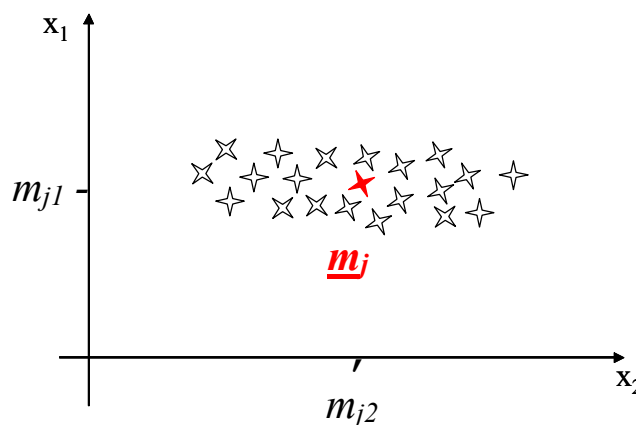


Agrupamentos - Resumo

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:

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

$$F_{in} = \sum_{\forall C_j} F_j \geq 0$$

F_{in} - à minimizar

Erro RMS de representação:

$$\sigma_j = \sqrt{\frac{F_j}{n_j}}$$

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}

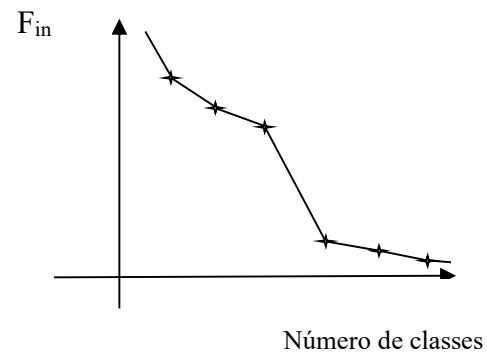
padrão das classes = baricentros $\underline{m}_j$

critério de pertinência

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

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

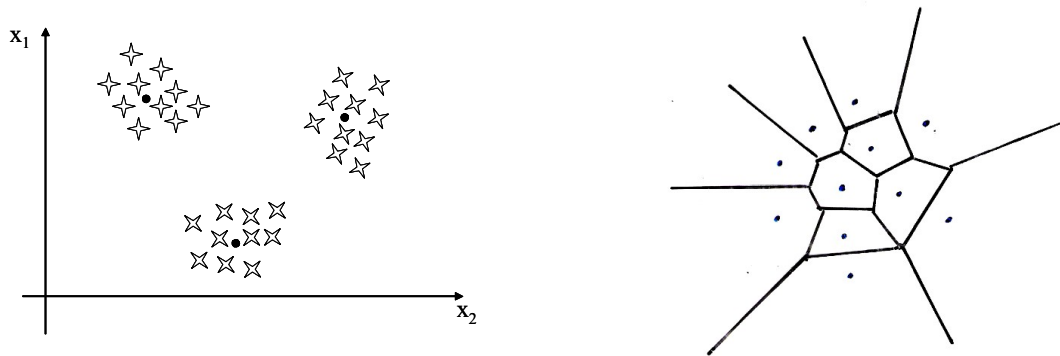
Iniciar com K-means

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

Classificação por Similaridade

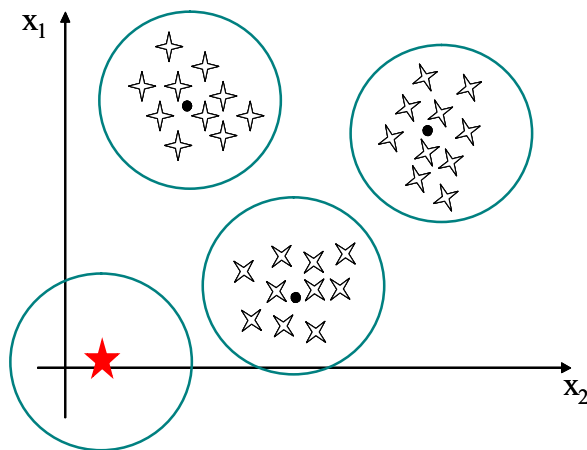
Critérios de pertinência

1 - Maior Similaridade - Padrão mais similar à entrada



$$\underline{x}_i \in C_i \quad \text{sse} \quad |\underline{x} - \underline{w}_i|^2 < |\underline{x} - \underline{w}_j|^2 \quad \forall j \neq i$$

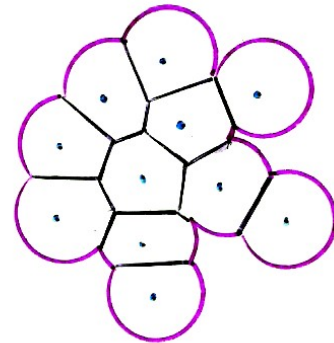
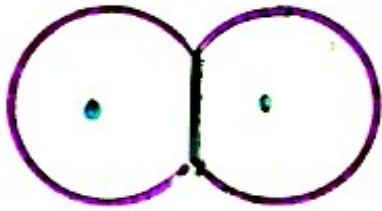
2 - Similaridade Mínima - Padrão que satisfaz uma similaridade mínima



$$\underline{x}_i \in C_i \quad \text{sse} \quad |\underline{x} - \underline{w}_i|^2 < r_0^2$$

r_0 – raio de similaridade

3 - Padrão mais similar & mínima similaridade atingida



$$\underline{x}_i \in C_i \quad \text{se} \quad \left| \underline{x} - \underline{w}_i \right|^2 < \left| \underline{x} - \underline{w}_j \right|^2 \quad \forall j \neq i$$

$$\text{e se} \quad \left| \underline{x} - \underline{w}_i \right|^2 < r_0^2$$