

## CPE 722 –

# Redes Neurais Não Supervisionadas e Agrupamentos

## Introdução aos Agrupamentos

**Agrupamentos**

**Clusterização**

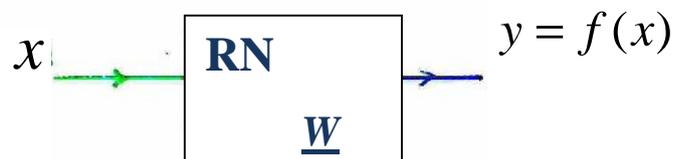
**Classificação**

## 1 - Máquinas

**Máquina, sistema**

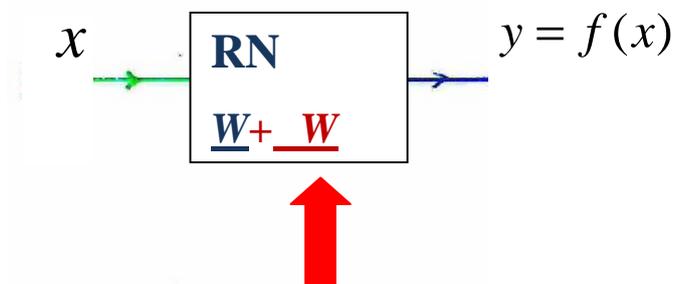
**ação** ➡ **reação**

**entrada** ➡ **saída**

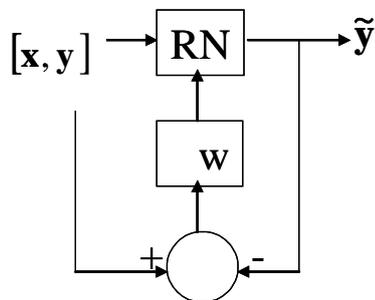


## 2 - Treinamento da máquina

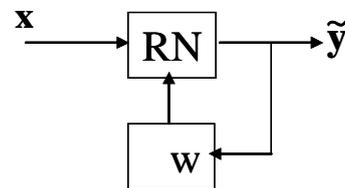
entrada → saída desejada



### - Tipo de treinamento supervisionado / não-supervisionado



Pode realizar mapeamentos arbitrários

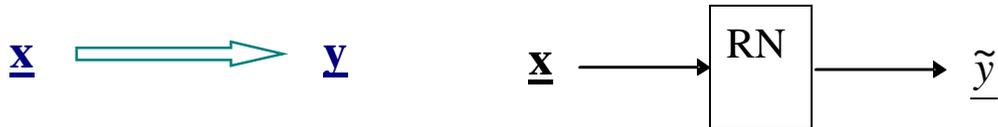


Necessita algum critério  
**Similaridade**



## 2 - Classificadores - Noções iniciais

### Máquinas matemáticas - Mapeadores



### Mapeadores lineares / não lineares

$$\underline{\tilde{y}} = \underline{M} \underline{x}$$

$$\underline{\tilde{y}} = \varphi(\underline{x})$$

### Tipos de mapeadores não lineares

#### Aproximadores

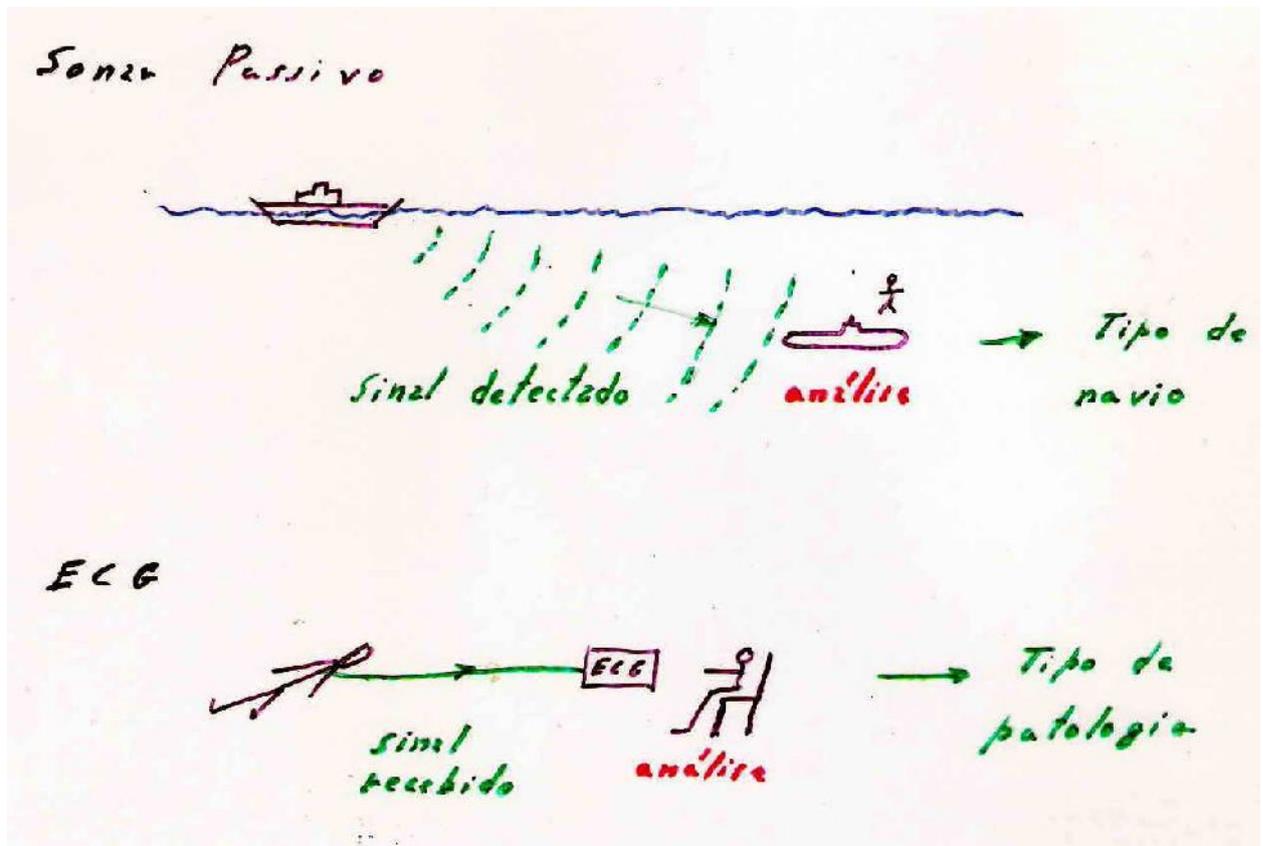
Associam saídas contínuas às entradas

Saídas contínuas  $y_i \in (-1, +1)$

#### Classificadores

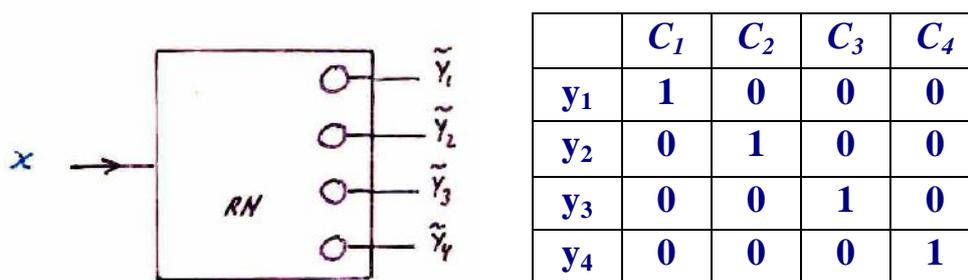
Associam saídas lógicas às entradas (classificam as entradas)

Saídas lógicas  $y_i \in \{0, 1\}$



## 2.1 Convenção / Estratégia:

**Codificação da Saída** (maximamente esparsa)



Cada saída separa a sua classe das demais, i.e.

classe versus não classe (um contra todos)

Separação de N classes



Separação de duas classes:

classe vs. não classe

### 3 - Tipos de classificação / Tipos de aprendizado

#### Classificadores:

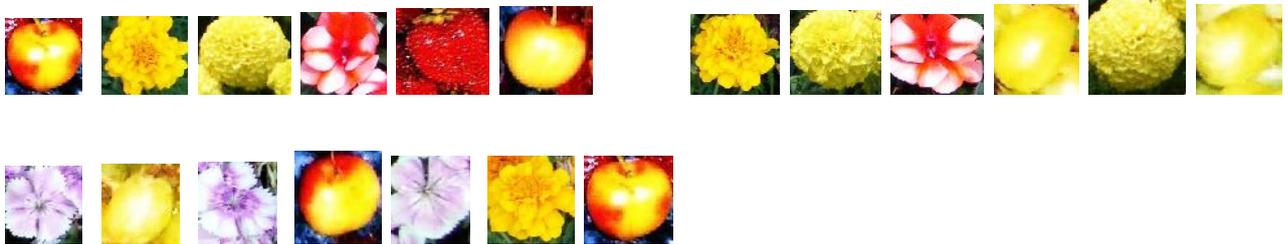


#### Critérios de classificação

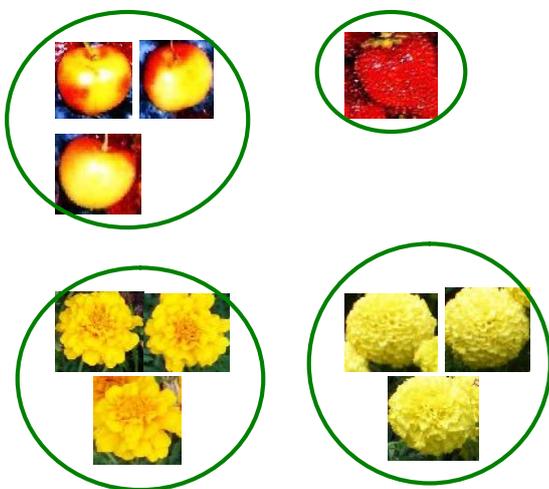
Critérios “não conhecidos”, arbitrários

Critérios conhecidos, determinados

#### Classificadores por similaridade



#### Agrupamentos por similaridade



#### Agrupamentos arbitrários



## 4 - Classificação, Agrupamento ou Clusterização por Similaridade

### Agrupar, “de forma natural”, conjuntos de dados com “similaridade interna”

Classes agrupam elementos ‘similares’ entre si ‘de forma natural’ mas

**O que significa “de forma natural ?”**

**O que significa “similaridade interna ?”**

**Quais são as classes ?**

### Classificadores supervisionados

As classes são informadas

**Ação: apenas classifica as entradas**

### Classificadores **não** supervisionados

As classes **não** são informadas

**Ações: determina as classes e**

**classifica as entradas**

**Sobrevivência vs.****Classificação por similaridade****Aprendizado não supervisionado****Plasticidade****Animais e vegetais venenosos ou comestíveis ?****4.1 – Dados reais vs. Representação matemática na****Classificação por Similaridade**

objetos físicos  $O_1$   $O_2$   $\longrightarrow$  objetos matemáticos  
vetores  $\underline{x}_1$   $\underline{x}_2$

similaridade física  $\longrightarrow$  similaridade matemática

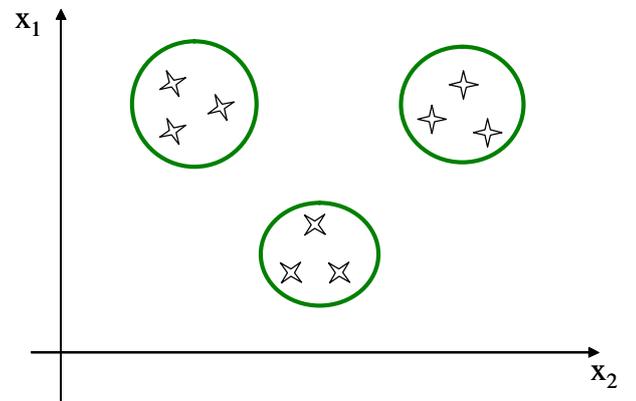
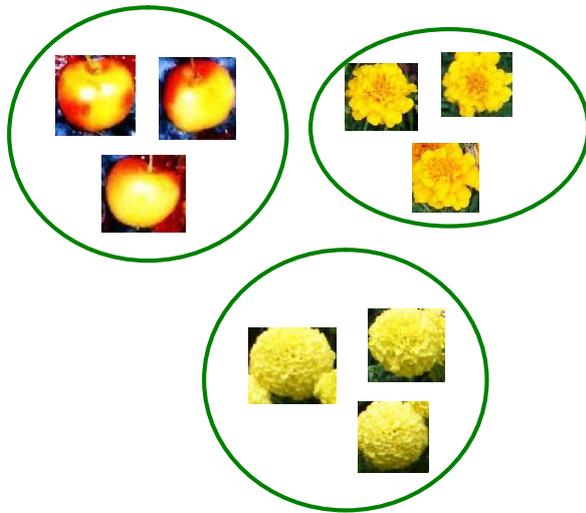
$$O_1 \approx O_2$$

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

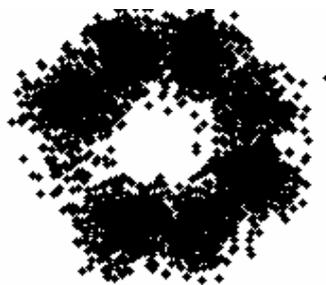
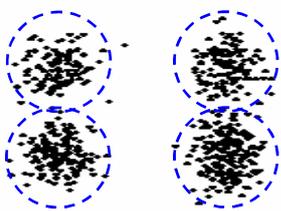
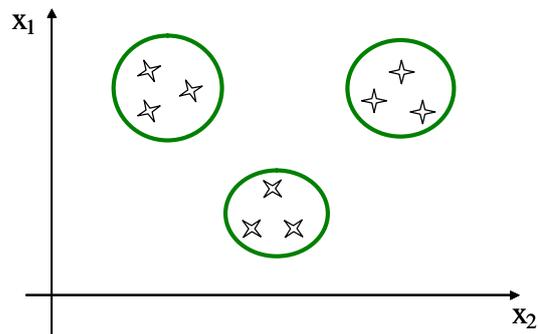
$$O_1 \approx O_2$$



$$\underline{x}_1 \cong \underline{x}_2$$



**Casos simples / complexos**



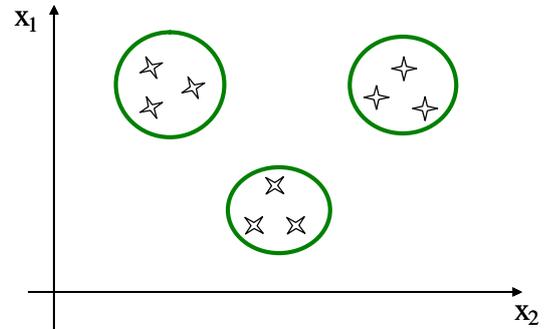
**Classificar por similaridade : alocar na mesma classe  
vetores que estão relativamente próximos entre si.**

**Como traduzir matematicamente**

**“relativamente próximos entre si ?“**

**Como determinar as classes ?**

**Como realizar a alocação ?**



## 5 - Características de um Agrupamento

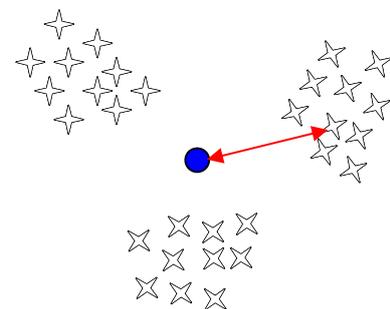
### 5.1 - Características do conjunto de todos os elementos

- Número total de elementos,  $n_0$
- Média, baricentro global

$$\vec{m}_0 = E_{\forall i} \vec{x}_i = \frac{1}{N} \sum_{i=1}^N \vec{x}_i$$

- Dispersão, dissimilaridade total

$$F_0 = \sum |\vec{x}_i - \vec{m}_0|^2$$



## Desvio Padrão Total

$$\sigma_0 = \sqrt{\frac{F_0}{n_0}} = \sqrt{\frac{1}{n_0} \sum |\vec{x}_i - \vec{m}_0|^2}$$

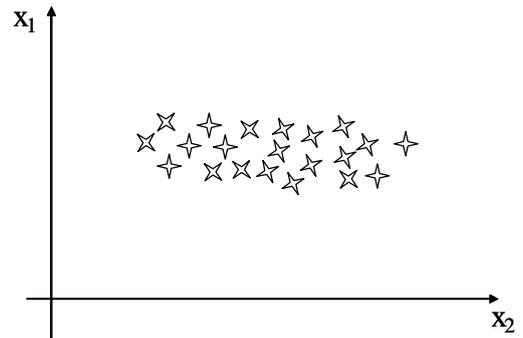
**$\underline{m}_0$ ,  $\sigma_0$  e  $F_0$  são independentes da clusterização**

## 5.2 Características de uma classe $C_j$

### Características intra classe

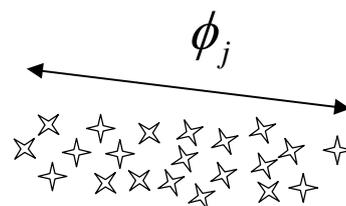
#### Classe $C_j$

- Número de elementos da classe  $n_j$



- Diâmetro da classe:

$$\phi_j = \underset{\forall \vec{x}_i, \vec{x}_k \in C_j}{\text{Max}} (|\vec{x}_i - \vec{x}_k|)$$



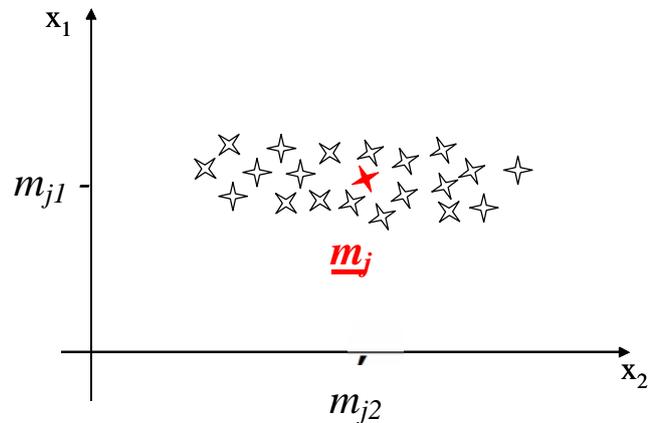
- **Baricentro (ou Média) da classe**

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

$$\vec{m}_j = [m_{j1} \dots m_{jk} \dots]^t$$

por componente k

$$m_{jk} = \mathbf{E}_{\forall \vec{x}_i \in C_j} x_{ik} = \frac{1}{n_j} \sum_{\forall \vec{x}_i \in C_j} x_{ik}$$

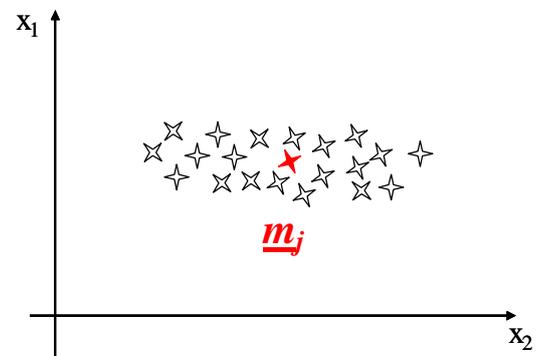


- **Variância intra classe,**

**Erro (médio quadrático) de representação da classe**

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

por componente k



$$\sigma_{jk}^2 = \mathbf{E}_{\forall \vec{x}_i \in C_j} (x_{ik} - m_{jk})^2 = \frac{1}{n_j} \sum_{\forall \vec{x}_i \in C_j} (x_{ik} - m_{jk})^2$$

$$\mathbf{e} \quad \sigma_j^2 = \sum_{\forall k} \sigma_{jk}^2$$

- **Erro RMS - Desvio Padrão da classe**

$$\sigma_j = \sqrt{\sigma_j^2} = \left[ \frac{1}{n_j} \sum_{\forall \vec{x} \in C_j} |\vec{x} - \vec{m}_j|^2 \right]^{1/2}$$

- **Padrão da Classe**

**Critério: representa os pontos da classe com o menor erro, minimiza a dispersão intra classe (variância) da classe j:**

$$\sigma_j^2 = \sum_{\forall k} \sigma_{jk}^2 \quad \sigma_{jk}^2 = \frac{1}{n_j} \sum_{\forall \bar{x} \in C_j} (x_{jk} - p_{jk})^2 = \mathbf{E}_{\forall \bar{x} \in C_j} (x_{jk} - p_{jk})^2$$

$$\frac{\partial \sigma_{jk}^2}{\partial p_{jk}} = 0$$

$$\frac{\partial \sigma_{jk}^2}{\partial p_{jk}} = \frac{\partial}{\partial p_{jk}} \frac{1}{n_j} \sum_{\forall \bar{x} \in C_j} (x_{jk} - p_{jk})^2 = \frac{\partial}{\partial p_{jk}} \frac{1}{n_j} \sum_{\forall \bar{x} \in C_j} (x_{jk}^2 - 2p_{jk}x_{jk} + p_{jk}^2)$$

$$= \frac{\partial}{\partial p_{jk}} \left[ \frac{1}{n_j} \sum_{\forall \bar{x} \in C_j} x_{jk}^2 - 2p_{jk} \frac{1}{n_j} \sum_{\forall \bar{x} \in C_j} x_{jk} + p_{jk}^2 \right] =$$

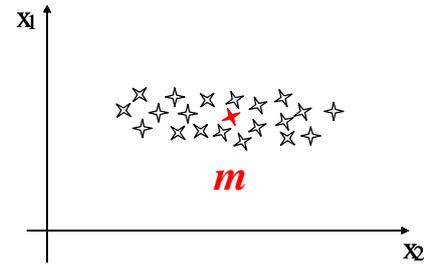
$$= -2m_{jk} + 2p_{jk} = 0 \quad \Rightarrow \quad p_{jk} = m_{jk} \quad \Rightarrow$$

$$\boxed{\vec{p}_j = \vec{m}_j}$$

**o padrão que minimiza a dispersão intra classe (ou erro de representação) de uma classe é o seu baricentro**

- **Dispersão total intra classe da classe  $C_j$ :**

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

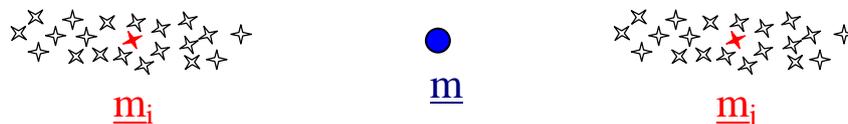


**Dispersão total intra classe para todas as classes:**

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

**$F_{in}$  - parâmetro à minimizar**

### 5.3 – Características (medidas) Inter Classes



- **Número de classes  $M$**
- **Dissimilaridade total inter classes**

$$F_{out} = \sum_{\forall j} n_j |\vec{m}_j - \vec{m}|^2 \geq 0$$

**$F_{out}$  - parâmetro à maximizar**

**Para um bom agrupamento  
escolher as classes de forma a**

**Minimizar a dispersão intra classe total  $F_{in}$**

**Maximizar a dissimilaridade inter classes total  $F_{out}$**

**Com alguma álgebra é possível mostrar que**

$$F_{in} + F_{out} = F_0 = \text{constante, independente da clusterização}$$

**Logo a clusterização que minimiza  $F_{in}$   
ao mesmo tempo maximiza  $F_{out}$ , i.e.,**

**basta que o processo de clusterização minimize  $F_{in}$**

**Obs 1:**

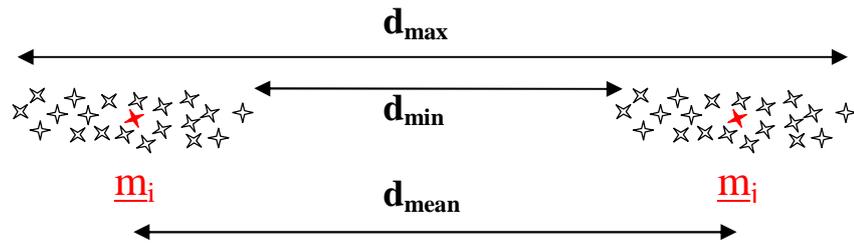
**Critério: usamos como medida inter classes a maximizar a**

**Dissimilaridade total inter classes**

$$F_{out} = \sum_{\forall j} n_j |\vec{m}_j - \vec{m}|^2 \geq 0$$

**mas existem outras medidas de dissimilaridade (distância, separação) inter classes:**

## Outras medidas de dissimilaridade (distância, separação) inter classes:



vizinho + próximo

$$d_{\min}(\mathcal{X}_i, \mathcal{X}_j) = \min_{\mathbf{x} \in \mathcal{X}_i, \mathbf{x}' \in \mathcal{X}_j} \|\mathbf{x} - \mathbf{x}'\|$$

vizinho + distante

$$d_{\max}(\mathcal{X}_i, \mathcal{X}_j) = \max_{\mathbf{x} \in \mathcal{X}_i, \mathbf{x}' \in \mathcal{X}_j} \|\mathbf{x} - \mathbf{x}'\|$$

distância média

$$d_{\text{avg}}(\mathcal{X}_i, \mathcal{X}_j) = \frac{1}{n_i n_j} \sum_{\mathbf{x} \in \mathcal{X}_i} \sum_{\mathbf{x}' \in \mathcal{X}_j} \|\mathbf{x} - \mathbf{x}'\|$$

distância entre médias

$$d_{\text{mean}}(\mathcal{X}_i, \mathcal{X}_j) = \|\mathbf{m}_i - \mathbf{m}_j\|.$$

(assim como existem também outras medidas de dispersão intra classe)

### Em resumo:

Para um bom agrupamento

escolher as classes de forma a

Minimizar a dispersão intra classe total  $\mathbf{F}_{\text{in}}$