Capítulo 2

Reconocimiento De Patrones

2.1. Generalidades

¿Qué es el Reconocimiento de Patrones?, ¿Qué debería ser?... Si fijamos nuestra atención en la palabra "Reconocimiento", vemos que el mismo es considerado como una función básica del ser humano así como de otros organismos vivientes. Por otra parte, cuando hablamos de un "Patrón" decimos que el mismo es la descripción de un objeto. Sería fácil enumerar ejemplos de nuestros permanentes actos de reconocimiento de objetos, una voz o un rostro familiares, gestos, formas, colores, etc. Un ser humano es un sofisticado sistema de información, y esto es debido en parte a su elevada capacidad de reconocimiento de objetos.

Podemos distinguir dos categorías de reconocimiento, a saber, reconocimiento de objetos concretos y reconocimiento de objetos abstractos.

En el primer caso se trata de reconocimiento perceptual, como por ejemplo una forma (patrón espacial) o una secuencia (patrón temporal).

En el segundo caso se trata de reconocimiento conceptual, tal como un viejo argumento o la solución de un problema.

El Reconocimiento de Patrones esta asociado al Reconocimiento Perceptual. [1]

A su vez el reconocimiento perceptual como atributo humano puede ser visto como un problema sicofisiológico basado en una relación entre una persona y un estímulo físico.

Cuando una persona percibe un patrón, realiza una inferencia inductiva y asocia esta percepción con algunos conceptos generales o pistas derivados de su experiencia pasada.

El reconocimiento humano es concebido como la estimación del parecido relativo con que los datos de entrada pueden ser asociados con un conjunto de poblaciones estadísticas conocidas, las que dependen de nuestra experiencia pasada y que forman las pistas e información *a priori* para el reconocimiento.

De este modo el problema del reconocimiento puede ser concebido como el de discriminar, clasificar o categorizar la información de entrada, no entre patrones individuales sino entre poblaciones, por medio de la búsqueda de *características* o atributos invariantes entre los miembros de una población.

En lo que sigue, se tratan aspectos relacionados con la informática e ingeniería sobre el diseño de sistemas de reconocimiento autónomos.

---

[1] Es interesante notar que lo que podríamos llamar "Reconocimiento Conceptual Artificial" esta usualmente basado en el empleo de funciones lógicas las cuales pueden ser materializadas por los métodos del Reconocimiento Perceptual simplemente por tratarse de funciones. Por lo tanto -al menos en teoría- los métodos del reconocimiento perceptual podrían aplicarse también al reconocimiento conceptual.

2.1.1. Aplicaciones

El Reconocimiento de Patrones provee una teoría general cuyas aplicaciones son múltiples. Algunas se enumeran en la tabla siguiente.

Tabla 2.1. Reconocimiento y Procesamiento de Patrones. Ejemplos de aplicación.

Tarea de clasificación	Datos de entrada	Salida
Reconocimiento de Caracteres	Señales ópticas o líneas	Nombre del carácter
Reconocimiento del habla	Ondas acústicas	Nombre de la palabra
Reconocimiento de voz	Voz	Nombre de quien habla
Predicción del clima	Mapas climáticos	Pronóstico climatológico
Diagnóstico médico	Síntomas	Enfermedad
Predicción de Acciones	Gráficos financieros	Predicción de alzas y bajas

2.1.2. Los Problemas a Resolver

El diseño de un sistema de reconocimiento automático involucra por lo general las tareas siguientes:

- Sensado
- Extracción de Características
- Clasificación

La primera se refiere a la representación de la información obtenida mediante algún tipo de sensor sobre los objetos a ser reconocidos. Cada cantidad medida describe una característica del objeto. Esto puede evidenciarse suponiendo por ejemplo que se obtiene información sobre caracteres alfanuméricos. En este caso puede considerarse un esquema de medición de grilla como el mostrado en el lado izquierdo de la figura 1.1.

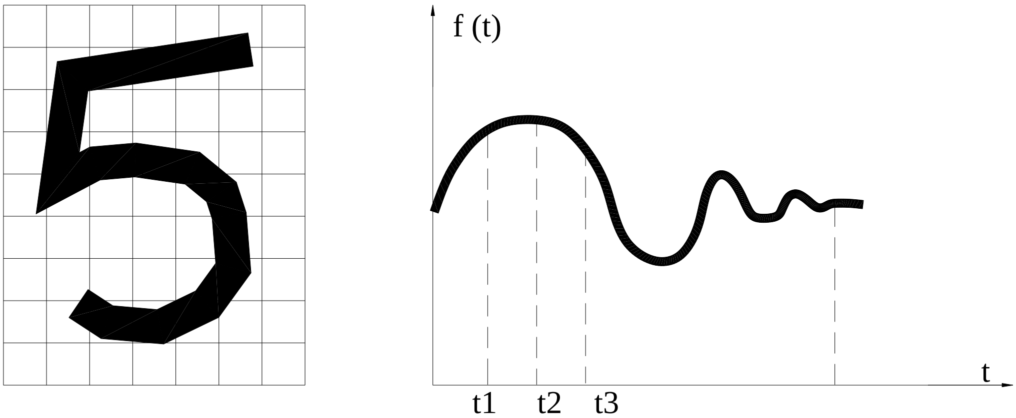


Fig. 2.1: Los objetos se representan mediante un vector patrón.

Esta grilla puede representarse mediante un vector patrón como sigue:

$$\mathbf{x} = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ .x_n \end{pmatrix} \qquad con \qquad x_i = 0 \qquad o \qquad x_i = 1$$

Donde por ejemplo se asigna a x_i el valor 1 si el elemento forma parte del carácter y 0 en caso contrario.

Un segundo ejemplo es una función continua de una variable t (Por ejemplo una señal acústica). Esta función es muestreada en puntos discretos $t_1, t_2, \ldots, t_n$ y puede conformarse el vector que representa al patrón tomando los valores de la función tomando $x_1 = f(t_1), x_2 = f(t_2), \ldots, x_3 = f(t_3)$

$$\mathbf{x} = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ .x_n \end{pmatrix} \qquad con \qquad x_i = f(t_i)$$

Los vectores patrones se indicarán en lo sucesivo por letras minúsculas en negrita, e.g. **x, y, z**. Estos son definidos como vectores columna pero pueden eventualmente ser expresados por su traspuesta $\mathbf{x} = (x_1, x_2, \ldots, x_n)'$

Toda la información medida disponible acerca de los patrones esta contenida en los vectores de patrones.

Cuando los vectores de patrones están formados por números reales es útil interpretarlos como puntos en el espacio Euclidiano n-dimensional

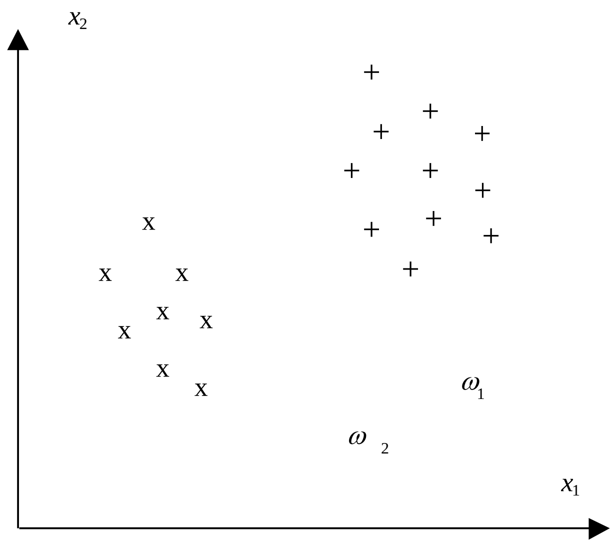


Figura 2.2: Agrupamientos de un conjunto de patrones que corresponden a dos clases.

En casos sencillos, el conjunto de puntos que corresponden a una misma clase resultan así agrupados en alguna región del espacio. A este agrupamiento suele llamárselo por su denominación en ingles, Cluster.
Esta situación se ejemplifica en la figura 2.2, la cual muestra dos clases ω_1 y ω_2 correspondientes a patrones de dimensión 2 y sus correspondientes agrupamientos (clusters). En casos prácticos, sin embargo estos agrupamientos rara vez se dan con tanta claridad como la mostrada en la figura.
En este caso podríamos suponer, por ejemplo, que se trata de patrones que describen jugadores de Básquet y Fútbol definidos por peso y altura respectivamente. Cada equipo constituiría entonces una clase a la que pertenecen sus jugadores. En general si representáramos su peso y altura en un sistema cartesiano los de Básquet aparecerían casi con seguridad arracimados en un cluster (Debido a que podemos esperar que todos sean muy altos).

El párrafo precedente plantea el problema de la obtención de los patrones, lo que da como resultado un conjunto de los mismos agrupados en racimos, y componiendo diferentes clases.

Un segundo problema en el área del Reconocimiento de Patrones, esta referido a la reducción de la dimensión de los vectores de los patrones. Se suele mencionar a esta etapa como preproceso o extracción de características.

El método analítico mas conocido se denomina “Análisis de las Componentes Principales”, o también “Transformación de los Ejes Principales” y en principio no será incluido en estas secciones. Lo que se logra con esto es reducir la dimensión de los patrones pero con la posibilidad de perder un bajo porcentaje de la información contenida en ellos. A su vez al poder emplear vectores de menor dimensión casi sin pérdida de información se disminuye el costo computacional lo que a veces es imprescindible por limitaciones tecnológicas.

El tercer y quizás mas crítico problema del Reconocimiento de Patrones, se refiere a la determinación de un procedimiento óptimo de clasificación.

Para lograr esto, se requiere diseñar una mecanismo que, dado un conjunto de patrones de los cuales no se conoce a priori la pertenencia a la clase de cada uno de ellos, permita determinar a que clase pertenece cada patrón.
Suponiendo el sistema diseñado para reconocer M clases diferentes, denominadas $w_1, w_2, \ldots, w_n$ puede considerarse al espacio de patrones compuesto por M regiones cada una de las cuales corresponde a los patrones de una clase.

La solución puede interpretarse como generar los límites de decisión entre las regiones. Estos pueden estar dados por las llamadas Funciones de Decisión o Funciones Discriminantes $d_1(\mathbf{x})$, $d_2(\mathbf{x})$,... $d_n(\mathbf{x})$, las que son funciones escalares de cada vector **x.**

Si una función tiene el mayor valor entre todas las demás, entonces el vector pertenece a la clase correspondiente a esa función, es decir si $d_i(\mathbf{x}) > d_j(\mathbf{x})$ para $i, j = 1,2,\ldots,M$, y $j \neq i$, entonces el patrón **x** pertenece a la clase w_i

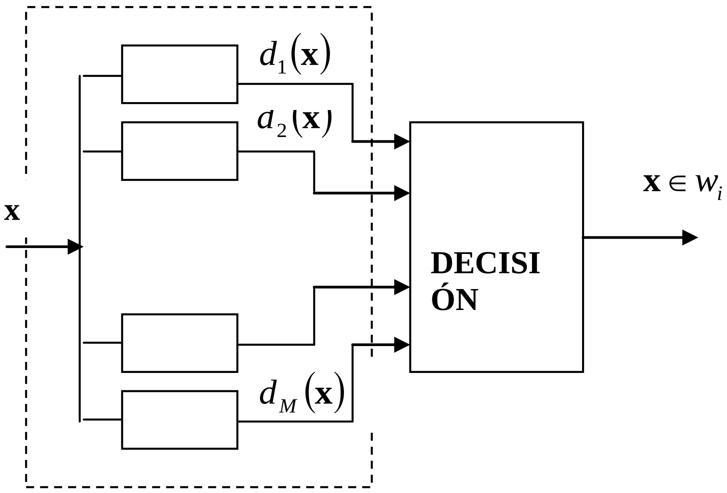


Fig. 2.3 Las funciones de decisión

Es decir que la pertenencia a una clase es determinada por el mayor valor de la función discriminante $d_i(\mathbf{x})$.

Tal sistema de clasificación es esquematizado en la figura incorporando un proceso de ejecución de decisiones.

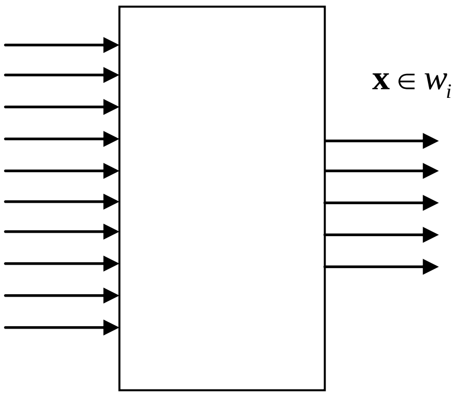


Fig. 2.4. El sistema de categorización representado como una caja negra con vectores de entrada y salida.

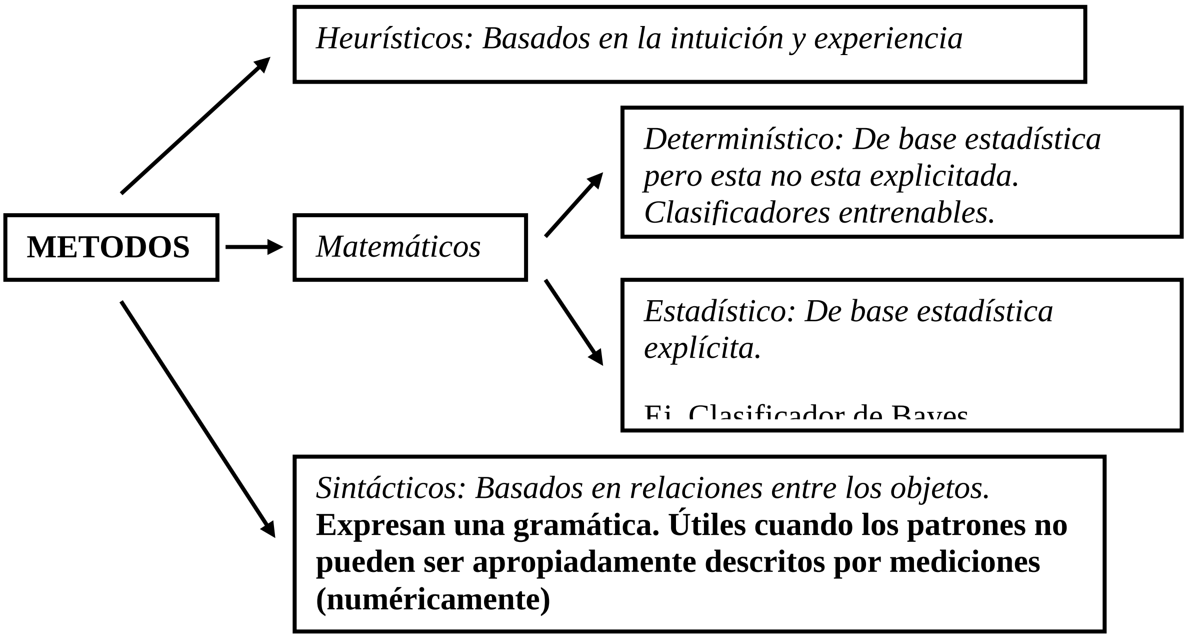
En la figura 2.3 se representa esquemáticamente lo que ocurre al emplear Funciones Discriminantes en la clasificación. Las Funciones Discriminantes están representadas individualmente, esto no debe interpretarse en el sentido de que realmente es posible obtener definiciones analíticas de cada una de estas funciones si no que de alguna manera, que se explicará mas adelante, se establecen las mismas.

Por otra parte muchas veces el proceso de establecer estas funciones se lleva a cabo en forma conjunta y simultánea y por lo tanto solo disponemos de un único sistema que generará las salidas de la manera deseada tal como se representa a la salida del bloque en línea de trazos (Fig. 2.3).

En este último caso solo tendremos una caja negra con vectores de entrada (Patrones a reconocer) y vectores de salida (Patrones que representan las clases de pertenencia de los patrones), lo que se muestra en la Fig. 2.4.

Este último sistema constituye el núcleo del sistema de reconocimiento. A los efectos de tener una visión lo mas general posible debemos considerar que el sistema debe hacerse cargo además de resolver los dos problemas anteriores (Sensado y Selección de Características) y clasificación.

El análisis de contexto es particularmente importante en temas tales como Reconocimiento del habla y reconocimiento de escritura manual.



A lo largo de estos capítulos veremos que si el sistema de reconocimiento puede autoajustar ciertos coeficientes internos que definen las funciones discriminantes estaremos en presencia de un sistema *adaptivo* o con capacidad de aprendizaje. También veremos que la teoría general de Optimización y Estimación de parámetros, (Cuyos fundamentos constituyen un vasto campo teórico) provee un enfoque relevante al problema del Reconocimiento de Patrones.

Por último podemos decir que si todas las representaciones posibles de aquel objeto que se desea clasificar, o un gigantesco número de ellas estuvieran disponibles, si toda la información contenida en el objeto pudiera ser extraída, y si el tiempo de desarrollo no tuviera límites seguramente un sistema de fuerza bruta podría resolver el problema, pero por lo general en la práctica las condiciones de trabajo distan mucho de ser estas. Las restricciones se dan por lo común en términos de costos, tiempo y tecnología existente.

2.1.3. Clasificación de los Métodos

En nuestro estudio, por razones de espacio, nos centraremos en comprender los métodos fundamentales. Como extracto de estas secciones podemos decir que, definiendo a una *clase* como una categoría determinada por algunos atributos comunes y un *patrón* como la descripción de cualquier miembro de una categoría que representa a una clase de patrones, decimos que la teoría del Reconocimiento de Patrones se ocupa de buscar la solución al problema general de reconocer miembros de una clase dada en un conjunto que contienen elementos de muchas clases diferentes.

2.1.4. Modelos de representación gráfica para el análisis

En general, todo el estudio sobre los sistemas de reconocimiento, se encara a partir de la representación de patrones en dos dimensiones, debido a la simplicidad del modelo y a la validez general (en n dimensiones) de los resultados que de el se deriven. Pero también es importante el hecho de que las personas podemos representar y visualizar claramente las situaciones, solo en el plano y en el espacio tridimensional.

A pesar de esto, a veces se realizan esfuerzos para visualizar resultados en mas dimensiones, por ejemplo, si estuviéramos en cuatro dimensiones, la cuarta dimensión podría ser representada asignando colores –en nuestro estudio a los patrones- que variarían según el espectro térmico, -de cálidos a fríos o viceversa-, eso nos permitiría ver quizás algo mejor las características de la información disponible.

Otro método auxiliar es representar un vector como los valores de un histograma, es decir, en ordenadas tenemos los valores que puede tomar la variable, y en abscisas los de cada coordenada.

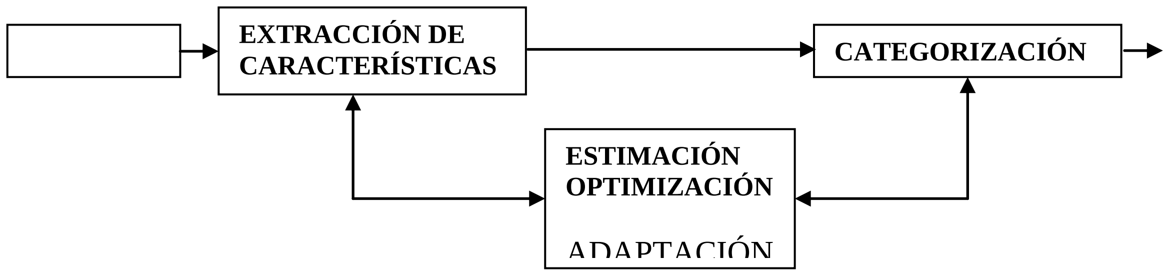


Fig. 2.6: Una visión global de un sistema de reconocimiento artificial

2.2. La Función de Decisión Lineal

La función principal de un sistema de reconocimiento de patrones es tomar decisiones acerca de la pertenencia a una clase dada de los patrones con que se lo confronta. Para lograr esto, es necesario establecer algunas reglas en que basar estas decisiones.

Una aproximación fundamental a este problema es el empleo de *funciones de decisión*. Como una introducción a este concepto relativamente simple, considérese la Fig. 6.2, donde se muestran dos clases hipotéticas. Puede verse que una línea puede separar las dos poblaciones del modelo convenientemente.

Considérese $d(\mathbf{x}) = w_1x_1 + w_2x_2 + w_3 = 0$ como la ecuación de una línea de separación donde ***w*** es el vector de parámetros y ***x*** es el vector de las variables de la entrada.

Cualquier modelo que pertenece a una clase ω_1 producirá una cantidad positiva $\mathbf{x}$ cuando sea sustituido en la ecuación del hiperplano $d(\mathbf{x})$. Similarmente $d(\mathbf{x})$ se hace negativo con la substitución de cualquier patrón de la otra clase ω_1. Por consiguiente esta línea $d(\mathbf{x})$ puede usarse como una función de decisión dado un patrón de clasificación desconocida $\mathbf{x}$, para que nosotros podemos decir que este pertenece a una determinada clase ω_1 si $d(\mathbf{x}) > 0$ o a ω_2 en caso contrario. Ninguno de estos conceptos necesita ser restringido a dos clases, solo se emplean dos clases por simplicidad.

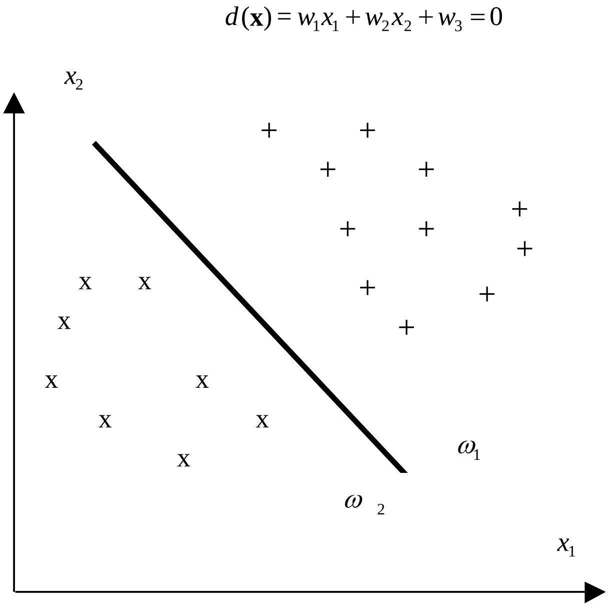


Figura 2.8: Función discriminante lineal.

Desgraciadamente, a menos que alguna información está disponible a priori, la única manera de establecer la efectividad de una función de decisión escogida es por ensayo directo, como se ira viendo a lo largo de este curso. Una vez que una cierta función (o funciones si se trata de más de dos clases) se ha seleccionado, el problema se reduce a la determinación de los coeficientes. Es posible usar los patrones de una muestra para determinar los coeficientes que caracterizan estas funciones.

Ejemplo:

Tomando,

$$w_1 = 1, w_2 = 1, w_3 = -4$$

Se pueden calcular los valores de

$d(\mathbf{x})$ para $x \in w_1$, $x \in w_2$ y $x \notin w_1, w_2$.

Este concepto básico de función discriminante se puede extender fácilmente a casos fronteras de decisión no lineales en un espacio euclidiano de dimensión finita n.

También es posible extenderlo a casos de clases no separables y que contengan múltiples subclases como veremos mas adelante.

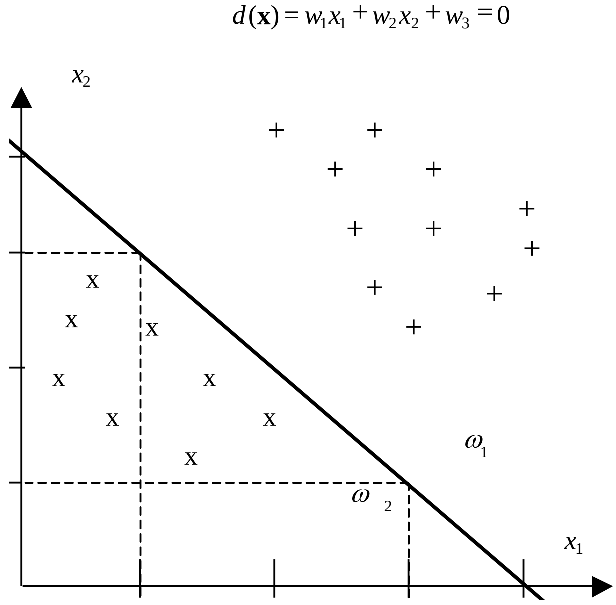


Figura 2.9: $d(\mathbf{x})$ con $w_1 = 1$, $w_2 = 1, w_3 = -4$

Por otra parte este concepto permite extender el tratamiento analítico del problema aun para aquellos casos en que no podemos visualizar el problema (dimensión n > 3).

2.2.1. Extensión al Espacio euclideano n-dimensional

El caso anterior puede ser fácilmente extendido a n-dimensiones definiendo una función de decisión lineal general de la forma:

$$d(\mathbf{x}) = w_1 x_1 + w_2 x_2 + \ldots w_n x_n + w_{n+1}$$

$$= \mathbf{w}_o^{'} \mathbf{x} + w_{n+1}$$

Donde el vector $\mathbf{w}_o = (w_1, w_2, \ldots, w_n)'$ es llamado *vector de pesos* o *vector de parámetros.*

Es además una convención muy difundida adicionar un 1 como última componente de todos los vectores patrones con lo que la igualdad anterior puede expresarse como

$$d(\mathbf{x}) = \mathbf{w}_o^{'} \mathbf{x}$$

Donde $\mathbf{x} = (x_1, x_2, \ldots, x_n, 1)'$ es llamado *vector patrón aumentado.*

Como la misma cantidad es agregada a todos los patrones de todas las clases en juego las propiedades geométricas entre las mismas se mantienen.

Dependiendo del contexto del problema y del sistema de clasificación se emplea el vector patrón aumentado o no, pero en cualquier caso siempre en la práctica se hace referencia al mismo simplemente como vector patrón.

2.2.2. Cálculo de la Matriz de Coeficientes

Los Vectores Patrones conocidos permiten calcular la Matriz de Coeficientes.

Asumiendo que se deban clasificar patrones que corresponden a dos clases ω_1 y ω_2, y que cada clase contiene dos patrones de dimensión dos, $\{\mathbf{x}_1^1, \mathbf{x}_2^1\}$ y $\{\mathbf{x}_1^2, \mathbf{x}_2^2\}$, donde los supra índices indican las clases ω_1 y ω_2, respectivamente. Si las clases son linealmente separables el problema es encontrar un vector $\mathbf{w} = w_1, w_2, w_3$ tal que

$$w_1 x_{11}{}^1 + w_2 x_{12}{}^1 + w_3 > 0$$
$$w_1 x_{21}{}^1 + w_2 x_{22}{}^1 + w_3 > 0$$
$$w_1 x_{11}{}^2 + w_2 x_{12}{}^2 + w_3 < 0$$
$$w_1 x_{21}{}^2 + w_2 x_{22}{}^2 + w_3 < 0$$

Es decir que $\mathbf{w}$ es la solución del sistema lineal determinado por todos los patrones de ambas clases.

Para hallar esta solución recordemos que el valor esperado de una variable aleatoria discreta se define como $E\{\mathbf{x}\} = \sum_i p_i x_i$ donde p_i es la probabilidad de que la variable aleatoria tome el valor x_i

Si $\mathbf{x}$ toma los valores 1,2,3,...,6 con probabilidad 1/6, $E\{\mathbf{x}\} = \frac{1}{6}(1+2+3+4+5+6) = 3.5$

Por otra parte debemos considerar que, al presentar un patrón x_i, obtendremos a la salida wx_i, que dependerá del valor de w, que deberá corresponder a un valor de salida y. En principio, con valores de w no optimizados, se verificara un error entre wx_i e y. Ese error se calcula como una distancia entre dos vectores y lo indicamos como $y\text{-}wx_i$. En el caso de tener varios vectores x_i e y_i lo indicamos en negrita $\mathbf{y - wx_i}$.

Para hallar la matriz de los coeficientes **w** que nos define la función discriminante, partimos de la aproximación del error cuadrático medio dada por:

$$S^2(\mathbf{w}) = E\left\{ \left| \mathbf{y} - \mathbf{w}^{\mathrm{T}}\mathbf{x}(\mathbf{v}) \right|^2 \right\}$$

Siendo **w** la matriz de coeficientes, $\mathbf{x}(\mathbf{v})$ el vector de patrones conocidos e $\mathbf{y}$ el vector de salida del clasificador. Existe un valor óptimo de la matriz de coeficientes **w** que minimiza este valor esperado del error:

$$S^2(\mathbf{w}) = E\left\{ \left| \mathbf{y} - \mathbf{w}^{\mathrm{T}}\mathbf{x}(\mathbf{v}) \right|^2 \right\} = \min(\mathbf{w}) = E\left\{ \sum (\mathbf{y} - \mathbf{w}^{\mathrm{T}}\mathbf{x}(\mathbf{v}))^2 \right\}$$

Derivando con respecto a **w** obtenemos **2 (y - w x) x** el coeficiente 2 se puede eliminar si se parte de la expresión de S^2 multiplicada por 1/2, lo que no altera el sistema, ya que al multiplicar todas las ecuaciones de un sistema lineal, este no se modifica.

de donde $\mathbf{x}\,\mathbf{y}^{\mathbf{t}} = \mathbf{x}\,\mathbf{x}^{\mathbf{t}}\,\mathbf{w}$

de donde podemos calcular el valor de **w** qué minimiza el error $S^2(\mathbf{w})$

$$\mathbf{w} = \left(\mathbf{x}\mathbf{x}^{\mathrm{T}}\right)^{-1}\mathbf{x}\mathbf{y}^{\mathrm{T}}$$

Donde, como resultado de la optimización de la matriz **w** a partir de la aplicación de la distancia cuadrática media, se obtiene un sistema de ecuaciónes lineales.
La Ec. (4.15) se corresponde a la definición de un vector como fila y su traspuesta como columna. En caso de definir originalmente vectores columna, la expresión es:

$$\mathbf{w} = \left(\mathbf{x}^{\mathrm{T}}\mathbf{x}\right)^{-1}\mathbf{y}^{\mathrm{T}}\mathbf{x}$$

El error mínimo puede darse también como una función de la matriz **w**. $S^2{}_{\min} = S^2(\mathbf{w}_{\mathbf{opt}})$

2.2.3. Ejemplo de Cálculo de la Matriz de Coeficientes

Consideramos el siguiente problema ya clásico en la literatura, relativo a la discriminación lineal de 2 clases de patrones bidimensionales. (Se requerirán al menos 3 vectores ya que no podemos tener menos ecuaciones que incógnitas)

Vectores de la Clase 1: (0 0), (0 1)
Vectores de la Clase 2: (1 0), (1 1)

Expandiendo los vectores:

$$\boldsymbol{X} = \begin{bmatrix} 0 & 0 & 1 \\ 0 & 1 & 1 \\ 1 & 0 & 1 \\ 1 & 1 & 1 \end{bmatrix} \qquad \boldsymbol{X'} = \begin{bmatrix} 0 & 0 & 1 & 1 \\ 0 & 1 & 0 & 1 \\ 1 & 1 & 1 & 1 \end{bmatrix} \qquad \boldsymbol{X'}\,\boldsymbol{X} = \begin{bmatrix} 0 & 0 & 1 & 1 \\ 0 & 1 & 0 & 1 \\ 1 & 1 & 1 & 1 \end{bmatrix} \begin{bmatrix} 0 & 0 & 1 \\ 0 & 1 & 1 \\ 1 & 0 & 1 \\ 1 & 1 & 1 \end{bmatrix} = \begin{bmatrix} 212 \\ 122 \\ 224 \end{bmatrix}$$

Calculamos la inversa:

$$\boldsymbol{X'}\,\boldsymbol{X} : \boldsymbol{I} = \begin{bmatrix} 2 & 1 & 2 & 1 & 0 & 0 \\ 1 & 2 & 2 & 0 & 1 & 0 \\ 2 & 2 & 4 & 0 & 0 & 1 \end{bmatrix}$$

Reemplazando la fila 1 por la 2

$$= \begin{bmatrix} 1 & 2 & 2 & 0 & 1 & 0 \\ 2 & 1 & 2 & 1 & 0 & 0 \\ 2 & 2 & 4 & 0 & 0 & 1 \end{bmatrix}$$

Restando a la fila 2 la fila 3

$$= \begin{bmatrix} 1 & 2 & 2 & 0 & 1 & 0 \\ 0 & -1 & -2 & 1 & 0 & -1 \\ 2 & 2 & 4 & 0 & 0 & 1 \end{bmatrix}$$

Multiplicando la fila 1 por 2 y restando a la fila 3

$$= \begin{bmatrix} 1 & 2 & 2 & 0 & 1 & 0 \\ 0 & -1 & -2 & 1 & 0 & -1 \\ 0 & -2 & 0 & 0 & -2 & 1 \end{bmatrix}$$

Sumando a la fila 1 la fila 3

$$= \begin{bmatrix} 1 & 0 & 2 & 0 & -1 & 1 \\ 0 & -1 & -2 & 1 & 0 & -1 \\ 0 & -2 & 0 & 0 & -2 & 1 \end{bmatrix}$$

Multiplicando la fila 2 por -1

$$= \begin{bmatrix} 1 & 0 & 2 & 0 & -1 & 1 \\ 0 & 1 & 2 & -1 & 0 & 1 \\ 0 & -2 & 0 & 0 & -2 & 1 \end{bmatrix}$$

Sumando a la fila 3 la 2 multiplicada por 2

$$= \begin{bmatrix} 1 & 0 & 2 & 0 & -1 & 1 \\ 0 & 1 & 2 & -1 & 0 & 1 \\ 0 & 0 & 4 & -2 & -2 & 3 \end{bmatrix}$$

Sumando a la fila 1 la 3 multiplicada por –1/2

$$= \begin{bmatrix} 1 & 0 & 0 & 1 & 0 & -1/2 \\ 0 & 1 & 2 & -1 & 0 & 1 \\ 0 & 0 & 4 & -2 & -2 & 3 \end{bmatrix}$$

Sumando a la fila 2 la 4 multiplicada por –1/2

$$= \begin{bmatrix} 1 & 0 & 0 & 1 & 0 & -1/2 \\ 0 & 1 & 0 & 0 & 1 & -1/2 \\ 0 & 0 & 4 & -2 & -2 & 3 \end{bmatrix}$$

Multiplicando por ¼ la fila 3

$$= \begin{bmatrix} 1 & 0 & 0 & 1 & 0 & -1/2 \\ 0 & 1 & 0 & 0 & 1 & -1/2 \\ 0 & 0 & 1 & -1/2 & -1/2 & 3/4 \end{bmatrix}$$

La inversa resulta:

$$\left(\boldsymbol{X}'\,\boldsymbol{X}\right)^{-1} = \begin{bmatrix} 1 & 0 & -1/2 \\ 0 & 1 & -1/2 \\ -1/2 & -1/2 & 3/4 \end{bmatrix}$$

Por otra parte:

$$\boldsymbol{X}'\boldsymbol{Y} = \begin{bmatrix} 0 & 0 & 1 & 1 \\ 0 & 1 & 0 & 1 \\ 1 & 1 & 1 & 1 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \\ -1 \\ -1 \end{bmatrix} = \begin{bmatrix} -2 \\ 0 \\ 0 \end{bmatrix}$$

De donde:

$$\boldsymbol{W} = \left(\boldsymbol{X}'\,\boldsymbol{X}\right)^{-1} \boldsymbol{X}'\boldsymbol{Y} = \begin{bmatrix} -2 \\ 0 \\ 1 \end{bmatrix}$$

Los valores de las salidas serán:

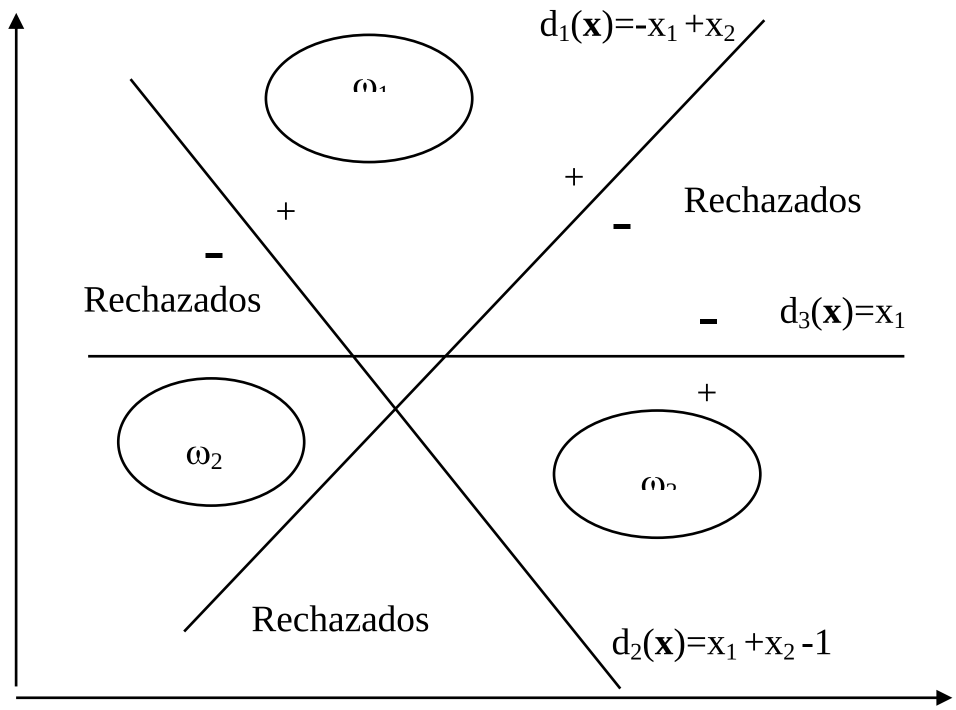
$$\boldsymbol{XW} = \begin{bmatrix} 0 & 0 & 1 \\ 0 & 1 & 1 \\ 1 & 0 & 1 \\ 1 & 1 & 1 \end{bmatrix} \begin{bmatrix} -2 \\ 0 \\ 1 \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \\ -1 \\ -1 \end{bmatrix}$$

Un concepto importante a tener presente es que de acuerdo a lo visto, un sistema de clasificación lineal para clases separables puede ser calculado matemáticamente en forma precisa.
Esta posibilidad es teóricamente aplicable a casos de mayor complejidad, tales como clasificación no lineal, no separable y con dimensión de los patrones mayor que dos.

Sin embargo como se verá enseguida los cálculos necesarios aumentan notablemente. De modo que si bien es teóricamente posible calcular un clasificador cualquiera, generalmente es impracticable.

Este problema da lugar a la aplicación de métodos numéricos de solución para el Clasificador Polinomial, a métodos recursivos, como las Redes Neurales, o a métodos de optimización mas elaborados como los Clasificadores de Margen Óptimo.

2.2.4. Extensión a Multiclases. Clasificación "Uno Contra Todos" (Piecewise / One Vs. All)



Extensión a Multiclases. Clasificación “Por Pares” (Pairwise)

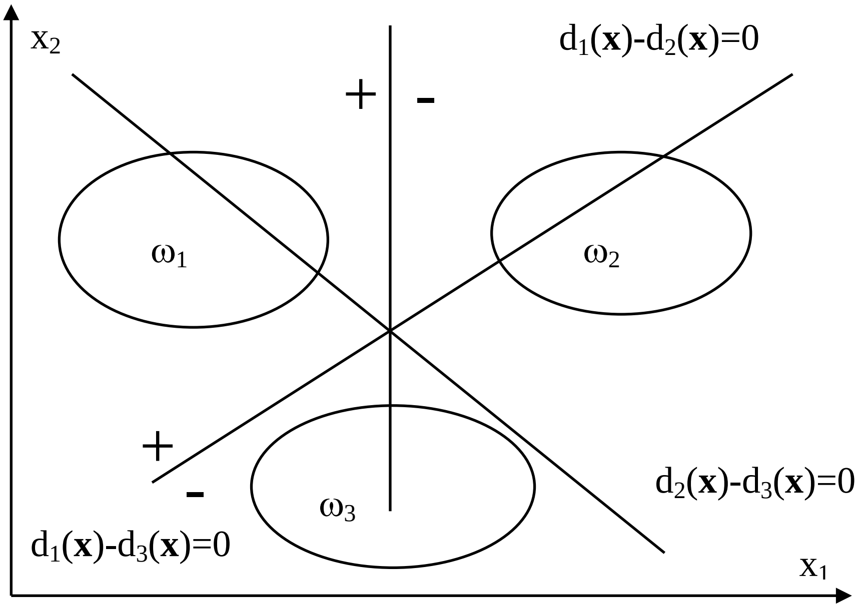
$d_1(\mathbf{x})=-x_1+x_2$

$d_2(\mathbf{x})=x_1+x_2-1$

$d_1(\mathbf{x})=-x_2$

$d_1(\mathbf{x})>d_2(\mathbf{x})$

$d_1(\mathbf{x})>d_3(\mathbf{x})$



Pasar a este último esquema, partiendo del anterior, permite mantener la validez del concepto mostrado en la figura 1.3 por el cual, al presentar un patrón al clasificador ya optimizado, una función de decisión resultará ganadora y determinará la pertenencia a la clase.

2.2.5. Empleo de Valores Normalizados

La salida del clasificador puede ser 0 o 1 en vez de > 0 y < 0.

Para lograr esto, las entradas y las salidas se computan normalizadas de modo que están siempre entre 0 y 1. Esto implica poder aplicar de forma general un programa que ejecuta algún algoritmo de clasificación a distintos problemas.

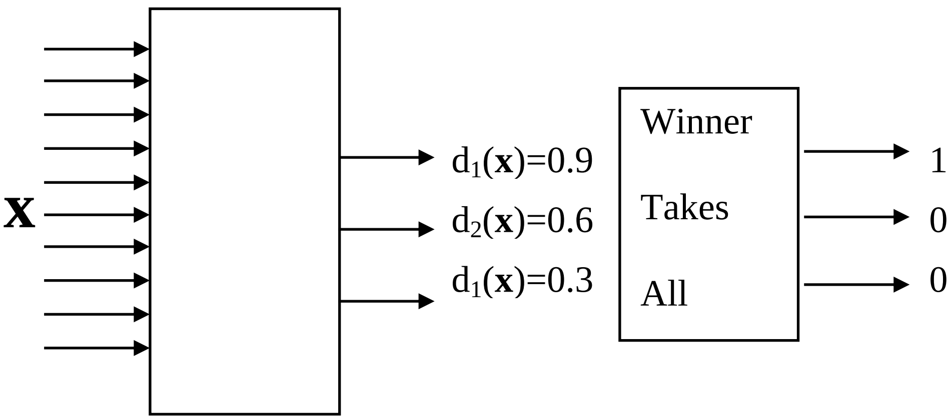


Figura 2.12: Los valores normalizados de la entrada generan valores de salida entre cero y uno, entre los cuales uno es el ganador

Los patrones son almacenados en forma normalizada en archivos de con la codificación de sus clases

(0.2 0.3 0.9)	100
(0.1 0.3 0.6)	100
(0.2 0.3 0.9)	010

2.2.6. Características de Problemas Reales

La situación que se encuentra usualmente en problemas reales se representa en la Fig 2.13, en la que no es posible separar las clases en una forma sencilla. Un caso muy frecuente y de mayor complejidad se tiene cuando los patrones de cada clase estan a su vez distribuidos en multiples clusters.

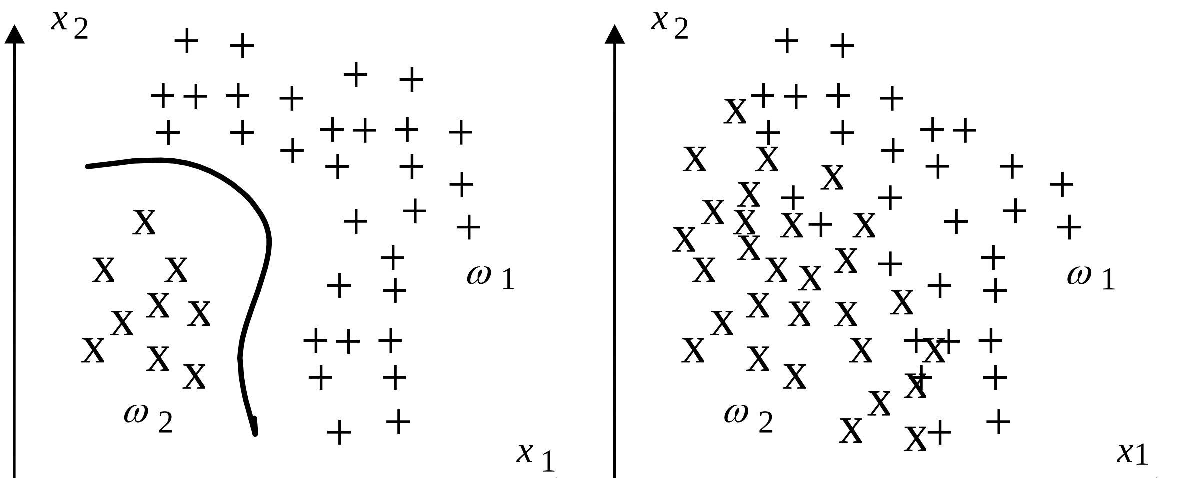


Figura 2.13: Representación de dos clusters de patrones bidimensionales correspondientes a dos clases. Caso en que son separables en forma no-lineal (izquierda) y caso no separable (derecha)

2.2.7. Evaluación de Resultados. Matriz de Confusión.

Fases de un clasificador:

1. Fase de Entrenamiento. La palabra entrenamiento esta mas asociada al caso de que el clasificador sea una red neural, como se vera mas adelante, pero aquí generalizamos su empleo. (Podemos hablar también de estimación, optimización, adaptación, aprendizaje, etc. Según el tipo de clasificador)

2. Fase de Prueba

Modos de entrenamiento:

1. Supervisado

2. No supervisado

Organización de los datos:

- Conjunto de entrenamiento
- Conjunto de prueba

Calculo de Coeficientes **W**

ESTIMACIÓN
OPTIMIZACIÓN
ADAPTACIÓN
APRENDIZAJE

Capacidad de Generalización de un clasificador

Matriz de confusión

	ω_1	ω_2	ω_3
ω_1	265	20	15
ω_2	19	276	5
ω_3	50	28	222

Factor de Clasificación
763 / 900 = 0.85

2.2.8. Funciones de Decisión Generalizadas.

Los límites de decisión pueden establecerse entre clases de modelos que no comparten patrones idénticos. La complejidad de estos límites va desde expresiones de carácter lineal a otras fuertemente no lineales, las que requieren un número grande de términos para su descripción. A menudo en aplicaciones prácticas las clases del modelo no son verdaderamente separables dentro de las limitaciones económicas o técnicas, y es deseable entonces buscar aproximaciones a las funciones de decisión.
Una manera conveniente de generalizar el concepto de función de decisión lineal es considerar las funciones de decisión de la forma:

$$d(\mathbf{x}) = w_1 f_1(\mathbf{x}) + w_2 f_2(\mathbf{x}) + ... + w_k f_k(\mathbf{x}) + w_{k+1} = \sum_{i=1}^{k+1} w_i f_i(\mathbf{x}), \qquad 4.1$$

donde $\{f_i(\mathbf{x})\},\ i = 1,2,...,k$ son funciones reales simples del patrón $\mathbf{x}$, y $f_{k+1}(\mathbf{x}) = 1$ donde k+1 es el numero de términos usados en la expansión. La última ecuación representa una infinita variedad de funciones de decisión, dependiendo de la elección de funciones $\{f_i(\mathbf{x})\}$ y de el número de términos usados en la expansión.
Considerando el hecho de que la Eq. 4.1 puede representar funciones de decisión muy complejas, es posible tratar dichas funciones como si ellas fueran lineales en virtud de una transformación. Para mostrar esto, se define el vector $\mathbf{x}^*$ cuyos componentes son las funciones $\{f_i(\mathbf{x})\}$, esto es,

$$\mathbf{x}^* = \begin{pmatrix} f_1(\mathbf{x}) \\ f_2(\mathbf{x}) \\ \vdots \\ f_k(\mathbf{x}) \\ 1 \end{pmatrix} \qquad 4.2$$

Empleando la Eq. 4.2, la Eq. 4.1 puede ser escrita como:

$$d(\mathbf{x}) = \mathbf{w}'\mathbf{x}^*, \qquad 4.3$$

donde

$$\mathbf{w} = (w_1, w_2, ..., w_k, w_{k+1})'$$

Una vez evaluadas, las funciones $\{f_i(\mathbf{x})\}$ no son nada más que un conjunto de valores numéricos, y $\mathbf{x}^*$ es simplemente un vector K-dimensional el cual ha sido aumentado en 1. Por consiguiente, la Eq. 4.3 representa una función lineal con respecto a los nuevos patrones $\mathbf{x}^*$. Claramente, si nosotros transformamos todos los patrones originales $\mathbf{x}$ en

los patrones $\mathbf{x}^*$ evaluando las funciones para todos los $\mathbf{x}$, se obtiene una representación lineal. Esto da que la posibilidad de restringir la discusión a la decisión lineal sin pérdida de generalidad.
Nada ha cambiado realmente en cuanto a que se conservan las propiedades no lineales de $\mathbf{x}$, sin embargo si se ha aumentado la dimensión de los patrones, pasando a un espacio de patrones de dimensión K > n.

Uno de los tipos mas comúnmente empleados de funciones de decisión generalizadas es aquel en el cual las funciones $\{f_i(\mathbf{x})\}$ están dadas en forma polinómica. Esta esquema de clasificación se explica en los párrafos siguientes.