

- Capacidad de aprendizaje.
- Capacidad de almacenamiento y recuperación de información.
- Almacenamiento distribuido
- Capacidad de relación
- Paralelismo
- Tolerancia a la falla
- Capacidad de generalización

### **2.1.3 La Neurona Artificial.**

El primer modelo de una neurona se dio alrededor de los años 40's, y fue llevado a cabo por Mc-Culloch y Pitts. El propósito del modelo es el siguiente: una neurona artificial forma una suma de los potenciales de entrada que llegan a ella (cada de estos potenciales es un valor numérico que representa el estado de la neurona que lo ha emitido) y se activa dependiendo del valor de esta suma de los potenciales. Si la suma excede a cierto umbral, la neurona es activada y transmite una repuesta, el cuál es el valor de activación. [Davaló & Naïm 92].

Esta neurona artificial es un autómata booleano, esto es, que sus entradas y sus salidas son valores booleanos (0, 1).

Escribimos:

$(e_i)_{i=1,n}$  son las entradas para la neurona artificial.

S es la Salida de la neurona artificial

$\beta$  es el umbral

$W_i$  son los pesos

$f$  es la función umbral mostrada en la figura 2.2.

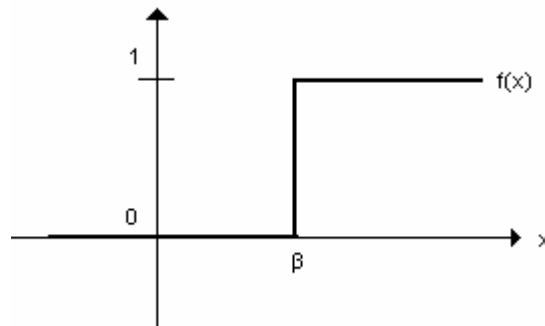


Figura 2.2. Función de activación [Haykin 94].

La Salida de la neurona artificial viene dada por:

$$S = f\left(\sum_{i=1,n} W_i e_i\right) \quad (2.1)$$

#### 2.1.4 Modelo General.

La estructura genérica de neurona artificial en el marco establecido por el grupo PDP (*Parallel Distributed Processing*) [Martí & Sanz 01]. Se denomina procesador elemental o neurona a un dispositivo simple de cálculo que, a partir de un vector de entrada procedente del exterior o de otras neuronas, proporciona una única respuesta o salida. Los elementos que constituyen la neurona de etiqueta  $i$  son los siguientes:

- Conjunto de entradas,  $x_j(t)$ .
- Pesos sinápticos de la neurona  $i$ ,  $w_{ij}$  que representa la intensidad de interacción entre cada neurona presináptica  $j$  y la neurona postsináptica  $i$ .
- Regla de propagación  $\sigma(w_{ij}, x_j(t))$ , que proporciona el valor del potencial postsináptico  $h_i(t) = \sigma(w_{ij}, x_j(t))$  de la neurona  $i$  en función de sus pesos y entradas.
- Función de activación  $f_i(a_i(t-1), h_i(t))$ , que proporciona el estado de activación actual  $a_i(t) = f_i(a_i(t-1), h_i(t))$  de la neurona  $i$ , en función de su estado anterior  $a_i(t-1)$  y de su potencial postsináptico actual.
- Función de salida  $F_i(a_i(t))$ , que proporciona la salida actual  $y_i(t) = F_i(a_i(t))$  de la neurona  $i$  en función de su estado de activación.

De manera general, la operación de la neurona  $i$  puede expresarse como

$$y_i(t) = F_i(f_i[a_i(t-1), \sigma_i(w_{ij}, x_j(t))]) \quad (2.2)$$

## 2.2 Red de Kohonen.

*Teuvo Kohonen*, profesor de la Facultad de Ciencias de la Información (Universidad de Helsinki), presentó en 1982 un modelo de red neuronal con capacidad para formar mapas de características de manera similar a como ocurre en el cerebro. Con este modelo intenta simular los mapas topológicos de los fenómenos sensoriales y motores existentes en el cerebro, a través de una organización matricial de neuronas artificiales. [Rogers 97]

Esta red de tipo Auto-Organizado, esto es, que se organizan por sí misma y está concebida para clasificar conjuntos de datos para los que no se conoce *a priori* ningún tipo de organización. La red, a partir de un proceso de autoorganización, proporciona un resultado,

el cual depende de la relación de similitud existente entre dichos patrones de entrada. El tipo de aprendizaje es no supervisado. Algunas de las características

- Los datos deben tener un grado de redundancia elevado para realizar su clasificación.
- La red divide el conjunto de datos en distintos subconjuntos (*clusters*), cada uno de los cuales agrupa a datos similares, con algún tipo de característica en común (*clustering*).
- El desarrollo de un método de *clustering* requiere elaborar alguna medida de la semejanza entre los datos (distancia euclidiana, Correlación, etc.).
- Cada *cluster* se representa por un prototipo: Elemento más representativo del *cluster*.

La estructura de la red esta formada solo por 2 capas o niveles.

Por otra parte la red de *Kohonen* también es una red competitiva. Esta idea se basa en la existencia de una cierta competitividad entre las neuronas de una cierta capa por la oportunidad de entrenarse. Esto, se refiere a que la neurona que produce la salida mayor se le considera Ganadora, y tiene la capacidad de inhibir a las otras neuronas (no presentan activación). Todo ello conlleva a que solamente los pesos de la neurona ganadora serán ajustados.

### **2.2.1 Arquitectura.**

En la arquitectura de la versión original LVQ (Learning Vector Quantization) del modelo Kohonen no existen conexiones hacia atrás. Se trata de una de las N neuronas de entrada y

M de salida. Cada una de las N neuronas de entrada se conecta a las M de salida a través de conexiones hacia adelante (*feedforward*).

Entre las neuronas de la capa de salida, puede decirse que existen conexiones laterales de inhibición (peso negativo) implícitas, pues aunque no estén conectadas (figura 2.3), cada una de las neuronas va a tener cierta influencia sobre sus vecinas. El valor que se asigne a los pesos de las conexiones hacia adelante entre las capas de entrada y salida ( $W_{ji}$ ) durante el proceso de aprendizaje de la red va a depender precisamente de esta interacción lateral.

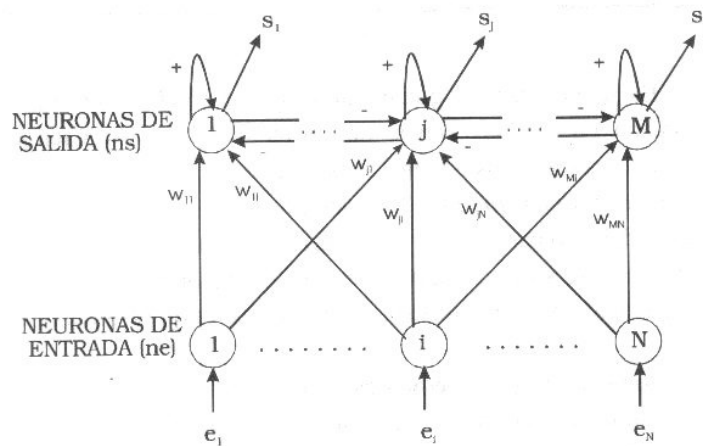


Figura 2.3. Diagrama de una Red LVQ [Hilera & Martínez00].

La influencia que una neurona ejerce sobre las demás es función de la distancia entre ellas, siendo muy pequeñas cuando están muy alejadas. Es frecuente que dicha influencia tenga la forma de un sombrero mexicano (figura 2.4).



Figura 2.4. Grafica del Sombrero Mexicano [Hilera & Martínez00].



Por otra parte, la versión del modelo denominado TPM (*Topology Preserving Map*) o SOFM (*Self-Organizing Feature Maps*) trata de establecer una correspondencia entre los datos de entrada y un espacio bidimensional de salida, creando mapas topológicos de dos dimensiones, de tal forma que ante datos de entrada con características comunes se deben activar neuronas situadas en próximas zonas de la capa de salida [Hilera & Martínez00].

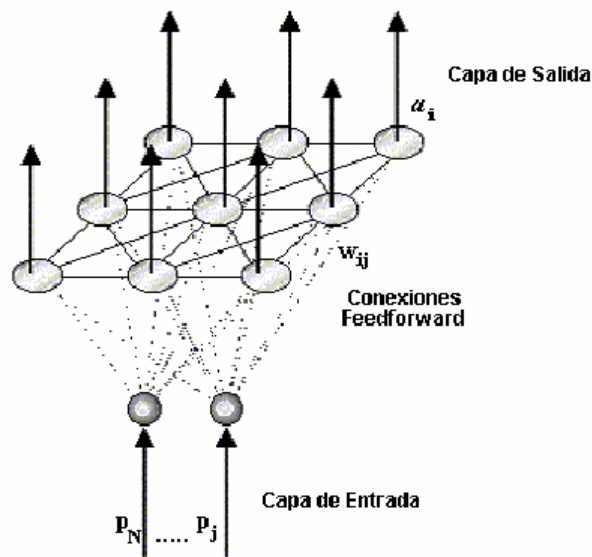


Figura 2.5 Diagrama de una Red SOM [Hilera & Martínez00].

La Figura 2.5, muestra una red SOM donde M neuronas de salida se disponen de forma bidimensional para representar los mapas de características.

Una limitación de la red es que algunas neuronas pueden no ser entrenadas. En otras palabras, pueden existir vectores de pesos muy distanciados de las entradas, con lo cual, nunca ganarán.

Cuando la medida de similitud que se utiliza para asignar un patrón de entrada a una determinada región es la distancia euclidiana, se produce un diagrama de *Voronoi*, el cual se define como.

### Diagrama de Voronoi

El conjunto de puntos de  $\mathbb{R}^n$  que están más cerca de  $y_i$ , que de los restantes prototipos forma un poliedro (polígono en el plano) que se denomina diagrama de *Voronoi* [Reinoso 02]. Este poliedro está limitado por los hiperplanos bisectores de los pares de puntos  $(y_i, y_j)$  para  $i \neq j$ . (ver figura 2.6).

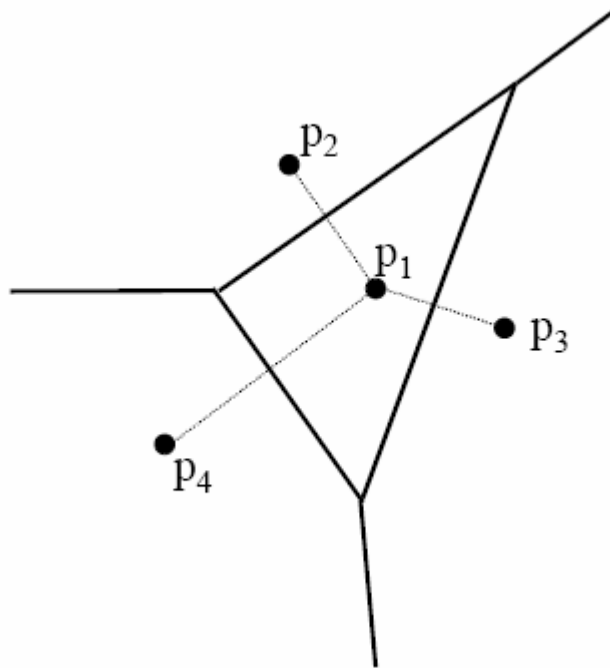


Figura 2.6. Diagrama de *Voronoi* [Reinoso 02].

#### 2.2.2 Algoritmo de Aprendizaje para la Red SOM [Martín & Sanz 01].

- 1) Inicialización de los pesos sinápticos  $w_{ijk}$ . Se puede partir en  $t = 0$  de diferentes configuraciones: pesos nulos, aleatorios de pequeño valor absoluto, o con un valor de partida predeterminado y se fija la zona inicial de vecindad entre las neuronas de salida.
- 2) En cada iteración, presentación de un patrón  $x(t)$  tomado de acuerdo con la función de distribución  $p(x)$  del espacio sensorial de entrada (en la muy habitual situación

de disponer solamente de un conjunto finito de patrones de entrenamiento, basta con tomar al azar uno de ellos y presentarlo a la red).

- 3) Cada neurona  $i \equiv (i, j)$  en paralelo del mapa se calcula la similitud entre su vector de pesos sinápticos  $w_{ij}$  y el actual vector de entradas  $x$ . Un criterio de medida de similitud muy utilizado es la distancia euclidiana:

$$d^2(w_{ij}, x) = \sum_{k=1}^n (w_{ijk} - x_k)^2 \quad (2.3)$$

- 4) Determinación de la neurona ganadora  $g \equiv (g1, g2)$ , cuya distancia sea la menor de todas.

- 5) Actualización de los pesos sinápticos de la neurona ganadora  $g = (g1, g2)$ , y los de sus neuronas vecinas (ver figura 2.7). Una característica importante es que ésta reduce el número de elementos que la conforman conforme avanza el tiempo (para la figura 2.7 mostrada se considera  $0 < t1 < t2$ ). La regla más empleada es

$$w_{ijk}(t+1) = w_{ijk}(t) + \alpha(t) \cdot h(|i-j|, t) \cdot (x_k(t) - w_{ijk}(t)) \quad (2.4)$$

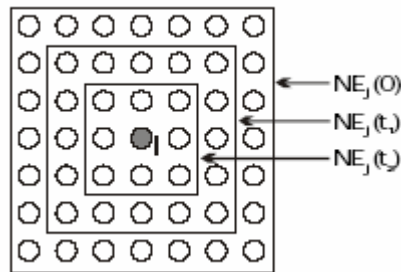


Figura 2.7 Vecindad para diferente  $t$  en una SOM [Hilera & Martínez00].

- 6) Si se ha alcanzado el número máximo de interacciones establecido, entonces el proceso de aprendizaje finaliza. En caso contrario, se vuelve al paso 2.

### **2.2.3 Aplicaciones de la Red de Kohonen.**

El modelo de *Kohonen* es muy utilizado en aplicaciones como: reconocimiento de patrones (voz, texto, imágenes, señales, etc.), codificación de datos y resolución de problemas de optimización como el agente viajero [Hilera & Martínez00]. Enseguida se describen algunas de sus aplicaciones asociadas en el reconocimiento de patrones.

#### **Reconocimiento de voz.**

El trabajo demuestra los resultados en el reconocimiento de voz de un individuo, tomando como patrones de a ser reconocidos las cifras decimales (0 - 9), y utilizando una red de *Kohonen*. Produce con 100 neuronas un resultado de reconocimiento del (65%), un aceptable comportamiento de la red neuronal de *Kohonen* frente la tarea de reconocimiento de un patrón tan estocástico como lo es la voz. Aunque su tasa de acierto dista aún mucho de las obtenidas mediante otros métodos, existen muchas variables en la red neuronal que pueden ser ajustadas para lograr mejores comportamientos de la misma; entre ellas, se pueden considerar:

- Número de neuronas de la red.
- Dimensión de los patrones y de dichas neuronas.
- Cantidad de locutores y de archivos sonoros utilizados en el entrenamiento.
- Método de codificación o mapeo elegido para la aplicación
- Valores de los coeficientes de aprendizaje y sintonización, y forma de decaimiento en el tiempo.

Todos estos factores interrelacionados contribuirían a mejorar la ejecución de la red [Merlo, Fernández, Caram, Priegue y García 97].

## Mapa fonético

Es un ejemplo muy conocido de aplicación de los SOM, en el que un mapa auto-organizado se entrena con los fonemas (en realidad con su transformada de *Fourier*) de un cierto idioma (en el trabajo original se utilizó el finlandés). Tras la fase de aprendizaje, los vectores de entrada son presentados a la red, obteniéndose como resultado final un mapa fonético (ver figura 2.8) en el que cada neurona se ha especializado en reconocer un tipo de fonema, y en el que sonidos similares o pequeñas variaciones quedan representados sobre neuronas vecinas [Martín & Sanz 01].

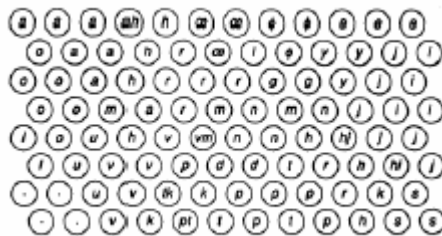


Figura 2.8 Mapa fonético [Martín & Sanz 01].

## El uso de Mapas Autoorganizativos en el Reconocimiento de clases en unidades de llanto Infantil.

Este trabajo utiliza algunos aspectos relevantes del llanto infantil extraídos del análisis acústico del mismo, con el propósito de su posible clasificación en normales y patológicos. Los parámetros empleados son todos cuantitativos e independientes de la valoración subjetiva del especialista médico. Estos parámetros son: Primera unidad de llanto, Segunda unidad de llanto, Segunda latencia, valor máximo del tono, valor mínimo del tono, valor promedio del tono, valor máximo del primer formante, valor mínimo del primer formante, valor promedio del primer formante, valor promedio de la primera unidad de llanto y valor

promedio de la segunda unidad de llanto. Para la clasificación se utiliza una Red Neuronal de *Kohonen* (Mapas Autoorganizativos). Se usó una muestra de 27 casos para entrenar la red y otra muestra de 5 casos para verificar el aprendizaje. Los resultados obtenidos son satisfactorios, pues la red entrenada pudo clasificar (diagnosticar) con acierto la totalidad de los nuevos casos presentados [Torre, Castillo, Cano y Escobedo 00]..

### 2.3 Clasificación por medio del Vecino Más Cercano.

Diremos que el vecino más cercano de un punto A dentro de un conjunto de puntos S será B si y solo si la distancia considerada entre dichos puntos es la mínima distancia existente entre el punto A y cualquier otro punto dentro del conjunto S:

$$\text{dist}(A,B) = \min \text{dist}(A,C). C \in \{S - A\} \quad (2.5)$$

Cabe señalar el hecho de que si un punto B es el vecino más cercano a otro punto A, no necesariamente habrá de ser A el vecino más cercano al punto B. Cuando un par de puntos A, B cumplen que A es el vecino más cercano a B y a su vez B es el vecino más cercano a A diremos que el par (A, B) es un par recíproco.

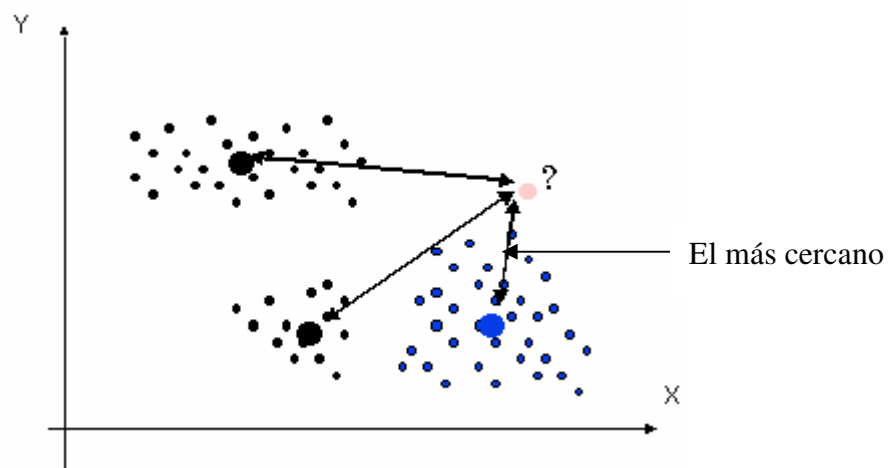


Figura 2.9 Representación del vecino más cercano [Cruz, Nieto, Tristacho 03].

Otra propiedad de esta clasificación es que un punto podrá ser el vecino más cercano de varios puntos dentro de un conjunto por lo que no podremos entender esta relación como una función.

### 2.3.1 Distancia Euclidiana.

El reconocimiento basado en la distancia euclidiana parte de dos hipótesis. En primer lugar, que las clases son de naturaleza determinística, y en segundo lugar, que toda la información necesaria y suficiente para su diseño se encuentra disponible *a priori*.

Se dice que una clase (Objetos predefinidos) es determinística cuando sus elementos se representan por un único vector, que es el prototipo de la clase.

Supóngase que existen  $N$  clases  $\alpha_1, \alpha_2, \dots, \alpha_N$  con sus respectivos prototipos,  $Z_1, Z_2, \dots, Z_N$ .

Sea  $X$  un vector de característica a clasificar; el reconocedor basado en la distancia asociará el vector  $X$  a la clase cuyo prototipo esté más próximo a él; es decir, aquél cuya distancia sea menor. En la figura 2.10 puede verse la representación de este concepto.

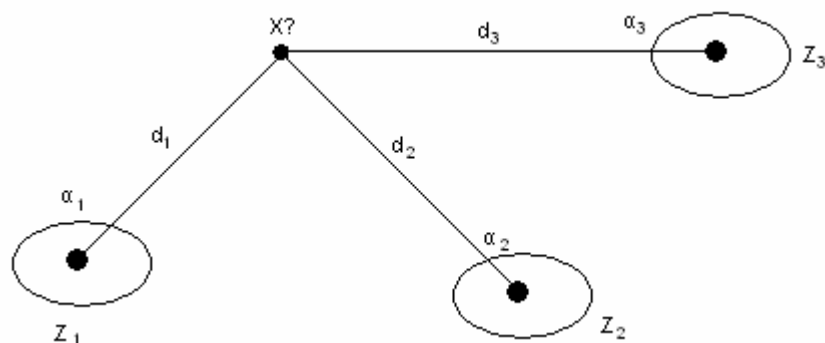


Figura 2.10. Ejemplo de distancia euclidiana [Maravall 93]

Se dijo anteriormente que para el diseño del reconocedor por la distancia se opera con todo el conocimiento *a priori* necesario y suficiente, esto significa que es preciso conocer previamente todos los prototipos de las clases.

La *distancia Euclídea (al cuadrado)* entre  $X_i$  y  $X_j$  se define como:

$$\delta_E^2(X_i, X_j) = \|X_i - X_j\|^2 = (X_i - X_j)^T (X_i - X_j) \quad (2.6)$$

### 2.3.2 Distancia Euclidiana respecto al Centro de Masa (Centroide).

Dado un conjunto de elementos de una cierta clase  $\alpha_i = \{X_{i1}, X_{i2}, \dots, X_{iP}\}$ , se suele tomar como prototipo su media ponderada:

$$Z_i = \frac{1}{P} \sum_{j=1}^P X_{ij} \quad (2.7)$$

La distancia euclidiana entre dos puntos  $i, j$ , que es la distancia entre ellos, está dada por:

$$d_E(i, j) = \sqrt{(X_{i1} - X_{j1})^2 + (X_{i2} - X_{j2})^2} \quad (2.8)$$

Así, la distancia entre un vector a clasificar  $X$  y el prototipo  $Z_i$  de una clase genérica  $\alpha_i$  viene dada por:

$$d_E(X, Z_i) = \|X - Z_i\| = \sqrt{(X - Z_i)^T \cdot (X - Z_i)} = \sqrt{\sum_{j=1}^n (x_j - z_{ij})^2} \quad (2.9)$$

En donde por T se indica, la transpuesta de un vector.

El modo de funcionamiento de un reconocedor euclidiano consiste en las siguientes etapas:

1. Determinación de los prototipos o centroides de las clases del universo del trabajo.



2. Dado un vector  $X$  a reconocer, calcular las distancias euclidianas de  $X$  a cada uno de los prototipos  $Z_1, Z_2, \dots, Z_N$ .
3. Asignar  $X$  a la clase  $\alpha_i$ , tal que la distancia  $d_E(X, Z_i)$  es mínima.

### 2.3.3 Algoritmo de agrupamiento *K-Means*.

Este es el algoritmo de agrupamiento más popular. También es llamado el *algoritmo de las medias móviles* porque en cada iteración se recalculan los centros de los agrupamientos. Por esta razón incorporamos el índice  $k$  a la notación que estamos empleando de manera que con  $S_i(k)$  indicamos el conjunto de patrones asociados al agrupamiento  $S_i$  en la iteración  $t$  y mediante  $Z_i(k)$  indicamos el valor de su centro en esa iteración.

Este algoritmo requiere un único parámetro,  $K$ , que es el número de agrupamientos que debe encontrar el algoritmo. Puede plantearse en tres pasos:

1. Inicialización.

Consiste en inicializar arbitrariamente los centros de los  $K$  grupos.

2. Asignación y actualización de los centros.

En este paso se asigna cada patrón al grupo más cercano y se recalculan los centros en base a esta asignación.

3. ¿Convergencia?

En el paso anterior algunos patrones pueden cambiar de agrupamiento y en consecuencia, los centros de éstos. Si esto ocurre, se trata de repetir el paso 2 hasta que no se modifiquen los centros. Cuando no exista modificaciones se considera que se ha encontrado una buena partición y se termina el agrupamiento.

El algoritmo correspondiente es el siguiente:

Algoritmo de las *K-Means*: *K-Means* (X, K) [Tau & Gonzalez 81]

Donde:

X Un conjunto N patrones  $\{X_1, \dots, X_N\}$

K  $\in$  N Numero de agrupaciones.

Salida:

$S_1, S_2, \dots, S_K$  K conjunto de patrones (agrupamiento).

$z_1, z_2, \dots, z_K$  Los centros de los K agrupamientos.

Algoritmo

Paso 1.

Seleccionar los K centros iniciales //Pueden seleccionarse los k primeros patrones o al azar

$z_i(1), i = 1, 2, \dots, K$

Paso 2.

Distribución de los patrones en los K agrupamientos usando la relación

$$x \in S_j(k) \text{ if } \|x - z_j(k)\| < \|x - z_i(k)\|$$

para toda  $i = 1, 2, \dots, K, i \neq j$ , donde  $S_j(k)$  denota el conjunto de patrones cuyo centro es  $z_j(k)$ .

Paso 3.

Con los resultados del paso 2, calcular los nuevos centros  $z_j(k+1), j = 1, 2, \dots, K$  usando

$$z_j(k+1) = \frac{1}{N_j} \sum_{x \in S_j(k)} x, j=1, 2, 3, \dots, K$$

donde  $N_j$  es el numero de patrones en  $S_j(k)$ .

Paso 4.

Si  $z_j(k+1) = z_j(k)$  para toda  $j = 1, 2, \dots, K$ , el algoritmo ha convergido y el proceso es terminado, en otro caso se regresa al Paso 2.