

## Estudio de segmentación de consumidores

**IN58B v1.1**

Ingeniería de Marketing

Marcel Goic F.

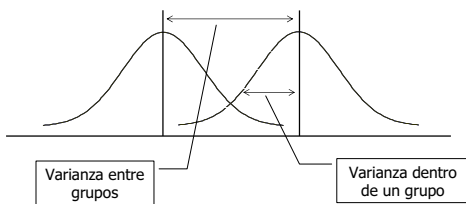
1

## Motivación

- Supongamos que tenemos que clasificar (segmentar) a los clientes de una multitienda.
  - ¿Qué variables utilizamos para clasificar?
  - ¿Qué criterios usamos para decir que dos clientes pertenecen al mismo grupo?
  - ¿Cuántos grupos debemos hacer?
- En general buscaremos grupos significativos, alcanzables, homogéneos y diferenciados de los demás.

2

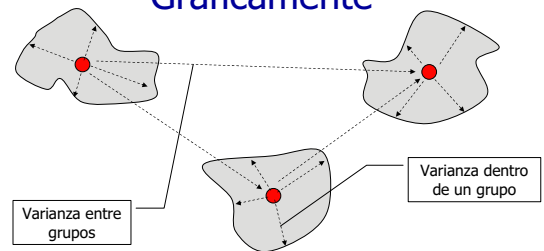
## Gráficamente



$$\max_{C \in \mathcal{C}} \left\{ \frac{\text{varianza entre grupos}}{\text{varianza en los grupos}} \right\} \quad C = \text{Conjunto de clusters posibles}$$

3

## Gráficamente



$$\max_{C \in \mathcal{C}} \left\{ \frac{\text{varianza entre grupos}}{\text{varianza en los grupos}} \right\} \quad C = \text{Conjunto de clusters posibles}$$

4

## Distinción preliminar

- La primera distinción que hay que hacer es entre técnica y aplicación:
  - **Técnica:** Herramientas estadísticas, algorítmicas, etc. para agrupar objetos (Análisis de conglomerados)
  - **Aplicación:** Consideraciones para la aplicación de técnicas de agrupación a problemas de marketing.
- Veremos primero las técnicas y luego explicaremos como aplicarlas a problemas de estructura de mercado:
  - Segmentación de clientes.
  - Clasificación de productos.

5

## Análisis de conglomerados

6

## Elementos generales

- Para poder agrupar objetos tenemos que definir:
  - **Medidas de asociación:** Corresponde a una indicación de que tan cercanos son dos objetos dada sus características.
  - **Métodos de asignación:** Corresponde a un esquema que indica como asignar los objetos a los grupos.

7

## Medidas de asociación

- Existen muchas medidas posibles y hay discusión respecto de la adecuación de cada una a cada problema.
- La medida de distancia adecuada dependerá del tipo de variable.
- En primera instancia distinguimos:
  - Medidas para variables continuas.
  - Medidas para variables discretas.

8

## Medidas de asociación continua

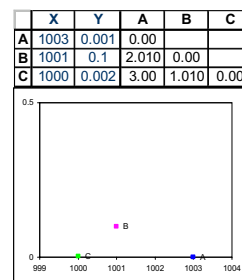
- Consideremos dos objetos  $i$  y  $j$  caracterizados por vectores  $x_i$  y  $x_j$  de  $A$  atributos.
- Medidas de distancia:

$$\left. \begin{array}{ll} \text{Euclideana} & d(i, j) = \sqrt{\sum_{a=1}^A (x_a^i - x_a^j)^2} \\ \text{Absoluta} & d(i, j) = \sum_{a=1}^A |x_a^i - x_a^j| \\ \text{Chebychev} & d(i, j) = \max_a \{x_a^i - x_a^j\} \end{array} \right\} \begin{array}{l} \text{Opción:} \\ w_a \text{ importancia} \\ \text{atributo } a \end{array}$$

9

## Respecto a las escalas

- ¿Cuál es el par de objetos mas cercanos? (distancia euclidiana).
- Para poder hacer cualquier comparación las variables deben estar en una misma escala de medida.



10

## Medidas de asociación discreta

- Consideremos dos objetos  $i$  y  $j$  caracterizados por vectores  $x_i$  y  $x_j$  de  $A$  atributos con valores en  $\{0,1\}$  (cumple o no cumple una condición)
- Sean:
  - $a$ : # atributos en que  $i$  y  $j$  cumplen
  - $b$ : # atributos en que  $i$  si cumple y  $j$  no cumple.
  - $c$ : # atributos en que  $i$  no cumple y  $j$  si cumple.
  - $d$ : # atributos en que  $i$  y  $j$  no cumplen.

$$\text{Jaccard} \quad d(i, j) = \frac{a}{a + b + c}$$

$$\text{Sokal-Sneath} \quad d(i, j) = \frac{a + d}{b + c}$$

11

## Métodos de asignación

- Corresponden a la lógica en que los objetos se asignan a cada grupo.
- Veremos dos tipos de métodos:
  - **Métodos jerárquicos:** Los objetos se agrupan (dividen) por partes hasta clasificar todos los objetos.
  - **Métodos de partición:** Se tiene un numero de grupos predefinidos y cada objeto se ubica en un grupo hasta alcanzar estabilidad.

12

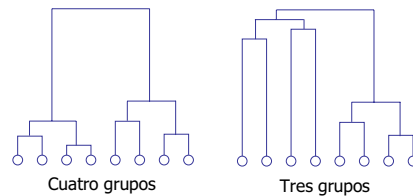
## Métodos Jerárquicos

- En cada iteración se determina un criterio de aglomeración (división) en cada iteración formando árboles o dendogramas.
- Las estructuras de árbol nos indican explícitamente la naturaleza de las relaciones entre objetos.
- Distinguimos dos tipo de métodos:
  - Aglomerativos:** Inicialmente cada objeto es un grupo y en cada iteración se juntan los grupos mas similares.
  - Divisivos:** Inicialmente se tiene un único grupo y en cada división se divide un grupo.

13

## Dendogramas

- Un dendograma es un árbol en el que el largo de las ramas está asociado inversamente a la fortaleza de la relación.



14

## Método Aglomerativo

- Esquema general algoritmo:
  - Cada objeto corresponde a un grupo.
  - Se juntan los dos grupos mas cercanos bajo algún criterio de cercanía entre grupos.
  - Los dos grupos recién unidos forman un único grupo.
  - Volver a 2 hasta formar un único grupo.
- El grupo de grupos puede definirse a posteriori.

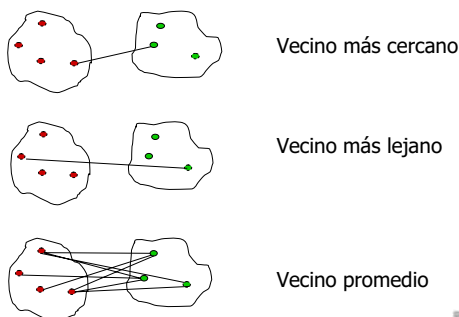
15

## Criterios de cercanías ente grupos

- Single linkage clustering (vecino más cercano):** Se juntan los grupos que verifican la mínima distancia entre objetos.
- Complete linkage clustering (vecino más lejano):** Se juntan los grupos que verifican la mínima distancia entre los objetos más distantes del grupo.
- Average linkage clustering (vecino promedio):** Se juntan los grupos que verifican la mínima distancia promedio entre grupos.
- Ward method (mínima perdida de información):** Los grupos se forman en base a los cambios en la suma de los errores cuadrados asociados con el agrupamiento de cualquier par de grupos.

16

## Criterios de cercanía



17

## Método de Ward

- En la iteración  $K$  tendremos  $K$  grupos cada uno de los cuales con  $n_k$  objetos.
- El objeto  $i$  del grupo  $k$  queda representado por un vector  $x_{ak}^i$ .
- En cada iteración ( $K$ ) buscamos la asignación que minimice la suma de los errores cuadrados ( $ESS_K$ )

$$ESS_k = \sum_{k=1}^K \left\{ \sum_{i=1}^{n_k} \sum_{a=1}^A \bar{x}_{ak}^i{}^2 - \frac{1}{n_k A} \left( \sum_{i=1}^{n_k} \bar{x}_{ak}^i \right)^2 \right\}$$

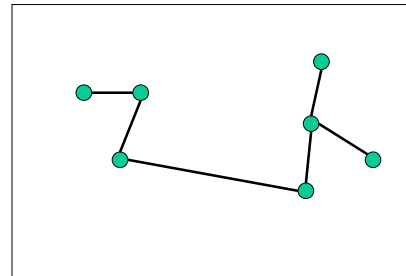
18

## Ejemplo método Ward

			x(AB)	x(AC)	x(AD)	x(AE)	x(BC)	x(BD)	x(BE)	x(CD)	x(CE)	x(DE)
Iteración 1	A	2										
	B	5	3.5	5.5	6	8.5	2	2	2	2	2	2
	C	9	3.5	5	5	5	7	7.5	10	5	5	5
	D	10	9	5.5	9	9	7	9	9	9.5	12	9
	E	15	10	10	6	10	10	7.5	10	9.5	10	12.5
			15	15	15	8.5	15	15	10	15	12	12.5
		ESS1	4.5	24.5	32	84.5	8	12.5	50	0.5	18	12.5
			x(CDA)	x(CDB)	x(CDE)	x(AB)	x(AE)	x(BE)				
Iteración 2	A	2										
	B	5	7	2	2	3.5	8.5	2				
	C	9	5	8	5	3.5	5	10				
	D	10	7	8	11.333	9.5	9.5	9.5				
	E	15	7	8	11.333	9.5	9.5	9.5				
		ESS2	38	14	20.667	5	85	50.5				

19

## Ejemplo gráfico



7 clusters  
6 clusters  
5 clusters  
4 clusters  
3 clusters  
2 clusters  
1 cluster

20

## Método Divisivo(1)

- En principio se puede proceder de manera similar que caso aglomerativo pero partiendo de un único conjunto y separando en vez de agrupar en cada iteración.
- Aparte de decidir que subgrupo separar debemos especificar con que criterio asignaremos cada objeto a cada subconjunto generado.

21

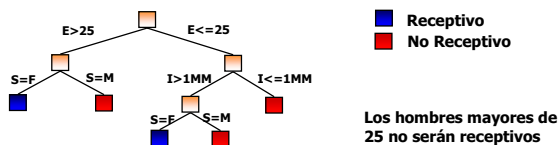
## Método Divisivo(2)

- Esquema general algoritmo:
  - Todos los objetos corresponde a un grupo.
  - Cada grupo se separa bajo algún criterio de maximización de varianza entre grupos.
  - Dividir cada uno de los grupos hasta que:
    - Todos los grupos sean tan homogéneos que no vale la pena seguir dividiendo.
    - Los grupos son tan pequeños que no vale la pena seguir dividiendo.
- El proceso genera junto con la estructura de árbol, **reglas explícitas** de clasificación para los objetos.

22

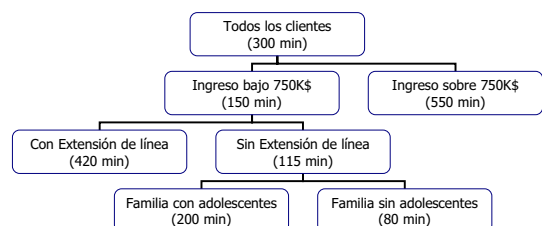
## Aplicación: árboles de clasificación

- Idea:** Dada una clasificación de objetos se busca determinar una estructura jerárquica de reglas que me permitan discriminar de la mejor manera posible entre las categorías.
- Ejemplo:** En las bases de datos de la empresa tengo una gran información de clientes y una categorización de si son Receptivos o no a la realización de una campaña de promoción.



23

## Esquemáticamente



24

## Criterios de división

- En general, se utiliza como método **supervisado**: tratamos de clasificar variables independientes de modo de discriminar mejor la variable dependiente.

- **Cart**: Cada nodo del árbol se divide para maximizar la *pureza* de los nodos hijos.
- **Discriminante**: Cada nodo del árbol se divide usando la mejor *función discriminante* lineal.

25

## Cart

- Sea  $i(N)$  la impureza del nodo  $N$ . En todos los casos queremos  $i(N)=0$  (todos los objetos con el mismo valor de la variable dependiente).
- Sea  $P(\omega_j)$  la fracción de los objetos que pertenecen a la categoría  $\omega_j$ .

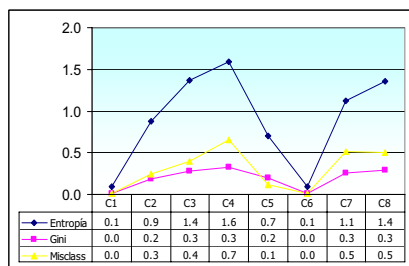
$$i(N) = -\sum_j P(\omega_j) \log_2 P(\omega_j) \quad \text{Entropía}$$

$$i(N) = \sum_{i \neq j} P(\omega_i) P(\omega_j) \quad \text{Gini}$$

$$i(N) = 1 - \max_j P(\omega_j) \quad \text{Missclasification}$$

26

## Medidas de impureza



27

## Discriminante\*

- Se plantea que existe una función discriminante lineal que permite separar adecuadamente cada una de las categorías.

$$D_j = \sum_k \alpha_{jk} x_{ik} \begin{cases} \leq d_j & \text{objeto } i \text{ en clase } j \\ > d_j & \text{objeto } i \text{ en clase } h \end{cases}$$

28

\* Volveremos sobre este punto al final del capítulo

## Métodos no Jerárquicos

- Algoritmos iterativos que en cada iteración ubican a los objetos en el grupo más cercano a él.
- El número de grupos está predefinido por lo que se recomienda probar para varios números de grupos.
- Al asignar todos los objetos en cada iteración, los algoritmos son mucho más rápidos para problemas con muchos objetos.
- Veremos dos tipos de métodos:
  - **K-Means**: Los objetos sólo pueden pertenecer a un único grupo.
  - **Fuzzy-C-Means**: Inicialmente se tiene un único grupo y en cada división se divide un grupo.

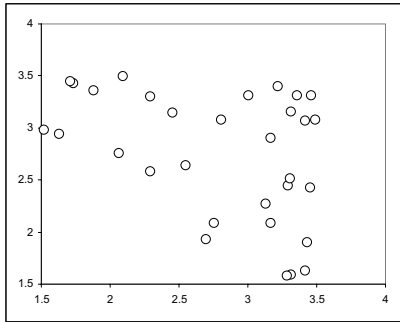
29

## K-means

- Entradas**
  - $X$  conjunto de  $N$  objetos
  - $K$  número de grupos
- Salidas**
  - $S_1, \dots, S_K$   $K$  grupos
  - $z_1, \dots, z_K$  Los centros de cada grupo
- Inicialización.**
  - $t = 0$
  - Elegir arbitrariamente  $z(t)$ .
- Asignación y actualización de centros.**
  - Asignar  $x^i$  al grupo más cercano para todo  $i=1 \dots N$ .
  - Recalcular  $z_j, j=1 \dots K$
  - $t=t+1$
- Criterio de parada.**
  - Si  $z_i(t) - z_i(t+1) < \epsilon$  para todo  $i$ , parar.

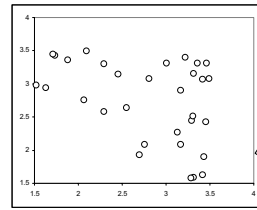
30

## Datos

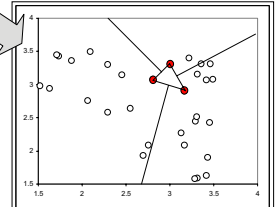


31

## Iteración 1

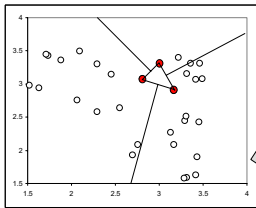


- Elegir los centros de los tres clusters aleatoriamente
- Localizar cada punto en su centro de cluster más cercano

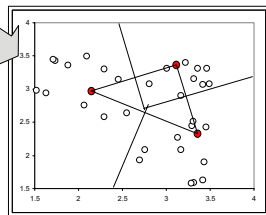


32

## Iteración 2

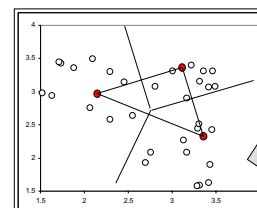


- Calcular nuevamente los centros de los clusters desde los centroides escogidos en la iteración 1.
- Localizar cada punto en el centroide que está más cerca a él.

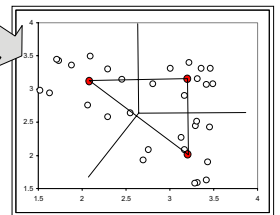


33

## Iteración 3

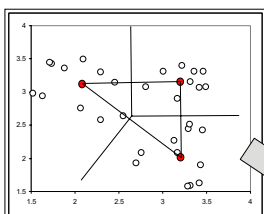


- Recalcular los centros como los centroides encontrados en la iteración 2.
- Localizar cada punto en el centroide que está más cerca a él.

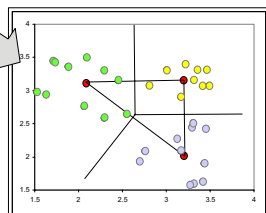


34

## Iteración 4 y etapa final



- recalcular los centros como los centroides de los clusters desde la iteración 3.
- Nada cambió!!
- Ok, está listo.



35

## Fuzzy-c-means

- Se define el grado de pertenencia del objeto  $i$  a la clase  $j$  ( $u_{ij}$ ) con  $j=1 \dots C$ . Se define además una medida vectorial de distancia  $d(x^i, x^j)$ .
- Inicialización.
  - $t = 0$
  - Elegir arbitrariamente  $u_{ij} \geq 0$  tal que  $\sum u_{ij} = 1$
- Asignación y actualización de centros.

$$z_j = \frac{\sum_{i=1}^N u_{ij} x^i}{\sum_{i=1}^N u_{ij}} \quad u_{ij} = \left\{ \frac{d(x^i, z_j)}{\sum_{h=1}^C \frac{d(x^i, z_h)}{d(x^i, z_h)^{2/(m-1)}}} \right\}^{-1}$$

- Criterio de parada.
  - Si  $z_j(t) - z_j(t+1) < \varepsilon$  para todo  $j$ , parar.
- $m$  es un parámetro de **difusividad** que va desde 1 (k-means) a  $\infty$  (completamente difusos).

36