
7. MÉTODOS MIXTOS DE CLASIFICACIÓN

7.1. Introducción

Los algoritmos de agregación en torno a «centros móviles» pueden operar sobre tablas de grandes dimensiones...

pero :

- ✧ producen particiones que dependen de los primeros centros de clases que hayan sido elegidos.
- ✧ sólo operan sobre un número de clases que debe ser definido *a priori*.

Los algoritmos de clasificación producen siempre el mismo resultado a partir de los mismos datos...

pero :

- ✧ se adaptan mal al tratamiento de tablas de grandes dimensiones.

Los **métodos de clasificación mixtos** combinan ambos procedimientos a fin de poder operar sobre tablas que pueden contener miles de unidades de observación.



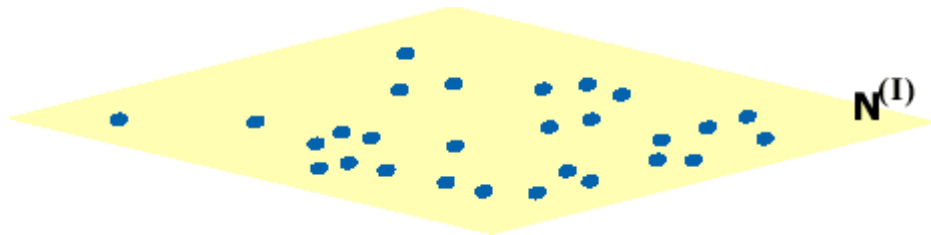
7.2. Funcionamiento de las clasificaciones mixtas

La clasificacion se realiza en cuatro etapas...

Primera etapa: Particion inicial de los n individuos de la tabla $T(n, p)$

Sea:

$n = 1000 \times x$: (x miles de individuos que componen la tabla $T(n, p)$)



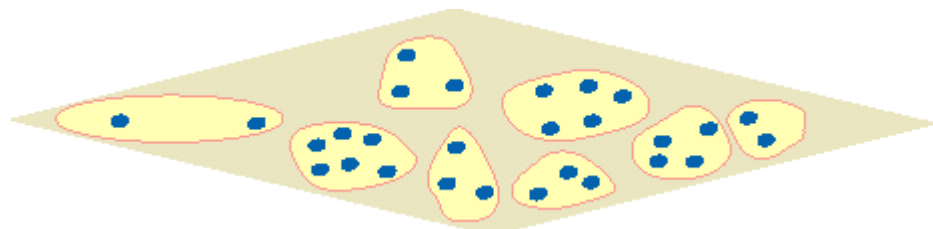
s : número de clases de la particion P_s que se desea construir a partir de la tabla $T(n, p)$

k : un número arbitrario de clases de una primera particion P_k de los n individuos de la tabla $T(n, p)$

$k = 10 \times y$: (del orden de algunas decenas), $k < n$
y $k > s$.

Procedimiento : Se emplea un algoritmo de agregación en torno a «centros móviles» para crear la partición P_k .

➡ Se determinan los «grupos estables» que resultan de la partición en k clases.



Resultado de la etapa : La partición de los n individuos en k clases homogéneas, de «grupos estables».

← **INDICE**

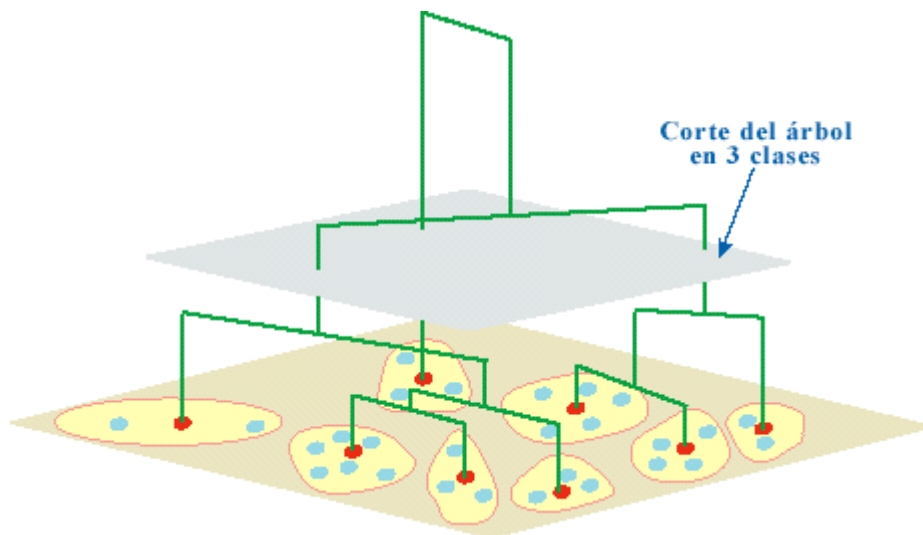


Segunda etapa : Clasificación jerárquica sobre los Centros de Gravedad de las k clases

Los elementos terminales del árbol de clasificación son las k clases definidas en la primera etapa.

- ➡ Se reagrupan las clases que hubieran sido fragmentadas por la partición realizada en la primera etapa.
- ➡ Se reúnen los elementos dispersos en torno a sus centros originales.
- ➡ Se elige - a partir del dendrograma - el número s de clases que permiten realizar una «buena» partición de los n individuos de la tabla $T(n, p)$.

Procedimiento : Se emplea el método de Ward.



Resultado de la etapa : Se establece el número s de clases de una «buena» partición.

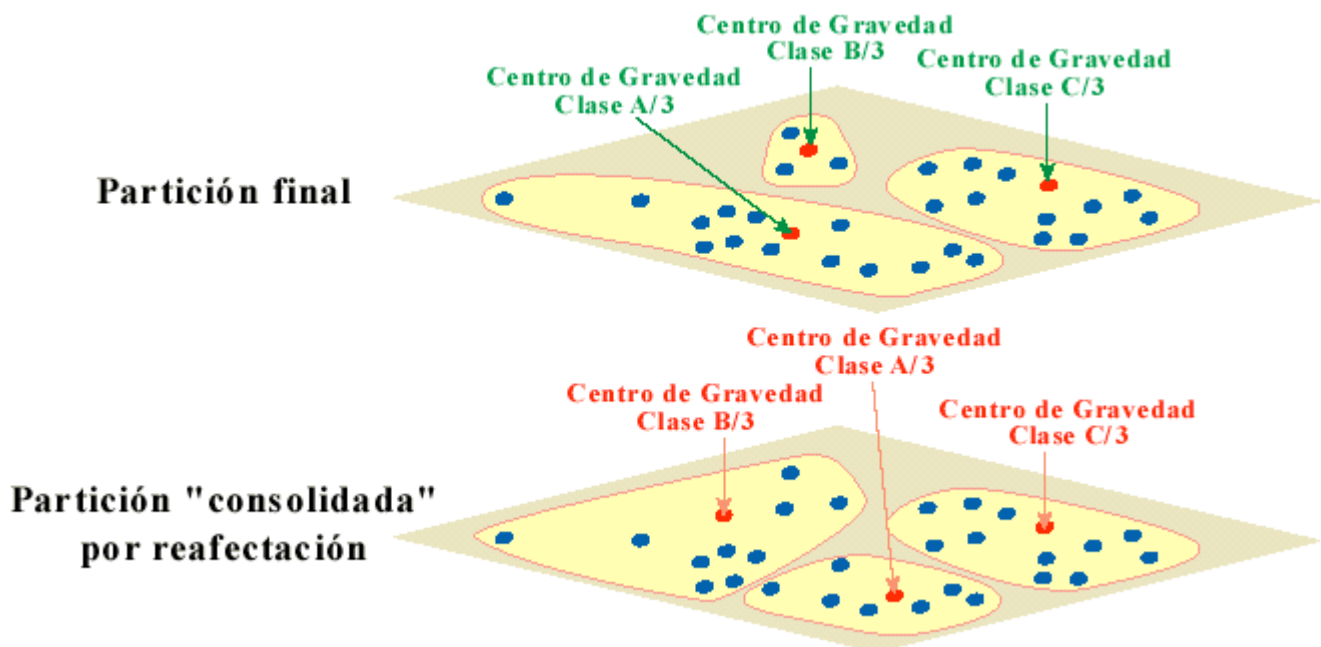
Tercera etapa : Construcción de la partición P_s

Los centros de clases de la partición inicial son determinados a partir del corte seleccionado en el árbol de clasificación.

En las iteraciones siguientes, los individuos son afectados en relación con el Centro de Gravedad más próximo, lo cual crea nuevas clases, a las cuales se les recalcula el Centro de Gravedad correspondiente y así siguiendo...

Procedimiento : Se emplea un algoritmo de agregación en torno a «centros móviles», optimizada por reafectación, para consolidar la partición P_s seleccionada originalmente sobre el dendrograma.

Resultado de la etapa : Se crea la partición P_s



Cuarta etapa : Descripción estadística de las clases

Se utilizan los procedimientos ya vistos de caracterización de las clases por medio del conjunto de variables (activas e ilustrativas) de la tabla $T(n, p)$.

