

Reconocimiento de objetos

Extracto de “Visión por Computador. Fundamentos y métodos”. Arturo de la Escalera. Prentice Hall, 2001

Copia para el alumno con fines didácticos

9. Reconocimiento de objetos

El reconocimiento de objetos es la última etapa dentro de un sistema de visión por computador. A partir de las características encontradas y de los posibles objetos que el conocimiento a priori del problema se espera que puedan aparecer, el sistema debe determinar qué objetos hay presentes en la imagen. Esta etapa es la de mayor grado de abstracción de todas las que forman un sistema de visión y a menudo se realizan fuertes simplificaciones para que funcione con éxito.

9.1 Estructura de un clasificador.

La estructura general puede observarse en la figura 9.1. Como se ha visto hasta ahora, a partir de la imagen original o procesada se obtienen una serie de características (color, textura, etc) y/o descriptores (momentos, descriptores de Fourier, etc) que definen cada objeto. Con ello se ha pretendido reducir el volumen de la información hasta hacerla manejable pero sin perder ninguna información vital o valiosa. La representación de los objetos por medio de estas características se habrá realizando mediante el análisis previo de otras imágenes en donde se habrá comprobado que realmente definen a los posibles objetos y los hacen distinguibles unos de otros. En el caso general se tendrán n valores numéricos que forman el vector de características representante de cada objeto extraído de la imagen.

Por otro lado, del análisis por parte de un operador humano de las imágenes de prueba se habrá obtenido una base de datos que contiene los modelos de los objetos que puedan aparecer en la imagen. El contenido de esta base de datos irá muy ligado al tipo de características que se van a obtener de la imagen. También con estas imágenes de prueba se realizarán una serie de hipótesis (número de objetos posibles, cómo se distribuyen los valores, etc.) que forman las hipótesis sobre los posibles candidatos que pueden aparecer en la imagen.

Por último se debe diseñar un clasificador que vista la información extraída de la imagen y los posibles candidatos clasifique el objeto.

Así, por ejemplo en la figura 9.2 se supone que la base de datos la constituyen dos elementos que quedan determinados por dos características. Con la formación de hipótesis se limita el espacio de búsqueda dentro de la base de datos. Por último con la información obtenida de ella y se elige uno de los posibles candidatos. Siguiendo con el ejemplo de la figura 9.2, ante un objeto cuyas características no coincidan exactamente con los valores ideales el reconocimiento de patrones asigna un elemento de la base de datos esas características obtenidas. Esto se alcanza encontrando una función, llamada función discriminante, que separe el espacio entre ellas dos. En el caso ideal esta función discriminante es una recta de ecuación:

$$fd = w_1x_1 + w_2x_2 + w_3$$

que en formato matricial puede expresarse:

$$fd = \begin{pmatrix} w_1 & w_2 & w_3 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ 1 \end{pmatrix} = W^T X$$

Si se aplica dicha función a un vector de características concreto su valor será positivo o negativo dependiendo de a que clase pertenece. Si se tienen N características el caso óptimo será encontrar $N-1$ funciones discriminantes y en el caso peor $\binom{N}{2}$ funciones.

Las características elegidas deben de cumplir cuatro propiedades:

1. Discriminación. Las características deben tomar valores significativamente distintos para las diferentes clase.
2. Fiabilidad. Las características deben tomar valores similares para todos los objetos de la misma clase.
3. Independencia. Las diversas características no deben estar correladas unas con otras ya que si no reflejarían la misma propiedad del objeto.
4. Número. El número de características debe ser el más pequeño posible ya que de lo contrario la complejidad aumenta considerablemente.

El reconocimiento de objetos tiene tres enfoques principales:

- Reconocimiento estadístico; **Error! Marcador no definido.** Se basa en la determinación y uso de funciones de probabilidad y es el que se explicará en este capítulo.
- Reconocimiento sintáctico. Se trata de analizar la estructura de los objetos (por ejemplo el esqueleto). Aunque mantiene su utilidad (por ejemplo para el reconocimiento de caracteres) su uso es cada vez es menor.
- Reconocimiento mediante redes neuronales. Constituye la técnica más reciente y trata de imitar el modelo y funcionamiento de los sistemas biológicos: cómo las redes de neuronas almacenan y manipulan la información.

9.2. Reconocimiento estadístico

Los clasificadores o reconocedores estadísticos usarán, como su nombre indican, la teoría de probabilidades para emparejar al objeto a clasificar con el de la base de datos cuya probabilidad sea mayor. Son métodos denominados *a priori* ya que se suponen el conocimiento de toda la información necesaria y suficiente para la clasificación, es decir el número de objetos posibles y el tipo y valores concretos de sus funciones de probabilidad.

9.2.1. El clasificador Bayesiano.

El clasificador Bayesiano es un enfoque estadístico para el reconocimiento de patrones. Supóngase que se tiene un caso en el que se quiere clasificar los objetos entre dos posibles. Tras los procedimientos explicados en los temas anteriores se llega a la conclusión que hay dos características C_1 y C_2 que definen bien cada uno de ellos y permiten diferenciarlos entre sí. Para comprobar ello se tomaría unas muestras de cada grupo de objetos, se obtendrían sus valores y se obtendría una figura parecida a la figura 9.3 izquierda. En ella se observa como las características separan bien las dos clases y además como los valores que se van a obtener tienen una cierta tolerancia o están entre cierto rango. El problema se complica si se tiene el caso de la figura 9.3 derecha ya que aunque todavía se observa una separación entre ambas clases ya no es tan sencillo

separarlas como antes. Una posible solución es buscar una nueva característica para obtener el caso anterior; si esto no fuese posible se puede usar un enfoque estadístico.

Para ello se suponen que los valores de las características siguen unas distribuciones de probabilidad que son conocidas. Supongamos que los valores siguen una distribución normal, es decir que se tiene un valor nominal (la media) y una tolerancia o desviación respecto a ese valor (la desviación típica). Para obtener estos valores se haría pasar una cierta cantidad de muestras por el sistema, se obtendría los valores numéricos de las características y por último su media y desviación típica. Se tendrían así las probabilidades de la características X para cada clase α_i $P(X|\alpha_i)$. Es decir la probabilidad de que el diámetro de un tornillo (por ejemplo) sea de 3 mm. Estas probabilidades se denominan *a priori*, ya que se han obtenido mediante la medición de objetos conocidos. Si ahora lo que se quiere es clasificar un objeto desconocido se compararían los valores de las probabilidades a priori y se asignaría el nuevo objeto a la clase que tuviera un valor más alto. Así en la figura 9.4 se tiene el caso de dos clases con una y dos características. La clasificación vendría dada por el valor más alto de las probabilidades.

Hasta ahora lo que se ha dicho es que:

$$P(\alpha_i | X) = P(X | \alpha_i)$$

Es decir que la probabilidad *a posteriori* $P(\alpha_i | X)$ es igual a la probabilidad *a priori*. Se denomina a posteriori porque lo que se quiere saber es qué probabilidad hay de que el objeto cuyo diámetro es de 3mm sea un tornillo.

Un nuevo factor es la relación entre el número de objetos de cada clase que pueden existir; esto es por ejemplo si por cada 3 tornillos hay una tuerca. Supongamos por un instante que no es posible realizar ninguna medición pero que a pesar de todo se tendría que realizar la clasificación. La única información posible sería ¿Qué tipo de objeto es el que se fabrica en mayor cantidad? Y asignar todos los objetos a esa clase. Esto es lo que se denomina $p(\alpha_i)$: la probabilidad a priori de que se presente un elemento de la clase α_i .

Juntando todo lo visto hasta ahora habrá que multiplicar las probabilidades a priori de que aparezca una pieza determinada y de que para ese tipo la característica tenga ese valor.

$$P(\alpha_i | X) = P(X | \alpha_i)p(\alpha_i)$$

Se llega así a la fórmula de Bayes:

$$P(\alpha_i | X) = \frac{P(X | \alpha_i)p(\alpha_i)}{p(X)}$$

donde el único termino nuevo es $p(X)$: probabilidad a priori de que se presente un elemento con vector de características igual a X y cuyo valor es la suma de las probabilidades a priori de todas las clases:

$$p(X) = \sum P(X | \alpha_i)p(\alpha_i)$$

De nuevo, el elemento de características X pertenecerá a la clase cuya sea máxima. En la figura 9.5 puede observarse las probabilidades a posteriori según $p(\alpha_i)$ sean:

$$\begin{array}{ll} p(\alpha_1) = 0.5 & p(\alpha_2) = 0.5 \\ p(\alpha_1) = 0.33 & p(\alpha_2) = 0.66 \\ p(\alpha_1) = 0.25 & p(\alpha_2) = 0.75 \end{array}$$

9.2.2. Decisión según el mínimo coste.

Aunque para algunos de los valores obtenidos de las características se estará bastante seguro de que la clasificación es correcta, habrá un rango de valores para los que la clasificación no es tan segura. ¿Qué ocurre entonces si el coste de equivocarse es mucho mayor que el de acertar? Puede tomarse por ejemplo el caso del control de calidad de una partida de piezas donde si se encuentra una defectuosa se devuelve toda la partida. En este tipo de casos el coste de acertar es muy inferior al de equivocarse y se preferirá pasarse por carta de menos que de más. Se tiene entonces un nuevo grupo de variables, λ_{ij} , que indican el coste de clasificar el objeto como clase α_i cuando es en realidad α_j . El coste por tanto de tomar una decisión es:

$$R(\alpha_i | X) = \sum_j \lambda_{ij} P(\alpha_j | X)$$

valor que se obtendrá para cada clase, eligiéndose el coste menor.

Por ejemplo para el caso de tener que decidir entre dos clases se tendría el conjunto de ecuaciones:

$$R(\alpha_1 | X) = \lambda_{11} P(\alpha_1 | X) + \lambda_{12} P(\alpha_2 | X)$$

$$R(\alpha_2 | X) = \lambda_{21} P(\alpha_1 | X) + \lambda_{22} P(\alpha_2 | X)$$

Se elegiría la clase primera si el riesgo es menor que elegir la segunda $R(\alpha_1 | X) < R(\alpha_2 | X)$, luego:

$$\lambda_{11} P(\alpha_1 | X) + \lambda_{12} P(\alpha_2 | X) < \lambda_{21} P(\alpha_1 | X) + \lambda_{22} P(\alpha_2 | X)$$

$$(\lambda_{12} - \lambda_{22}) P(\alpha_2 | X) < (\lambda_{21} - \lambda_{11}) P(\alpha_1 | X)$$

ya que el coste de equivocarse suele ser mayor que el de acertar. Sustituyendo las probabilidades a posteriori por la fórmula de Bayes vista anteriormente, lleva a:

$$(\lambda_{12} - \lambda_{22}) \frac{P(X | \alpha_2) p(\alpha_2)}{p(X)} < (\lambda_{21} - \lambda_{11}) \frac{P(X | \alpha_1) p(\alpha_1)}{p(X)}$$

$$\frac{(\lambda_{12} - \lambda_{22}) p(\alpha_2)}{(\lambda_{21} - \lambda_{11}) p(\alpha_1)} < \frac{P(X | \alpha_1)}{P(X | \alpha_2)}$$

9.2.3. Funciones discriminantes.

Algunas de las fórmulas vistas con anterioridad pueden ser tomadas como función discriminante. El argumento de estas funciones debe ser el vector de características y debe cumplir que sea máxima para la clase a la que pertenece el objeto a clasificar. Así, si denominamos $g_i(X)$ a la función discriminante, se tienen:

$$g_i(X) = P(\alpha_i | X)$$

$$g_i(X) = -R(\alpha_i | X)$$

ya que en este último caso al buscar el valor mínimo, si tomamos su valor negativo será el máximo. Si se eligiese la primera opción, cualquiera de las siguientes funciones discriminantes daría el mismo resultado.

$$g_i(X) = P(\alpha_i | X)$$

$$g_i(X) = \frac{P(X | \alpha_i) p(\alpha_i)}{p(X)}$$

$$g_i(X) = P(X | \alpha_i) p(\alpha_i)$$

$$g_i(X) = \log P(X | \alpha_i) + \log p(\alpha_i)$$

Las funciones discriminantes crean regiones de decisión, en donde su valor es mayor que para el resto de las clases, $g_i(X) > g_j(X)$. Esto es lo mismo que decir que cualquier objeto cuyo vector de características tenga un valor determinado corresponde al objeto clase i . Los puntos de paso de una zona de decisión a otra son las fronteras de decisión.

En el caso de tener que decidir entre dos clases se toma una nueva función discriminante:

$$g(X) = g_1(X) - g_2(X)$$

donde ahora dependerá del signo el que se asocie el objeto a la clase uno o dos.

9.2.4. La función de densidad normal.

La función de densidad normal, para el caso unidimensional, viene definida por la fórmula:

$$p(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left[-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right]$$

donde:

$$\mu = E[x] = \int_{-\infty}^{\infty} xp(x)dx$$

es el valor de la media, y:

$$\sigma^2 = E[(x-\mu)^2] = \int_{-\infty}^{\infty} (x-\mu)^2 p(x)dx$$

el de la varianza. Para el caso de tener varias variables:

$$p(X) = \frac{1}{(2\pi)^{\frac{n}{2}} |\Sigma|^{\frac{1}{2}}} \exp\left[-\frac{1}{2} (X-\mu)^t \Sigma^{-1} (X-\mu)\right]$$

donde X es el vector de características de dimensión n , μ el vector de medias y Σ la matriz de covarianzas (simétrica y semidefinida positiva). Donde:

$$\begin{aligned} \mu &= E[X] \\ \Sigma &= E[(X-\mu)(X-\mu)^t] \end{aligned}$$

siendo cada término individual:

$$\mu_i = E[x_i] \quad \sigma_{ij} = E[(x_i - \mu_i)(x_j - \mu_j)]$$

Mientras que el vector de medias da información del punto donde se agrupan los datos, la matriz de covarianzas indica la forma en la que se agrupan entorno a ese valor o valores. El lugar geométrico de los puntos para los que el valor de la distribución normal es constante son hiperelipsoides para los que el valor $(X-\mu)^t \Sigma^{-1} (X-\mu)$ es constante. Los ejes principales del elipsoide vienen dados por los autovectores de la matriz de covarianzas y su importancia por los autovalores. Para indicar la distancia de un valor de las características respecto a la media se utiliza la distancia de Mahalanobis:

$$r^2 = (X-\mu)^t \Sigma^{-1} (X-\mu)$$

que es la distancia euclídea normalizada por la dispersión de los datos. Con ella se penaliza más la dispersión de los datos respecto del valor medio conforme su desviación típica sea menor.

9.2.5. Funciones discriminantes para funciones de densidad normales.

Anteriormente se indicó que una función discriminante posible era:

$$g_i(X) = \log P(X | \alpha_i) + \log p(\alpha_i)$$

Si se considera ahora que la probabilidad vienen indicada por la función de densidad normal, se tiene:

$$g_i(X) = -\frac{1}{2}(X - \mu_i)' \Sigma^{-1}(X - \mu_i) - \frac{n}{2} \log 2\pi - \frac{1}{2} \log |\Sigma_i| + \log P(\alpha_i)$$

A continuación se tratarán algunos casos especiales que permitirán simplificar esta función discriminante.

Caso en que las características sean independientes y las desviaciones típicas iguales.

Si las características elegidas son estadísticamente independientes el valor de las covarianzas es cero; si además las varianzas tienen el mismo valor:

$$\Sigma = \sigma I$$

Los hiperelipsoides se transforman ahora en hiperesferas (figura 9.6). La función discriminante es ahora:

$$g_i(X) = -\frac{(X - \mu_i)'(X - \mu_i)}{2\sigma^2} - \frac{n}{2} \log 2\pi - n \log \sigma + \log P(\alpha_i)$$

Como $-\frac{n}{2} \log 2\pi - n \log \sigma$ es una constante no influye en la función discriminante por lo que se puede simplificar a :

$$g_i(X) = -\frac{(X - \mu_i)'(X - \mu_i)}{2\sigma^2} + \log P(\alpha_i)$$

Si las probabilidades de cada clase fueran iguales, $P(\alpha_i)$ también podría despreciarse, quedando entonces:

$$g_i(X) = -\frac{(X - \mu_i)'(X - \mu_i)}{2\sigma^2}$$

Este caso especial se denomina clasificador por mínima distancia ya que el valor de $(X - \mu_i)'(X - \mu_i)$ es el que importa es al ser las varianzas iguales por lo que la función discriminante puede reescribirse:

$$g'_i(X) = -(X - \mu_i)'(X - \mu_i)$$

Este método también puede considerarse como el caso en que no se tenga información de cómo es la distribución probabilística de las variables y solo se conozcan los valores de los vectores prototipos de cada clase (lo que en parte se dice al considerar el caso en que las características son independientes y de igual desviación típica). El criterio entonces es medir la distancia a cada prototipo y elegir aquel cuya distancia sea más pequeña. Es por ello que también recibe el nombre de clasificador mediante el vecino más próximo

El reconocimiento mediante la distancia euclídea es *a priori* como el bayesiano pero a diferencia de él es determinista, ya que se considera que los diversos patrones tomarán valores fijos. Esto es imposible en la práctica, pero lo que se considera aquí es que existen unos valores nominales y que los verdaderos patrones no se alejarán mucho de ellos. Si esto no fuera así habría que ir a un reconocedor estadístico.

Geométricamente la función discriminante sería la perpendicular a la línea que une los dos vectores de características y que pasa por el punto medio (figura 9.6) ya que si se desarrolla la función discriminante:

$$g_i(X) = -\frac{X^t X - 2\mu_i^t X + \mu_i^t \mu_i}{2\sigma^2} + \log P(\alpha_i)$$

como $X^t X$ es igual para todos los términos no influye, quedando

$$g_i(X) = \frac{2\mu_i^t X - \mu_i^t \mu_i}{2\sigma^2} + \log P(\alpha_i)$$

que puede rescribirse como:

$$g_i(X) = A_i^t X + A_{i0}$$

siendo:

$$A_i = \frac{\mu_i}{\sigma^2}$$

$$A_{i0} = -\frac{\mu_i^t \mu_i}{2\sigma^2} + \log P(\alpha_i)$$

ecuación de un hiperplano.

Las superficies de decisión vienen determinadas por:

$$g_i(X) = g_j(X)$$

que da la ecuación:

$$A^t (X - X_0) = 0$$

siendo:

$$A = \mu_i - \mu_j$$

$$X_0 = \frac{\mu_i + \mu_j}{2} - \frac{\sigma^2}{\|\mu_i - \mu_j\|^2} \log \frac{P(\alpha_i)}{P(\alpha_j)} (\mu_i - \mu_j)$$

hiperplano que pasa por X_0 y ortogonal al vector A , el vector que une las medias. Si las probabilidades son iguales, $P(\alpha_i) = P(\alpha_j)$, el vector pasa por el punto central.

Una variación, denominada clasificador mediante el N vecino más próximo, se presenta en la figura 9.7. Ahora cada prototipo no viene representado por un único valor sino por un conjunto de ellos. Dado un objeto nuevo a clasificar, se calcularían todas las distancias a los valores almacenados en la base de datos y se clasificaría de acuerdo a la clase a la que pertenezca el objeto con mínima distancia.

Caso de que las covarianzas sean idénticas

En este caso las variables no son independientes, por lo que los valores que no estén en las diagonales de las matrices de covarianzas dejan de ser nulos (existe una correlación significativa entre las características elegidas). Se sigue suponiendo, sin embargo, que los valores son los mismos para cada clase:

$$\Sigma_i = \Sigma$$

Ahora se tiene hiperelipsoides. De la ecuación de las funciones discriminantes para el caso de funciones de densidad normales:

$$g_i(X) = -\frac{1}{2} (X - \mu_i)^t \Sigma^{-1} (X - \mu_i) - \frac{n}{2} \log 2\pi - \frac{1}{2} \log |\Sigma_i| + \log P(\alpha_i)$$

al ser Σ_i la misma queda:

$$g_i(X) = -\frac{1}{2}(X - \mu_i)' \Sigma^{-1}(X - \mu_i) + \log P(\alpha_i)$$

pudiéndose ignorar $\log P(\alpha_i)$ si las probabilidades a priori son la misma para cada clase. En este caso se mide la distancia de Mahalanobis a cada vector media y se toma el valor más pequeño (figura 9.8). Así, por ejemplo si se tuviese:

$$\Sigma_i = \Sigma = \begin{pmatrix} 1/2 & 0 \\ 0 & 1/4 \end{pmatrix} \Rightarrow \Sigma^{-1} = \begin{pmatrix} 2 & 0 \\ 0 & 4 \end{pmatrix}$$

$$\mu_1 = \begin{pmatrix} 0 \\ 3 \end{pmatrix} \quad \mu_2 = \begin{pmatrix} 5 \\ 2 \end{pmatrix} \quad \mu_3 = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$$

$$P(\alpha_1) = P(\alpha_2) = P(\alpha_3)$$

La función discriminante para la primera clase es:

$$g_1(X) = (X - \mu_1)' \Sigma^{-1}(X - \mu_1) = (x_1 - 0, x_2 - 3) \begin{pmatrix} 2 & 0 \\ 0 & 4 \end{pmatrix} \begin{pmatrix} x_1 - 0 \\ x_2 - 3 \end{pmatrix}$$

$$g_1(X) = (2x_1, 4x_2 - 12) \begin{pmatrix} x_1 - 0 \\ x_2 - 3 \end{pmatrix} = 2x_1^2 + 4(x_2 - 3)^2$$

y de forma análoga:

$$g_2(X) = 2(x_1 - 5)^2 + 4(x_2 - 2)^2$$

$$g_3(X) = 2(x_1 - 1)^2 + 4x_2^2$$

Las fronteras de las regiones de decisión vendrían dadas por:

$$g_1(X) = g_2(X) = 2x_1^2 + 4(x_2 - 3)^2 = 2(x_1 - 5)^2 + 4(x_2 - 2)^2$$

ecuación del plano P_{12} :

$$g_1(X) = g_2(X) = 2x_1^2 + 4(x_2 - 3)^2 = 2(x_1 - 5)^2 + 4(x_2 - 2)^2 \Rightarrow 10x_1 - 4x_2 - 15 = 0$$

análogamente, P_{23} :

$$g_2(X) = g_3(X) \Rightarrow -x_1 - x_2 + 4 = 0$$

y P_{13}

$$g_1(X) = g_3(X) \Rightarrow 2x_1 - 12x_2 + 17 = 0$$

Ejemplo que se muestra gráficamente en la figura 9.9.

Si se tiene en cuenta que del término $g_i(X) = (X - \mu_i)' \Sigma^{-1}(X - \mu_i)$, $X' \Sigma^{-1} X$ es igual para todos, queda la siguiente ecuación lineal si se elimina:

$$g_i(X) = A_i' X + A_{i0}$$

siendo:

$$A_i = \Sigma^{-1} \mu_i$$

$$A_{i0} = -\frac{\mu_i' \Sigma^{-1} \mu_i}{2} + \log P(\alpha_i)$$

Las superficies de decisión vienen determinadas por:

$$g_i(X) = g_j(X)$$

que da la ecuación:

$$A' (X - X_0) = 0$$

siendo:

$$A = \Sigma^{-1}(\mu_i - \mu_j)$$

$$X_0 = \frac{\mu_i + \mu_j}{2} - \frac{\sigma^2}{(\mu_i - \mu_j)' \Sigma^{-1} (\mu_i - \mu_j)} \log \frac{P(\alpha_i)}{P(\alpha_j)} (\mu_i - \mu_j)$$

Caso general.

En el caso general, Σ , puede tomar cualquier valor. La función discriminante quedará ahora:

$$g_i(X) = -\frac{1}{2}(X - \mu_i)' \Sigma^{-1}(X - \mu_i) - \frac{1}{2} \log |\Sigma_i| + \log P(\alpha_i)$$

que puede reescribirse:

$$g_i(X) = X' A_i X + B_i' X + C_i$$

ecuación de una cuadrática y cuyos términos son:

$$A_i = -\frac{1}{2} \Sigma_i^{-1}$$

$$B_i = \Sigma_i^{-1} \mu_i$$

$$C_i = -\frac{1}{2} \mu_i' \Sigma_i^{-1} \mu_i - \frac{1}{2} \log |\Sigma_i| + \log P(\alpha_i)$$

Las superficies de decisión son hipercuadráticas. Los casos representados en la figura 9.10 suponen que las covarianzas entre las dos características son nulas. En el caso a) la varianza de la clase 1 es más pequeña que la de clase 2. Si las varianzas para las dos características son iguales las superficies de decisión son circunferencias. Si una varianza es mayor se tienen elipses (caso b). En c) ambas clases tienen la misma varianza para la característica 2, pero mucho menor en la clase 1 para la característica 1 que la clase 2. Las superficies de decisión son parábolas. Esta cambia a hiperparábola si se aumenta la varianza para la característica 1 para la clase 2. La hiperparábola degenera en dos rectas en el caso e).

9.3 Agrupamiento de datos.

Hasta ahora los métodos vistos se basaban en el conocimiento de todos los posibles objetos en la imagen. Pero supóngase que se encuentra un elemento como el descrito en la figura 9.11. Cualquiera de los dos métodos anteriores lo clasificaría como perteneciente a una clase aunque está muy alejado de ellas. Una posible solución es poner un límite. Así si se aplicase la distancia euclídea se podría limitar el máximo que se acepta y si el enfoque fuera estadístico la mínima probabilidad aceptable. Otro enfoque es el denominado agrupamiento de datos (*clustering* en la literatura anglosajona) en el que los objetos se agrupan por sí solos.

Al algoritmo de agrupación se le presentan los vectores de características de los objetos a clasificar y el algoritmo de forma autónoma agrupa los objetos en clases. Lógicamente serán usados cuando no hay un conocimiento a priori del número y tipo de clases.

9.3.1. Algoritmo de las distancias encadenadas.

El algoritmo sigue los siguientes pasos:

1. Se tiene una serie de objetos a clasificar, O_1, O_2, \dots, O_n y se escoge uno al azar, O_i .

2. A continuación se realiza la cadena siendo O_i el primer elemento, el segundo el elemento más cercano a él, el tercero el más cercano al segundo y así sucesivamente (de ahí el nombre de distancias encadenadas).
3. El siguiente paso es el establecimiento de las clases. Para ello se analizan los valores de las distancias euclídeas entre cada elemento y el elemento siguiente. Si una distancia y la siguiente se diferencian más de un umbral dado se considera que existe una nueva clase. Si no, se pasa a analizar la distancia siguiente.

La representación gráfica de este método puede verse en la figura 9.12.

9.3.2. Algoritmo Max-Min.

El algoritmo sigue los siguientes pasos:

1. De nuevo se tiene una serie de objetos a clasificar, O_1, O_2, \dots, O_n y se escoge uno al azar, O_i , que se elegirá como primera clase.
2. A continuación se obtiene la distancia euclídea del resto de los elementos respecto a O_i . Para determinar si existe una segunda clase se elige el elemento que está más lejano, y si la distancia que los separa es mayor que un valor prefijado se considera como segunda clase.
3. Para ver si son necesarias más clases o no se realizaría una primera clasificación de los $N-2$ objetos restantes a las dos clases conocidas. Para ello se obtienen las parejas de distancias entre los $N-2$ puntos restantes y las dos clases anteriores. Para cada elemento se toma la mínima de las dos. De entre ellas se elige la máxima, que si es mayor a un tanto por ciento de la distancia entre las dos clases se elegirá a dicho elemento como prototipo de una tercera (de aquí el nombre de algoritmo Min-Max).
4. El proceso se vuelve a realizar solo que tomando ahora tres distancias y obteniéndose la cuarta clase.
5. Cuando no se obtengan más clases se ordenan el resto de los objetos según el criterio de mínima distancia.

La representación gráfica de este método puede verse en la figura 9.13.

9.3.3. Algoritmo K-medias.

Este algoritmo parte del conocimiento a priori del número de clases pero no de sus características.

1. De entre la serie a clasificar, O_1, O_2, \dots, O_n se escogen, de forma arbitraria, tantos elementos como número de clases y se considera que constituyen los centroides (o valores prototipo) de cada clase.
2. El resto de los elementos se asignan a cada clase siguiendo el criterio de mínima distancia a los centroides antes elegidos.
3. Se recalculan los centroides de cada clase. Para ello se toma la media de todos los valores dentro de cada clase.
4. Se vuelven a asignar, ahora todos los elementos, a cada clase con el criterio de mínima distancia.
5. Se vuelven a calcular los centroides. Si no varían se considera que el algoritmo a terminado, si no, se vuelve a repetir el paso anterior.

Una serie de factores influyen el comportamiento del algoritmo: el número prefijado de clases, los centroides iniciales, la distribución geométrica de los valores. Por ello no es extraño que se tengan que hacer varias suposiciones hasta encontrar una que resuelva el problema.

9.3.4. Algoritmo ISODATA.

El algoritmo ISODATA modifica el caso anterior permitiendo que aparezcan o desaparezcan el número a priori de clases.

1. De entre la serie a clasificar, O_1, O_2, \dots, O_n se escogen K elementos que constituyen los centroides de cada clase.
2. Se aplica el algoritmo K-medias visto anteriormente pero con la siguientes variaciones:
 - Se eliminan las clases con un número inferior de elementos a un valor prefijado de antemano.
 - Se vuelven a calcular los centroides de cada clase.
3. Se calcula la distancia media para cada clase.
4. Se obtiene la distancia media de todas las clases.
5. Se estudian posibles uniones (hay menos clases de las que se pensaba). Para ello se comprueba si el número de clases es mayor del doble de las esperadas. Si es así se pasa a la unión, si no se estudian posibles divisiones (hay más de las prefijadas).
6. Divisiones de clases.
 - Se calculan las desviaciones típicas para cada una de las características de cada clase.
 - Para cada clase se averigua cual característica tiene una desviación típica mayor.
 - Para que una clase se divida en dos debe cumplirse:
 - La desviación típica es mayor que un valor prefijado.
 - El número de elementos es mayor del doble de elementos mínimo.
 - El número de clases es menor de la mitad de clases esperada.
 - La dispersión media de esa clase es mayor que la media de todas las clases.

El proceso de división consiste en la obtención de dos nuevos centroides. Para ello se obtienen los dos elementos más alejados del centroide y los dos nuevos son la media entre ellos y el centroide antiguo.

7. Si ha existido alguna división se vuelve a empezar con el algoritmo K-medias.
8. Uniones de clase.
 - Se calculan todas las distancias entre parejas de centroides.
 - Todas las distancias se comparan con un valor, y las que sean menores se ordenan de menor a mayor.
 - Empezando por la distancia menor se van tomando las clases y si ninguna se ha fusionado en esa iteración se obtiene el nuevo centroide, media de todos los elementos.
9. Se vuelve a realizar el algoritmo K-medias a menos que los centroides no varíen o el número de iteraciones supere un valor prefijado.

9.3.5. Algoritmo Vector Quantization.

Este método también intenta superar la limitación impuesta en el algoritmo de las k-medias de conocer a priori el número de clases. De hecho, se comienza con ninguna sola clase asignada. Los pasos a repetir son:

1. Para cada patrón se calcula su distancia euclídea con todos los centroides existentes. Para el primer elemento, él mismo constituirá el primer centroide.
2. Se toma el centroide más cercano y se compara su distancia con un umbral prefijado, que indicará la tolerancia para las características.
3. Si la distancia es menor que el umbral el elemento se asocia a esa clase. Se recalcula entonces el centroide obteniendo la media de todos los elementos.
4. Si la distancia es mayor que la tolerancia está indicando que se necesita una nueva clase por lo que se crea siendo el valor del centroide el del elemento.

Los inconvenientes de este método son:

- Depende del orden en el que se presentan los elementos.
- El valor de la tolerancia se define antes de empezar la clasificación. Por lo que se deberá probar con varias.

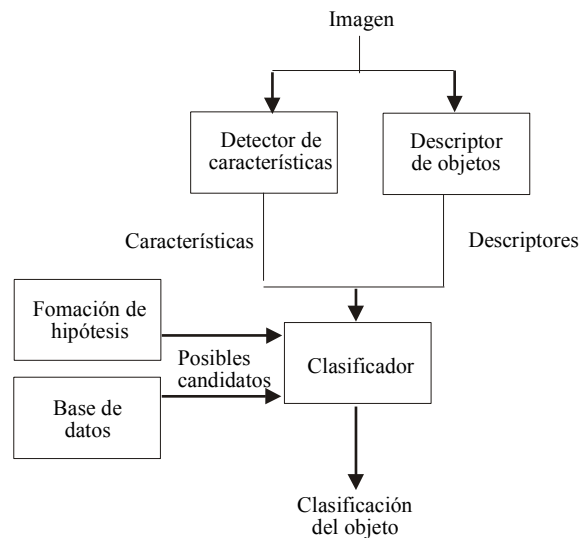


Figura 9.1. Esquema general del reconocimiento de objetos.

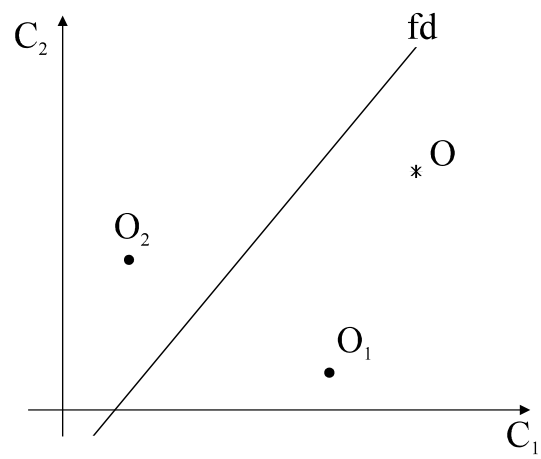


Figura.9.2 Función discriminante y vectores prototipo

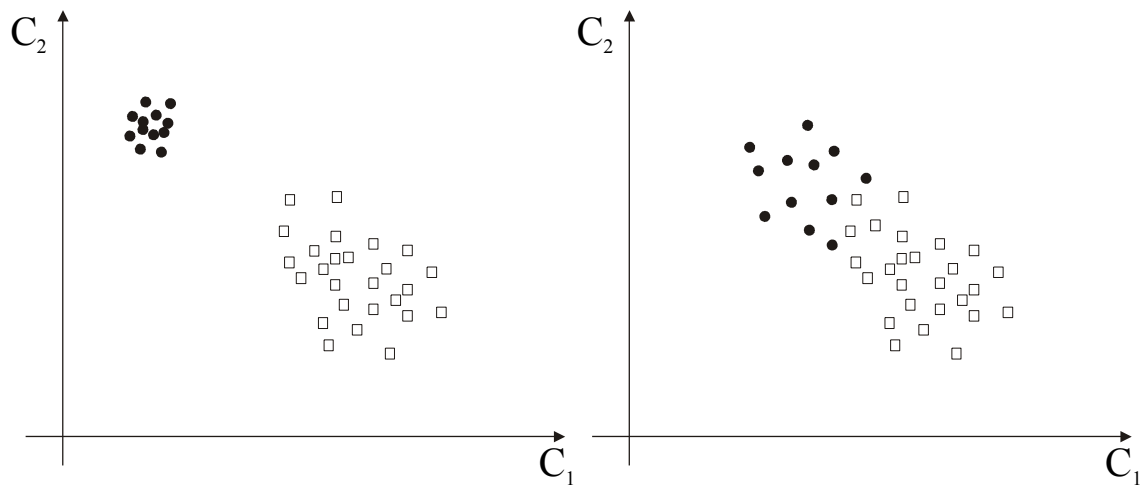


Figura 9.3. Enfoque estadístico del reconocimiento de patrones.

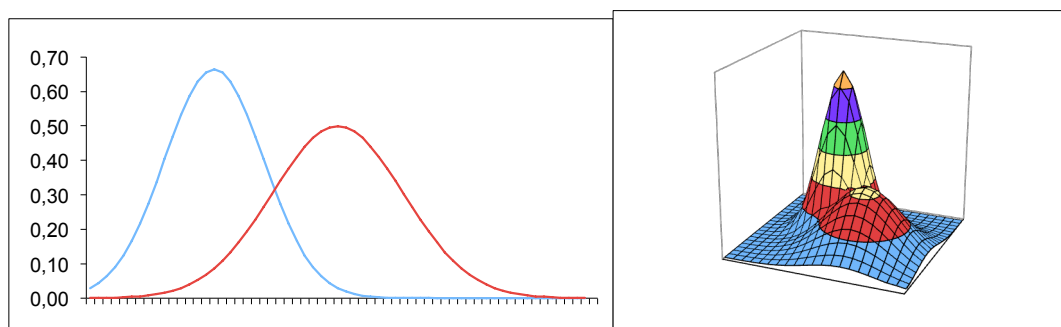


Figura 9.4 Probabilidades a priori

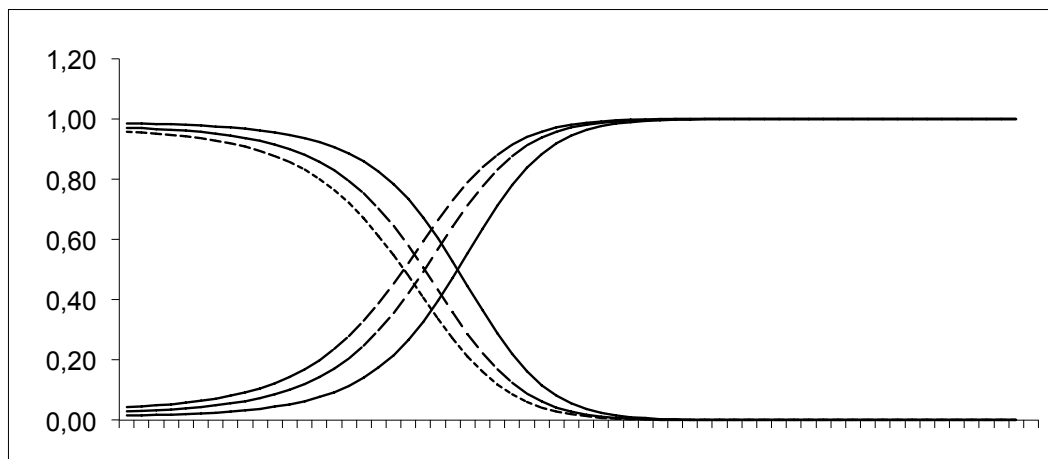


Figura 9.5 Probabilidades a posteriori.

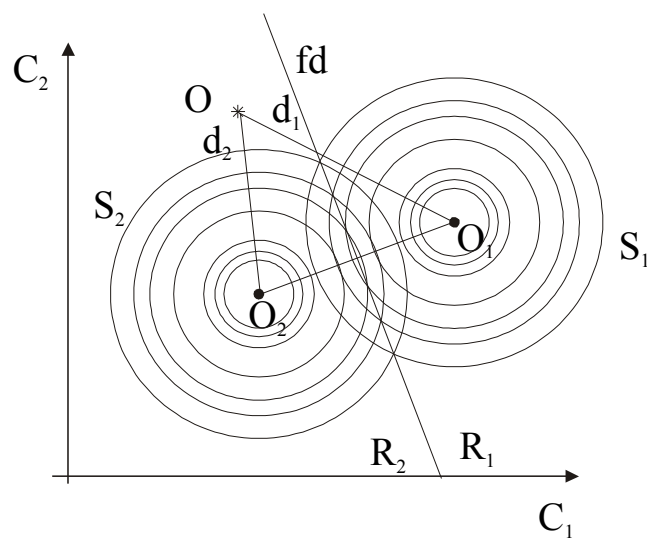


Figura 9.6. Clasificador de la mínima distancia

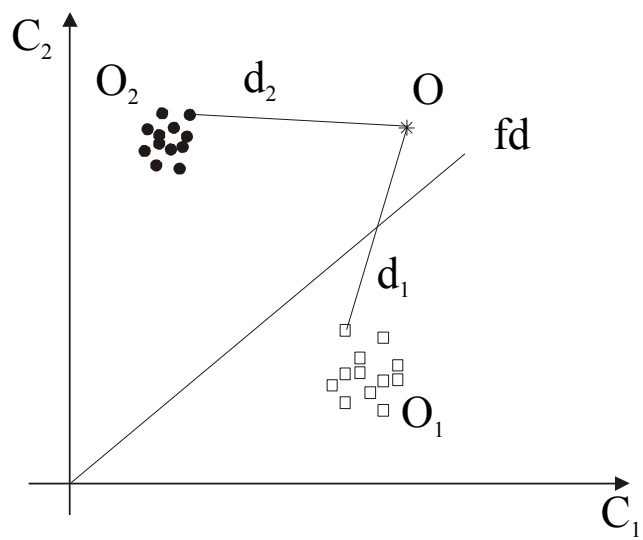


Figura 9.7. Clasificador mediante el N-vecino más próximo.

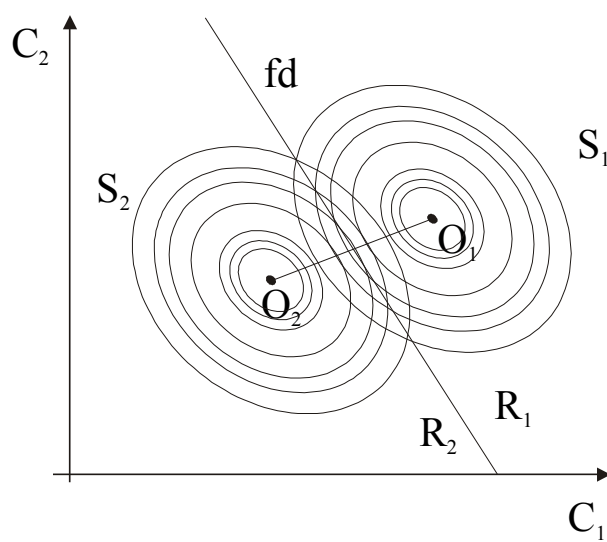


Figura 9.8. Clasificador mediante la mínima distancia de Mahalanobis

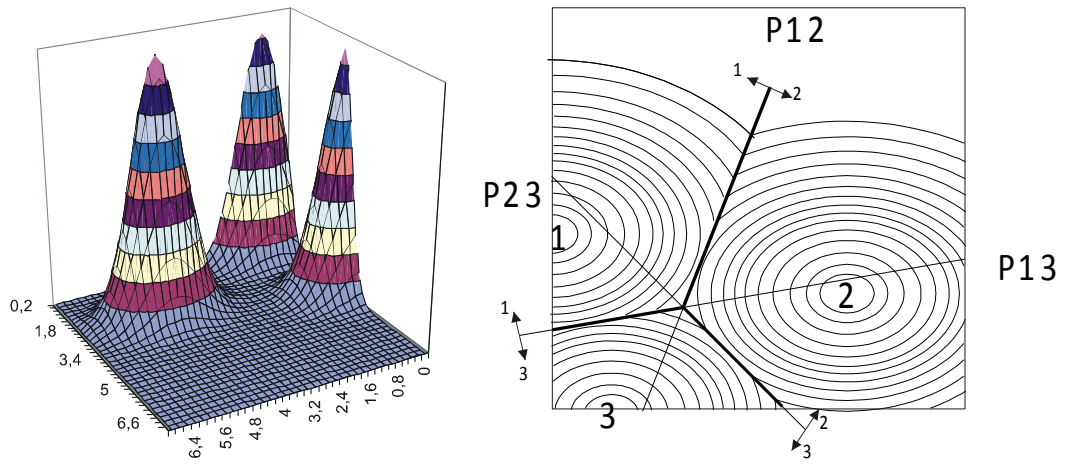


Figura 9.9. Ejemplo de clasificador mediante la mínima distancia de Mahalanobis

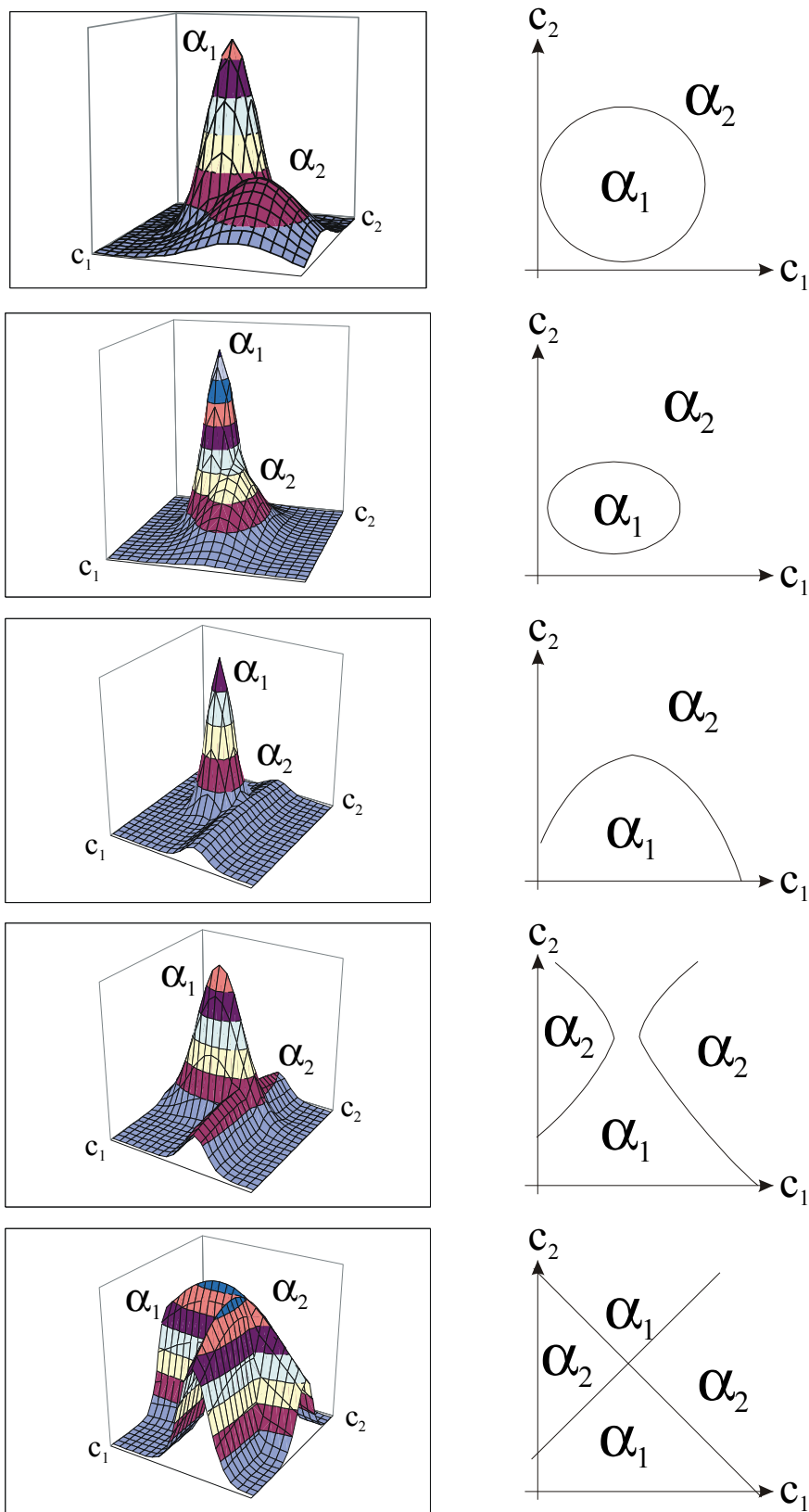


Figura 9.10. Superficies de decisión para el caso general

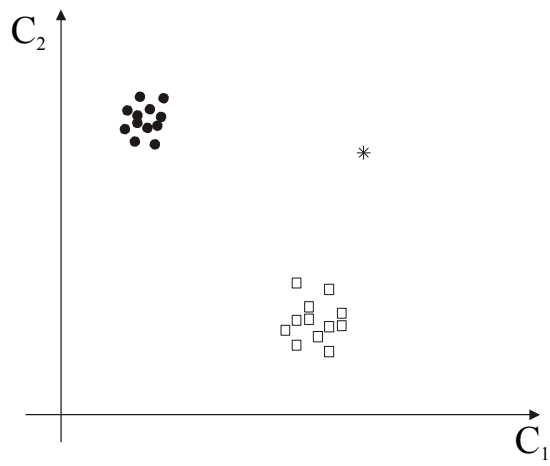


Figura 9.11. Aparición de una nueva clase.

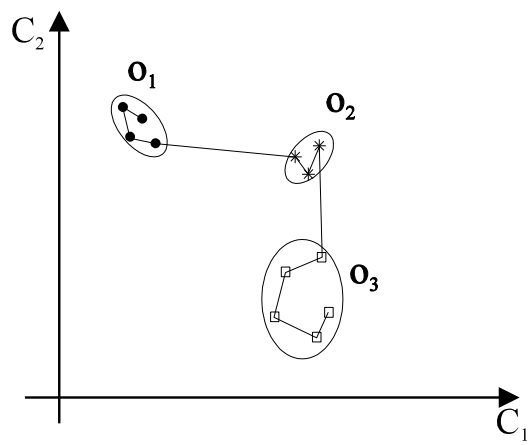


Figura 9.12. Algoritmo de las distancias encadenadas.

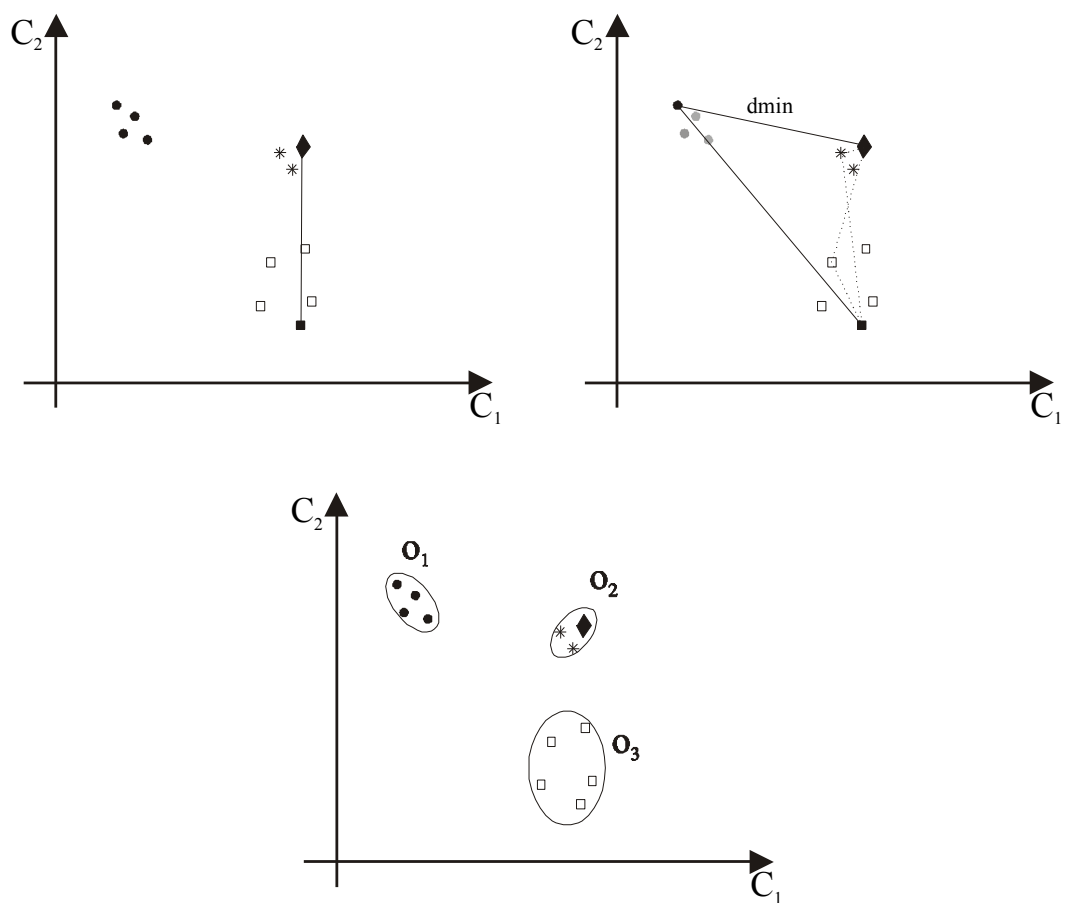


Figura 9.13 Algoritmo min-max