



Instituto Federal de Educação, Ciência e Tecnologia do Ceará
PPGER – PPGCC

Aula 10: Introdução à Machine Learning

Visão Computacional

Prof. Dr. Pedro Pedrosa Rebouças Filho

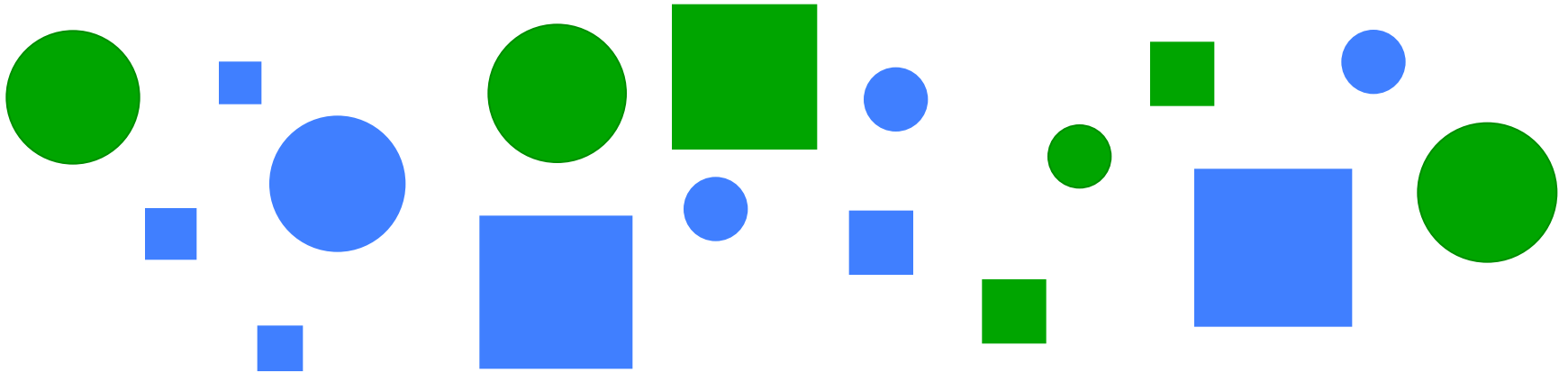
pedrosarf@ifce.edu.br

professorpedrosa.com

Agrupamento e Classificação de Padrões

Reconhecimento de Padrões

Agrupamento

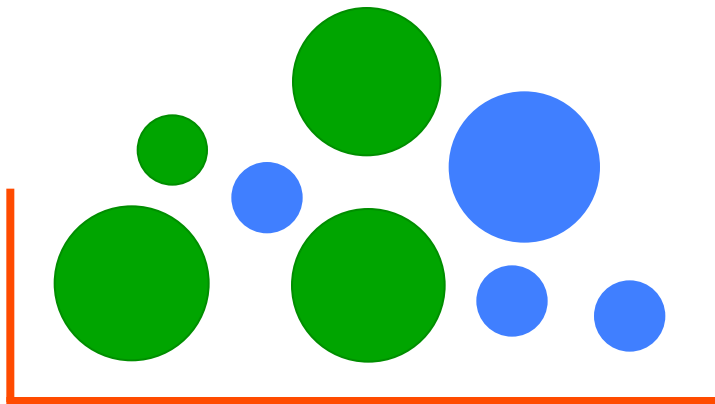


Agrupamento:
Característica = Número de Vértices

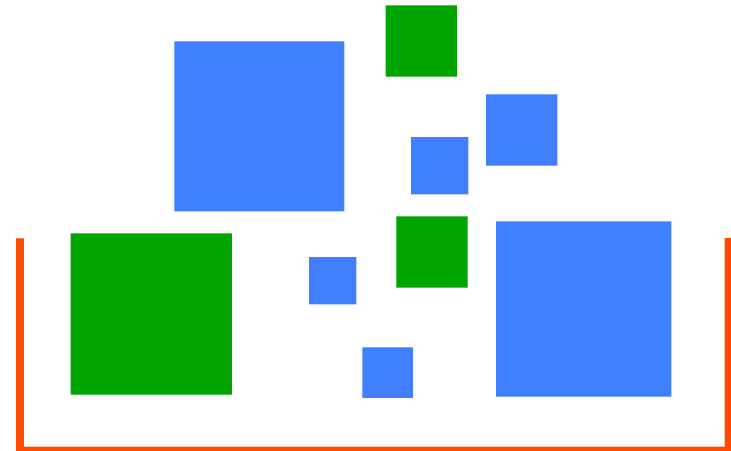


Agrupamento

Agrupamento:
Característica = Número de Vértices



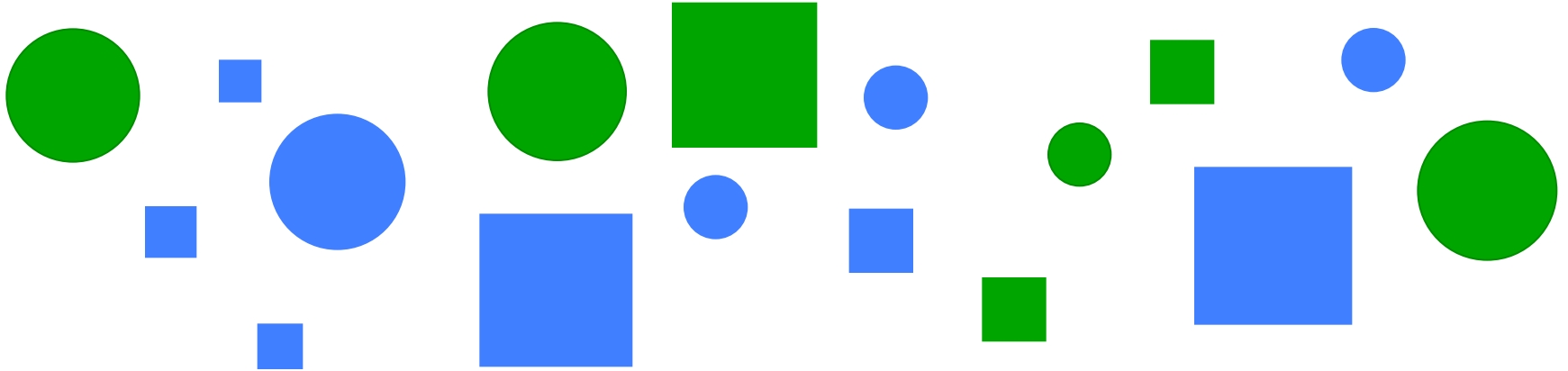
0 vértices



4 vértices



Agrupamento

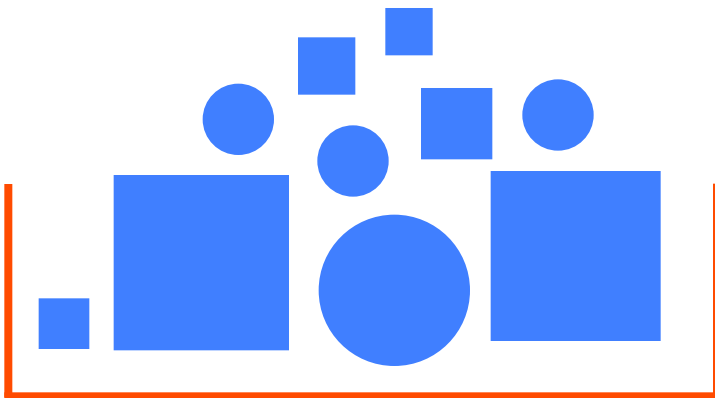


Agrupamento:
Característica = Cor (Comprimento de Onda)

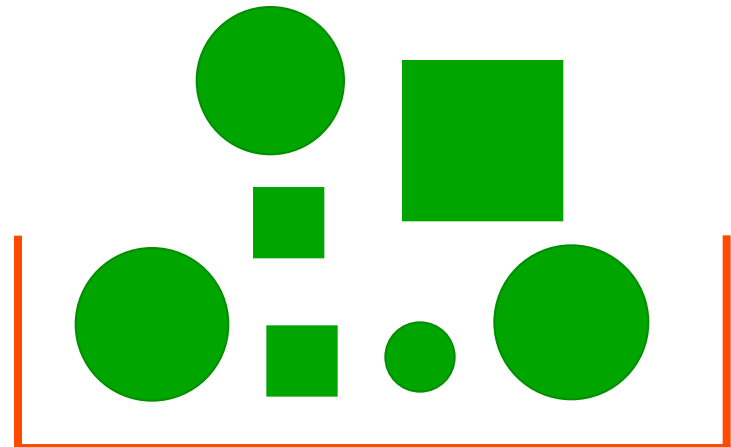


Agrupamento

Agrupamento:
Característica = Cor (Comprimento de Onda)



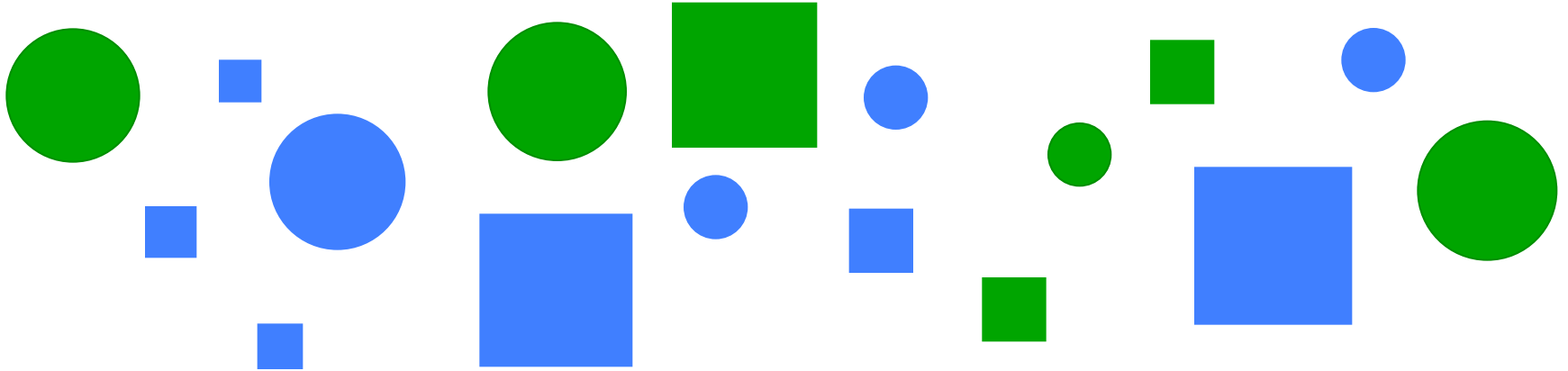
$\lambda = 470 \text{ nm}$



$\lambda = 550 \text{ nm}$



Agrupamento

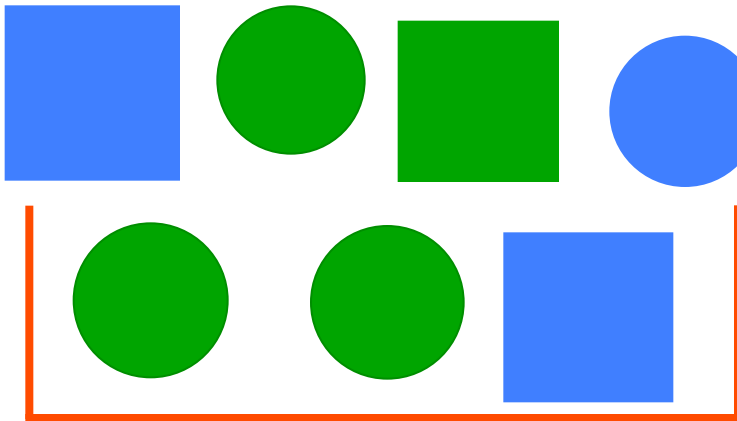


Agrupamento:
Característica = Área

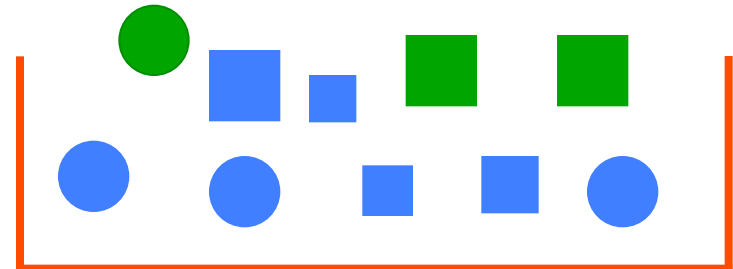


Agrupamento

**Agrupamento:
Característica = Área**



0 vértices



4 vértices



Agrupamento x Classificação

- Agrupamento:
 - Geração de grupos utilizando atributos pré-estabelecidos
 - Atributos = Propriedades = Descritores
 - Grupos gerados são definidos como **Padrões de Classificação**
 - Esta etapa também é conhecida como **Treinamento**
 - Os grupos gerados estão associados ao Aprendizado do problema
- Classificação:
 - Define a qual grupo um objeto desconhecido pertence
 - Utiliza os mesmos atributos do agrupamento
 - Utiliza métricas para medir a similaridade dos **novos objetos** com os grupos já formados
 - Distância euclidiana, manhatan, entre outras...
 - Esta etapa é o que define esta aula, que é o **Reconhecimento de Padrões**

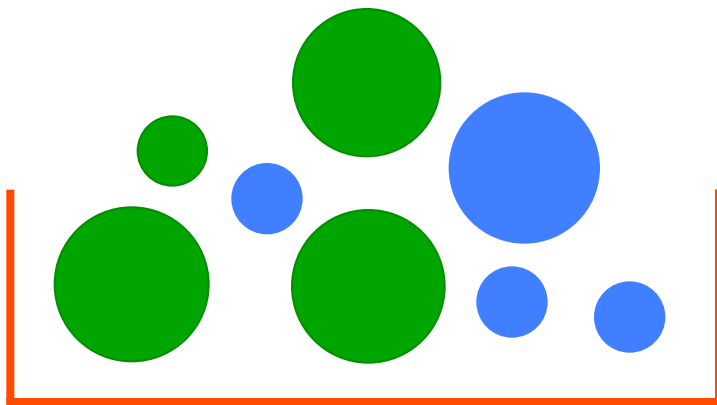


Classificação

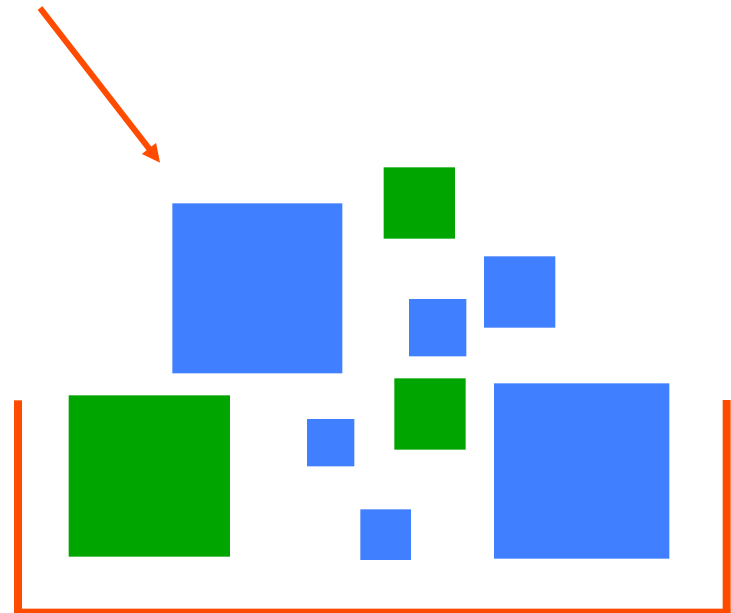
Novo objeto



Quantos vértices?



0 vértices



4 vértices

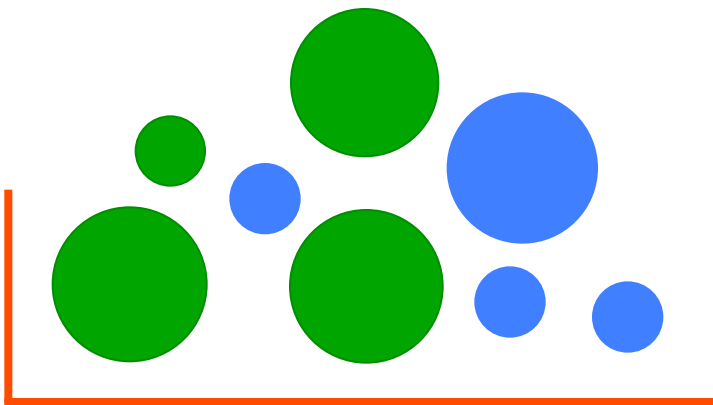


Classificação

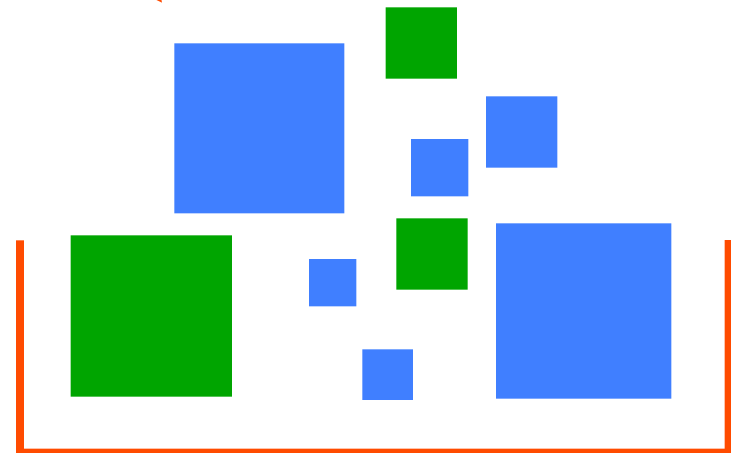
Novo objeto



Quantos vértices?



0 vértices



4 vértices

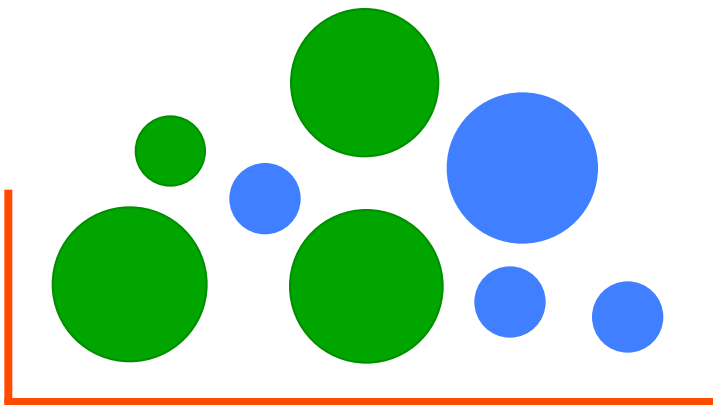
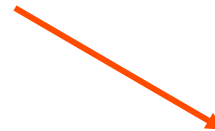


Classificação

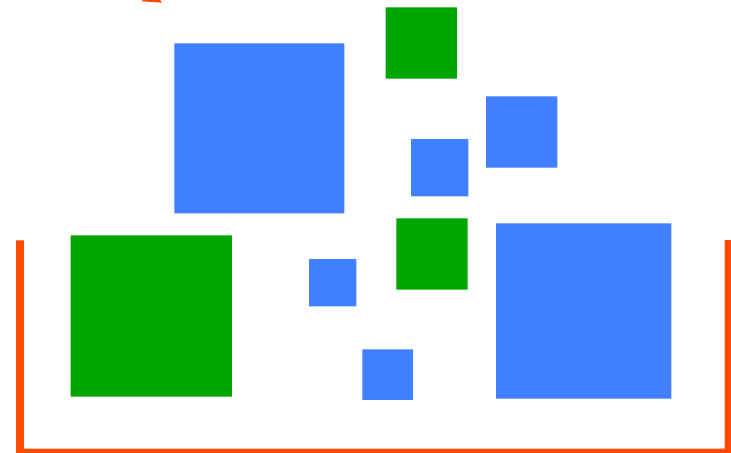
Novo objeto



Quantos vértices?



0 vértices



4 vértices

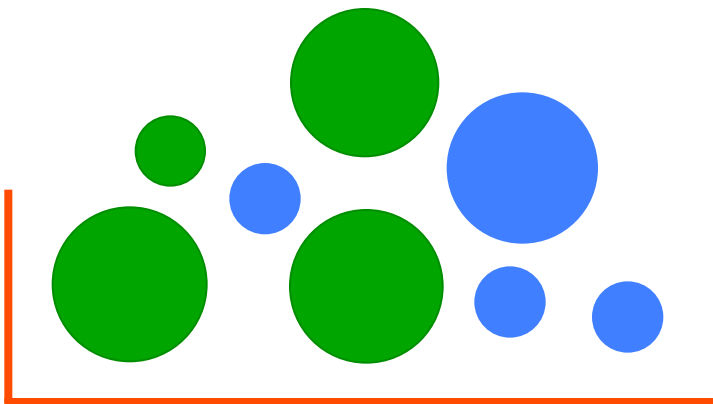
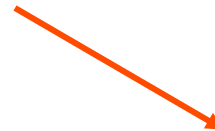


Classificação

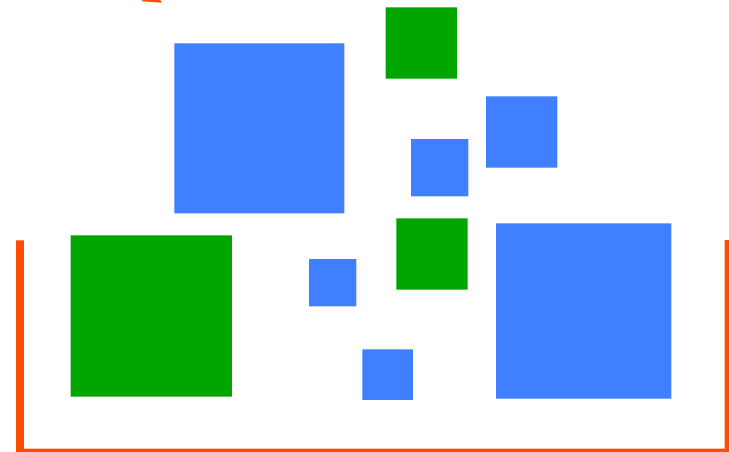
Novo objeto



Quantos vértices?



0 vértices



4 vértices

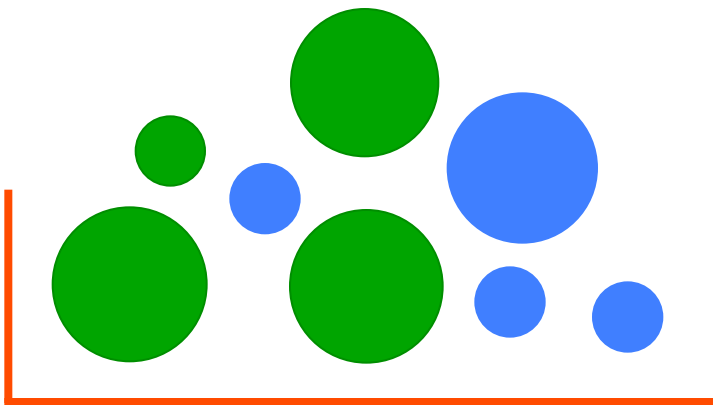
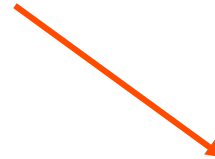


Classificação

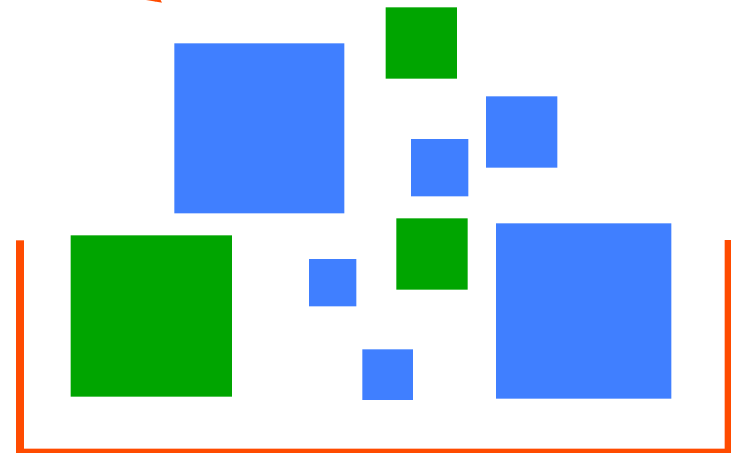
Novo objeto



Quantos vértices?



0 vértices



4 vértices

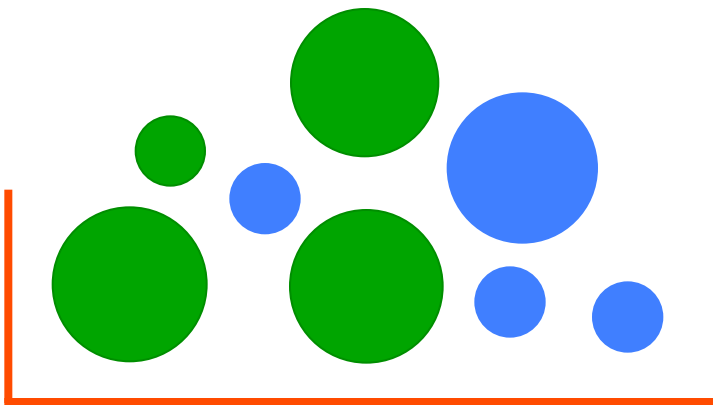


Classificação

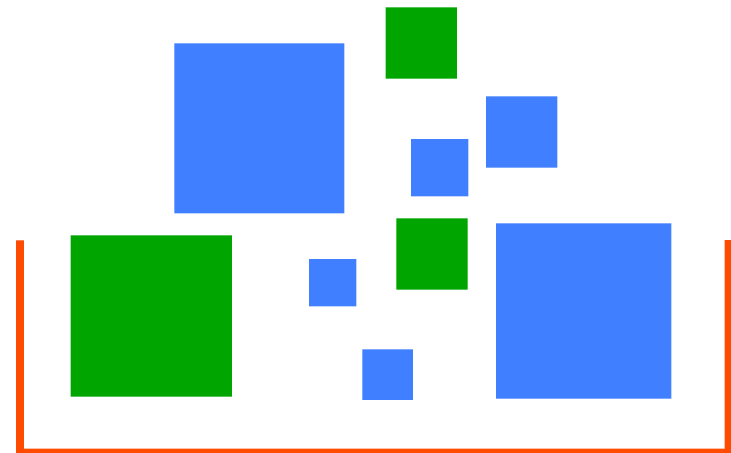
Novo objeto



Quantos vértices?



0 vértices



4 vértices



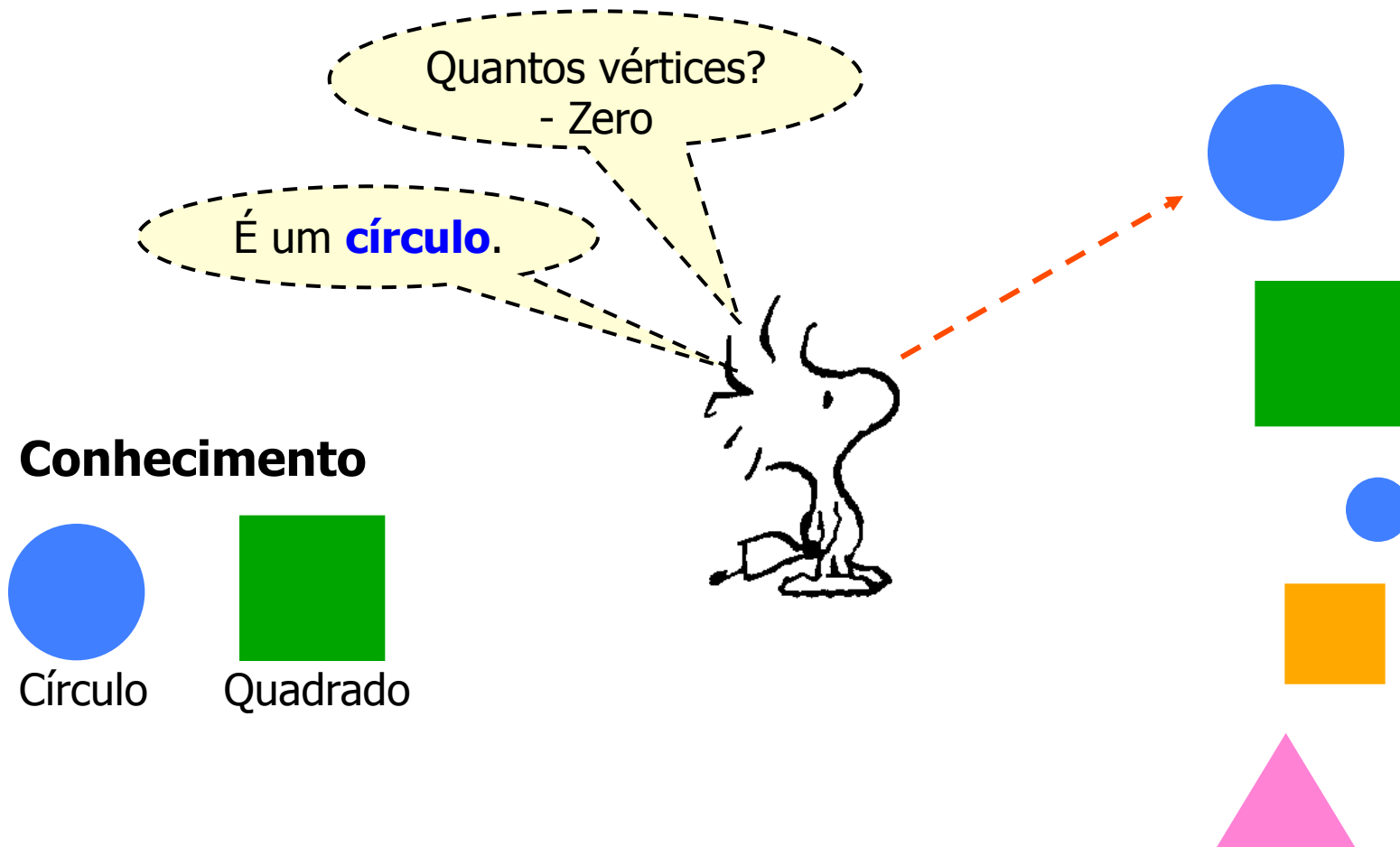
Problemas no Agrupamento

Agrupamento e
Classificação de Padrões

Reconhecimento de Padrões

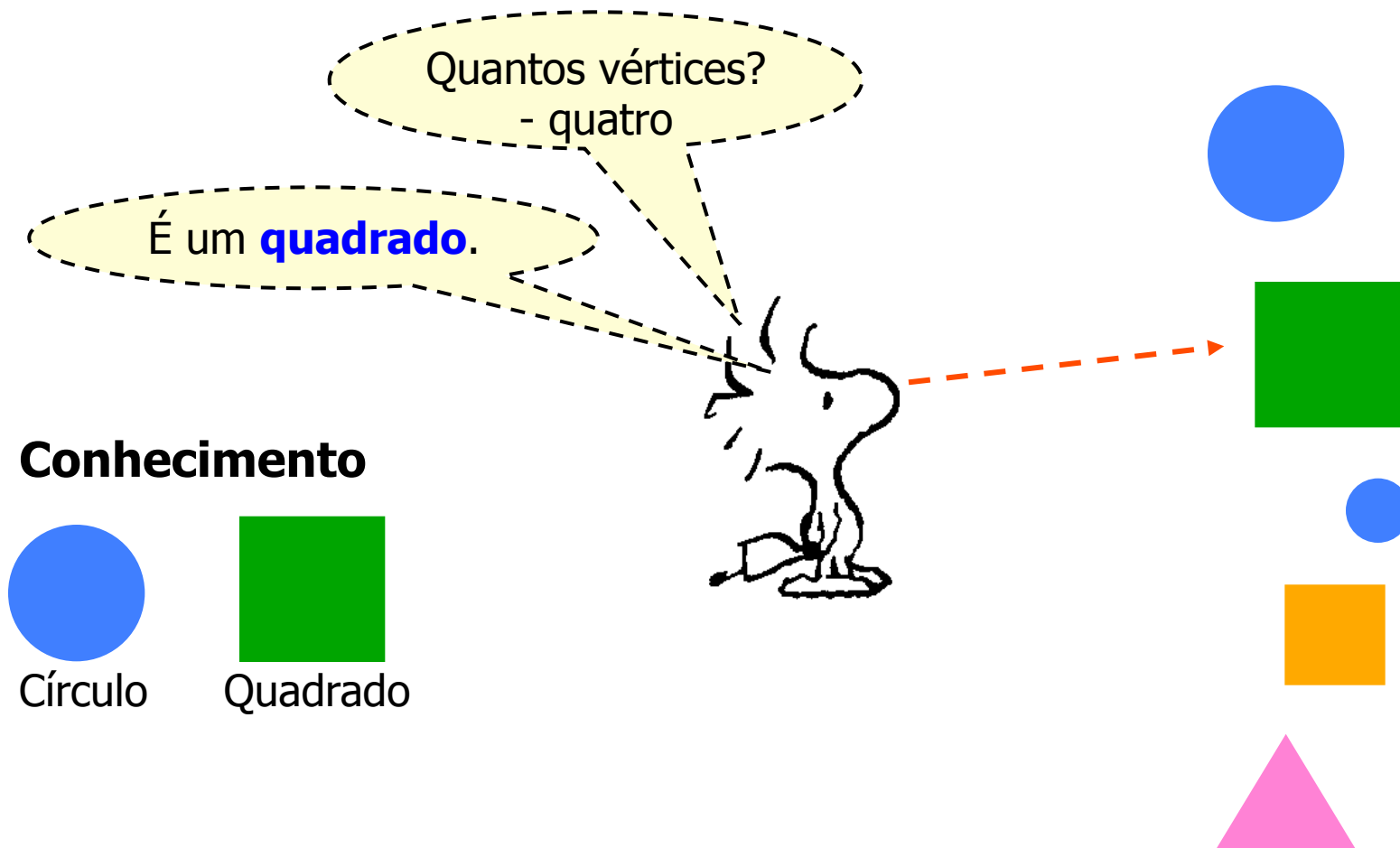
Reconhecimento de Padrões

Problemas na classificação



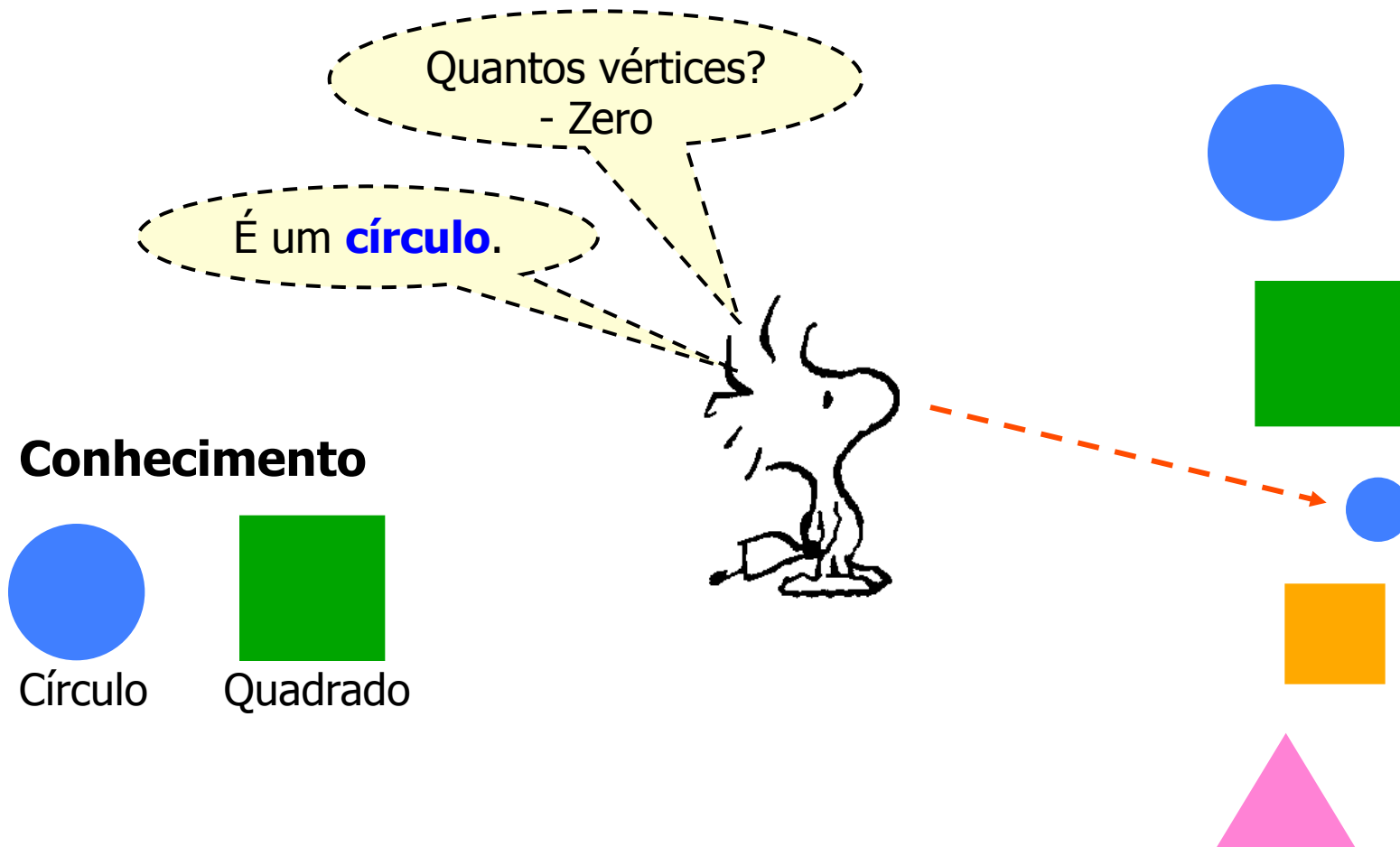
Reconhecimento de Padrões

Problemas na classificação



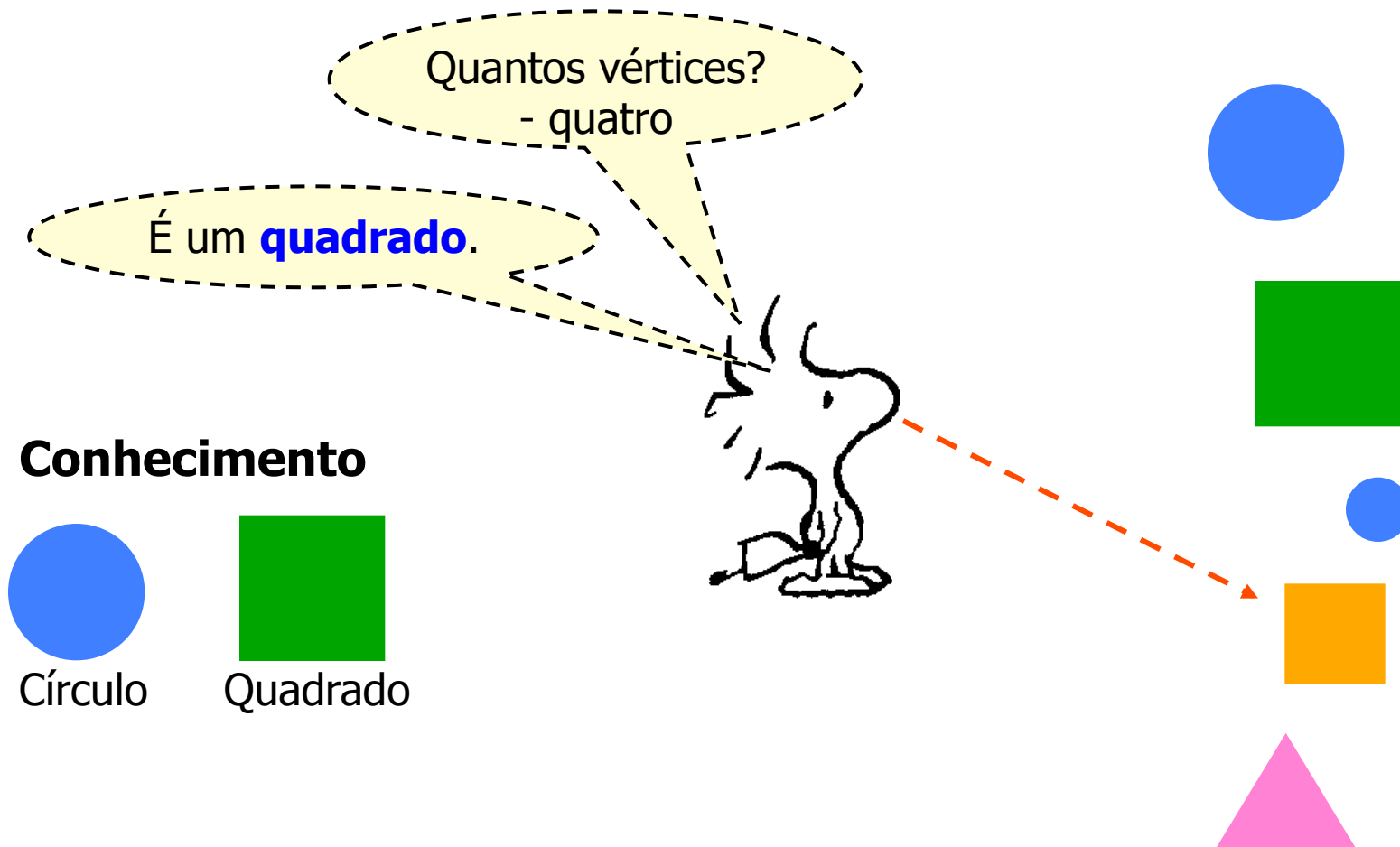
Reconhecimento de Padrões

Problemas na classificação



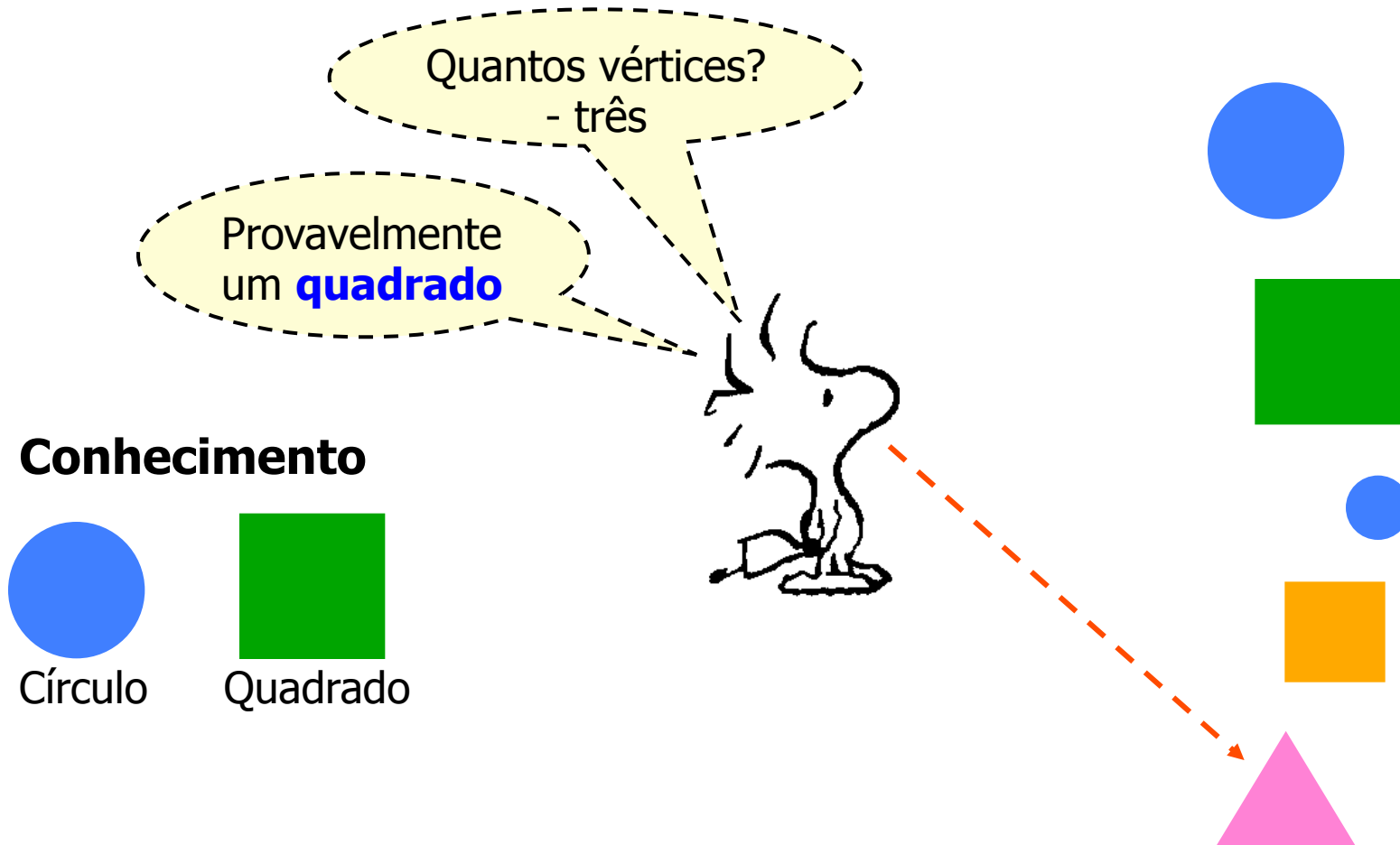
Reconhecimento de Padrões

Problemas na classificação



Reconhecimento de Padrões

Problemas na classificação



Tipos de Classificação

Reconhecimento de Padrões

Tipos de Classificação

- Classificação **supervisionada**
 - Aprendizado com informações à priori
 - Amostras de treinamento são classificadas
 - Número de Classes é conhecido
 - Aprendizado por **Exemplo**
- Clusterização **Não supervisionada**
 - Aprendizado sem amostras classificadas
 - Conhecido como clusterização
 - Aprendizado por **Observação**



Medidas de similaridade

Reconhecimento de Padrões

Medidas de Similaridade

- Distância euclidiana
- Distância de Manhattan
- A maioria dos algoritmos de análise de agrupamento têm como base estas medidas de dissimilaridade;
- Quanto maior for a medida de dissimilaridade menor será a semelhança entre os indivíduos.



Distância euclidiana

- A distância euclidiana entre os indivíduos a e b é dada analiticamente por:

$$d_{ab} = \left[\sum_{j=1}^p (X_{aj} - X_{bj})^2 \right]^{1/2}$$

$p = 1, 2, \dots, j;$

X_{aj} = valor da variável j para o indivíduo a ;

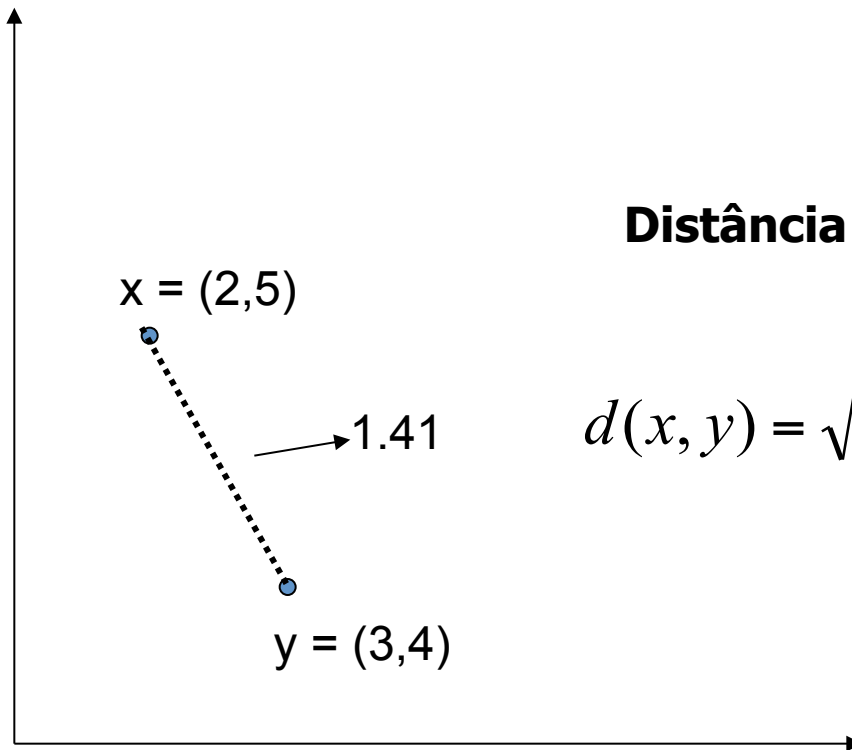
X_{bj} = valor da variável j para o indivíduo b .



Distância euclidiana

- Definição
- Exemplo:

$$d(x, y) = \sqrt{\sum_{i=1}^n (x_i - y_i)^2}$$



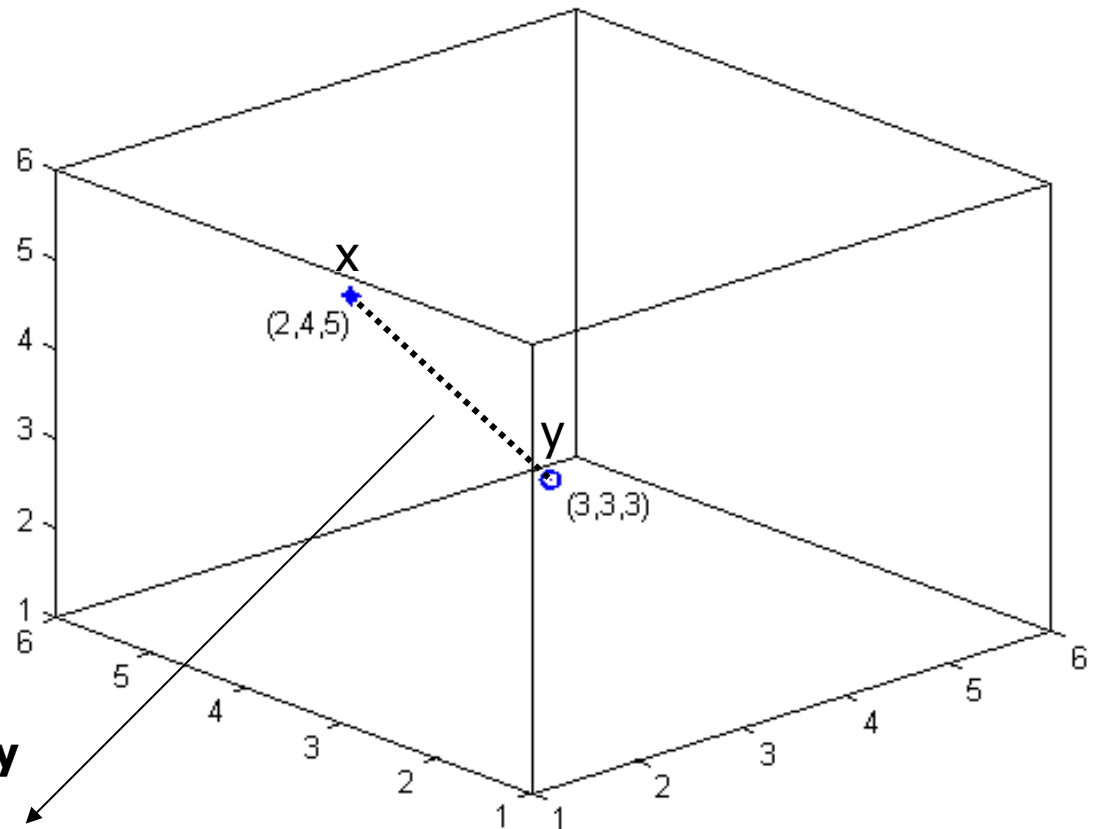
Distância entre x e y com duas variáveis

$$d(x, y) = \sqrt{(2 - 3)^2 + (5 - 4)^2} = \sqrt{2} = 1.41$$

Distância euclidiana

- Definição
- Exemplo:

$$d(x, y) = \sqrt{\sum_{i=1}^n (x_i - y_i)^2}$$



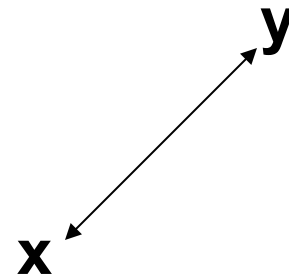
Distância entre x e y
com **três variáveis**

$$d(x, y) = \sqrt{(2-3)^2 + (4-3)^2 + (5-3)^2} = \sqrt{6} = 2.44$$

Calculando Distâncias

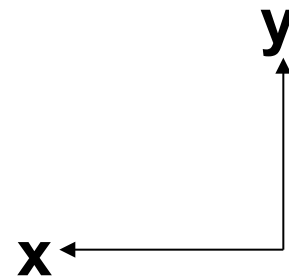
- Distância Euclidiana

$$d = \sqrt{\sum_{i=1}^n (x_i - y_i)^2}$$



- Manhattan (City Block)

$$d = \sum_{i=1}^n |x_i - y_i|$$



Calculando Distâncias

- Minkowski
 - Parâmetro r
 - $r = 2$, distância Euclidiana
 - $r = 1$, City Block

$$d = \left(\sum_{i=1}^n (x_i - y_i)^r \right)^{1/r}$$



Classificação Supervisionada

Método KNN

Reconhecimento de Padrões

K-Vizinho mais próximos

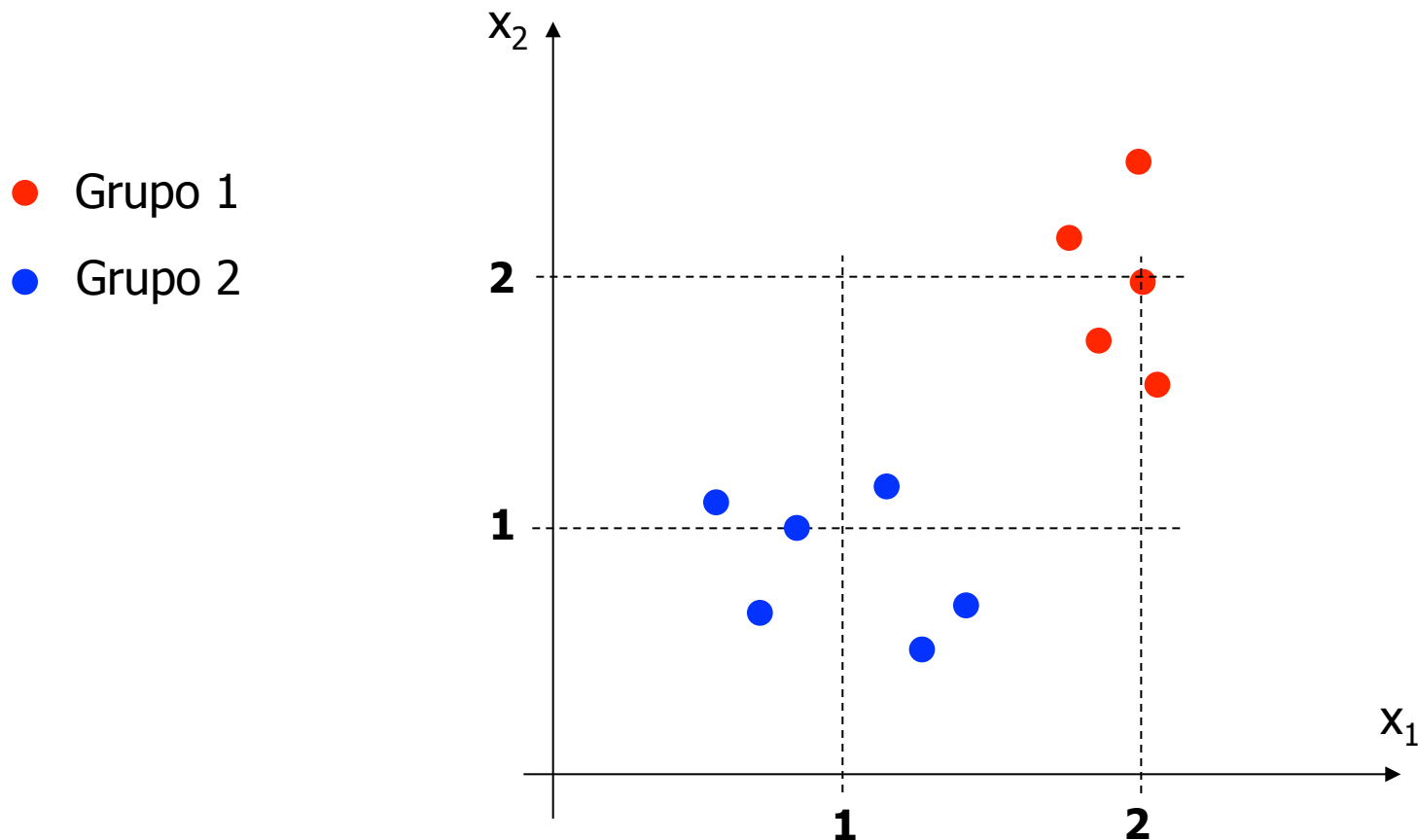
- Consiste em calcular a similaridade entre um novo objeto e todos os objetos presentes no banco de treinamento
- Considere
- Algoritmo de treinamento
 - Todos os exemplos classificados são adicionados no banco de análise
 - Treinamento rápido
- Algoritmo de Classificação
 - Calcular a distância do novo objeto para todos os objetos do banco de treinamento
 - O objeto que apresentar a menor distância será usado no resultado, pois o grupo a qual pertence o objeto mais próximo será o resultado da classificação
 - Classificação lenta, ficando ainda mais lenta quando existe muitos objetos no banco de treinamento.



K-Vizinho mais próximos

Treinamento

Adicione todos os objetos conhecidos ao banco de treinamento (Conhecimento)

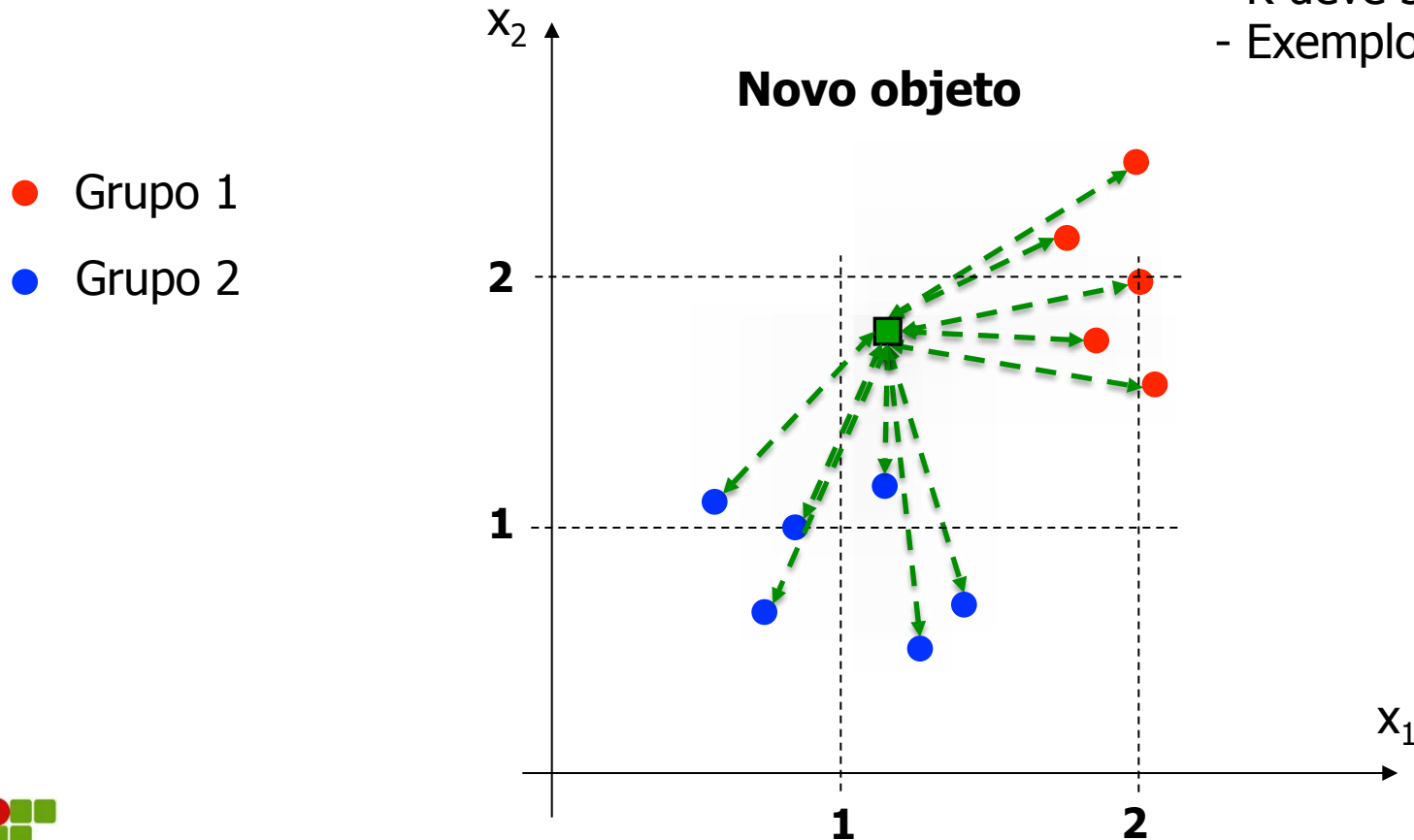


K-Vizinho mais próximos

Reconhecimento

- 1-Computar distância para todos do banco de conhecimento
- 2-Encontrar os k vizinhos mais próximos

- K deve ser ímpar (1,3,5,...)
- Exemplo do **k = 1**



K-Vizinho mais próximos

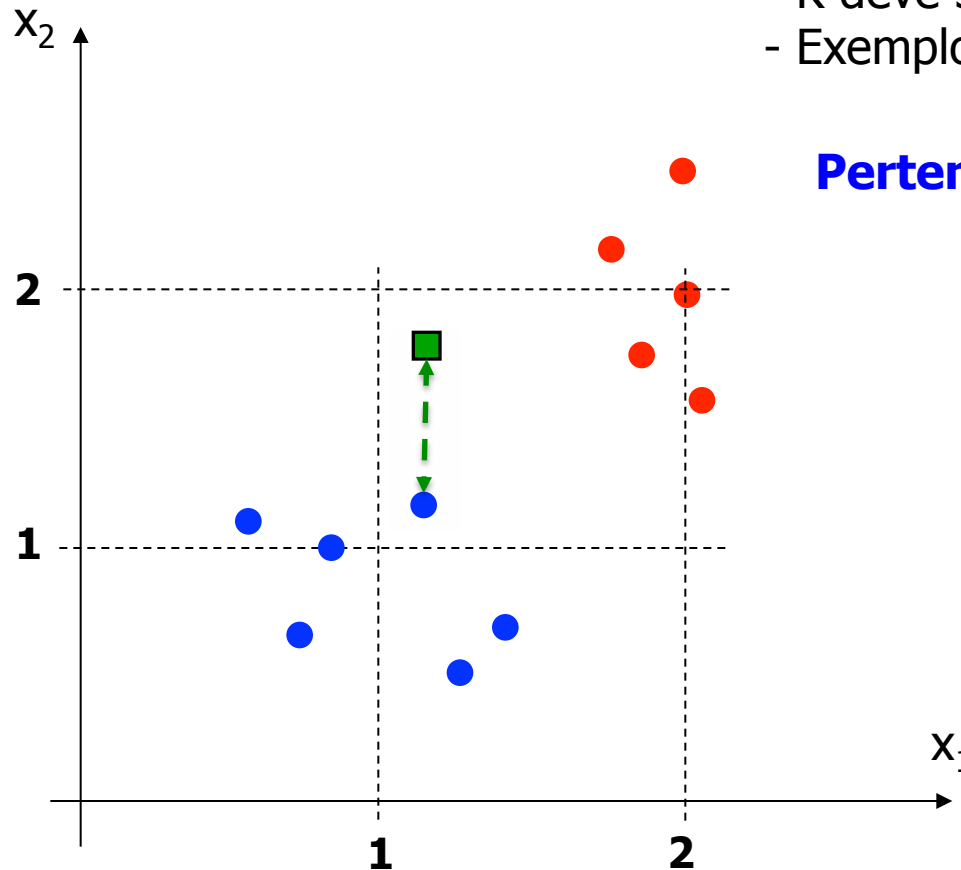
Reconhecimento

- 1-Computar distância para todos do banco de conhecimento
- 2-Encontrar os k vizinhos mais próximos

- K deve ser ímpar (1,3,5,...)
- Exemplo do **k = 1**

- Grupo 1
- Grupo 2

Pertence ao Grupo 2

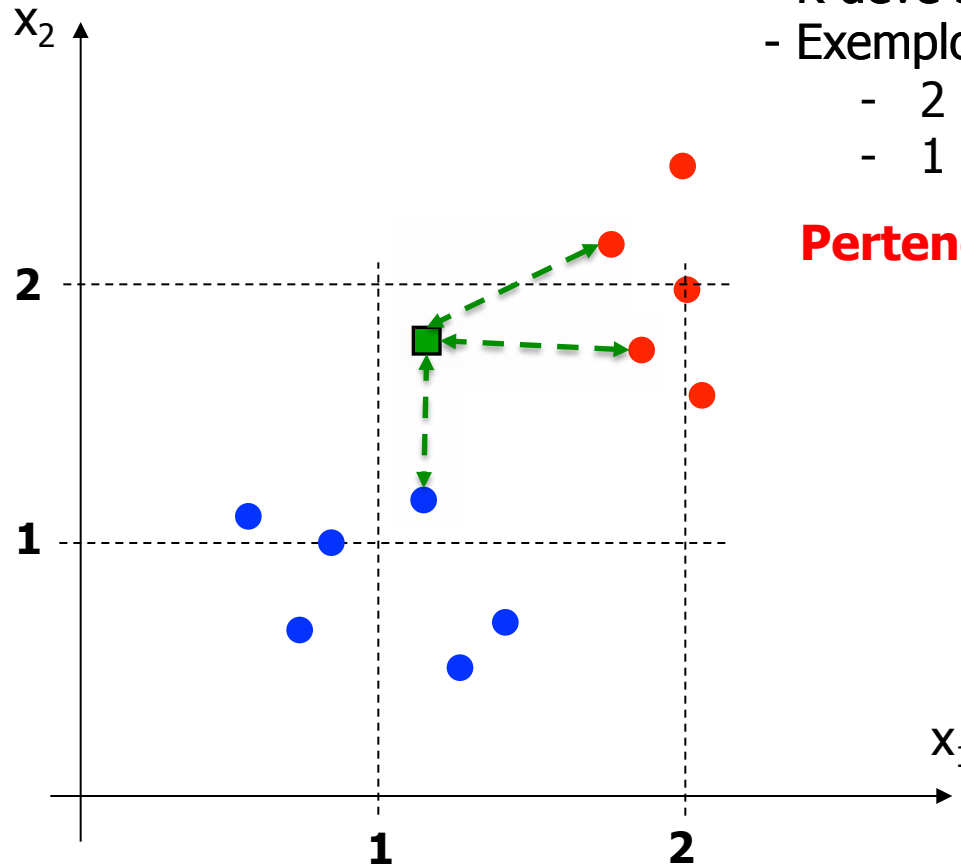


K-Vizinho mais próximos

Reconhecimento

- 1-Computar distância para todos do banco de conhecimento
- 2-Encontrar os k vizinhos mais próximos

- Grupo 1
- Grupo 2



- K deve ser ímpar (1,3,5,...)
- Exemplo do **k = 3**
 - 2 do grupo 1
 - 1 do grupo 2

Pertence ao Grupo 1

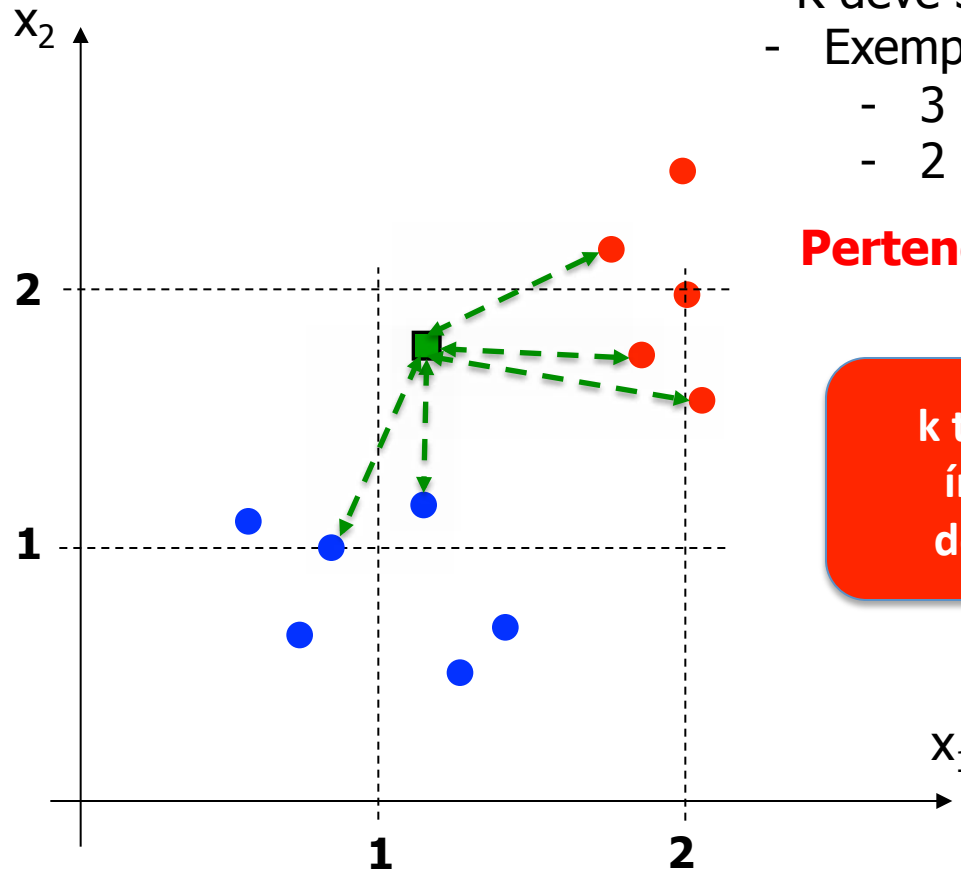


K-Vizinho mais próximos

Reconhecimento

- 1-Computar distância para todos do banco de conhecimento
- 2-Encontrar os k vizinhos mais próximos

- Grupo 1
- Grupo 2



- K deve ser ímpar (1,3,5,...)
- Exemplo do **k = 5**
 - 3 do grupo 1
 - 2 do grupo 2

Pertence ao Grupo 1

k tem que ser
ímpar para
desempatar



Classificação Supervisionada

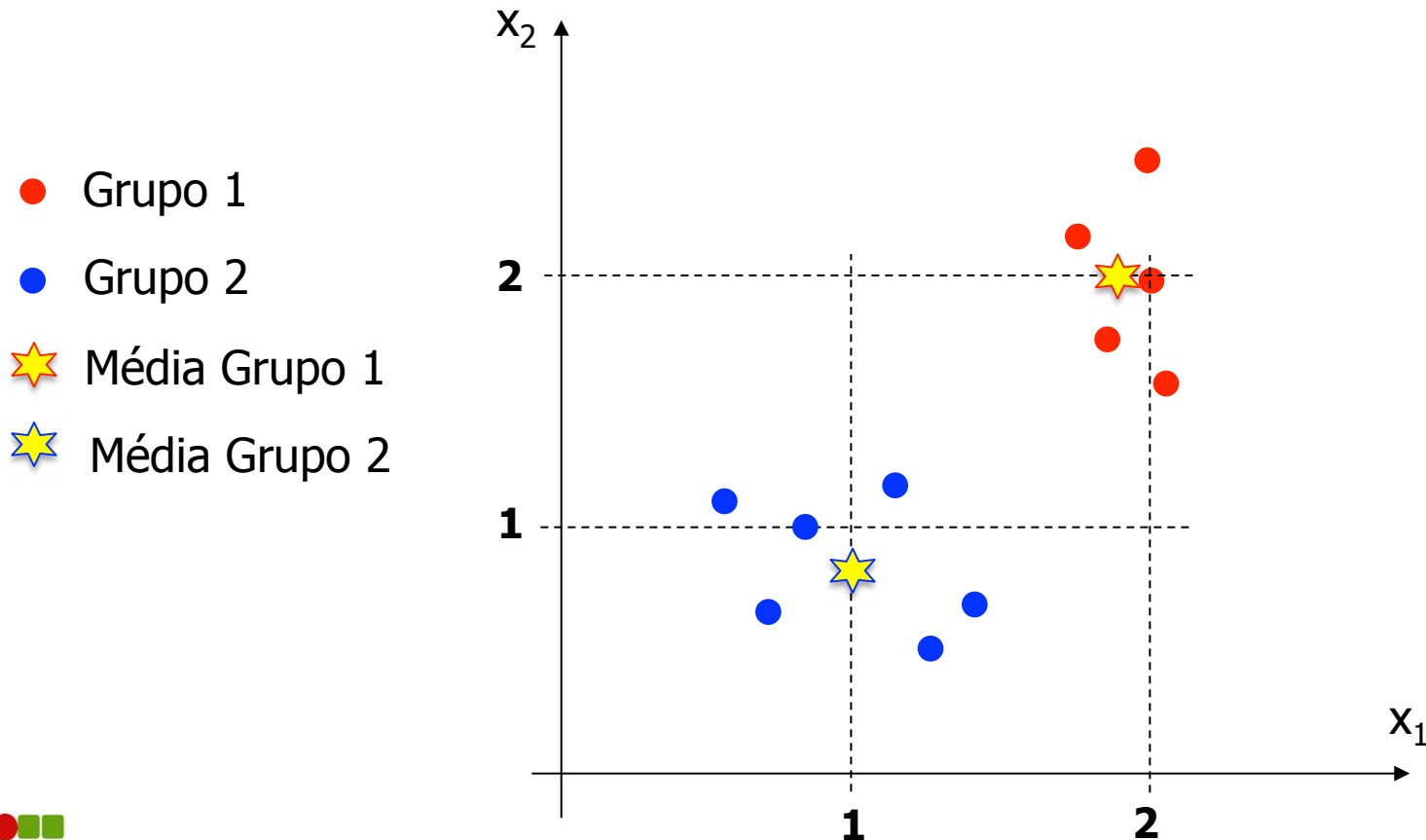
Método Menor Distância Centróides

Reconhecimento de Padrões

Método Menor Distância Centróides

Treinamento

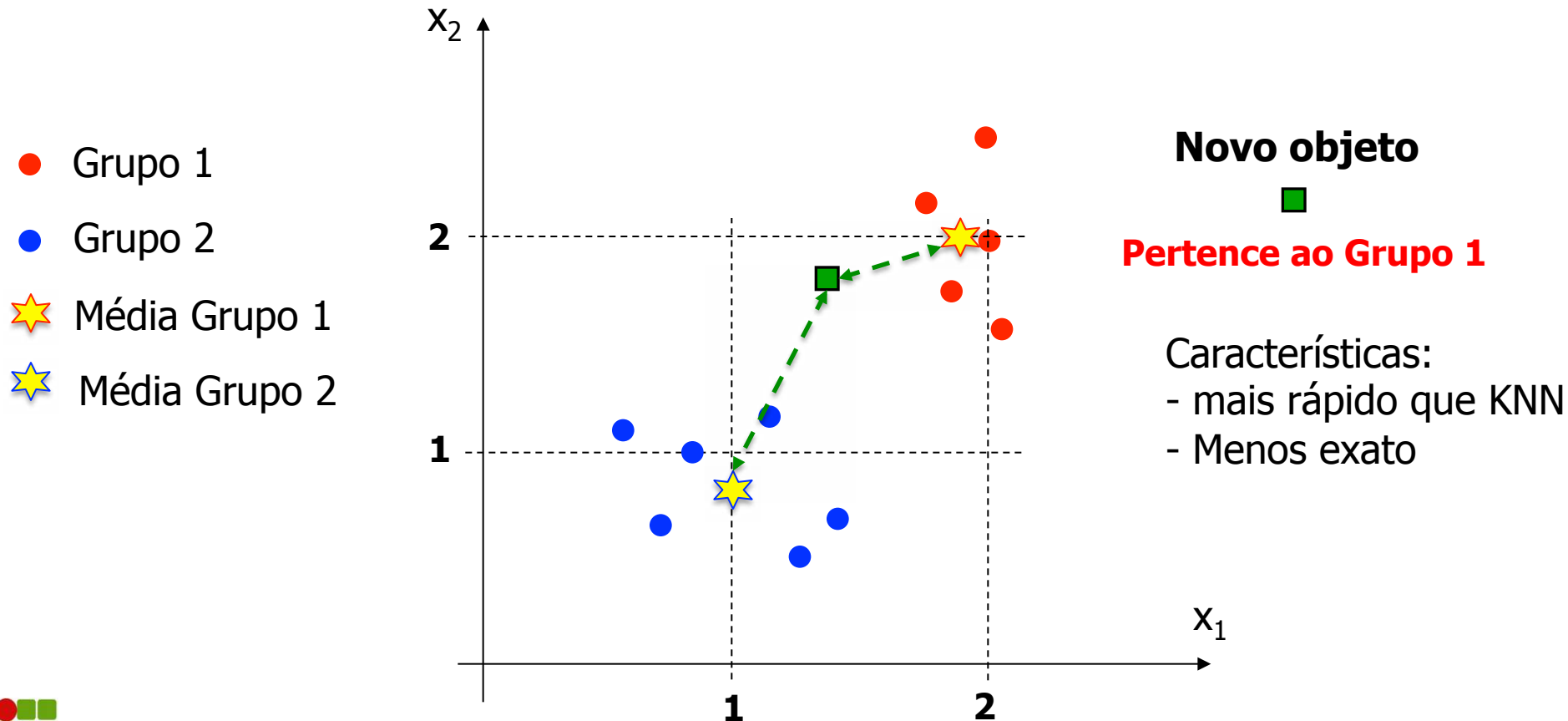
Calcular a média de todos os parâmetros usando todo banco de dados (Conhecimento)



Método Menor Distância Centróides

Reconhecimento

- 1-Computar distância para todos os centróides obtidos no treinamento
- 2-Encontrar o grupo que possui o centróide mais próximo do novo objeto



Classificação **Não** Supervisionada

Método K-médias

Reconhecimento de Padrões

Tipos de Técnicas de Clusterização

- **Particionamento**

- K-means;
- K-fuzzy;
- Outros

Particional e baseada em protótipos.

Encontra um número k de clusters (k é fornecido pelo usuário) que são representados por seus centróides.



Método K-médias

- Análise de agrupamento ou Cluster analysis:
- Sequência de regras (algoritmo) para agrupar objetos sem inferência de probabilidade a priori dos grupos.
- Técnica utilizada em classificadores denominados de ‘não supervisionados’.
- Definição formal:
 - “Dado um conjunto de ‘n’ unidades amostrais (tratamentos, objetos, indivíduos, ...), os quais são medidos segundo ‘p’ variáveis, obter um algoritmo que possibilite reunir os indivíduos, tal que exista homogeneidade dentro do grupo e heterogeneidade entre grupos (Regazzi, 2000).”



Método K-médias

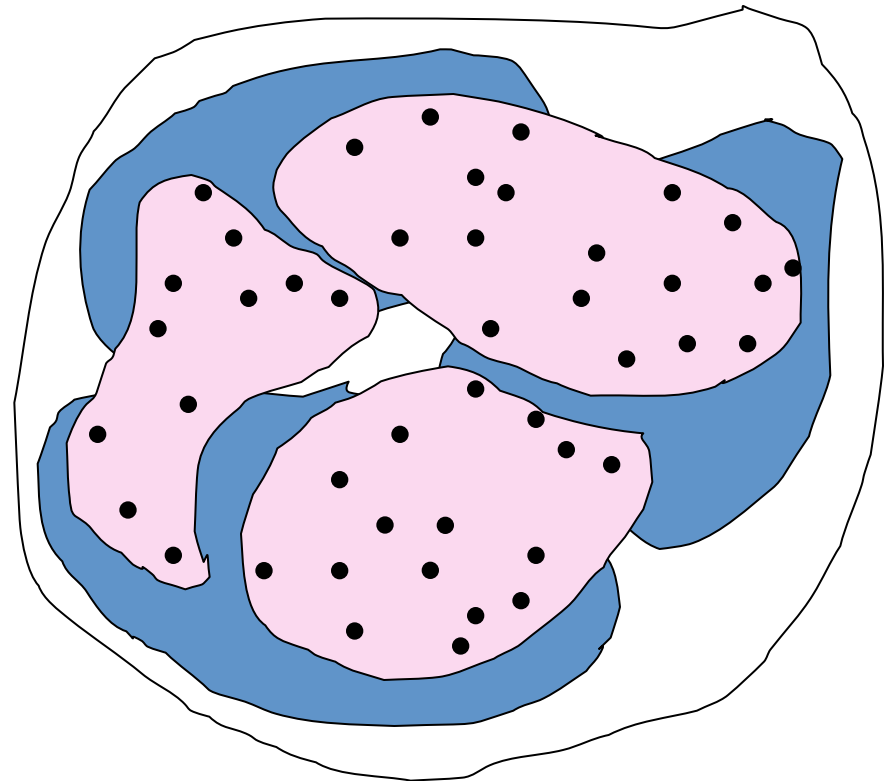
- Também chamado de método do encadeamento simples “single linkage method”.
- Neste método calcula-se a matriz de distâncias entre os ‘n’ indivíduos da população, em seguida os indivíduos mais próximos são agrupados.



Método K-médias

- BD com n amostras
- K = número de clusters desejado (parâmetro)
- $K \leq n$

Exemplo com $k=3$

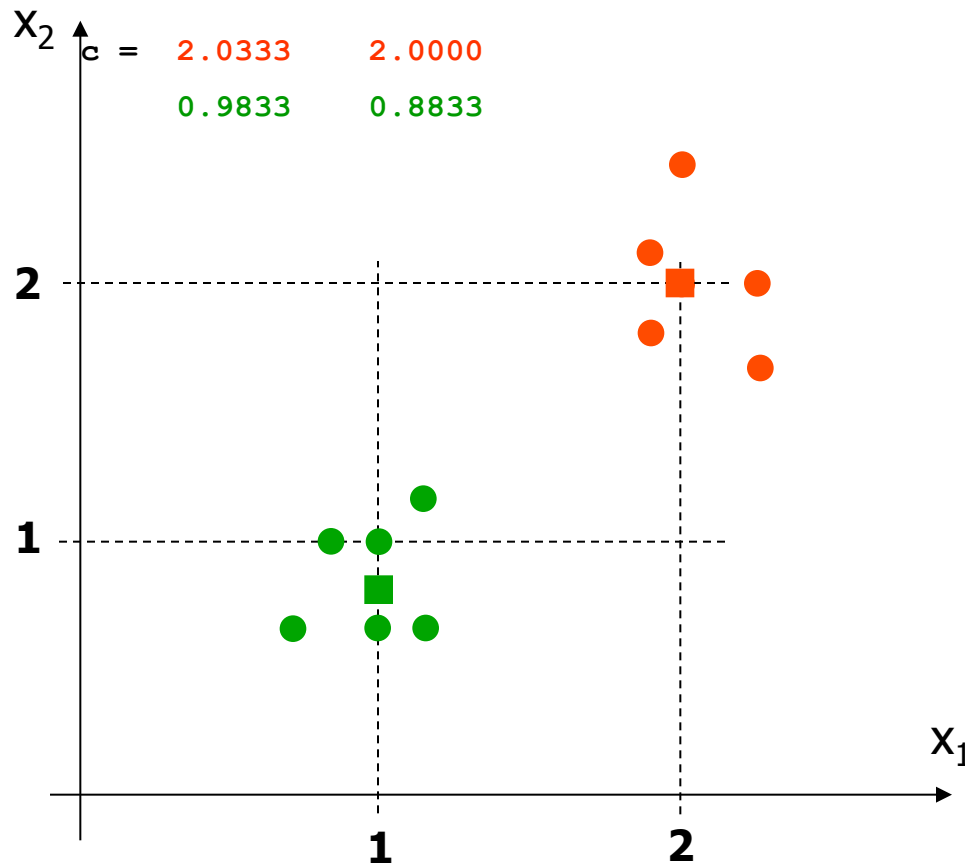


Método K-médias

Exemplo

Treinamento

K-means



Dados:

1.0	1.0
1.0	0.7
1.2	0.7
1.2	1.2
0.8	1.0
0.7	0.7
2.0	2.0
1.8	1.8
1.8	2.1
2.3	2.0
2.3	1.7
2.0	2.4

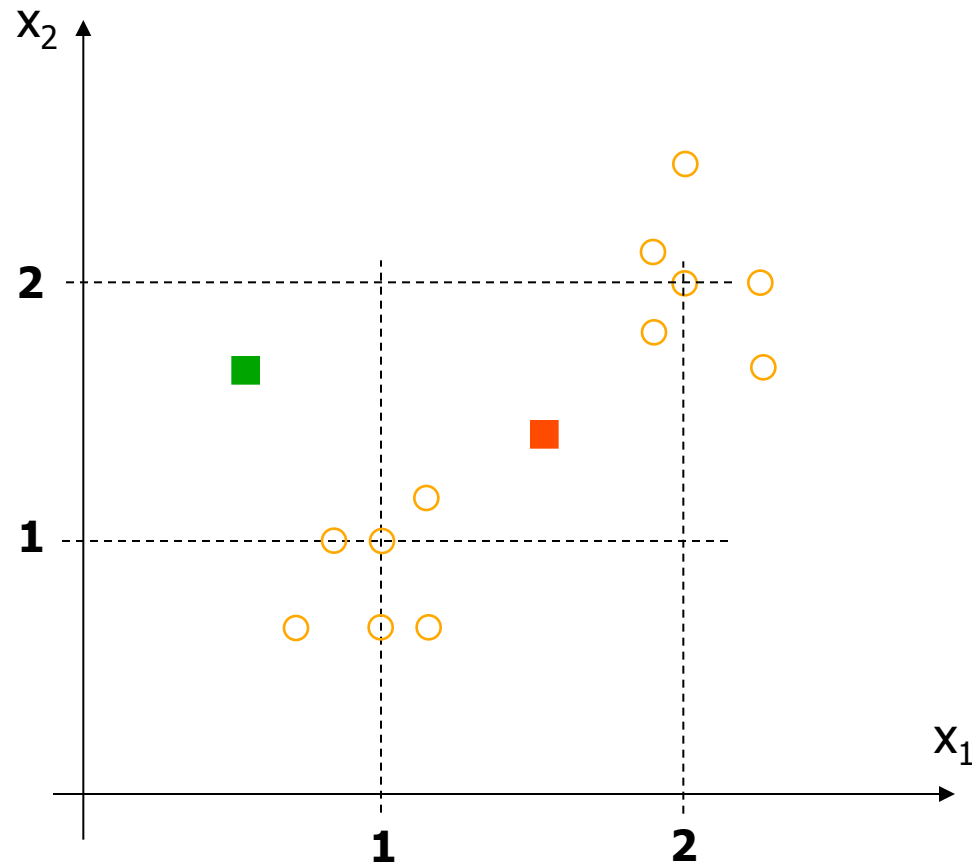


Método K-médias

Exemplo

Treinamento

K-means

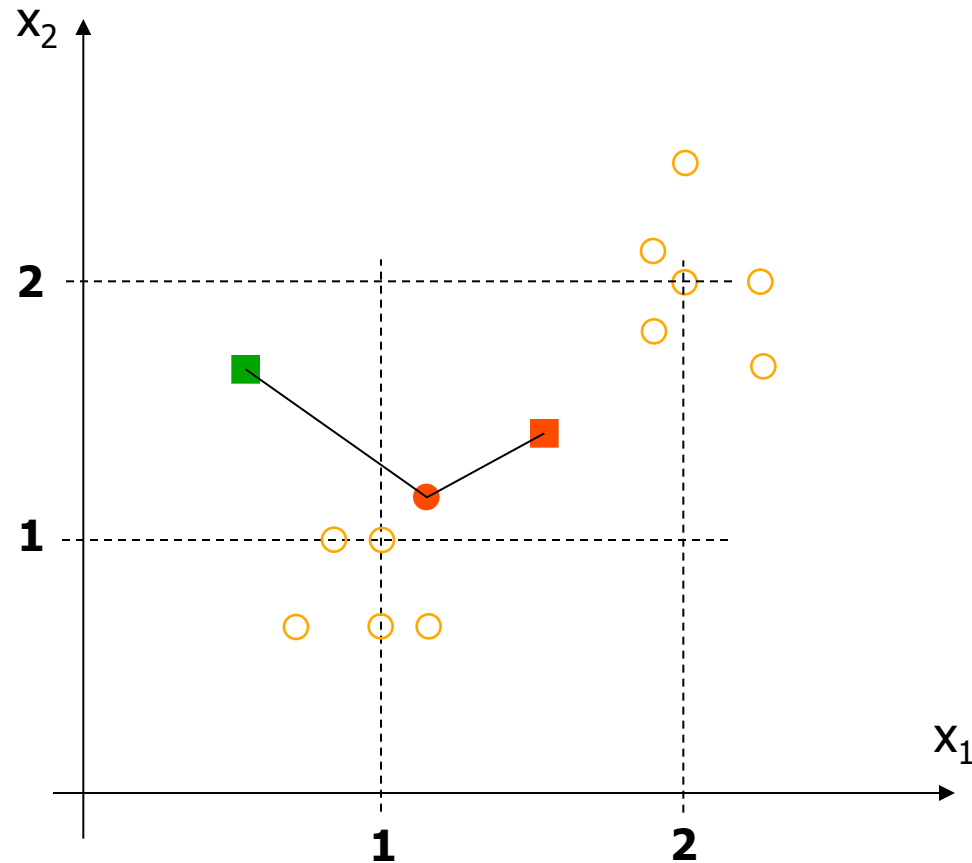


Método K-médias

Exemplo

Treinamento

K-means

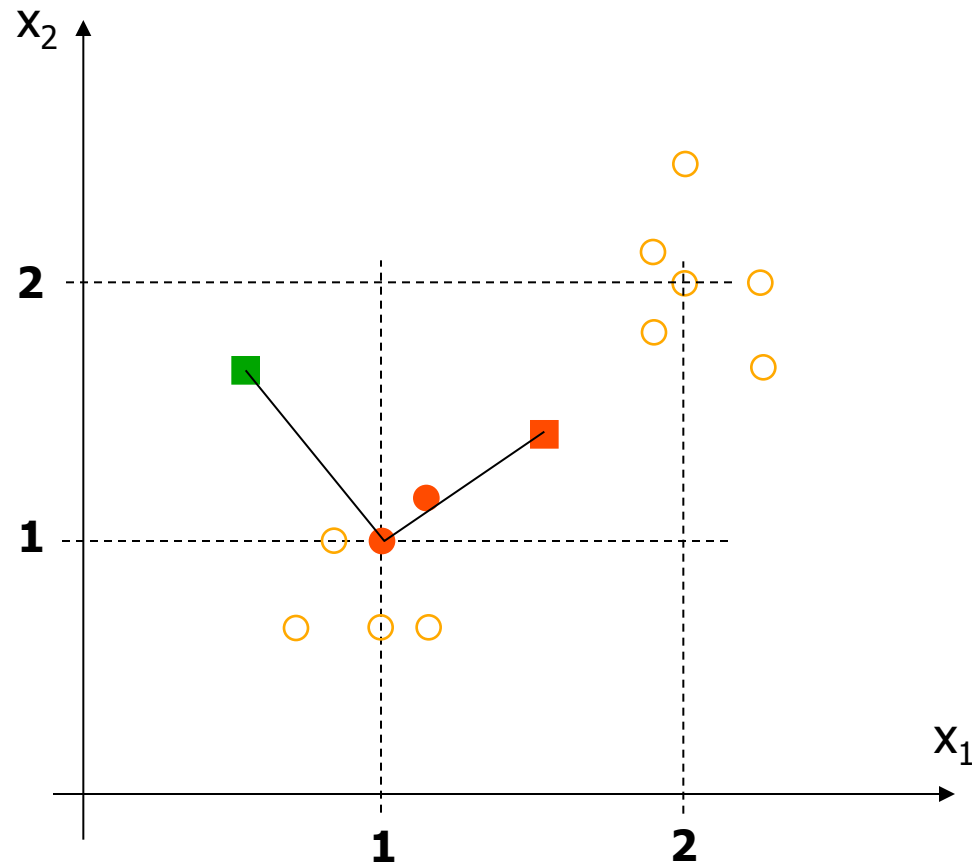


Método K-médias

Exemplo

Treinamento

K-means

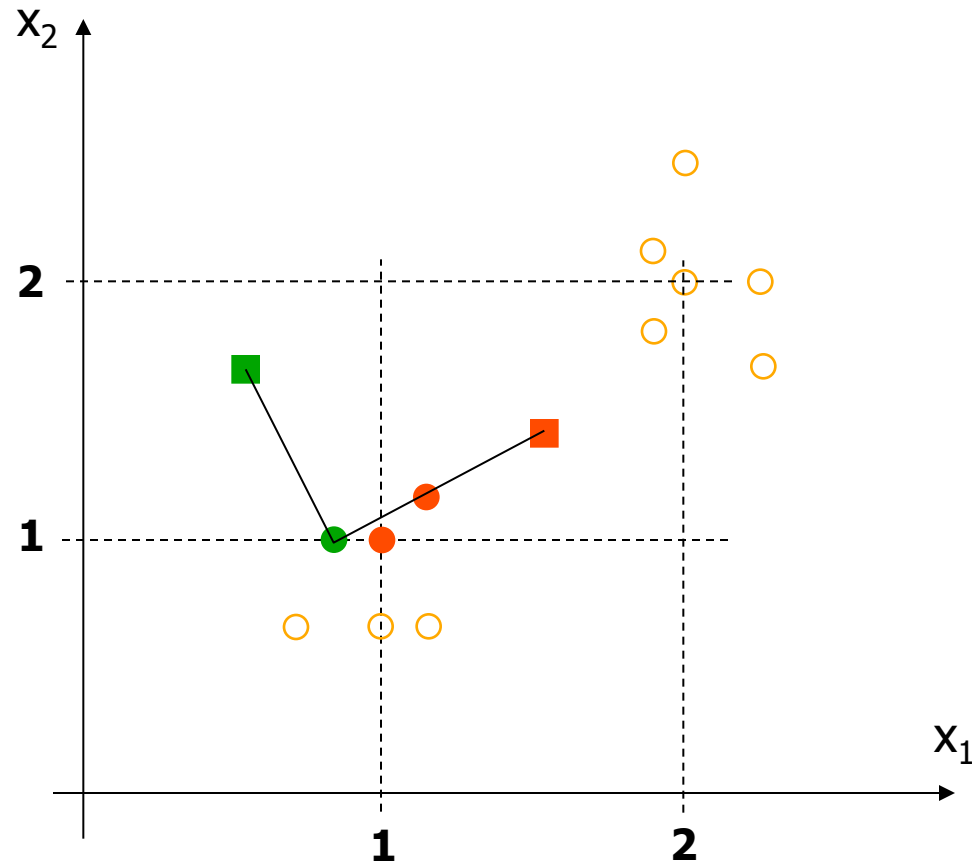


Método K-médias

Exemplo

Treinamento

K-means

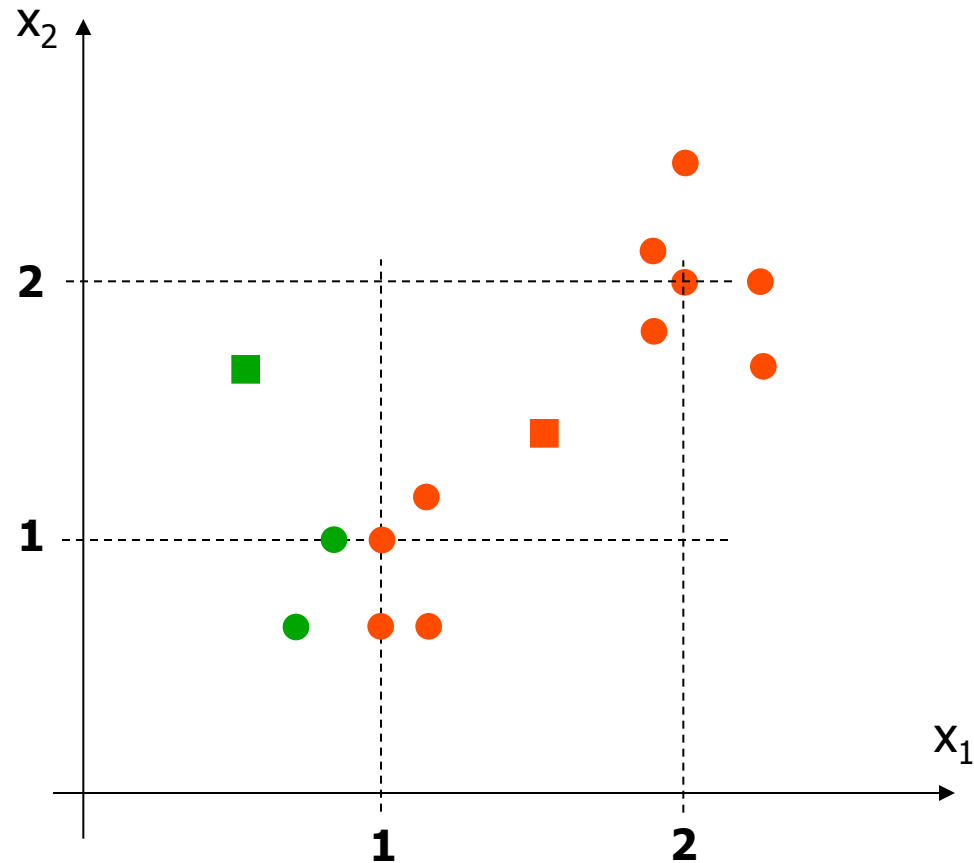


Método K-médias

Exemplo

Treinamento

K-means

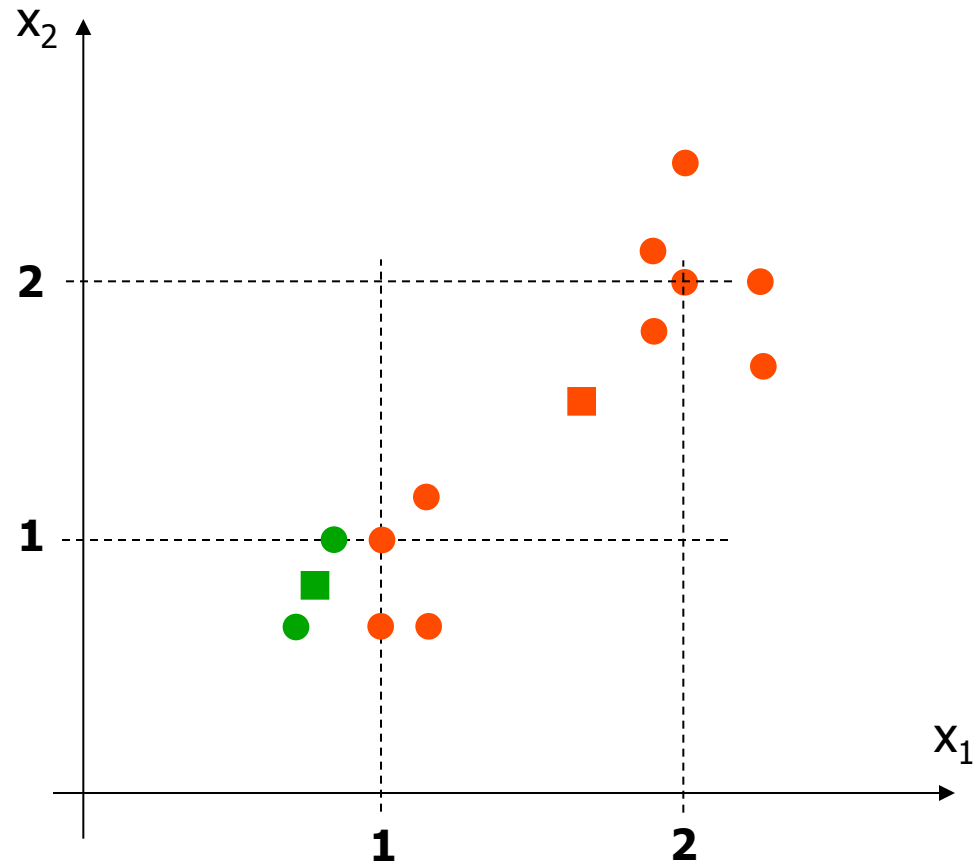


Método K-médias

Exemplo

Treinamento

K-means

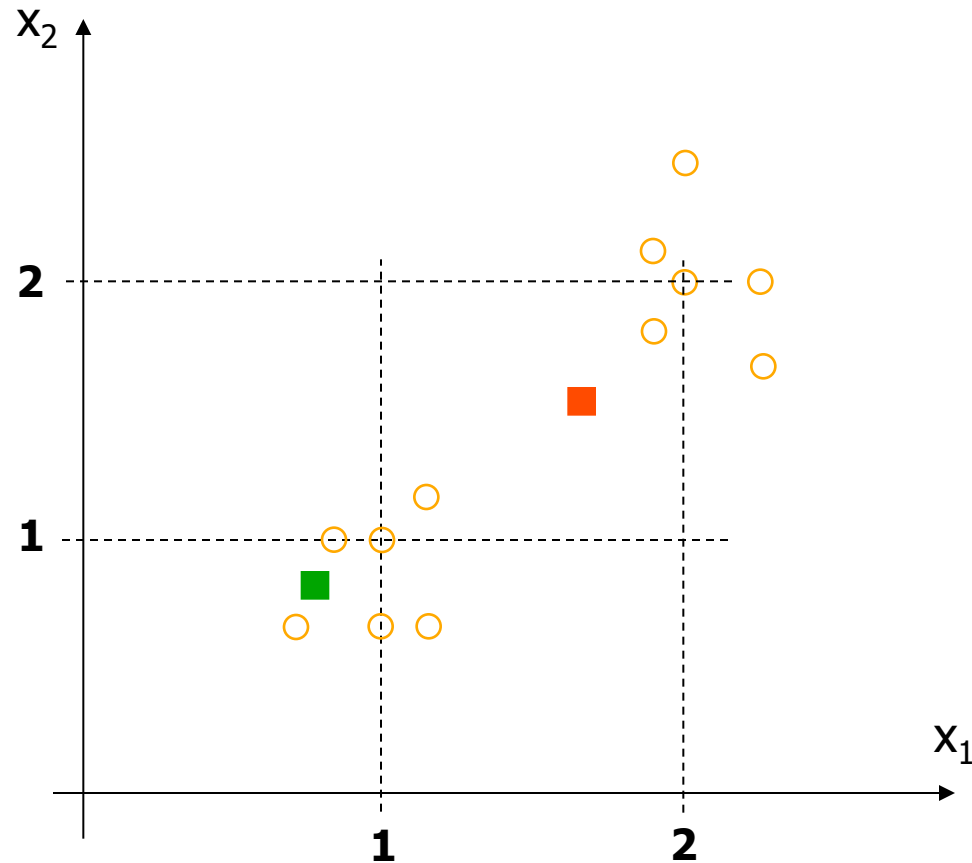


Método K-médias

Exemplo

Treinamento

K-means

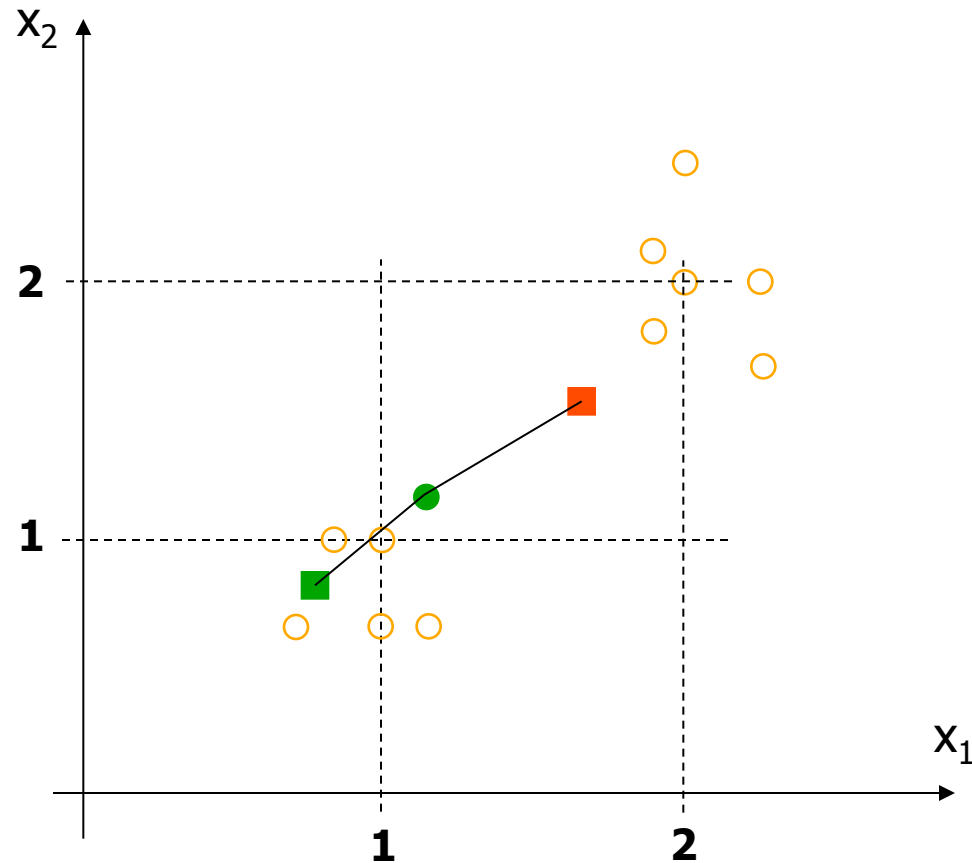


Método K-médias

Exemplo

Treinamento

K-means

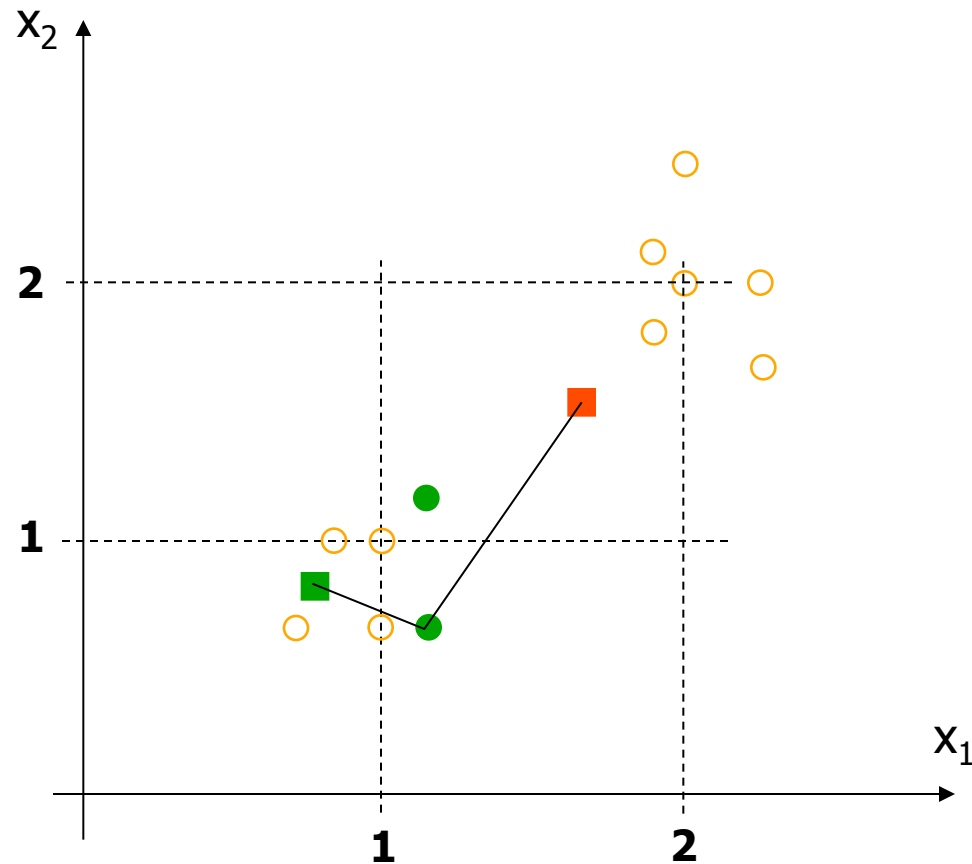


Método K-médias

Exemplo

Treinamento

K-means

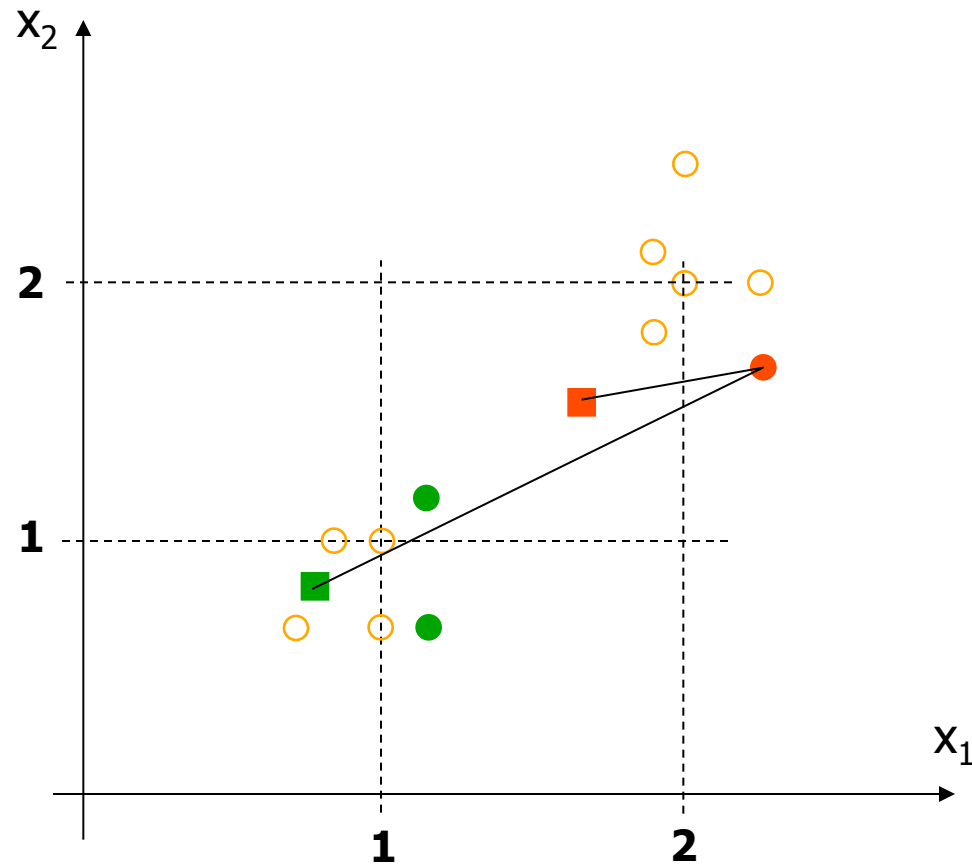


Método K-médias

Exemplo

Treinamento

K-means

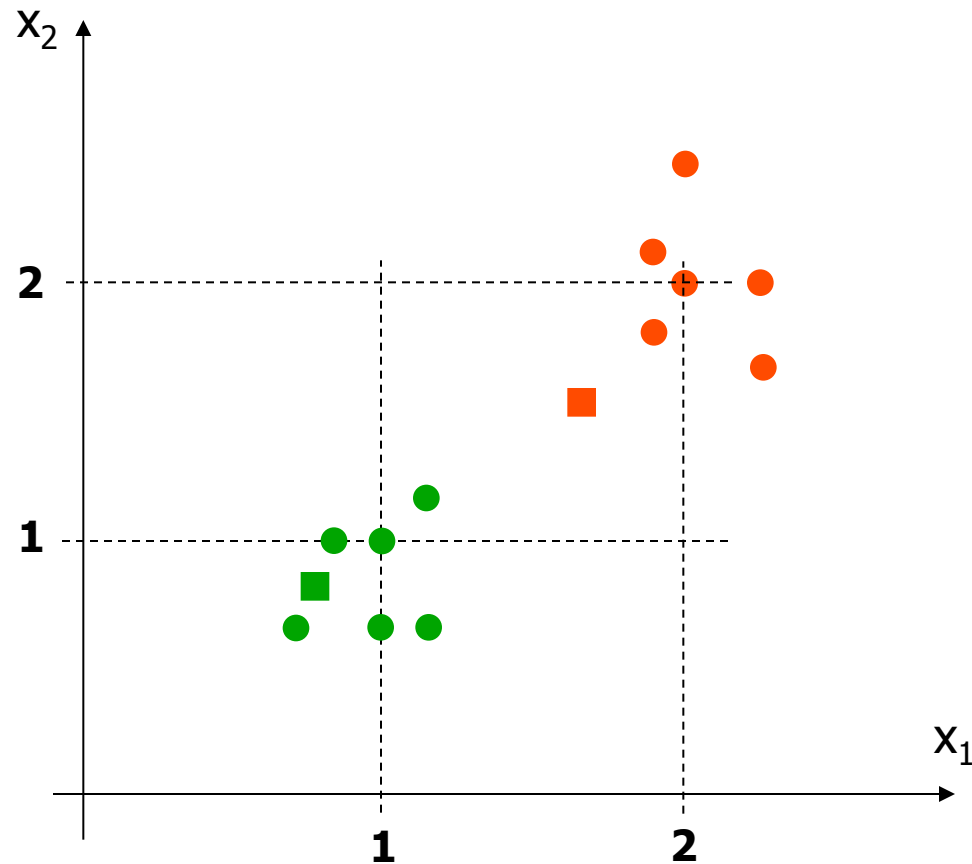


Método K-médias

Exemplo

Treinamento

K-means

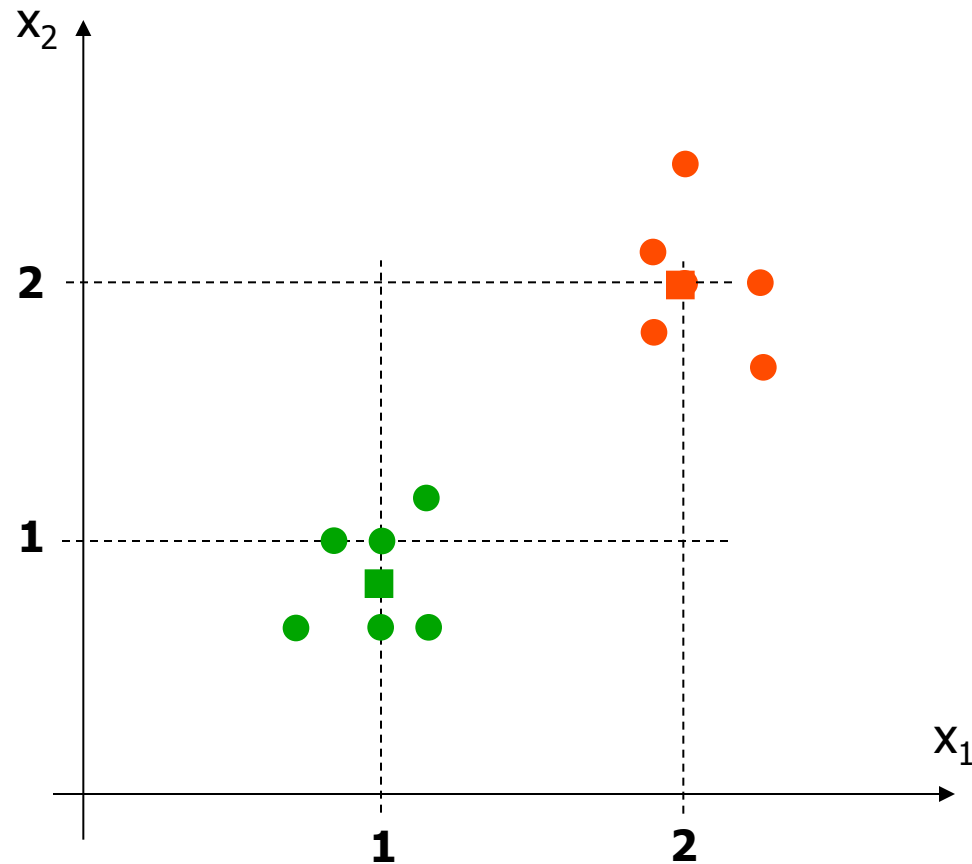


Método K-médias

Exemplo

Treinamento

K-means

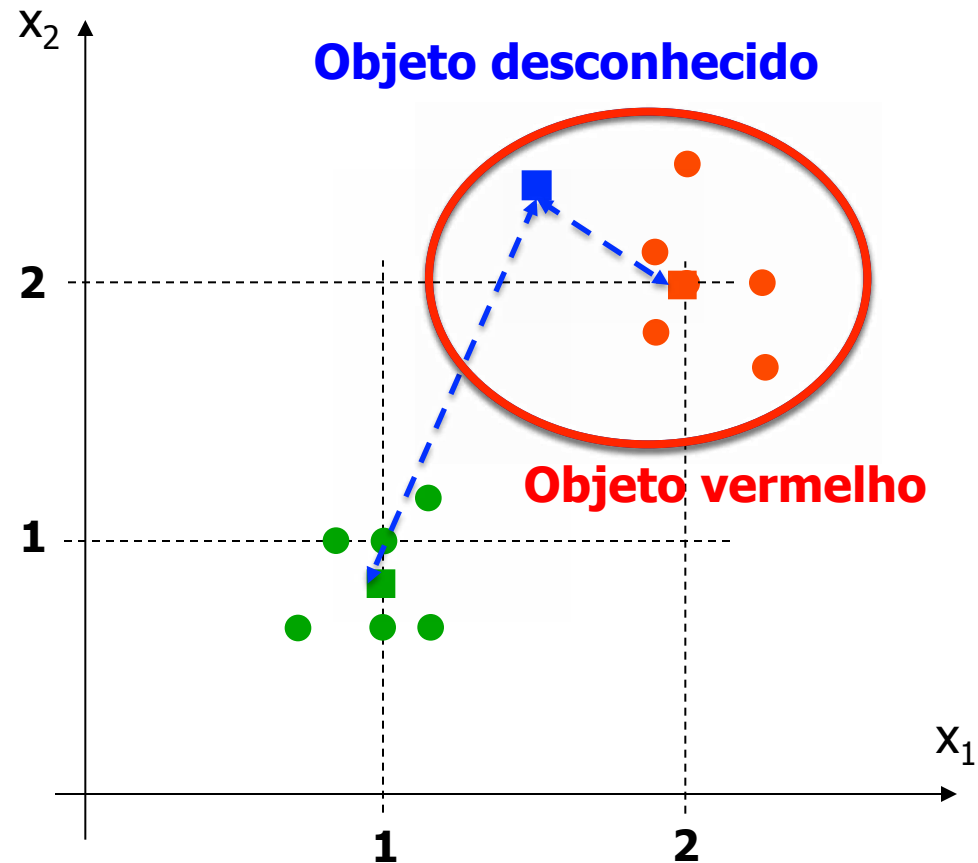


Método K-médias

Exemplo

K-means

Reconhecimento



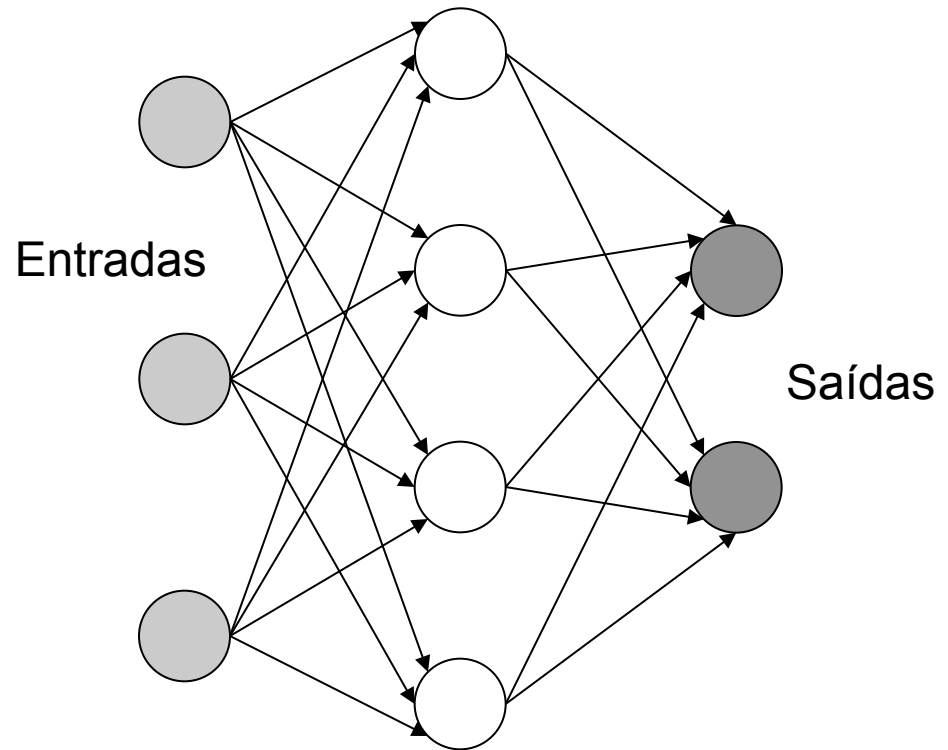
Classificação Supervisionada

Redes Neurais Artificiais (RNA)

Reconhecimento de Padrões

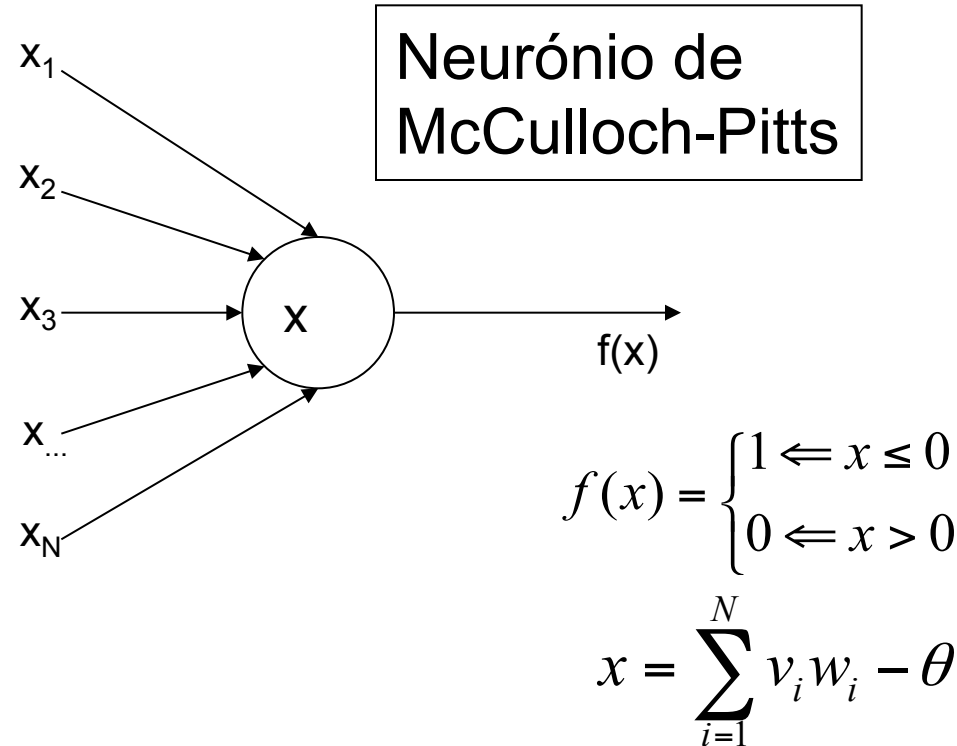
Redes neuronais

- Constituídas por elementos simples.
 - Neurónios.
- Elevado grau de conectividade.
 - Sistema complexo!
- Inspiração biológica.
 - Cérebro humano.



Neurônios - Perceptron

- N entradas.
- 1 saída.
- Faz um a soma ponderada das entradas.
 - Pesos
- Limiar de disparo.
 - *Threshold*



Feed-forward networks

- Fase de Treino
 - Rede ‘alimentada’ com dados pré-annotados.
 - Auto-aprendizagem dos pesos.
 - Auto-aprendizagem do limiar de decisão.
- *Backpropagation*
- Classificação
 - Entrada: *Vector de características* desconhecido.
 - Saídas: Neurónios disparam caso a classificação seja positiva.

Custo computacional elevado!

- Muitas outras formas de treinamento surgiram depois.
- Muitas outras redes surgiram em seguida: RBF, SVM, OPF, entre outros.



Classificação Supervisionada

Classificador Bayesiano

Reconhecimento de Padrões

Classificador Bayesiano

- Baseia-se em conceitos de probabilidade e estatística
- Utiliza a regra de Bayes

$$P(y = Y_k | \mathbf{x}) = \frac{P(\mathbf{x} | y = Y_k) P(y = Y_k)}{\sum_{j=1}^K P(\mathbf{x} | y = Y_j) P(y = Y_j)}$$

Diagram illustrating the components of Bayes' theorem:

- probabilidade a posteriori** (posterior probability) points to $P(y = Y_k | \mathbf{x})$.
- probabilidade condicional** (conditional probability) points to $P(\mathbf{x} | y = Y_k)$.
- probabilidade a priori** (prior probability) points to $P(y = Y_k)$.
- fator de normalização** (normalization factor) points to the denominator $\sum_{j=1}^K P(\mathbf{x} | y = Y_j) P(y = Y_j)$.

- Probabilidade **a priori**
 - Frequência de amostras de uma classe
- Probabilidade **condicional**
 - Função de Densidade de Probabilidade (FDP) dos dados



Classificador Bayesiano

- As distribuições podem ser estimadas utilizando os dados de treinamento.
- A classe \mathbf{y} atribuída a nova mostra \mathbf{x} é dada por

$$\hat{y} = \arg \max_{Y_k} P(y = Y_k | \mathbf{x}), \quad k \in \{1, \dots, K\}$$

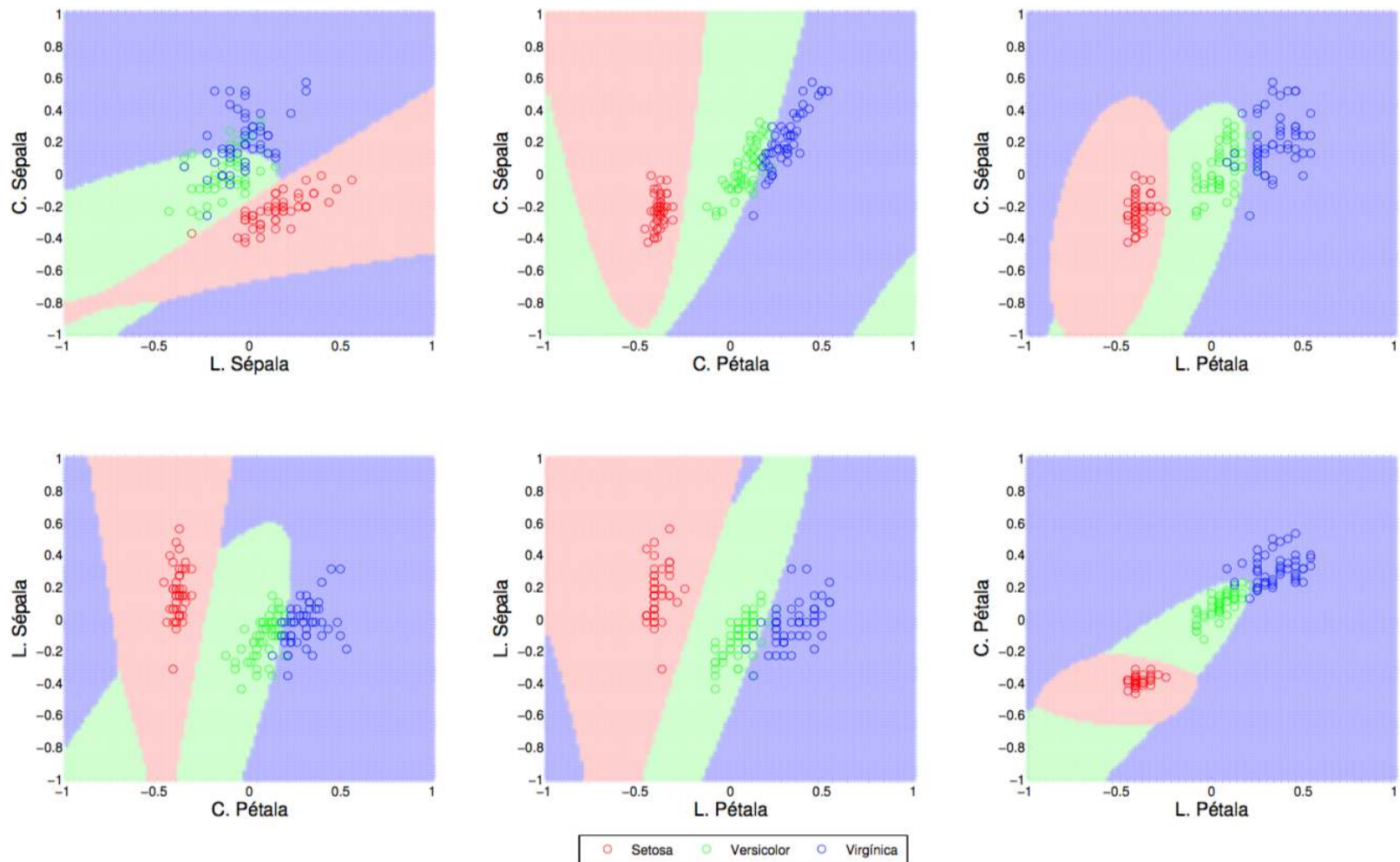
- Uma das FDPs mais utilizadas é a Gaussiana ou Normal

$$P(\mathbf{x} | y = Y_k) = \frac{1}{(2\pi)^{\frac{D}{2}} |\boldsymbol{\Sigma}_k|^{\frac{1}{2}}} \exp \left\{ -\frac{1}{2} (\mathbf{x} - \boldsymbol{\mu}_k)^T \boldsymbol{\Sigma}_k^{-1} (\mathbf{x} - \boldsymbol{\mu}_k) \right\}$$

- $\boldsymbol{\mu}_k$ (vetor de média), $\boldsymbol{\Sigma}_k$ (matriz de covariância), $|\boldsymbol{\Sigma}_k|$ e $\boldsymbol{\Sigma}_k^{-1}$ (determinante e inversa)



Classificador Bayesiano



Superfície de decisão do classificador Bayesiano na base de dados Iris



Classificação Supervisionada

Máquina de Vetores de Suporte

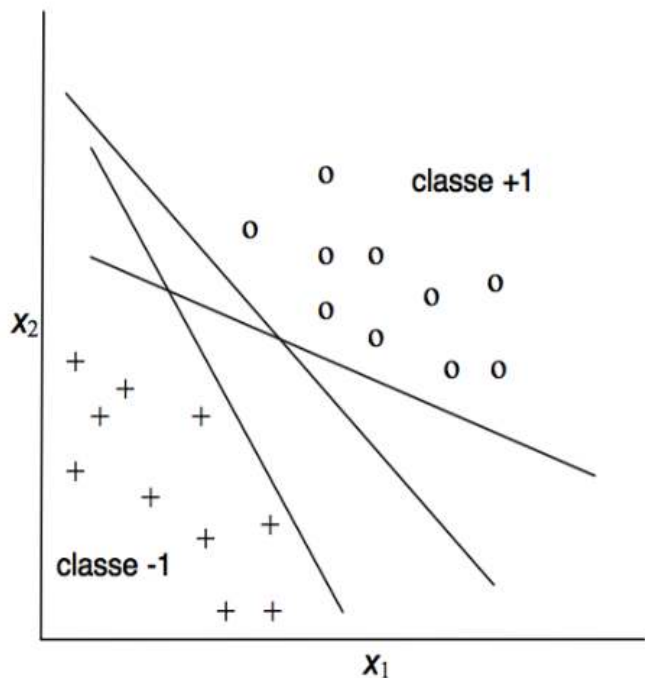
Reconhecimento de Padrões

Máquina de Vetores de Suporte

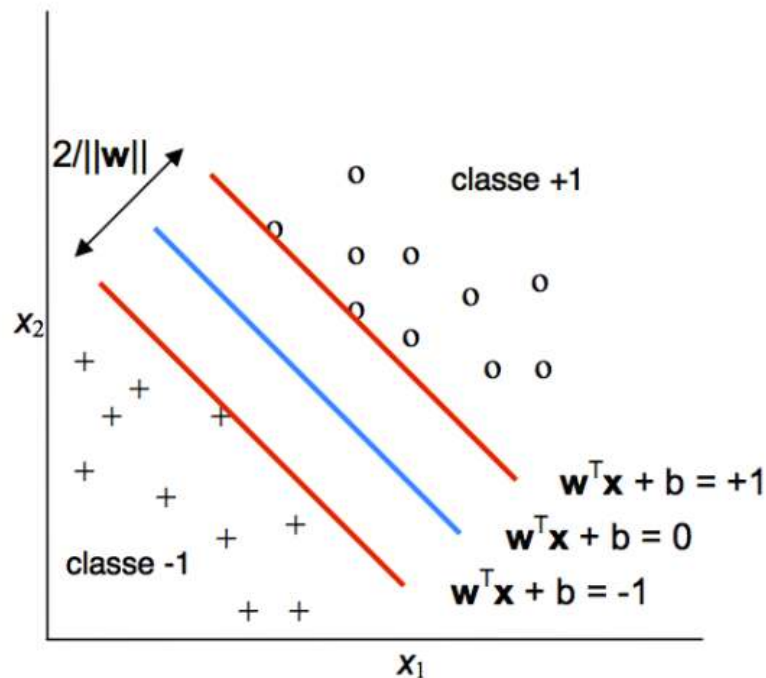
- *Support Vector Machines* (SVMs)
- Baseia-se na Teoria de Aprendizado Estatístico, desenvolvida por Vapnik (1998)
 - Série de princípios para obtenção de classificadores com boa generalização
- Técnicas tradicionais minimizam o risco empírico (erro no treino)
- SVMs minimizam **também o risco estrutural** (erro no teste)
- Buscam um hiperplano de separação ótima



Máquina de Vetores de Suporte



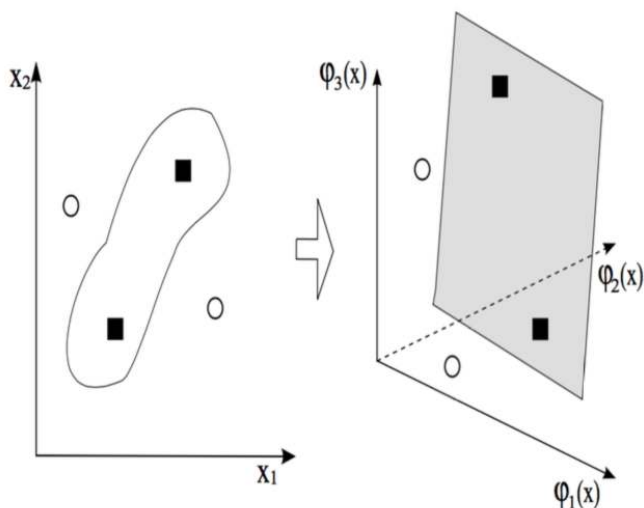
(a) Diversos hiperplanos capazes de resolver o problema de classificação binária.



(b) Margem (vermelho) e hiperplano (azul) de separação ótima.

Máquina de Vetores de Suporte

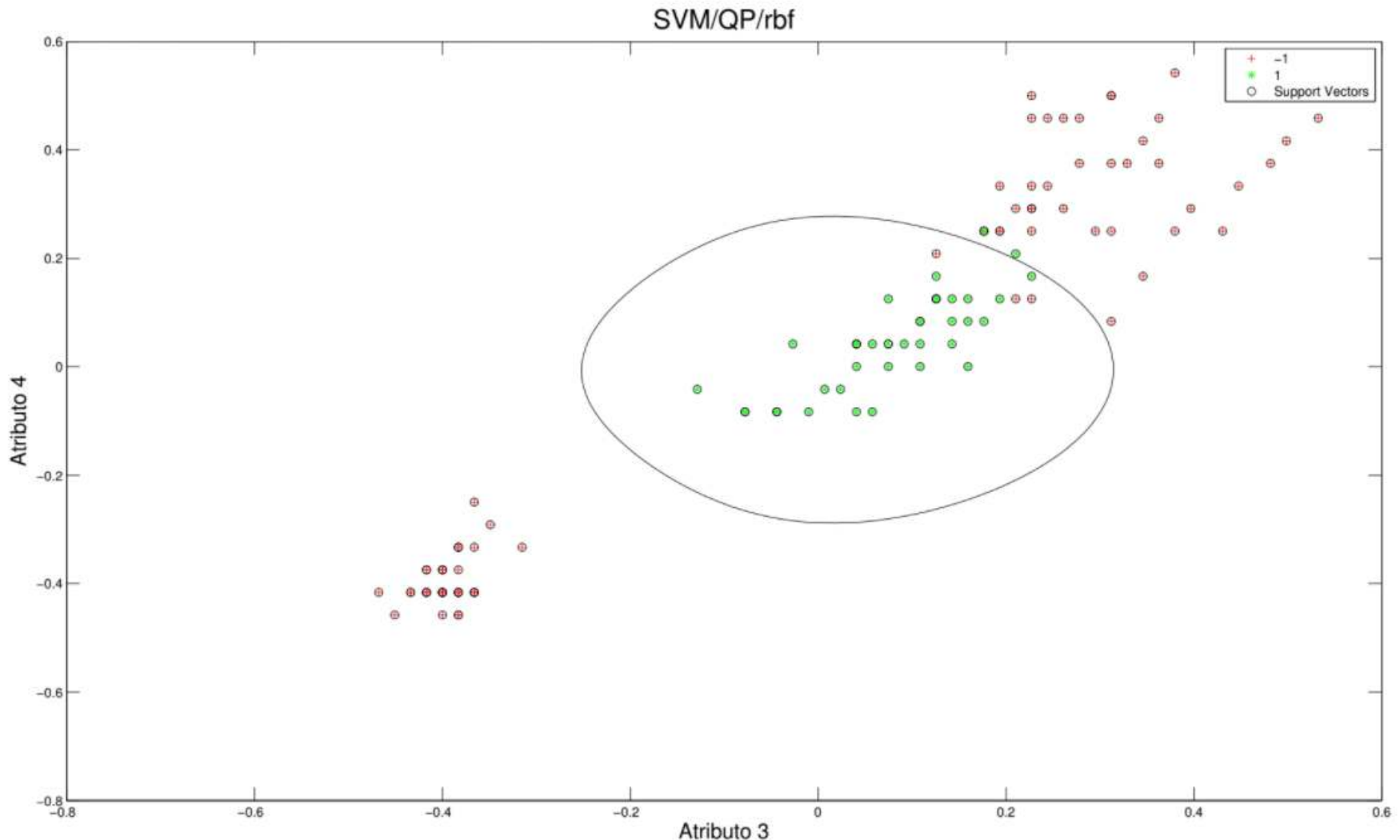
- O SVM é proposto inicialmente para problemas lineares
- Truque do *kernel*
 - Mapea o espaço dos dados para outro onde podem ser linearmente separáveis



Mapeamento de dados para um espaço de características de mais alta dimensão através do truque do *kernel* (ROCHA NETO, 2011).

<i>Kernel</i>	Descrição
Linear	$K(\mathbf{x}, \mathbf{x}_i) = \mathbf{x}_i^T \cdot \mathbf{x}$
Polinomial	$K(\mathbf{x}, \mathbf{x}_i) = (\mathbf{x}_i^T \cdot \mathbf{x} + 1)^d$, em que d é o grau do polinômio.
Gaussiano (RBF)	$K(\mathbf{x}, \mathbf{x}_i) = \exp\{-\ \mathbf{x} - \mathbf{x}_i\ ^2/\sigma^2\}$, em que σ é uma constante.

Máquina de Vetores de Suporte



Avaliação de classificadores

Reconhecimento de Padrões

Avaliação de Classificadores

- True Positive (TP) = hit
- True Negative (TN) = correct rejection
- False Positive (FP) = false alarm, Type I error
- False Negative (FN) = miss, Type II error
- Sensitivity or True Positive Rate (TPR) = hit rate, recall : $TPR = TP / (TP + FN)$
- False Positive Rate (FPR) = false alarm rate, fall-out : $FPR = FP / (TP + FN)$
- Accuracy (ACC) : $ACC = (TP + TN) / (TP + FN + FP + TN)$
- Specificity (SPC) = True Negative Rate : $SPC = TN / (FP + TN) = 1 - FPR$
- Positive Predictive Value (PPV) = Precision : $PPV = TP / (TP + FP)$
- Negative Predictive Value (NPV) : $NPV = TN / (TN + FN)$
- False Discovery Rate (FDR) : $FDR = FP / (FP + TP)$
- Matthews Correlation Coefficient (MCC) :

$$MCC = (TP * TN - FP * FN) / \text{SQRT}((TP + FP) * (TP + FN) * (TN + FP) * (TN + FN))$$



Tabela de Contingência ou Matriz de Confusão

		Valor Verdadeiro	
		P	N
Resultado da Classificação	P	Positivo V	Positivo F
	N	Negativo F	Negativo V



Encaminhamentos

- Dúvidas?
- Próximo assunto
 - Lista 3 disponível hoje
 - Trabalho Final

