UNIVERSIDADE ESTADUAL PAULISTA
"JÚLIO DE MESQUITA FILHO"
Campus de São José dos Campos
Instituto de Ciência e Tecnologia

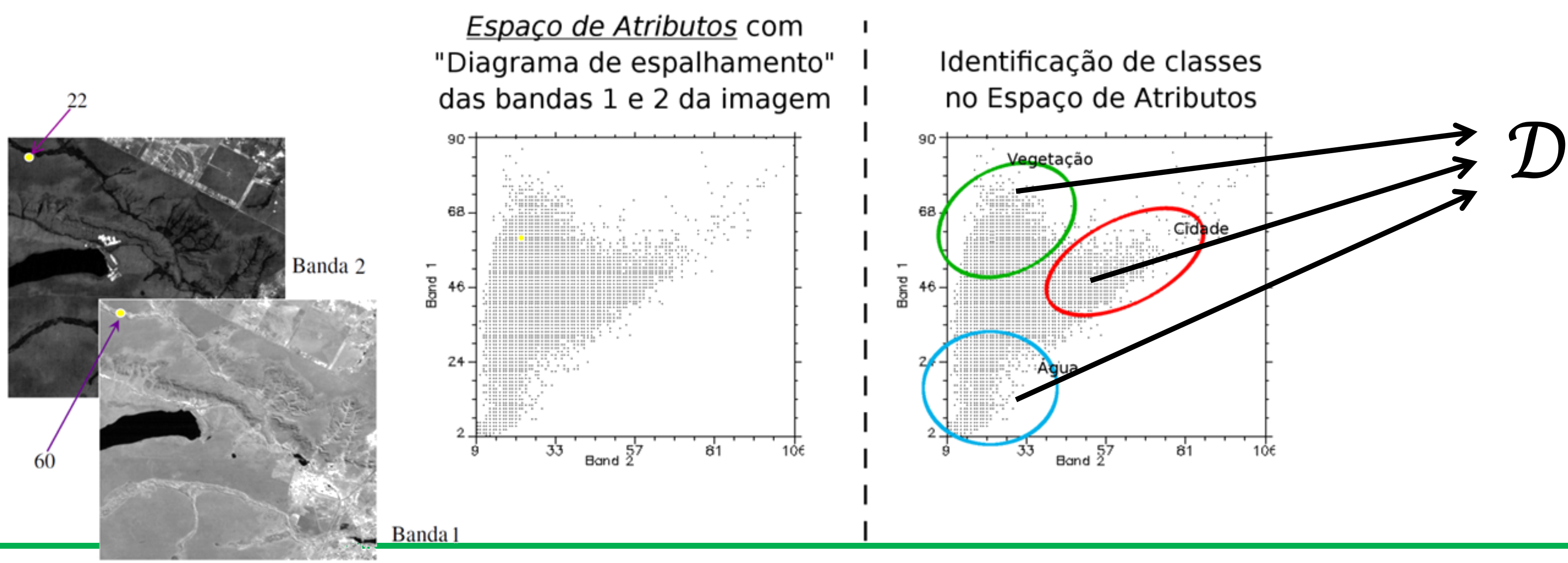
# Classificação Supervisionada

Prof. Dr. Rogério Galante Negri

# Recordando definições

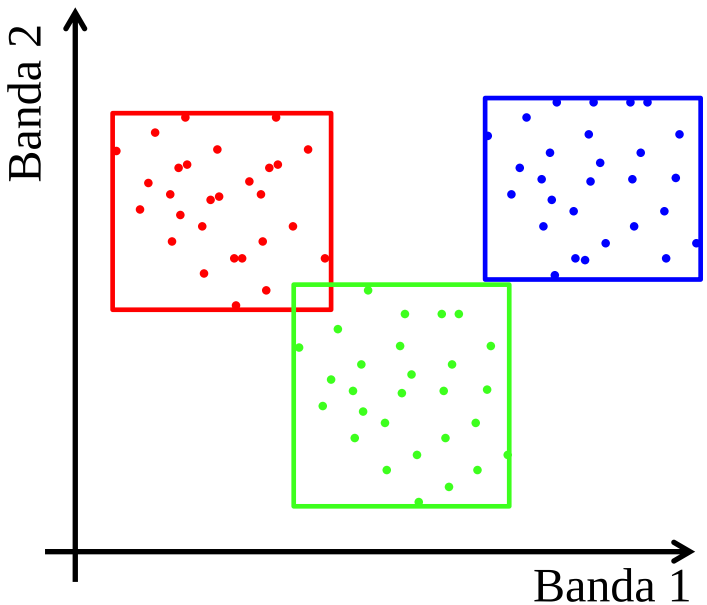
- $\mathcal{I}$ uma imagem definida sobre um reticulado $\mathcal{S} \in \mathbb{N}^2$
- $\mathcal{I}(s) = \mathbf{x}$ denota que o pixel $s \in \mathcal{S}$ possui vetor de atributos $\mathbf{x} \in \mathcal{X}$
- $F: \mathcal{X} \rightarrow \mathcal{Y}$ é um <u>classificador</u> que associa elementos de $\mathcal{X}$ a um indicador de classe em $\mathcal{Y} = \{1, 2, \dots c\}$
- Classificadores de aprendizado <u>supervisionado</u> estimam $F$ via um conjunto de treinamento $\mathcal{D} = \{(\mathbf{x}_i, y_i) \in \mathcal{X} \times \mathcal{Y} : i = 1, \dots, m\}$
- Seja $\Omega = \{\omega_1, \omega_2, \dots, \omega_c\}$ o conjunto de classes, o valor do indicador de classe $y_i$ determina a classe ao qual $\mathbf{x}_i$ é associado:

$$\text{Se } y_i = j, \text{ então } \mathbf{x}_i \text{ está associado a } \omega_j$$

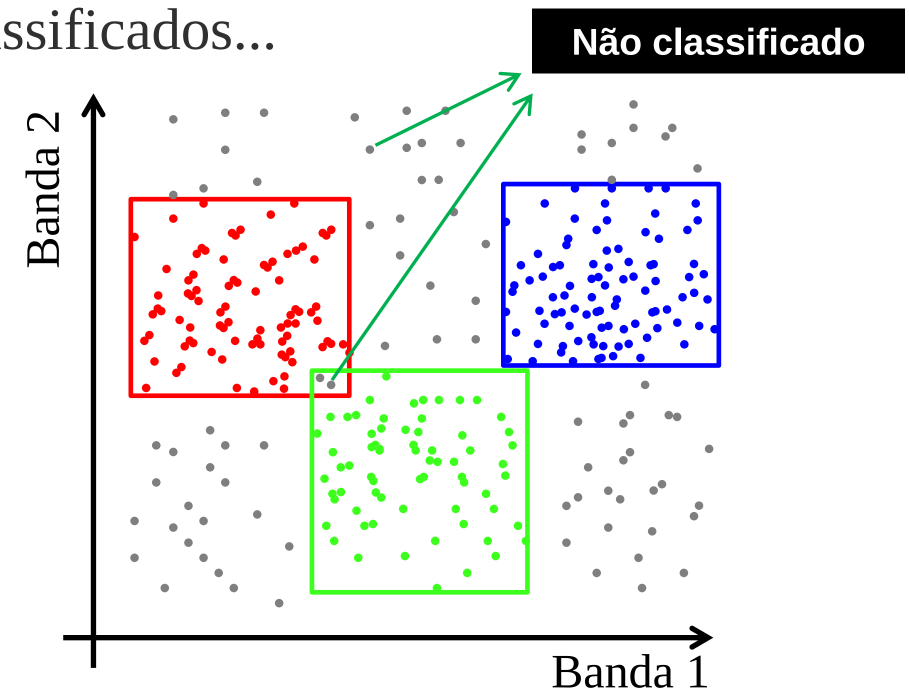


# Paralelepípedo

- Uma das mais simples técnicas de classificação
- Consiste em definir “paralelepípedos”
  - Os limites inferior/superior de cada classe é estimado via $\mathcal{D}$
  - O paralelepípedo possui tantas dimensões quantas bandas na imagem
  - Vários padrões (pixels) não são classificados...




Definição dos paralelepípedos a partir de $\mathcal{D}$




Classificação dos pixels de $\mathcal{I}$ a partir dos paralelepípedos definidos

# Classificador de Mínima Distância

- Muito simples também
- Associa cada pixel da imagem à classe mais próxima
- Mas... pixel é um <u>ponto</u> e a classe é um <u>conjunto</u>
- Distância de <u>Mahalanobis</u> fornece a medida "ponto-conjunto"

$$D_j(\mathbf{x}) = \sqrt{\left(\mathbf{x} - \mu_j\right)^T \Sigma_j^{-1} \left(\mathbf{x} - \mu_j\right)}$$

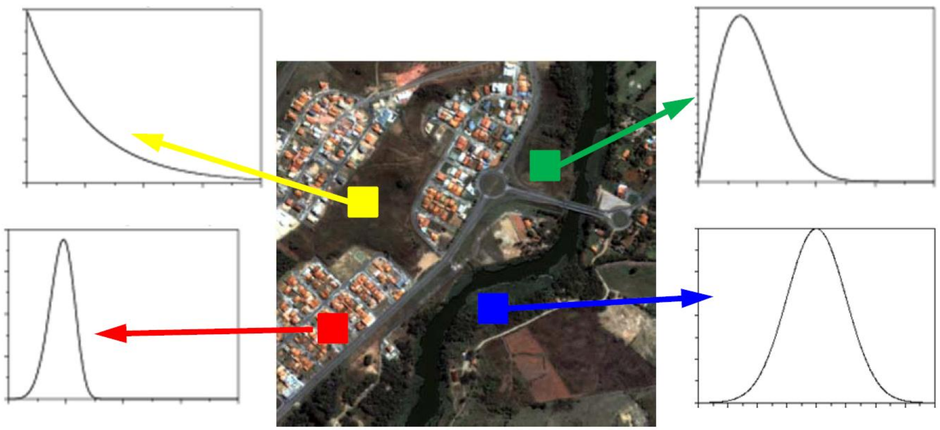
- Dado $\mathcal{D}$, para cada $\omega_j$, calcula-se $\mu_j$ e $\Sigma_j$ (vetor média e mat. covar.)

- Para cada $s_i \in \mathcal{S}$ tal que $\mathcal{I}(s_i) = \mathbf{x}_i$ é obtida a classificação via:

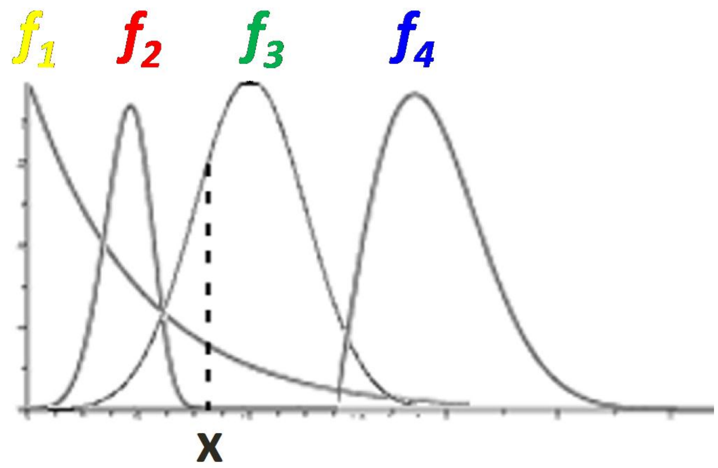
$$\mathcal{C}(s_i) = \arg \min_{j=1,\ldots,c} D_j(\mathbf{x}_i)$$

# Cla. de Máxima Verossimilhança

- Associa cada pixel da imagem à classe “máxima verossimilhança”
- Máxima Verossimilhança ~ Maior Probabilidade
- Dado $\mathcal{D}$, são estimadas f.d.p. $f_j$ para cada uma das $c$ classes



$\mathcal{D}$ é composto pelas diferentes amostras

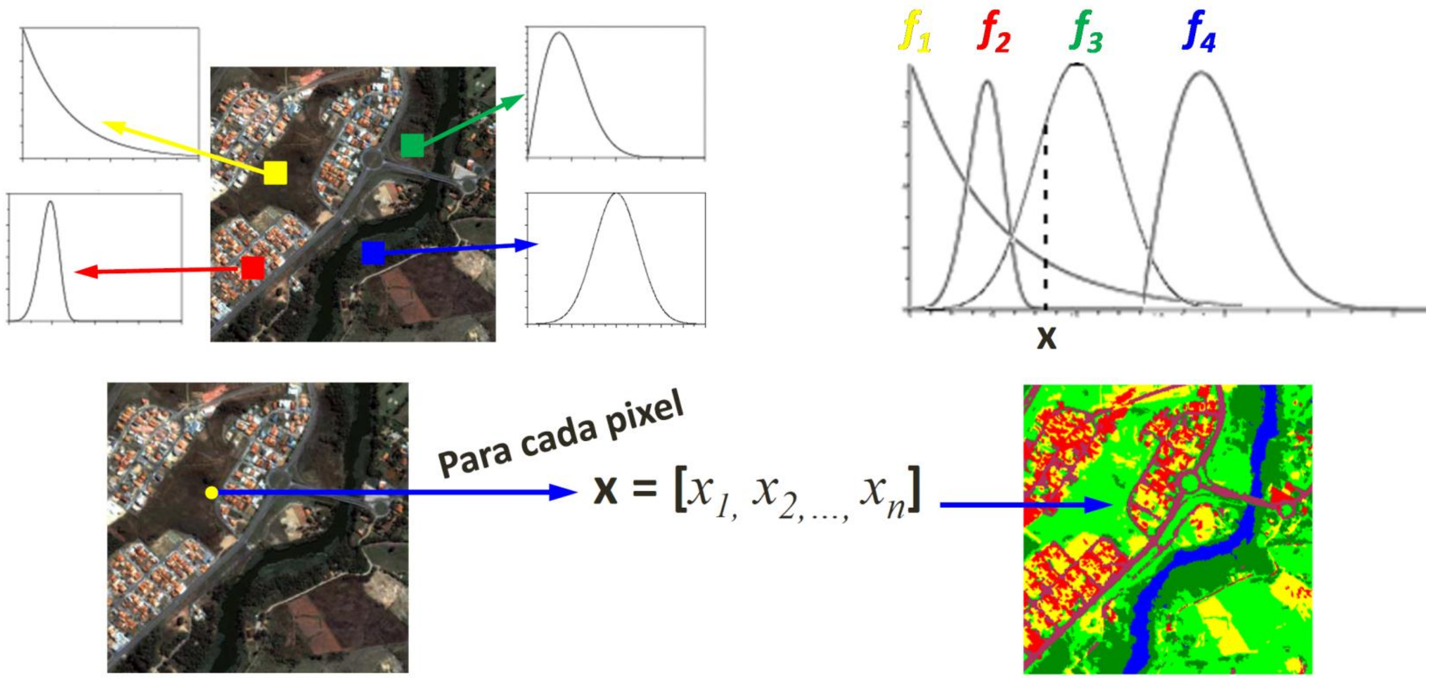



Cada amostra é usada para estimar a f.d.p. da respectiva classe

# Cla. de Máxima Verossimilhança

- Usualmente, a f.d.p. é Gaussiana Multivariada (basta estimar $\mu_j$ e $\Sigma_j$)

$$f_j(\mathbf{x}) = f\left(\mathbf{x}; \mu_j, \Sigma_j\right) = \cdots$$

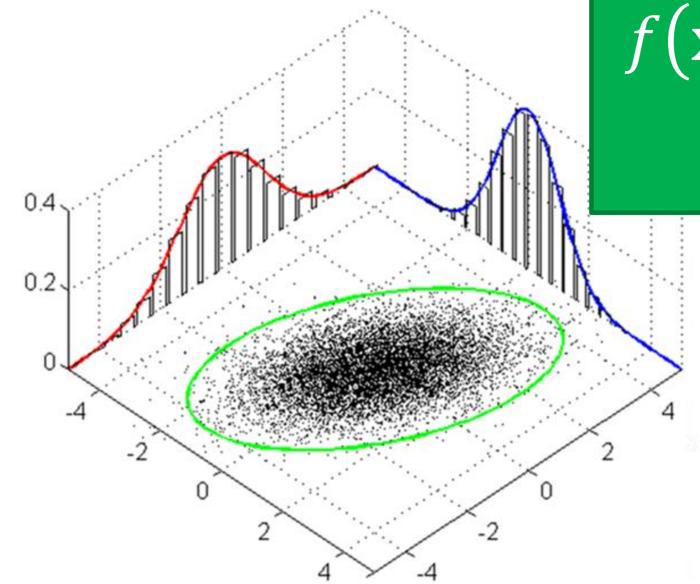


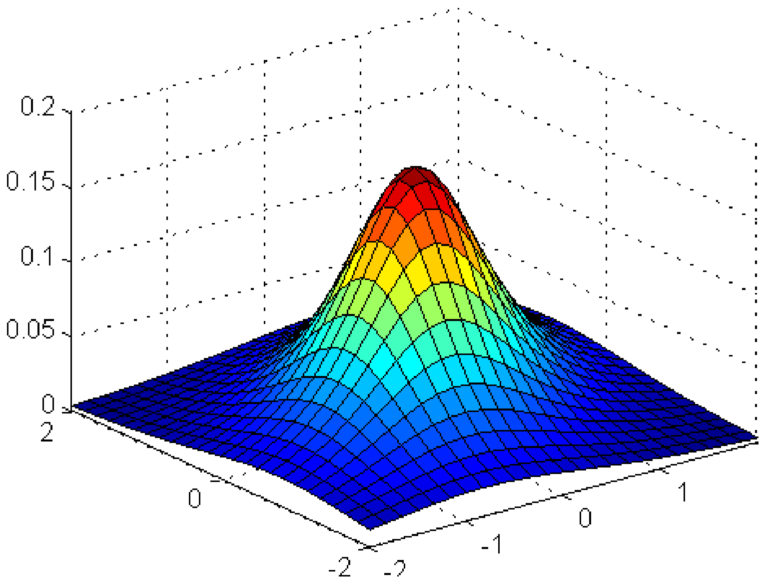

- De forma simplificada, a classificação é dada pela regra de decisão:

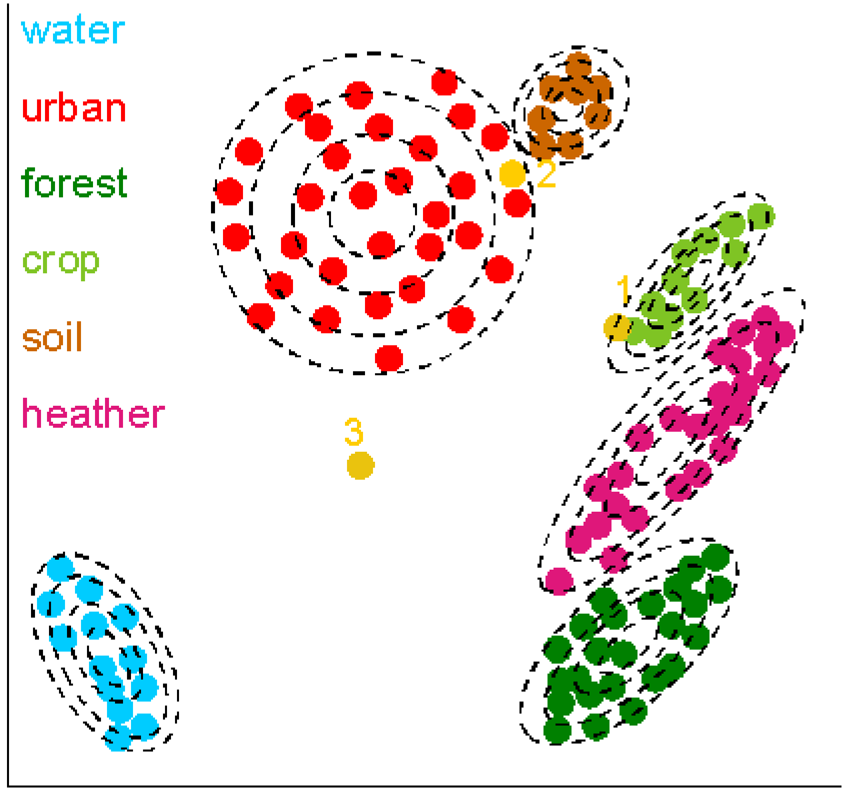
$$x_i \mapsto \omega_j \Leftrightarrow \arg \max_{j=1,\ldots,c} f_j(x)$$

# Cla. de Máxima Verossimilhança

$$f\left(\mathbf{x}; \mu_j, \Sigma_j\right) = \frac{1}{\sqrt{(2\pi)^k |\Sigma_j|}} e^{-\frac{1}{2}(\mathbf{x}-\mu_j)^T \Sigma_j^{-1}(\mathbf{x}-\mu_j)}$$



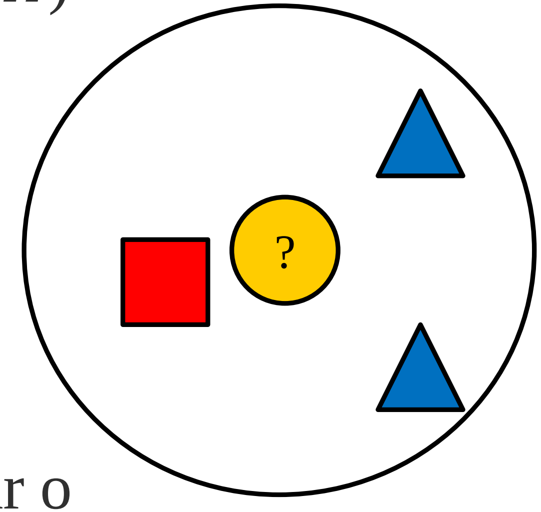




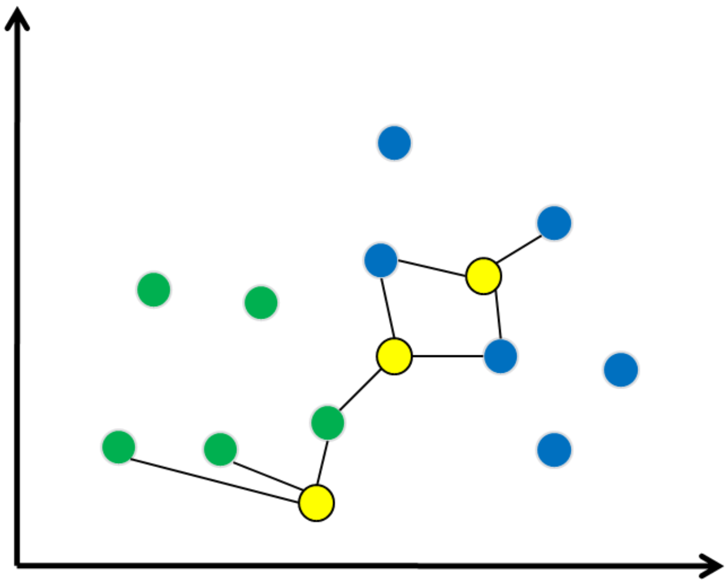
# K Vizinhos Mais Próximos

- Realiza a classificação segundo exemplos (conjunto de treinamento) mais próximos [segundo uma métrica] no espaço de atributos

i. Definir o número de vizinhos mais próximos ($K$)
ii. Calcular a distância entre o exemplo desconhecido e os exemplos do conjunto de treino
iii. Identificar os $K$ vizinhos mais próximos
iv. Utilizar o rótulo da classe dos vizinhos mais próximos que é mais frequente para determinar o rótulo da classe do exemplo desconhecido



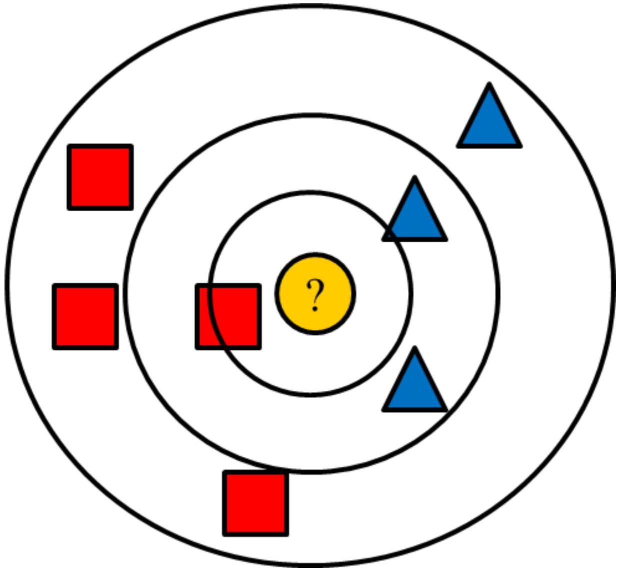
# K Vizinhos Mais Próximos



Ação dos vizinhos na classificação

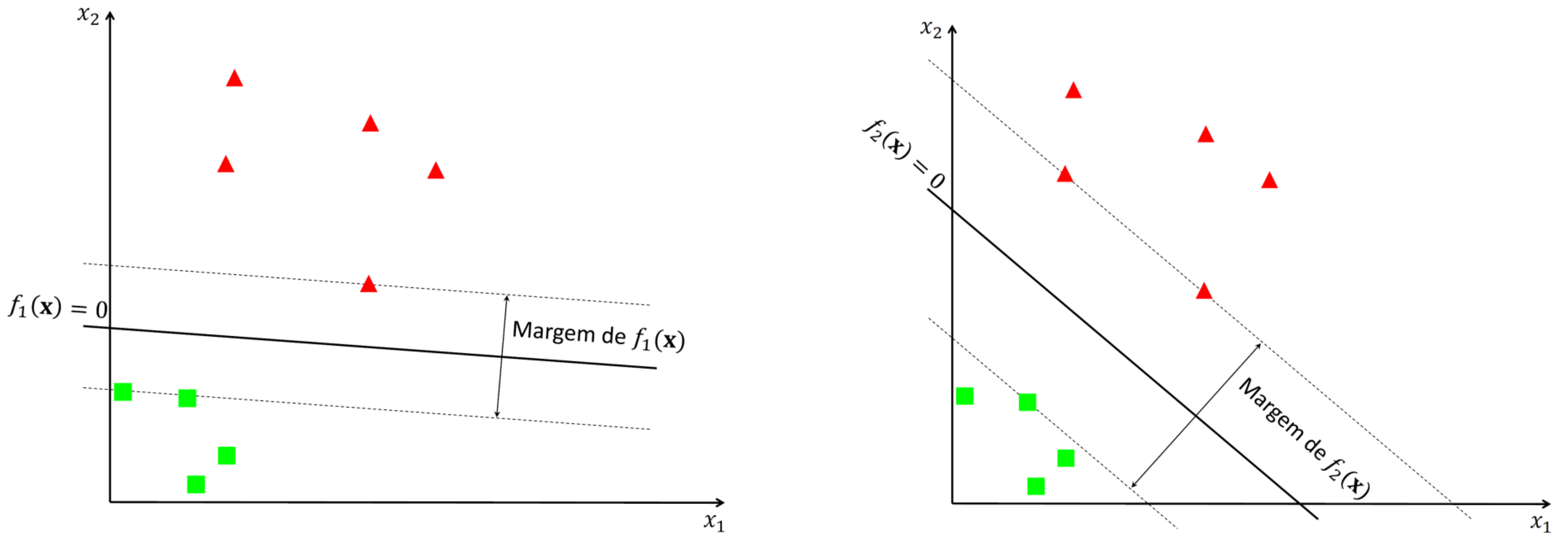
Vizinhos no Espaço de Atributos!

- **K = 1:** Pertence a classe de quadrados
- **K = 3:** Pertence a classe de triângulos
- **K = 7:** Pertence a classe de quadrados



# Máquina de Vetores Suporte

- SVM – Support Vector Machine
- Estado da arte em RP
- Excelente capacidade de generalização
- Independente da distribuição dos dados
- Classifica padrões a partir de hiperplanos cuja margem de separação entre os padrões de treinamento é máxima
- Hiperplano equivale ao lugar geométrico onde $f\ (\mathbf{x}) = \langle \mathbf{w}, \mathbf{x} \rangle + b$ é nula



# SVM

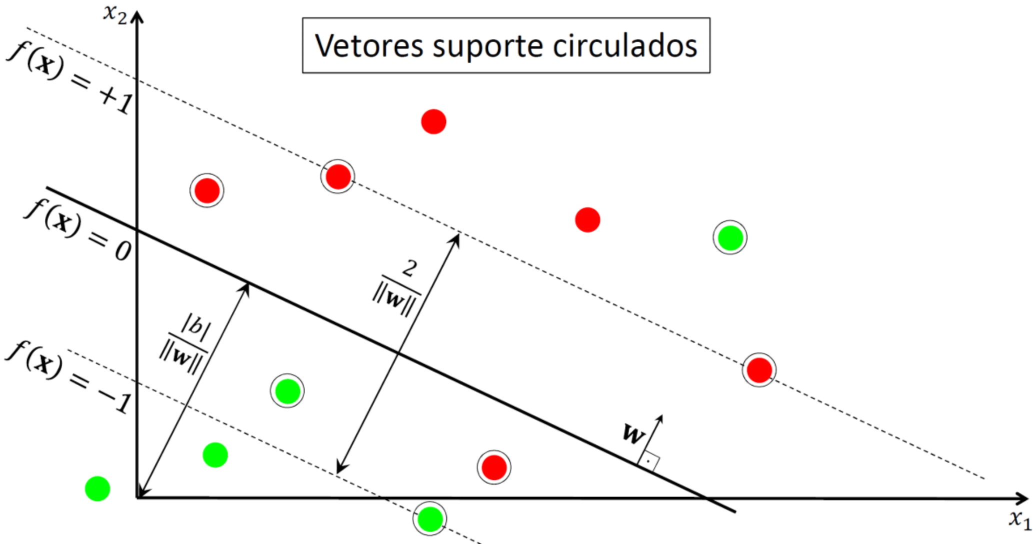
- Dado $\mathcal{D}$, o hiperplano é obtido com a resolução de:

$$\max \sum_{i=1}^{m} \lambda_i - \frac{1}{2}\sum_{i=1}^{m}\sum_{j=1}^{m} \lambda_i \lambda_j y_i y_j \langle \mathbf{x}_i, \mathbf{x}_j \rangle$$

$$\text{Sujeito a} \sum_{i=1}^{m} \lambda_i y_i = 0,\ 0 \le \lambda_i \le C$$

- Sendo $SV = \{x_i \mid \lambda_i \neq 0; i = 1, \dots, m\}$

$$\mathbf{w} = \sum_{\mathbf{x}_i \in \mathrm{SV}} \lambda_i y_i \mathbf{x}_i \qquad b = \frac{1}{\#SV} \sum_{\mathbf{x}_i \in SV} \left(y_i - \sum_{\mathbf{x}_j \in \mathrm{SV}} \lambda_j y_j \langle \mathbf{x}_i, \mathbf{x}_j \rangle\right)$$

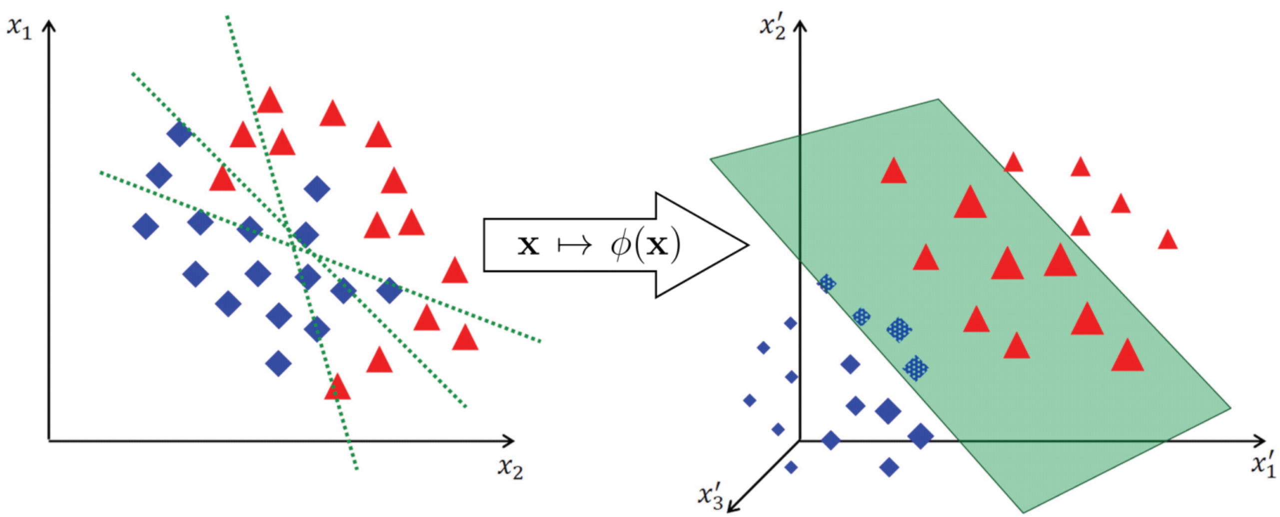



$$F(\mathrm{x}) = sgn(f(\mathrm{x}))$$

$$F(\mathrm{x}) = \begin{cases} +1 \Rightarrow (\mathbf{x}, \omega_1) \\ -1 \Rightarrow (\mathbf{x}, \omega_2) \end{cases}$$

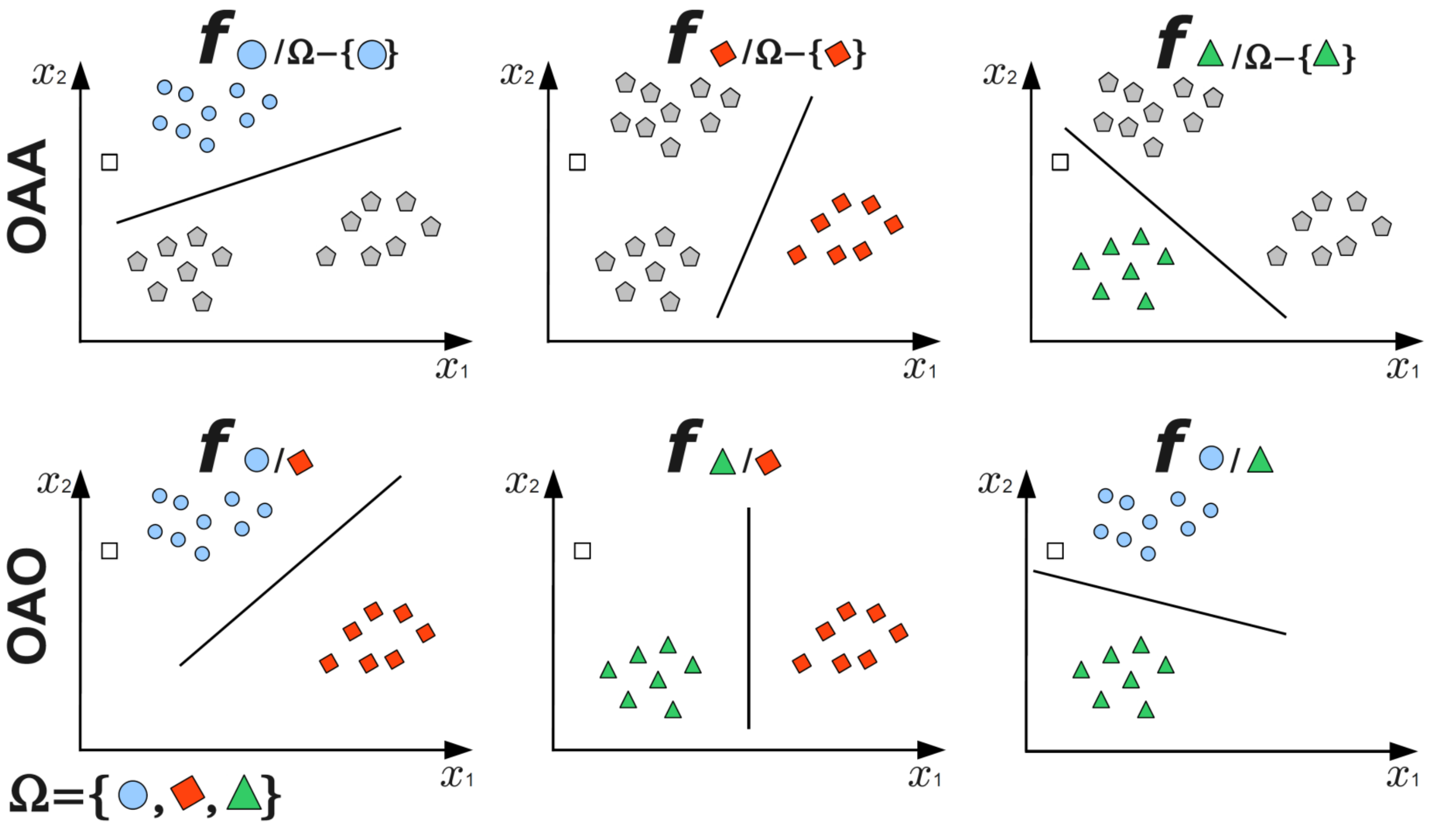
# Funções Kernel

- Substitui o produto interno $\langle x_i, x_j \rangle$ de ★ por $K(x_i, x_j) = \langle \varphi(x_i), \varphi(x_j) \rangle$
- $\varphi(\cdot)$ mapeia os padrões para um novo espaço onde a separação pode ser linear
- Geralmente $K$ é definida sem o conhecimento de $\varphi$
- Exemplos:
  - Kernel Linear: $K(x_i, x_j) = \langle x_i, x_j \rangle$
  - Kernel de Base Radial: $K(x_i, x_j) = e^{-\gamma \|x_i - x_j\|^2}; \gamma \in \mathbb{R}_+^*$



# Estratégias Multiclasses

- SVM possibilita a separação apenas entre duas classes
- Problemas onde são discriminadas um número maior de classes é necessário o emprego de Estratégias Multiclasses
- Dois exemplos de estratégias são:
    - Um-Contra-Todos (OAA)
    - Um-Contra-Um (OAO)



# Bibliografia da aula


- ...
- Lima, E. S. Notas de aula: INF 1771 – Inteligência Artificial - Aula 15 – K-Nearest Neighbor (KNN), 2012.