

4. ALGUNOS PROCEDIMIENTOS DE PARTICIÓN

4.1. Funcionamiento de los procedimientos de agregación en torno a «centros móviles»

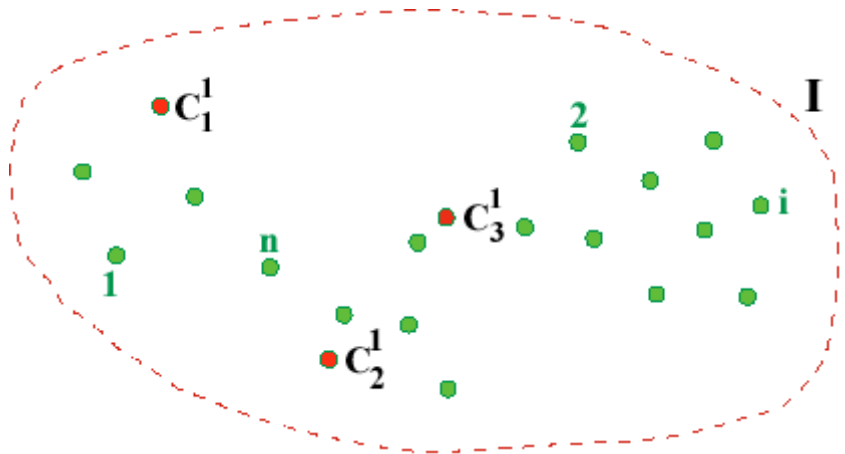
- ✓ Sea I un conjunto de n individuos caracterizados por p variables.
- ✓ La nube de puntos $N^I \in \mathbb{R}^p$. En \mathbb{R}^p está definida la distancia d apropiada.

Objetivo : Con los n individuos se desea «construir» q clases.

Primera etapa : Se eligen q «centros provisorios» de las q clases : $(C_1^1, C_2^1, \dots, C_q^1)$.

Ejemplo :

Para $q = 3$, los «centros provisorios» (C_1^1, C_2^1, C_3^1) fueron seleccionados al azar entre los n individuos observados.



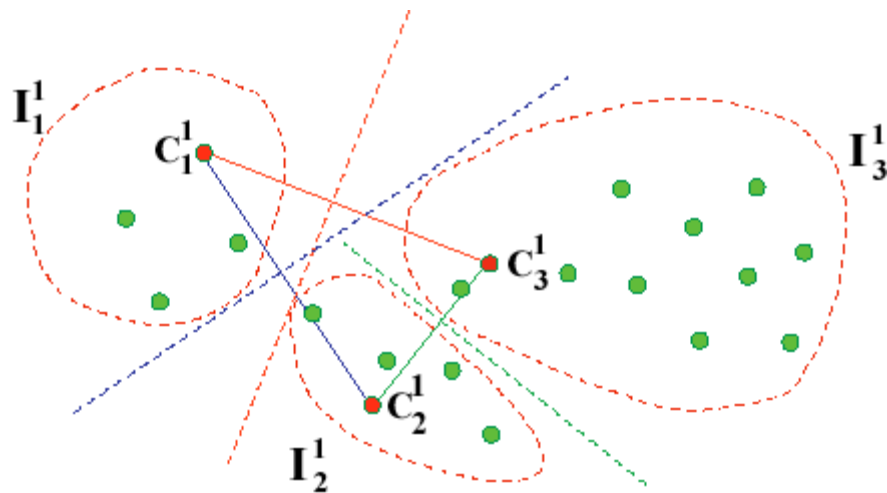
Los «centros provisorios» (C_1^1, C_2^1, C_3^1) inducen la primera partición P_1 de I en q clases : (I_1^1, I_2^1, I_3^1) .

El individuo $i \in I_k^1 \quad \forall k = 1, \dots, q$ si i está más próximo de

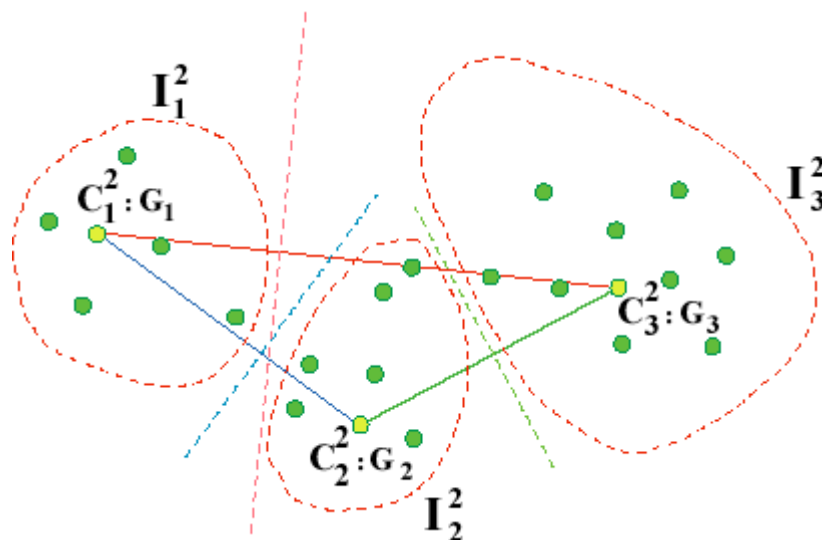
C_k^1 que de todo otro «centro provisorio».



Partición P_1 (I_1^1, I_2^1, I_3^1) inducida por
los «centros provisionarios» (C_1^1, C_2^1, C_3^1)



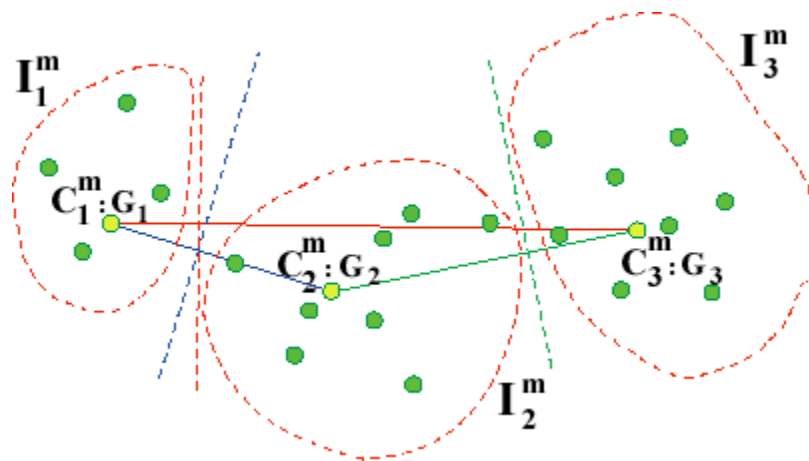
Segunda etapa : Se determinan q centros de las nuevas
clases ($C_1^2, C_2^2, \dots, C_q^2$), tomando los
«Centros de Gravedad» de las mismas.



Los nuevos centros de clases (C_1^2, C_2^2, C_3^2) inducen una nueva
partición P_2 de I en q clases : (I_1^2, I_2^2, I_3^2).



m-ésima etapa : Se determinan **q** nuevos centros $(C_1^m, C_2^m, \dots, C_q^m)$ de las clases resultantes de la etapa anterior $(I_1^{m-1}, I_2^{m-1}, \dots, I_q^{m-1})$, tomando los «Centros de Gravedad» de las mismas.



Los nuevos centros de clases (C_1^m, C_2^m, C_3^m) inducen una nueva partición P_m de I en q clases : (I_1^m, I_2^m, I_3^m) .

✓ El procedimiento se estabiliza necesariamente cuando dos etapas sucesivas conducen a la misma partición :

- sea porque la varianza intraclases queda constante,
- sea porque se realizó un número máximo de iteraciones fijado *a priori*.

✓ La partición obtenida depende de la elección inicial de los «centros provisionales».



4.2. Decrecimiento de la varianza intra-clase en cada iteración del procedimiento de agregación en torno a «centros móviles»

Los n individuos a clasificar poseen una masa p_i , siendo $\sum_i p_i = 1$.

Sea $d_{(i,C_k^m)}^2$ el cuadrado de la distancia al centro del individuo i de la clase k , en la etapa m .

Sea $V(m)$ la varianza intraclases en la etapa m .

$$V(m) = \sum_{k=1}^q \left(\sum_{i \in I_k^m} p_i d_{(i,C_k^m)}^2 \right)$$

Definiendo la cantidad «criterio» $v(m)$:

$$v(m) = \sum_{k=1}^q \left(\sum_{i \in I_k^m} p_i d_{(i,C_k^m)}^2 \right)$$

Se verifica que : $v(m) \geq V(m) \geq v(m+1)$

Por el teorema de Huygens...

$$v(m) = V(m) + \sum_{k=1}^q p_k d_{(C_k^m, C_k^{m+1})}^2 \Rightarrow v(m) \geq V(m).$$

Por otra parte, $v(m+1) = \sum_{k=1}^q \left(\sum_{i \in I_k^{m+1}} p_i d_{(i,C_k^{m+1})}^2 \right)$ y como I_k^{m+1} es el conjunto de puntos más próximos de C_k^{m+1} , las distancias de los i puntos con respecto a ese centro deben ser menores que con respecto a C_k^m . De modo que $\Rightarrow V(m) \geq v(m+1)$.

4.3. Algunas variantes de los procedimientos de partición por agregación en torno a «centros móviles»

a) Método de K-means

- ✓ Comienza con un sorteo pseudo-aleatorio de los «centros provisionales» iniciales.
- ✓ En cada afectación de un individuo en una clase se modifica la posición del Centro de Gravedad de la clase de afectación (**recentrado** de la clase).
- ✓ En una sola iteración, este método puede dar una partición de buena calidad. Pero la partición obtenida depende del orden de los individuos en la tabla $T(n, p)$.

b) Método de las «nubes dinámicas»

- ✓ Las clases no están caracterizadas por sus Centro de Gravedad sino por un cierto número de individuos a clasificar que constituyen los «**nódulos**» de las clases.
- ✓ El recentrado de las clases en torno a sus «nódulos» respectivos se hace cuando todos los individuos han sido afectados a una clase.

c) Método «Isodata»

- ✓ Se imponen condiciones suplementarias de afectación de los individuos en las clases, para impedir la formación de clases con efectivos muy bajos o de diámetro muy grande.

d) Método de los «individuos típicos»

INDICE

- ✓ En lugar de tomar como referencia el centro de gravedad de una clase (individuo promedio de la misma, no observado), se selecciona, en cada etapa, un «individuo típico» (entre los individuos a clasificar), el cual constituye el «centro» de cada clase en formación.



4.4. Ventajas e inconvenientes de los procedimientos de agregación en torno a «centros móviles»

😊 Ventajas

- ☞ Esos métodos tienden a maximizar los momentos de inercia interclases (clases bien diferenciadas).
- ☞ Simplicidad de cálculo :
 - ➡ Se trabaja con la **tabla $T(n, p)$** y una **tabla de centros de gravedad** o de nódulos de clases, cuyo tamaño depende del número de clases elegido.
 - ➡ Las distancias a calcular lo son solamente entre los **n** individuos y los **centros de gravedad**.

☹ Inconveniente

- ☞ Los resultados dependen de la elección de la partición inicial... **Los algoritmos convergen en «óptimos locales»**.
 - ➡ Solución «grupos estables» :
 1. Con un número dado de clases, se efectúa una partición inicial aleatoria seguida del algoritmo de partición.
 2. Se recomienza el proceso con otras afectaciones aleatorias seguidas de sendas aplicaciones del algoritmo de partición.La composición de las clases no será la misma al cabo de diferentes ensayos. Ciertos individuos serán afectados, en todos los ensayos, en la misma clase.
Estos nódulos de individuos «estables» constituyen las **«formas fuertes»** : grupos homogéneos cuya existencia no depende del número de clases elegido.

- ➡ Solución «métodos mixtos» :

Ver sección § 7. *Métodos mixtos de clasificación*

INDICE

