

Escola Livre de IA

Inteligência Artificial ao alcance de todos

www.escolalivre-ia.com.br

Algoritmos de Classificação: Visão Geral + KNN

Texto



O que vamos aprender hoje?

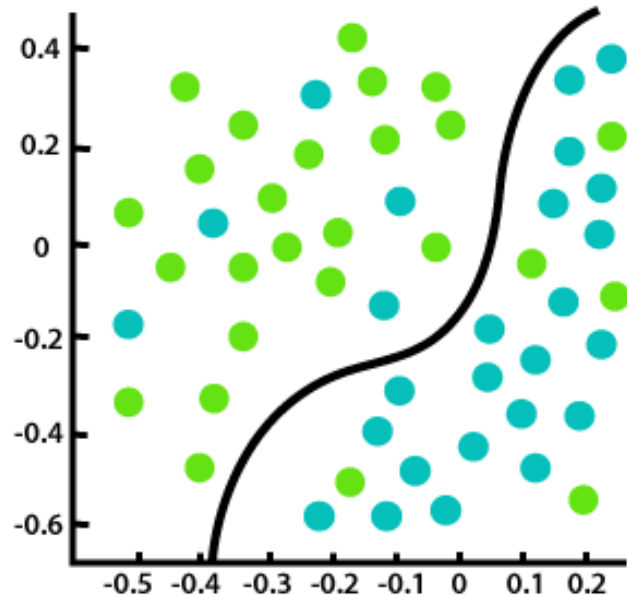
1. O que são algoritmos de classificação?
2. Quais as diferenças entre modelos de classificação e regressão?
3. Quais são os principais tipos de modelos de classificação?
4. Quais métricas usamos para avaliar modelos de classificação?
5. Como funciona o modelo KNN – K Nearest Neighbors



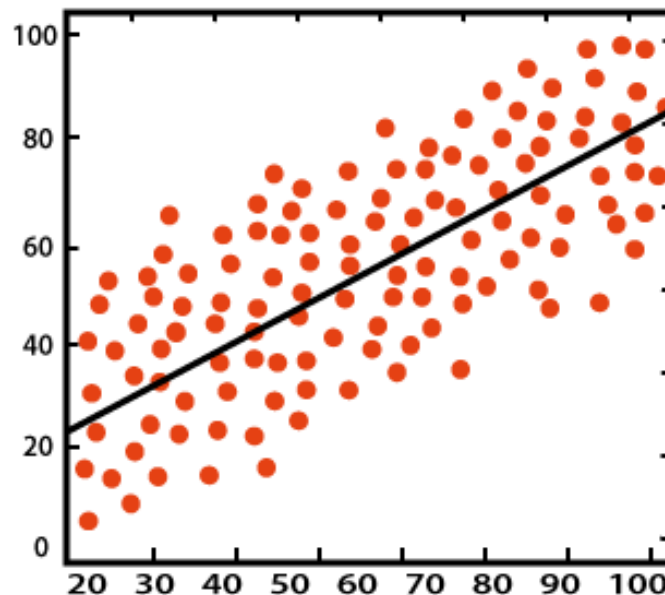


Modelos Classificação

Classificação x Regressão



Classificação

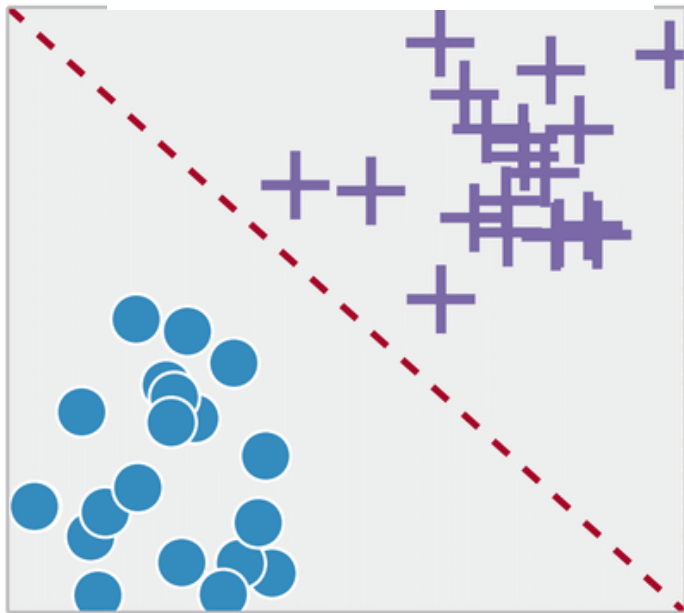


Regressão

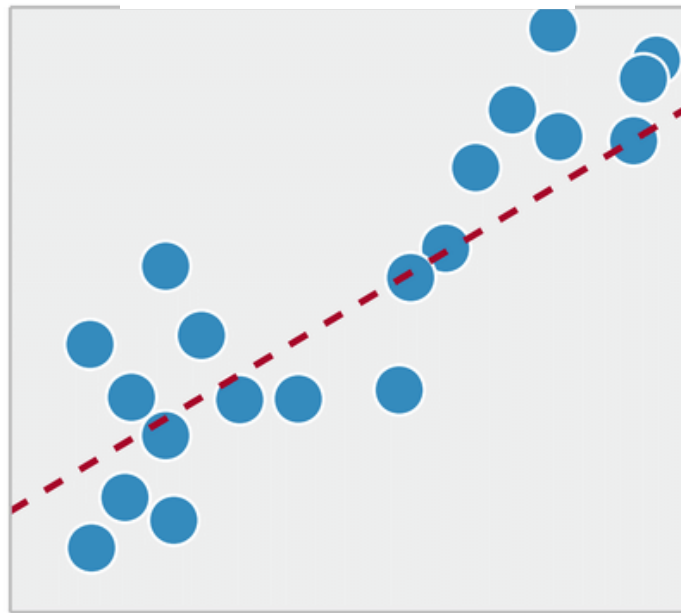


Classificação x Regressão

Classificação



Regressão



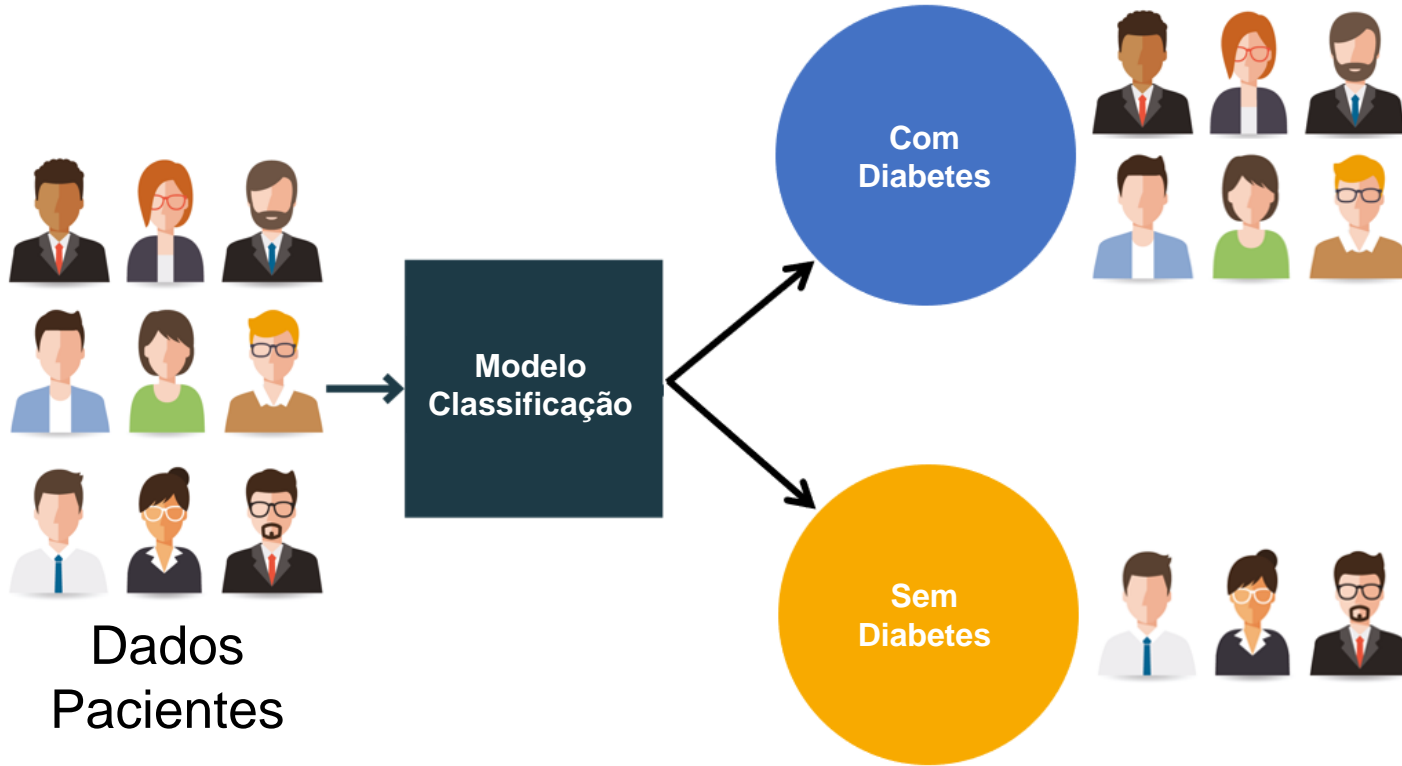
Exemplos Classificação



OU



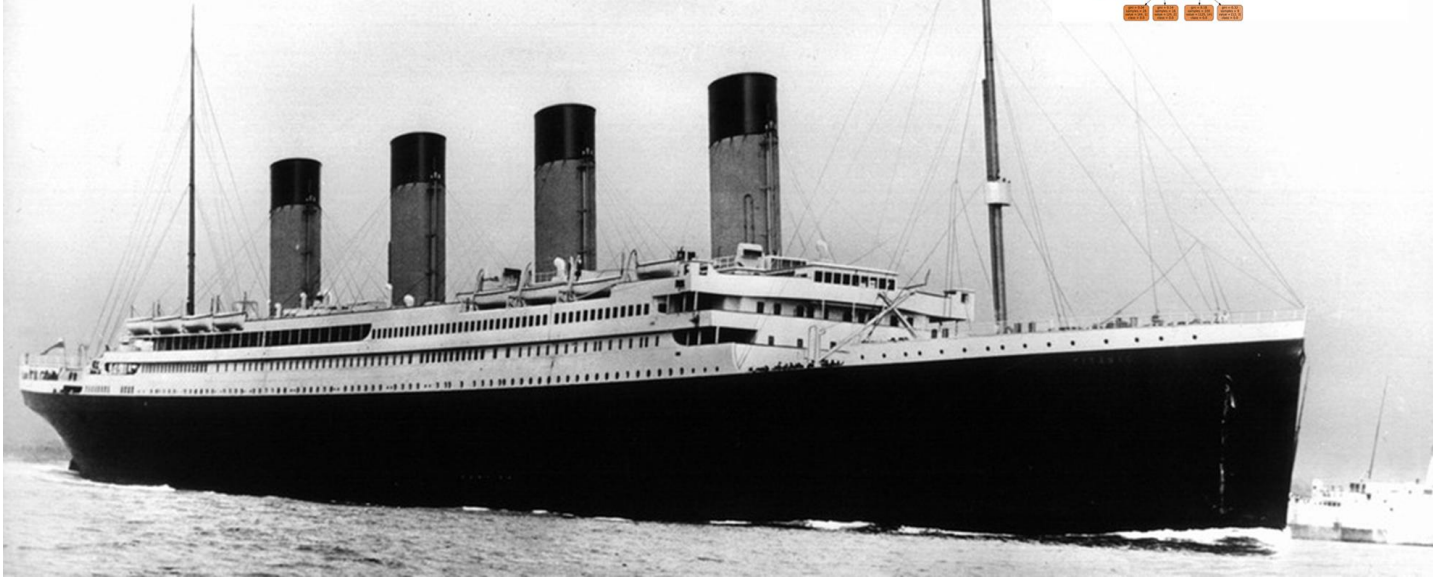
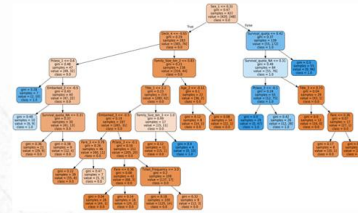
Exemplos Classificação



Exemplos Classificação

Machine Learning with the Titanic Dataset

An end-to-end guide to predict the Survival of Titanic passenger



<https://towardsdatascience.com/machine-learning-with-the-titanic-dataset-7f6909e58280>



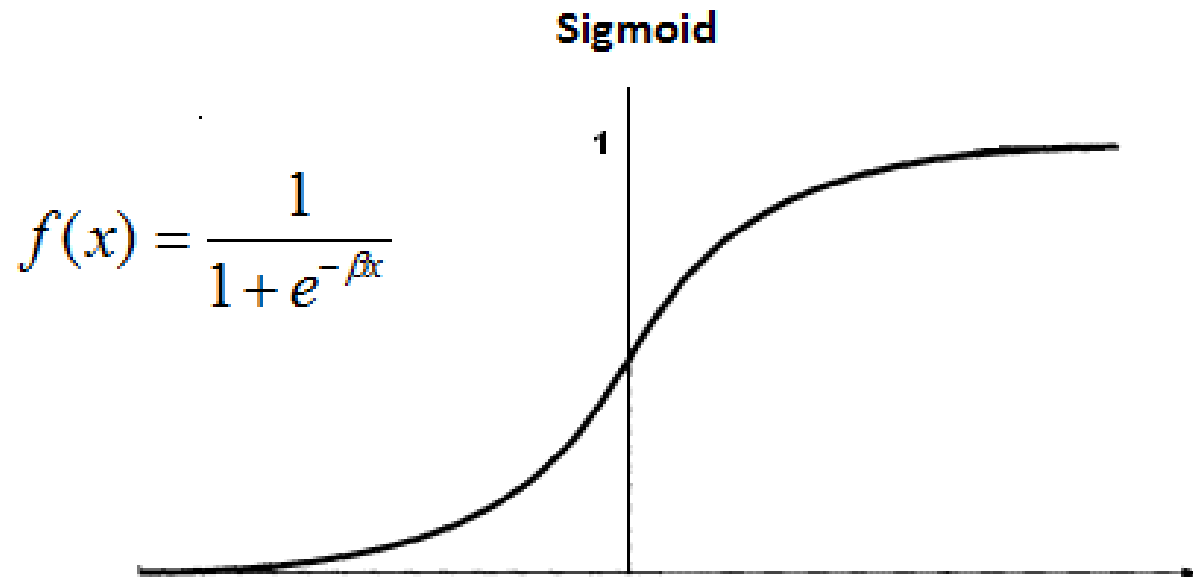


Tipos Modelos Classificação

1. Aprendizado por minimização dos erros (*error-based learning*)
 - **Regressão Logística***
 - **Suport Vector Machines (SVM)**
2. Aprendizado por probabilidades (*probability-based learning*)
 - **Naive Bayes**
3. Aprendizado por ganho de informação (*information-based learning*)
 - **Árvores de Decisão (Decision Trees)**
4. Aprendizado por similaridade (*similarity-based learning*)
 - **K-Nearest Neighbors (KNN)**



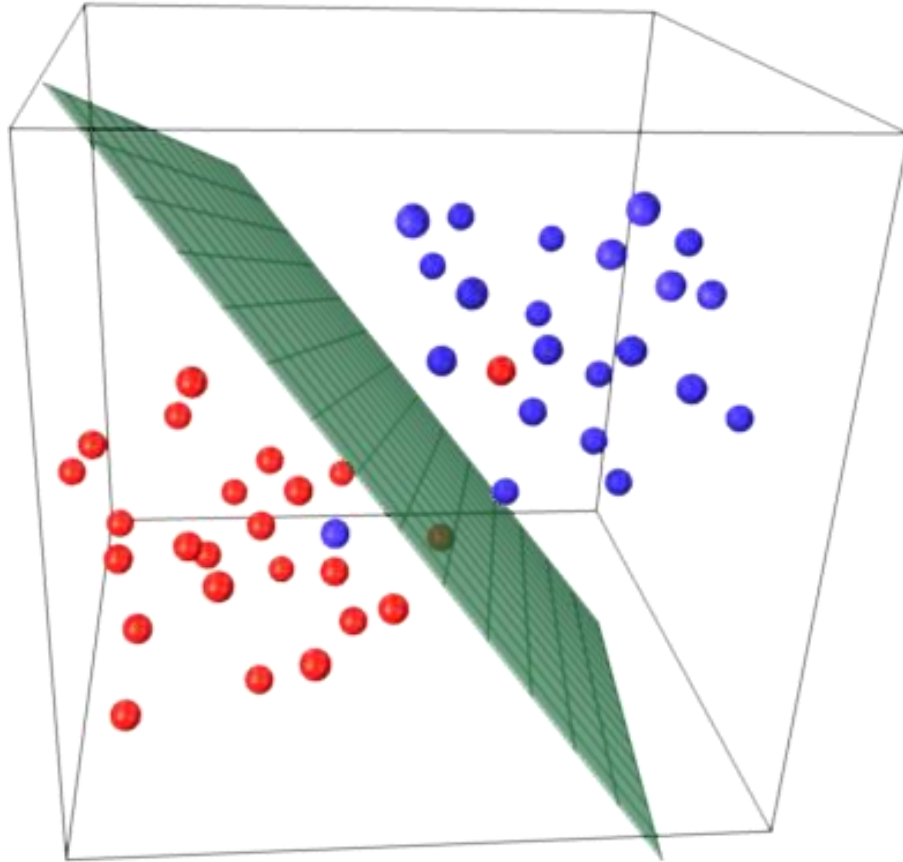
Regressão Logística*



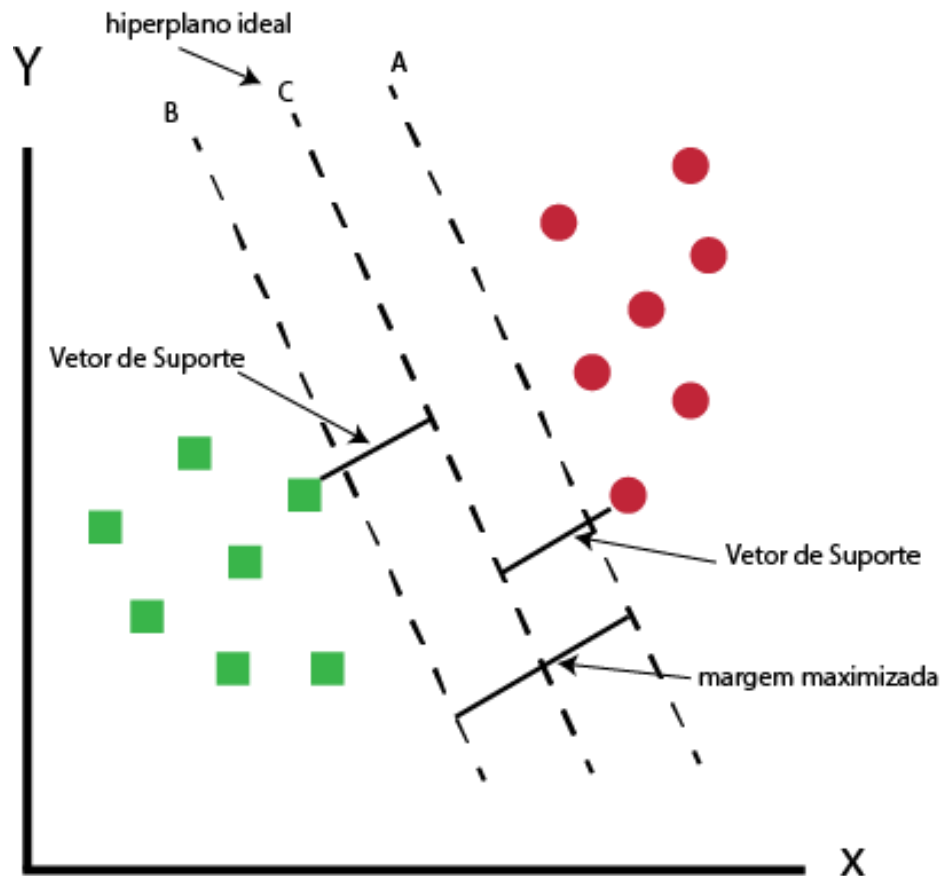
*a rigor não é um classificador, mas pode ser utilizado para classificação se utilizando de um limiar (*threshold*) para determinar a qual classe pertence.



SVM – Support Vector Machines



SVM – Support Vector Machines



Classificador Naive Bayes



Thomas Bayes
1702 - 1761

$$P(c | x) = \frac{P(x | c)P(c)}{P(x)}$$

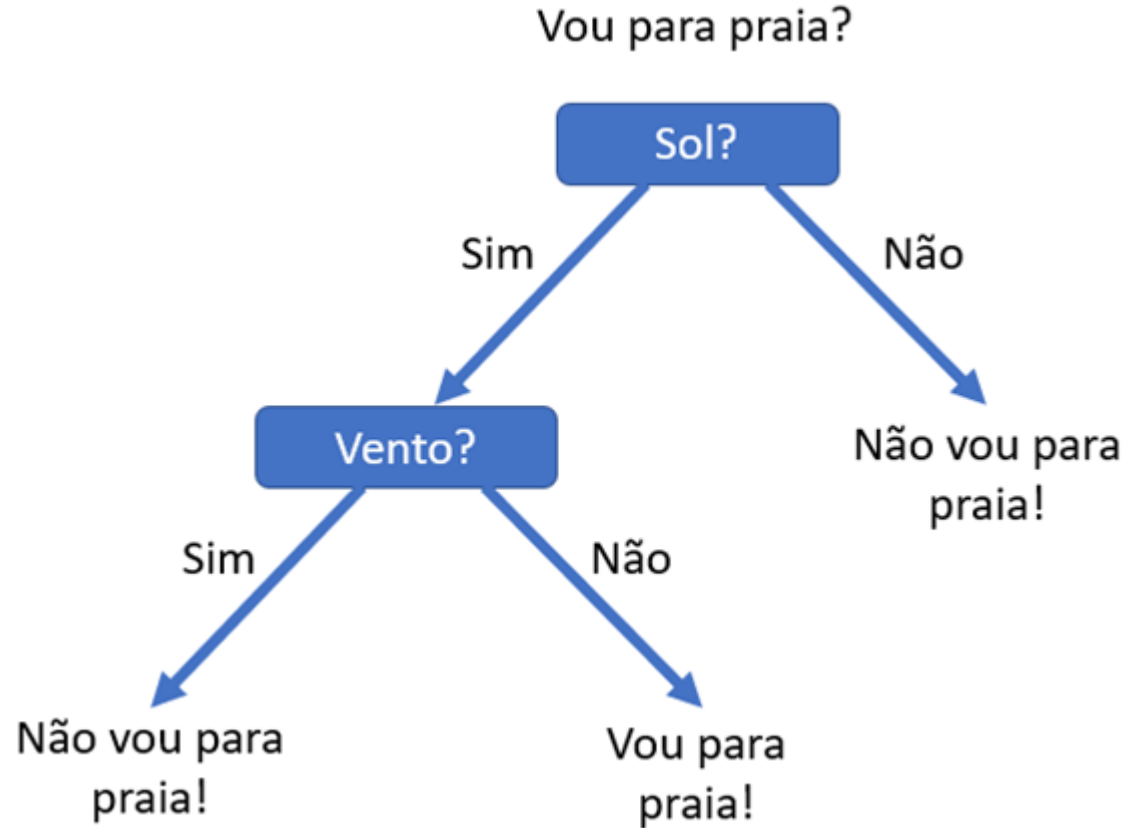
Diagram illustrating the components of the Naive Bayes formula:

- $P(c | x)$ is labeled "Probabilidade posterior" (Posterior Probability).
- $P(x | c)$ is labeled "Probabilidade" (Likelihood).
- $P(c)$ is labeled "Probabilidade original da Classe" (Prior Probability).
- $P(x)$ is labeled "Preditor da probabilidade posterior" (Predictor of the posterior probability).

$$P(c | X) = P(x_1 | c) \times P(x_2 | c) \times \dots \times P(x_n | c) \times P(c)$$



Árvores de Decisão



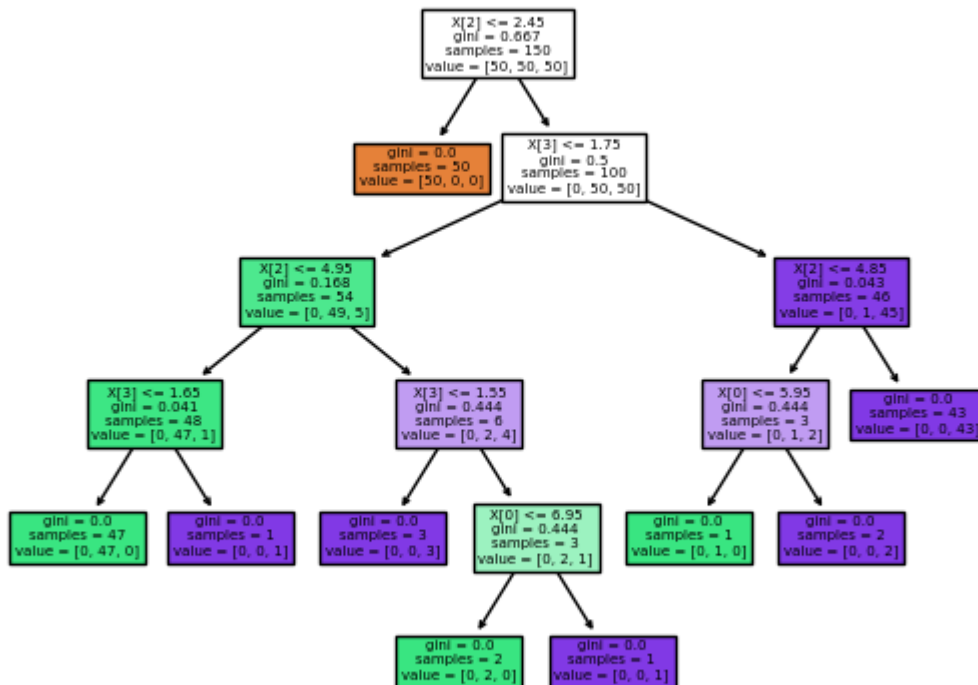
Thomas Bayes
1702 - 1761



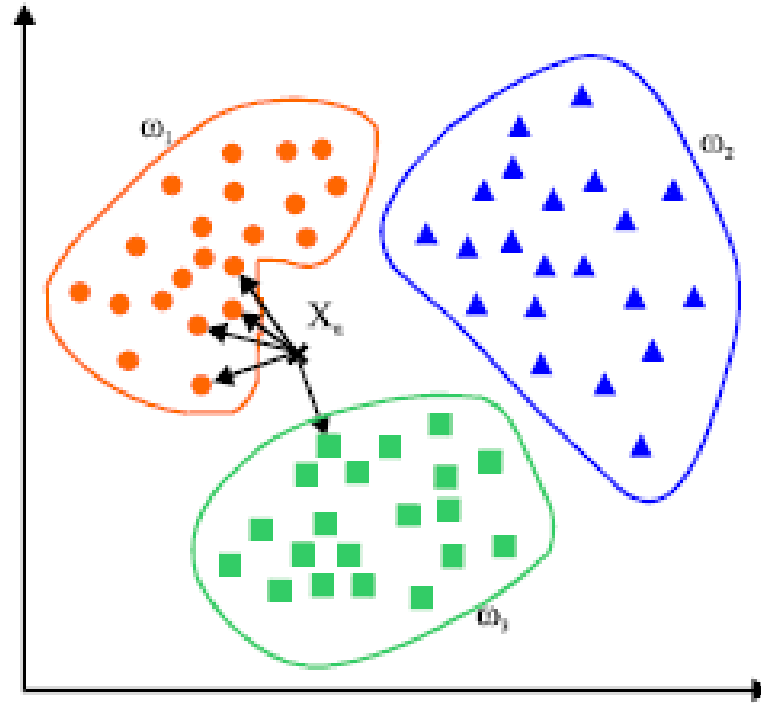
Árvores de Decisão



Thomas Bayes
1702 - 1761



K-Nearest Neighbors



1. Aprendizado por minimização dos erros (*error-based learning*)
 - **Regressão Logística***
 - **Suport Vector Machines (SVM)**
2. Aprendizado por probabilidades (*probability-based learning*)
 - **Naive Bayes**
3. Aprendizado por ganho de informação (*information-based learning*)
 - **Árvores de Decisão (Decision Trees)**
4. Aprendizado por similaridade (*similarity-based learning*)
 - **K-Nearest Neighbors (KNN)**





Métricas Modelos Classificação

Matriz de Confusão

		P R E D I T O	
		👍 POSITIVO	👎 NEGATIVO
R E A L	👍 POSITIVO	✅ 👍 TP verdadeiro positivo	❌ 👎 FN falso negativo
	👎 NEGATIVO	❌ 👍 FP falso positivo	✅ 👎 TN verdadeiro negativo



Matriz de Confusão

		Classe esperada	
		Gato	Não é gato
Classe prevista	Gato	25 Verdadeiro Positivo	10 Falso Positivo
	Não é gato	25 Falso Negativo	40 Verdadeiro Negativo



Matriz de Confusão

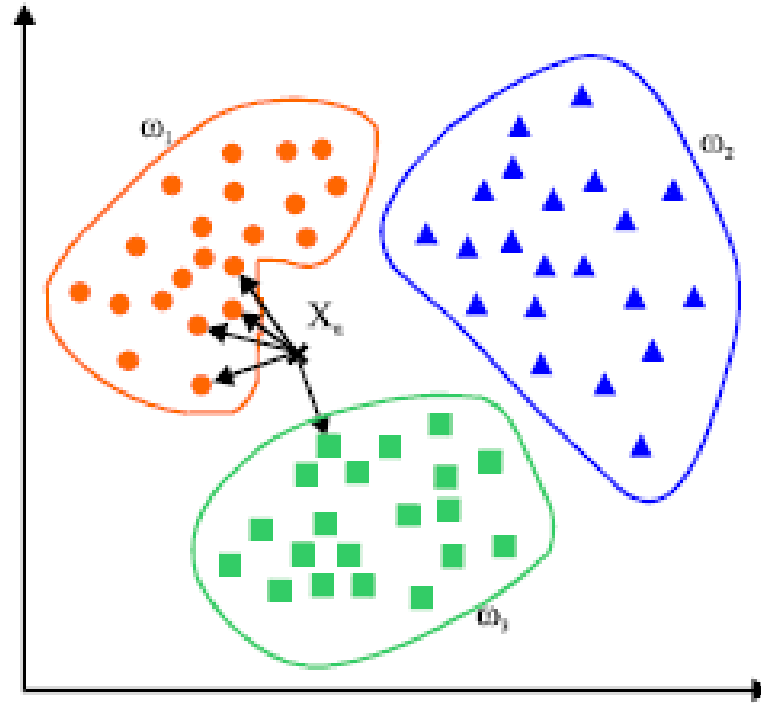
$$\text{Acurácia} = \frac{\checkmark \text{👍 TP} + \checkmark \text{👎 TN}}{\checkmark \text{👍 TP} + \checkmark \text{👎 TN} + \text{✗} \text{👍 FP} + \text{✗} \text{👎 FN}}$$





KNN – K-Nearest Neighbors

K-Nearest Neighbors

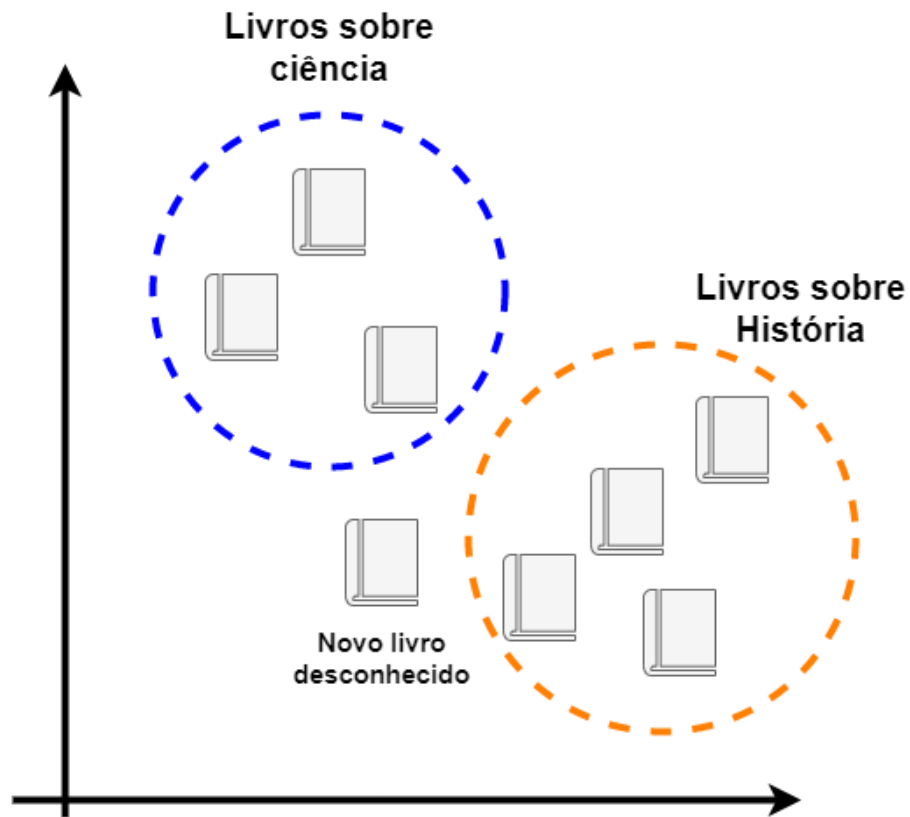


**DIGA-ME COM QUEM ANDAS, QUE EU TE
DIREI QUEM TU ES.**

Diga me com quem tu andas...



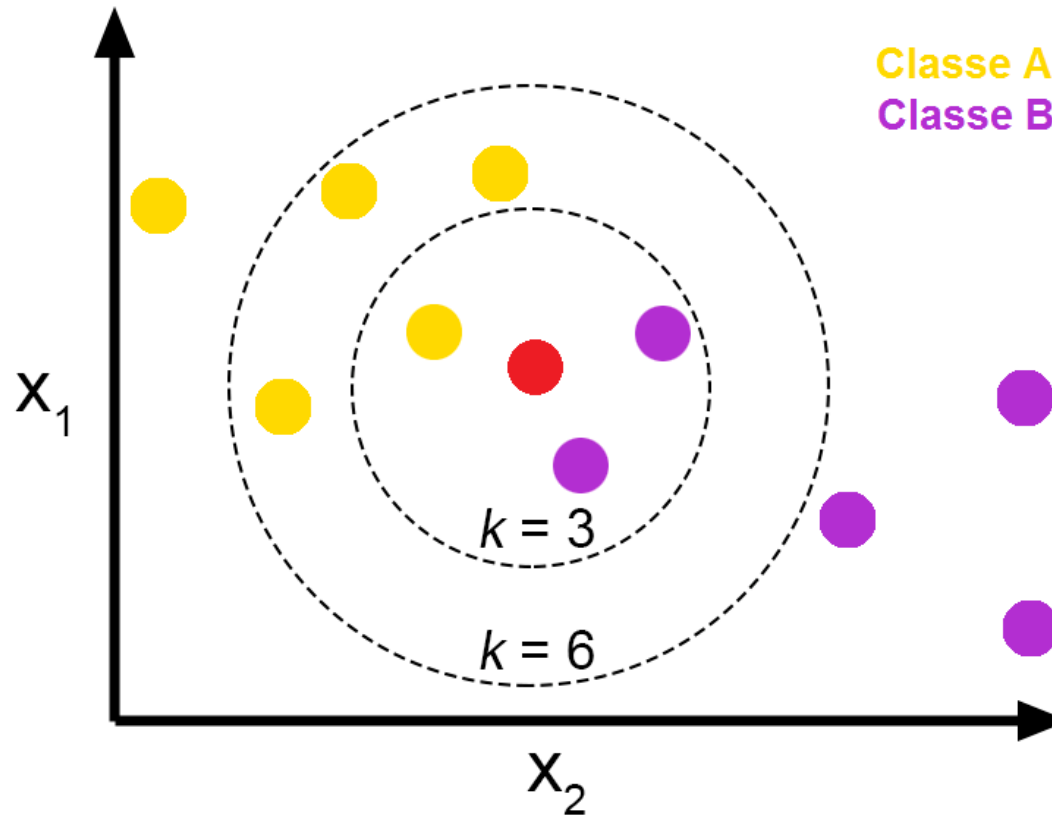
Diga me com quem tu andas...



Algoritmo K-Vizinhos Mais Próximos



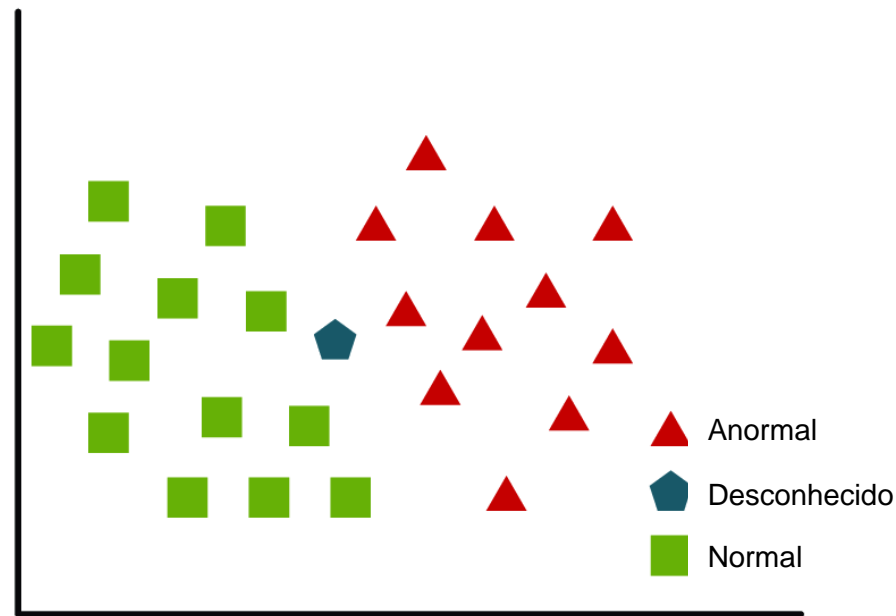
Número K - Quantos vizinhos?



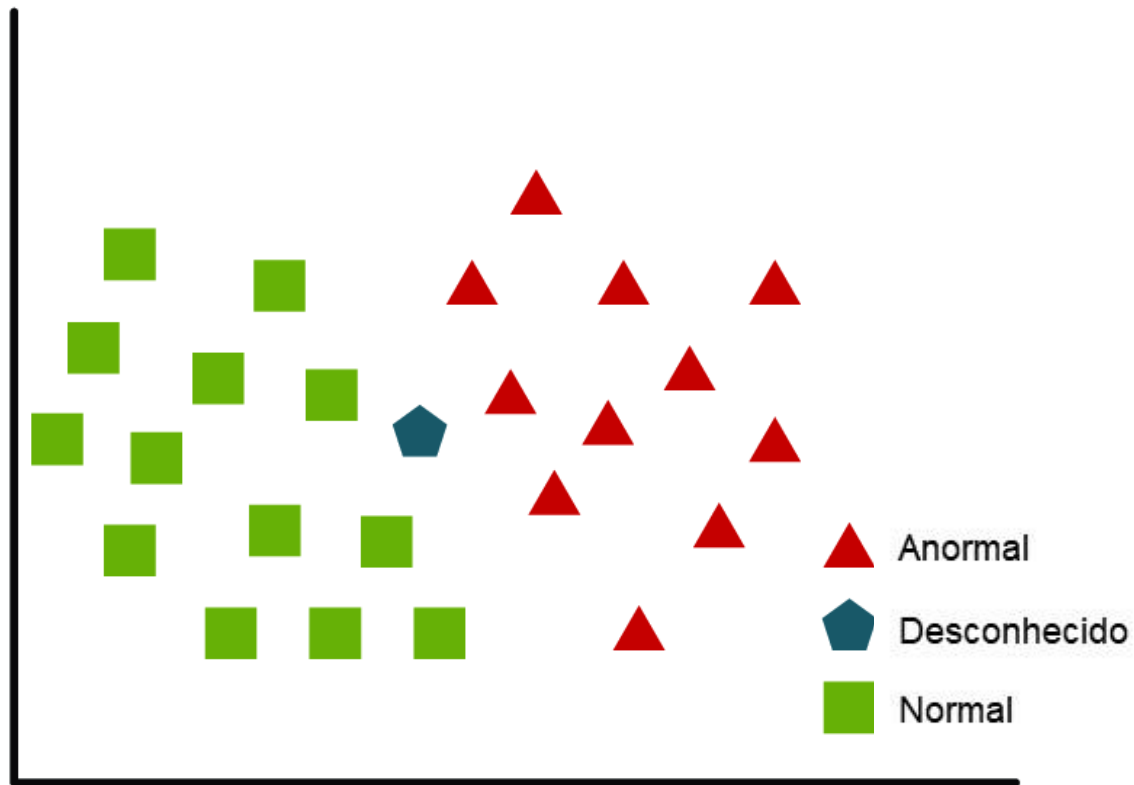
Algoritmo KNN

PARA TODAS NOVAS OBSERVAÇÕES

- PASSO 1: Dado uma nova observação: escolha o número K de vizinhos
- PASSO 2: Selecione os K vizinhos mais próximos, de acordo com a métrica escolhida
- PASSO 3: Destes K vizinhos, conte a quantidade de cada categoria (target)
- PASSO 4: Atribua à nova observação a categoria com maior quantidade no ponto anterior



Algoritmo KNN



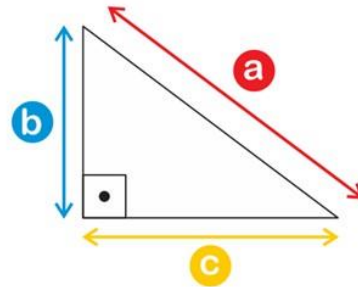
Como calculamos a distância?



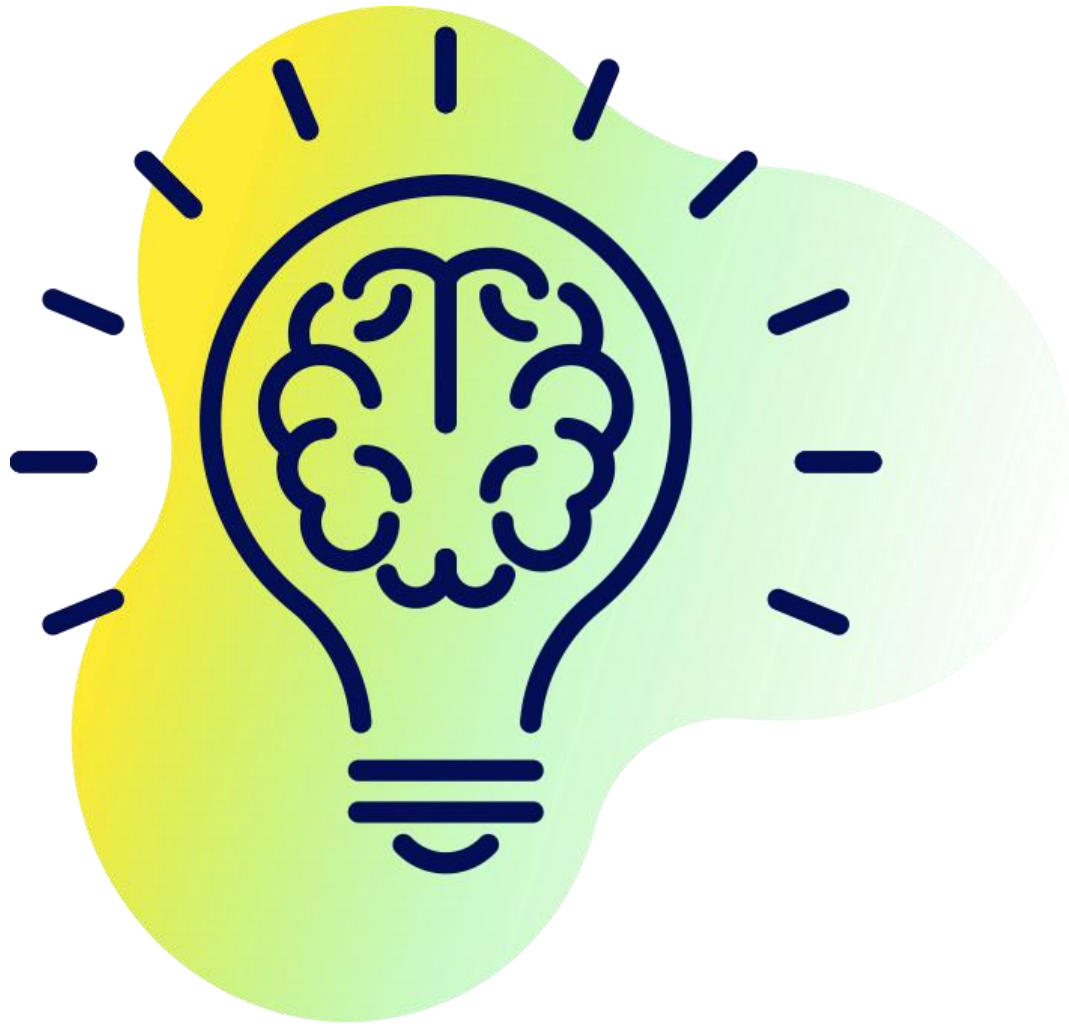
Como calculamos a distância?

Distância Euclidiana

$$\text{Euclidean}(A, B) = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2}$$



$$a^2 = b^2 + c^2$$



Obrigado!