

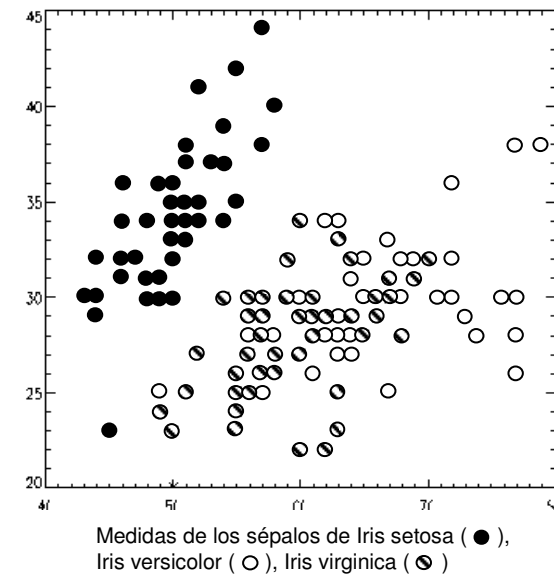
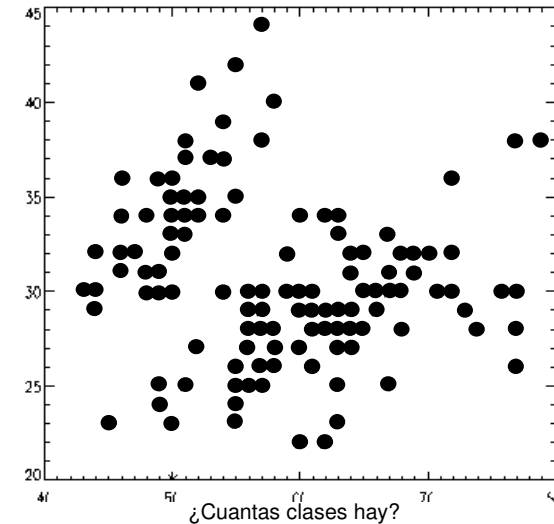
Introducción

- **Objetivo:**
 - Dado un problema de R.P. para el que se posee un conjunto representativo $H=\{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_N\}$ de muestras no etiquetadas, definir las clases a las que pertenecen las muestras presentes en H .
- **Información Disponible:**
 - Un conjunto H de muestras no etiquetadas que deben poseer las siguientes características:
 - › La cardinalidad debe ser H grande.
 - › Todas las clases han de estar representadas en H
 - › Pueden definirse subconjuntos de H en los que las muestras se agrupan de forma natural.
- **Aplicabilidad**
 - Se dispone de muestras pero se carece de sus etiquetas.
 - Las características de las clases varían con el tiempo y el sistema debe adaptarse reestimándolas.

Metodologías

- **Algunas Metodologías**
 - Paramétrica
 - › Estimación a partir de mezclas de funciones de densidad.
 - No Paramétrica
 - › Basado en la partición de los datos en subconjuntos mediante algún criterio de agrupamiento
- **Ejemplo**

Aunque los elementos marcados con ● se agrupan de forma natural, no es sencillo elegir agrupaciones naturales para los elementos marcados con ○, ⊙.



Mezclas de Densidades

- **Información Disponible**

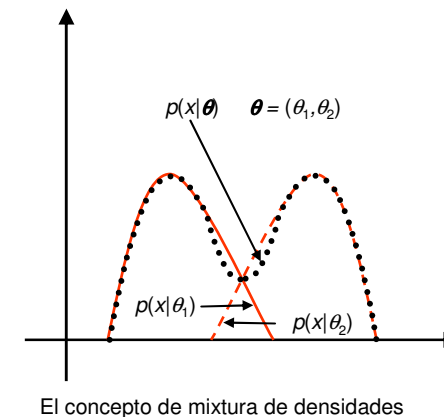
- El número de clases c .
- La forma de $p(\mathbf{x}|w_k, \theta_k)$, donde θ_k representa un vector de parámetros para la clase w_k .
- El conjunto de entrenamiento $H = \{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_N\}$

- **Asunciones**

- Todas las muestras en H están generadas por la mezcla de funciones de densidad:

$$p(\mathbf{x} | \theta) = \sum_{k=1}^c p(\mathbf{x} | w_k, \theta_k) P(w_k) \quad \theta = (\theta_1, \theta_2, \dots, \theta_c)$$

Esto implica que el modelo de generación de datos es “Elegir una clase con probabilidad $P(w_k)$ y generar un elemento a partir de su función de densidad $p(\mathbf{x}|w_k, \theta_k)$ ”



Estimación de los parámetros

- **Estimación por máxima verosimilitud**

- Para estimar $\theta = (\theta_1, \theta_2, \dots, \theta_c)$ y $P(w_k)$ por máxima verosimilitud se debe resolver las ecuaciones:

$$P(w_k) = \frac{1}{N} \sum_{n=1}^N P(w_k | \mathbf{x}_n, \theta)$$

$$\sum_{n=1}^N P(w_k | \mathbf{x}_n, \theta) \nabla_{\theta_k} \ln(p(\mathbf{x}_n | w_k, \theta_k)) = 0 \quad k = 1, 2, \dots, c$$

$$P(w_k | \mathbf{x}_n, \theta) = \frac{p(\mathbf{x}_n | w_k, \theta_k) P(w_k)}{\sum_{k=1}^c p(\mathbf{x}_n | w_k, \theta_k) P(w_k)}$$

Estas ecuaciones son en general complejas de resolver para una distribución arbitraria.

- **El caso gaussiano**

- Si las distribuciones de las distintas clases son gaussianas es posible aplicar el método de Maximización de la Esperanza (ME) para obtener un esquema iterativo de resolución.

Ejemplo de estimación en mezclas

• Vectores de medias desconocidas

- En este caso el vector de parámetros $\theta = (\theta_1, \theta_2, \dots, \theta_c)$ está compuesto por el vector de medias $\mu = (\mu_1, \mu_2, \dots, \mu_c)$ que se calcula de forma iterativa a partir de una estimación inicial $\mu_k^{(0)}$, $k=1, 2, \dots, c$ como:

$$\mu_k^{(r+1)} = \frac{\sum_{n=1}^N P(w_k | \mathbf{x}_n, \boldsymbol{\mu}^{(r)}) \mathbf{x}_n}{\sum_{n=1}^N P(w_k | \mathbf{x}_n, \boldsymbol{\mu}^{(r)})}, \quad k = 1, 2, \dots, c \quad \boldsymbol{\mu} = (\mu_1, \mu_2, \dots, \mu_c)$$

- El esquema iterativo no tiene por qué converger a la solución óptima. Es necesario además darle una buena estimación inicial.

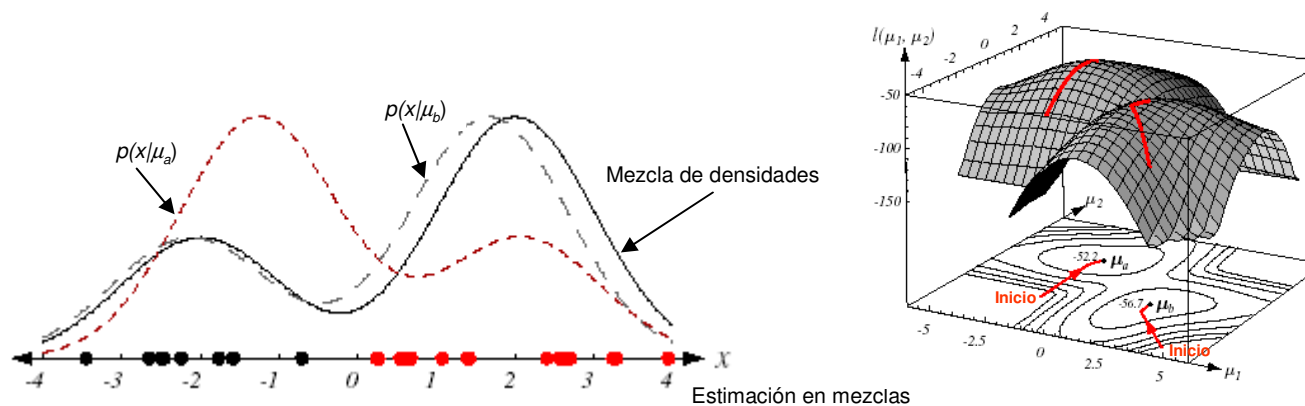


Gráfico de: Richard O. Duda, Peter E. Hart, and David G. Stork, Pattern Classification. Copyright (c) 2001 por John Wiley & Sons, Inc.

El Algoritmo Isodata (k -medias)

- **Metodología**

- Se basa en una aproximación de las ecuaciones del caso gaussiano basada en:

Si \mathbf{x}_n es próximo a μ_k $P(w_k|\mathbf{x}_n, \mu_k) \approx 1$ si \mathbf{x}_n es lejano a μ_k $P(w_k|\mathbf{x}_n, \mu_k) \approx 0$.

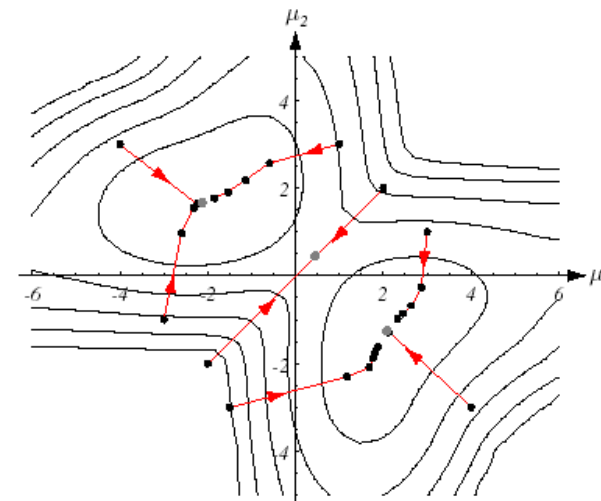
Esto lleva al siguiente algoritmo aproximado:

Paso 0: Elegir valores iniciales para las medias $\mu_1, \mu_2, \dots, \mu_c$

Paso 1: Clasificar las N muestras
asignándolas a las de su
media más próxima

Paso 2: Recalcular las medias como
el valor medio de las muestras
en su clase.

Paso 3: Si se da la condición de
convergencia parar, en
otro caso ir al Paso 2.



Ejemplo del algoritmo Isodata

Gráfico de: Richard O. Duda, Peter E. Hart, and David G. Stork, Pattern Classification. Copyright (c) 2001 por John Wiley & Sons, Inc.

Métodos no Paramétricos

- **Están basados en agrupar los datos mediante algún criterio de proximidad.**
- **Por tanto, es necesario definir:**
 - La medida de proximidad
 - › Indica cuan “similares” o “distintos” pueden considerarse dos vectores de características
 - El criterio de agrupamiento
 - › Que se define generalmente mediante alguna función de error que mide la “calidad” de los agrupamientos
 - El algoritmo de agrupamiento
 - › Que se encarga de obtener el óptimo global de la función de error o en su defecto un óptimo local
 - El método de validación de los resultados
 - › Que se encarga de verificar la validez de los resultados generalmente mediante tests estadísticos

El Cálculo del Agrupamiento Óptimo

- **La Medida de Proximidad**
 - Dado el conjunto de muestras $H=\{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_N\}$, se define una medida $d(\mathbf{x}_i, \mathbf{x}_j)$ que tiene una magnitud pequeña si \mathbf{x}_i y \mathbf{x}_j son similares y grande si \mathbf{x}_i y \mathbf{x}_j son distintos
- **El Criterio de Agrupamiento**
 - Dada entonces una partición $P=\{H_1, H_2, \dots, H_c\}$ del conjunto H se construye una función $E(P)$ que mida la bondad de la partición P en base a $d(\mathbf{x}_i, \mathbf{x}_j)$. Los criterios de agrupamiento suelen estar basados en que $d(\mathbf{x}_i, \mathbf{x}_j)$ debe ser pequeña si \mathbf{x}_i y \mathbf{x}_j pertenecen al mismo agrupamiento y grande en caso contrario.
- **El Algoritmo de Optimización**
 - Para resolver el problema de agrupamiento se busca la partición óptima que minimice $E(P)$. Sin embargo el problema es muy complejo: Dado un conjunto con N datos el número de posibles particiones en c clases es del orden de $c^N/c!$.
Esto da para $N=100$ y $c=10$ del orden de 3×10^{93} particiones

Estrategias de Agrupamiento

- **Algunas estrategias**

- Iterativas

- Basado en optimizar la función $E(P)$ de forma iterativa

- › Ventajas

- Se puede modelizar su comportamiento

- › Inconvenientes

- Dependencia del punto inicial donde comienza la iteración.

- Tiende a quedarse atrapado en óptimos locales

- Jerárquicas

- Basado en organizar los agrupamientos de forma jerárquica.

- Se dividen en aglomerativas y divisivas

- › Ventajas

- No dependen del punto inicial

- › Inconvenientes

- Una vez realizada una asignación de una muestra a un agrupamiento no puede deshacerse la acción.

Agrupamiento Básico Iterativo mediante Error Cuadrático Medio (ABIECM)

- **Medida de proximidad**

- La distancia euclídea al cuadrado: $d(\mathbf{x}_i, \mathbf{x}_j) = \|\mathbf{x}_i - \mathbf{x}_j\|^2$

- **Criterio de Agrupamiento**

Suma de Dispersiones
 Dispersión en el agrupamiento k :
 Suma de las distancias de los elementos
 del agrupamiento k a su media

$$E_{ECM}(P) = \sum_{k=1}^c \sum_{\mathbf{x}_n \in H_k} \|\mathbf{x}_n - \mathbf{m}_k\|^2, \quad \mathbf{m}_k = \frac{1}{N_k} \sum_{\mathbf{x}_n \in H_k} \mathbf{x}_n, \quad N_k = |H_k|$$

- **Algoritmo de Agrupamiento basado en:**

- Dada una partición $P = \{H_1, H_2, \dots, H_c\}$, mover un elemento \mathbf{x}_n de un agrupamiento H_k al H_l , creando una partición P' y buscando que $E_{ECM}(P') < E_{ECM}(P)$.

- **Aplicabilidad**

- Cuando los agrupamientos están bien separados y hay poca variación en su número de elementos

Algoritmo ABIECM

- Selección de la Partición**

- Dada una partición $P=\{H_1, H_2, \dots, H_c\}$, al mover un elemento \mathbf{x}_n de un agrupamiento H_k a H_l la variación en E_{ECM} es:

$$E_{ECM}(P') = E_{ECM}(P) + \frac{N_l}{N_l + 1} \|\mathbf{x}_n - \mathbf{m}_l\|^2 - \frac{N_k}{N_k - 1} \|\mathbf{x}_n - \mathbf{m}_k\|^2$$

- Descripción del Algoritmo**

Paso 0 Elegir una partición inicial $P=\{H_1, H_2, \dots, H_c\}$, de las N muestras en agrupamientos. Calcular las medias \mathbf{m}_k y $E_{ECM}(P)$.

Paso 1 Seleccionar un candidato \mathbf{x} de H_k

Paso 2 Si $N_k=1$ ir al Paso 4
 en otro caso calcular: $\rho_l = \begin{cases} \frac{N_l}{N_l + 1} \|\mathbf{x} - \mathbf{m}_l\|^2 & l \neq k \\ \frac{N_l}{N_l - 1} \|\mathbf{x} - \mathbf{m}_l\|^2 & l = k \end{cases}$
 $\rho_i = \min \rho_l$

Paso 3 Si $i \neq k$ Cambiar a \mathbf{x} a H_i . Recalcular $E_{ECM}(P)$, \mathbf{m}_i , \mathbf{m}_k

Paso 4 Si $E_{ECM}(P)$ no ha cambiado en N intentos parar
 en otro caso ir al Paso 1

- Convergencia:** A un óptimo local en un n° finito de pasos

Agrupamiento Jerárquico Aglomerativo Mínimo Cuadrático (AJAEMC)

- **Medida de proximidad**

- La distancia euclídea al cuadrado: $d(\mathbf{x}_i, \mathbf{x}_j) = \|\mathbf{x}_i - \mathbf{x}_j\|^2$

- **Criterio de Agrupamiento**

$$E_{ECM}(P) = \sum_{k=1}^c \sum_{\mathbf{x}_n \in H_k} \|\mathbf{x}_n - \mathbf{m}_k\|^2, \quad \mathbf{m}_k = \frac{1}{N_k} \sum_{\mathbf{x}_n \in H_k} \mathbf{x}_n, \quad N_k = |H_k|$$

- **Algoritmo de Agrupamiento basado en:**

- Dada una partición $P = \{H_1, H_2, \dots, H_c\}$, unir dos agrupamientos H_k y H_j , creando una partición P' y buscando que $E_{ECM}(P') < E_{ECM}(P)$.

- **Aplicabilidad**

- Da buenos resultados cuando los agrupamientos están bien separados y no se conoce una buena partición inicial.

Algoritmo AJAEMC

- **Selección de la Partición**

- Dada una partición $P=\{H_1, H_2, \dots, H_c\}$, al unir dos agrupamientos H_k y H_l la variación en E_{ECM} es:

$$E_{ECM}(P') = E_{ECM}(P) + \frac{N_k N_l}{N_k + N_l} \|\mathbf{m}_k - \mathbf{m}_l\|^2$$

- **Descripción del Algoritmo**

Paso 0 Hacer \underline{c} (número actual de agrupamientos)= N . Construir una partición inicial $P=\{H_1, H_2, \dots, H_N\}$, con $H_n=\{\mathbf{x}_n\}$. Construir las medias $\mathbf{m}_n=\mathbf{x}_n$.

Paso 1 Si $\underline{c} = c$ (número deseado de agrupamientos). Parar

Paso 2 Encontrar los agrupamientos H_k, H_l para los que es mínimo:

$$\Delta E_{ECM} = \frac{N_k N_l}{N_k + N_l} \|\mathbf{m}_k - \mathbf{m}_l\|^2$$

Paso 3 Hacer $H_k = H_k \cup H_l$. Borrar H_l . Recalcular \mathbf{m}_k, N_k .
Decrementar \underline{c} una unidad

Paso 4 Ir al Paso 1

- **Convergencia:** En un n° finito de pasos