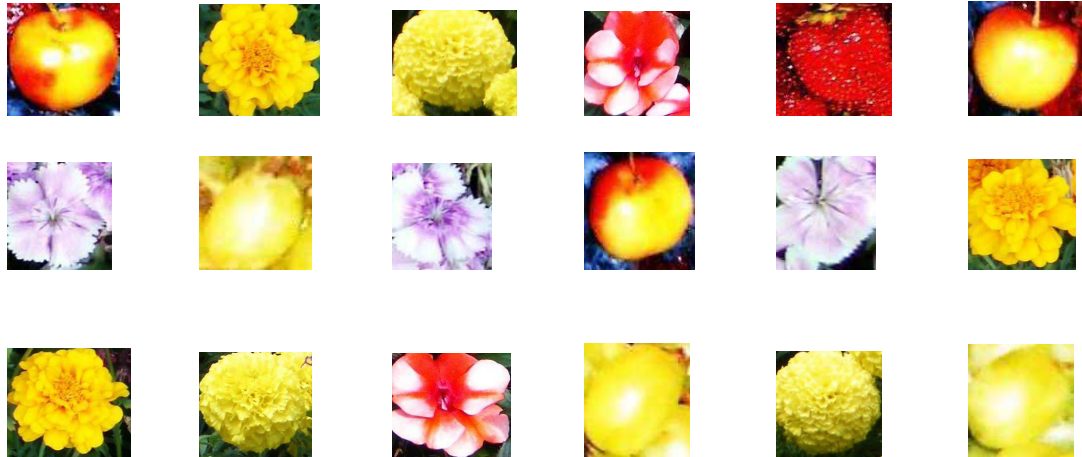


Classificação por Similaridade

Classes agrupam elementos “similares” (**mas não iguais**) entre si



Classificação por Similaridade –

Critérios de pertinência à uma classe

objetos físicos $\mathcal{O}_1 \mathcal{O}_2 \rightarrow$ objetos matemáticos $\underline{x}_1$

$\underline{x}_2$

similaridade física $\rightarrow$ similaridade matemática

$$\mathcal{O}_1 \approx \mathcal{O}_2$$

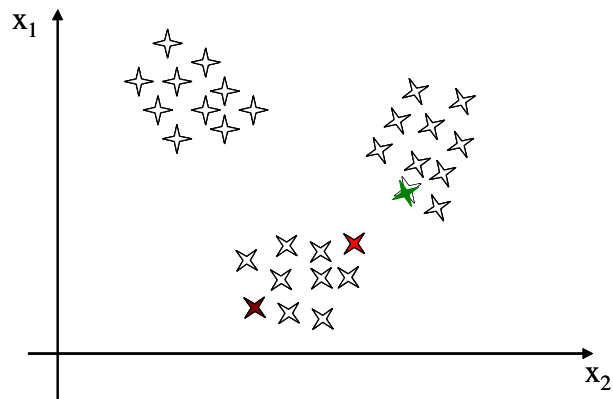
$$\underline{x}_1 \cong \underline{x}_2 \quad \text{ou}$$

$$|\underline{x}_1 - \underline{x}_2| \ll$$

Classificadores por similaridade

Critérios de Pertinência à uma Classe

Critério básico



Dois elementos pertencem à mesma classe se estão próximos entre si

Critério: k vizinhos mais próximos.

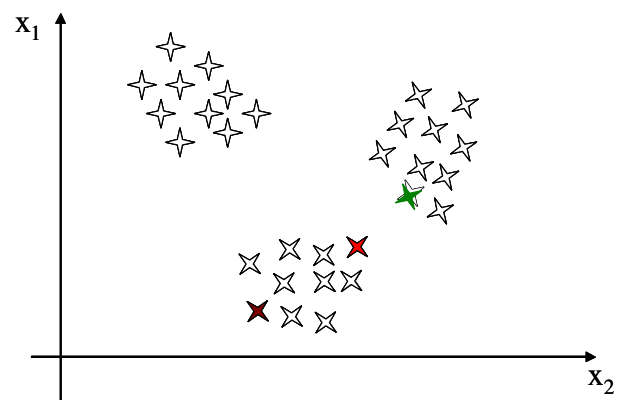
Considere os k elementos mais próximos da entrada que se pretende classificar. A entrada pertence à classe à qual pertencem a maioria dos k vizinhos.

Critério simplificado: vizinho mais próximo.

Se $k=1$ uma entrada pertence a uma classe se seu vizinho mais próximo é um elemento desta classe

pouco prático

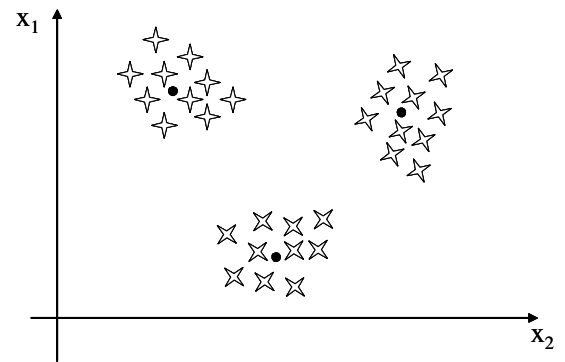
a descrição de uma classe exige a utilização de todos seus elementos (domínio da classe)



Critérios baseados no Padrão de classe

Padrão $\underline{w}_i$ da classe C_i

$$\underline{w}_i = E_{\forall \underline{x}_j \in C_i} \underline{x}_j = \frac{1}{N_i} \sum_{\forall \underline{x}_j \in C_i, j=1}^{N_i} \underline{x}_j$$



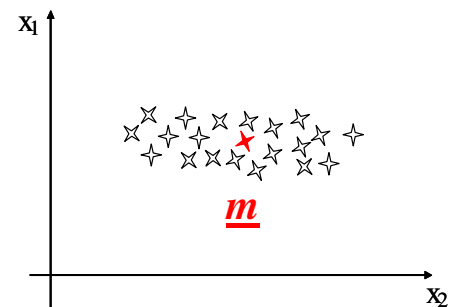
a **média** ou **baricentro** é escolhido como padrão da classe porque
minimiza a **dissimilaridade**
e o **erro de representação**
da classe pelo seu padrão

Dispersão média intra classe ou erro (médio quadrático) de representação

da classe C_j para o seu padrão $\vec{p}_j$

$$\sigma_j^2 = E_{\forall \vec{x} \in C_j} |\vec{x}_i - \vec{m}_j|^2 = \frac{1}{n_j} \sum_{\forall \vec{x} \in C_j} |\vec{x}_i - \vec{m}_j|^2$$

por componente k $\sigma_{jk}^2 = \frac{1}{n_j} \sum_{\forall \vec{x} \in C_j} (x_{jk} - p_{jk})^2$

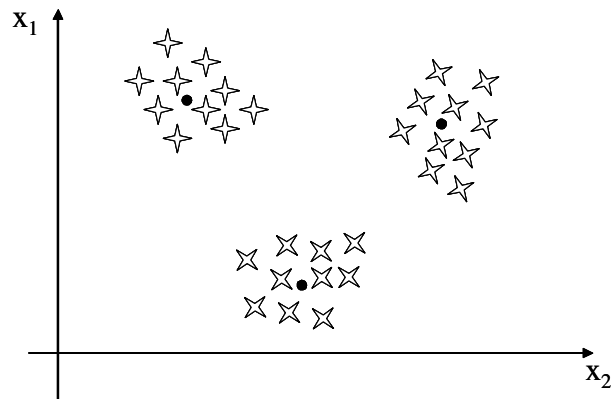


Como visto anteriormente $\frac{\partial \sigma_{jk}^2}{\partial p_{jk}} = 0 \Rightarrow p_{jk} = m_{jk} \Rightarrow \boxed{\vec{p}_j = \vec{m}_j}$

o padrão que minimiza a dispersão intra classe - e o erro de representação da classe pelo seu padrão - é o seu baricentro $\vec{m}_j$

1 - Critério de pertinência 1 : Maior Similaridade

Padrão mais similar à entrada



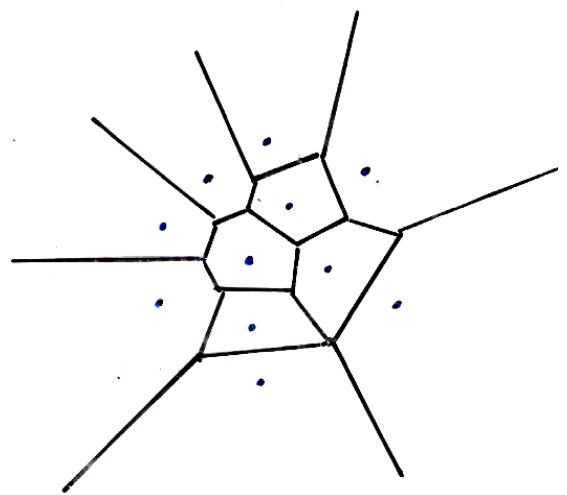
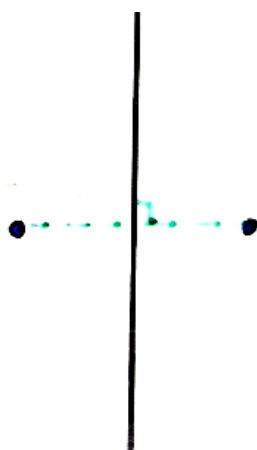
$$\underline{x}_i \in C_i \quad \text{sse} \quad |\underline{x} - \underline{w}_i|^2 < |\underline{x} - \underline{w}_j|^2 \quad \forall j \neq i$$

eq (1)

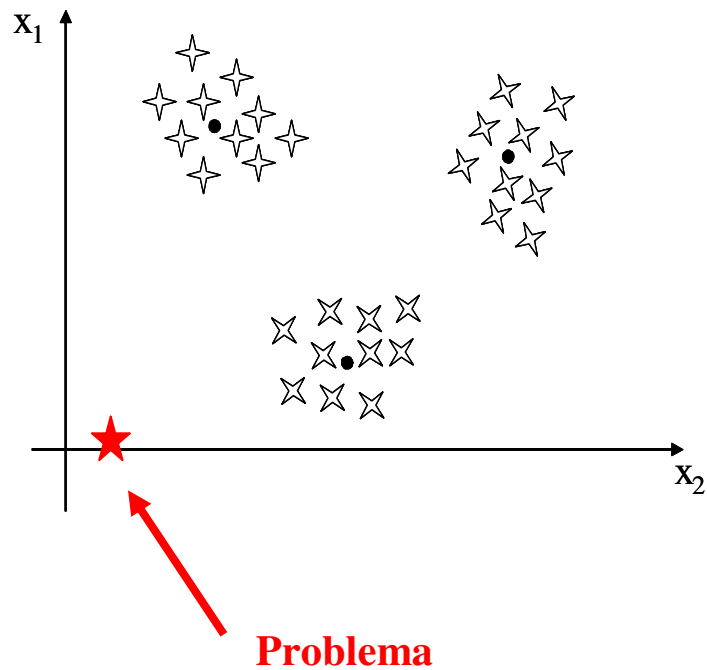
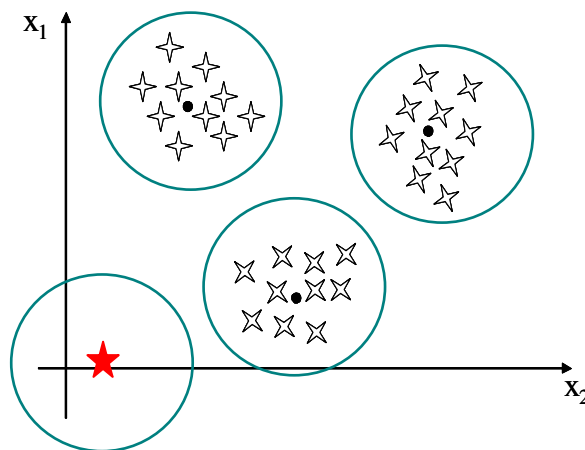
porque minimiza a dispersão intra-classes total

Separadores para o critério 1

- hiperplanos



Tecelagem (diagrama) de Voronoi

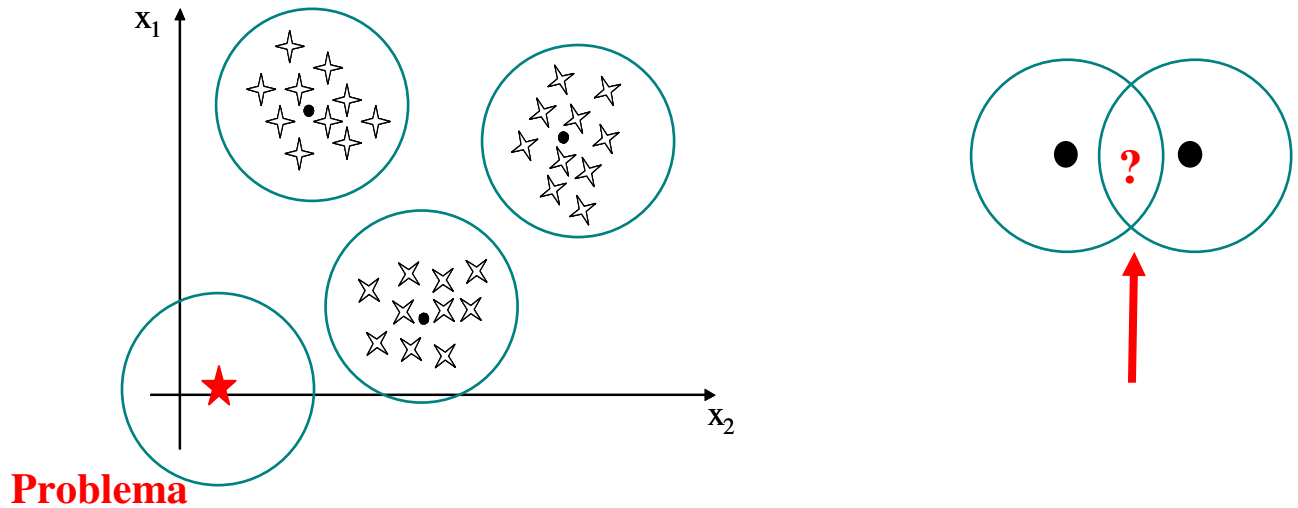
Problema:**2 - Critério de pertinência 2: Similaridade Mínima****Padrão que satisfaz uma similaridade mínima**

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

 r_0 – raio de similaridade

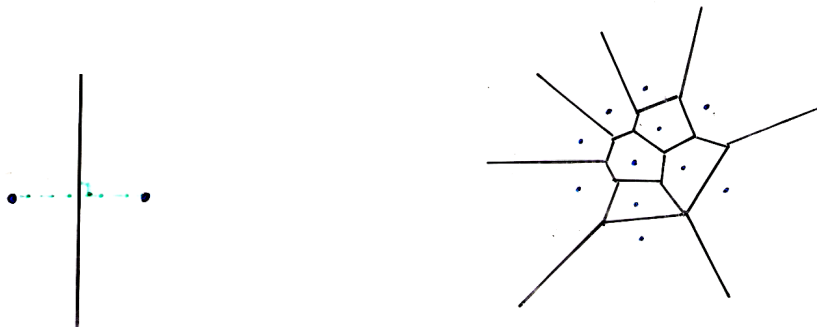
Separadores para o critério 2

- hiperesferas



Separadores

Critério 1 - padrão mais similar (mais próximo) a entrada Eq(1)



Critério 2 - mínima similaridade exigida Eq (2)



3 - Critérios 1 e 2, $Eq(1) + Eq(2)$

padrão mais similar & mínima similaridade atingida

