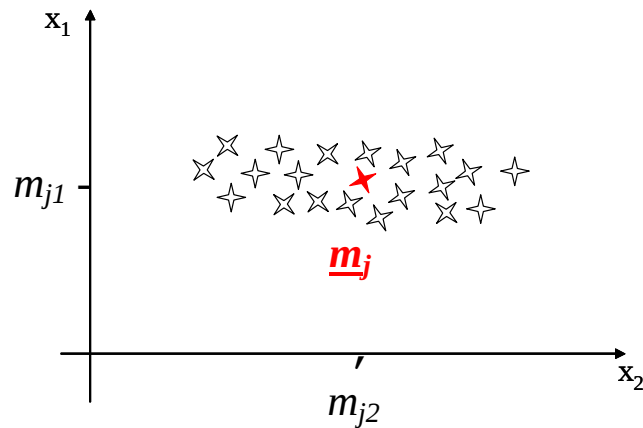


Características das classes

Centro, Baricentro ou Padrão da classe

$$\vec{m}_j = E_{\forall \vec{x}_i \in C_j} \vec{x}_i = \frac{1}{n_j} \sum_{\forall \vec{x}_i \in C_j} \vec{x}_i$$



Dispersão total intra-classe:

$$F_j = \sum_{\forall \vec{x} \in C_j} |\vec{x} - \vec{m}_j|^2 \geq 0 \quad F_{in} = \sum_{\forall C_j} F_j \geq 0$$

F_{in} - à minimizar

Agrupamento por similaridade

Agrupar, “de forma natural”,

conjuntos de dados com “similaridade interna”

minimizar F_{in} (para maximizar a similaridade interna)

utilizar um número “adequado” de classes

(para obter um agrupamento natural)

minimizar F_{in}

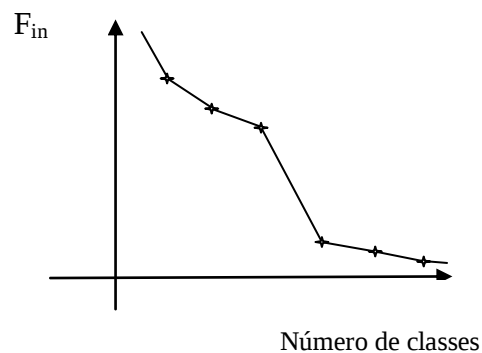
padrão das classes = baricentros $\underline{m}_j$

critério de pertinência

$$\vec{x}_i \in C_j \quad \text{se} \quad \left| \vec{x}_i - \vec{m}_j \right| < \left| \vec{x}_i - \vec{m}_k \right| \quad \forall k \neq j$$

número “adequado” de classes

variação brusca de F_i



Processos de clusterização:

K-means

Inicialização: Arbitre K centros

Loop

Reclassificação das entradas

Atualização dos centros de classe

Quando nenhum elemento trocar de classe: Fim.

Algoritmos Hierárquicos: dendogramas, árvores

Algoritmo Aglomerativo, Associativo:

**Junte classes com centros próximos
até F_{in} crescer repentinamente**

Algoritmo Divisivo, Dissociativo:

**Divida o cluster que reduz F_{in} o máximo possível
até F_{in} decrescer repentinamente**

Algoritmos Mistos

Iniciar com K-means

Associar / Dissociar buscando uma variação brusca de F_{in}