

---

# Aprendizagem de Máquina

Programa de Pós-Graduação em Engenharia Elétrica e de Computação (PPGEEC)

Introdução à Ciências de Dados - UFC *Campus* Sobral – 2023.1

Andressa Gomes Moreira – [andressagomes@alu.ufc.br](mailto:andressagomes@alu.ufc.br)



# Sumário

---

1. Algoritmos de Classificação
2. Avaliação dos Classificadores
3. Referências

# Algoritmos de Classificação

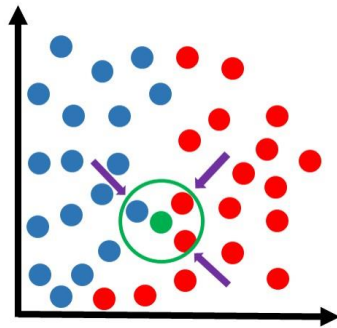
- **Árvores de Decisão:** Modelos de aprendizagem de máquina que representam uma sequência de regras de decisão em uma estrutura de árvore, permitindo a classificação ou regressão de dados de entrada com base em suas características.

**Árvore de Decisão para Jogar Tênis**

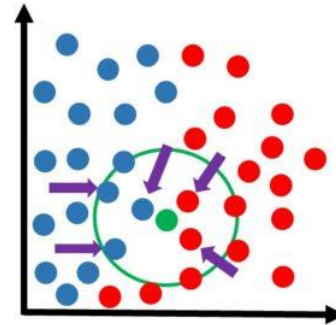
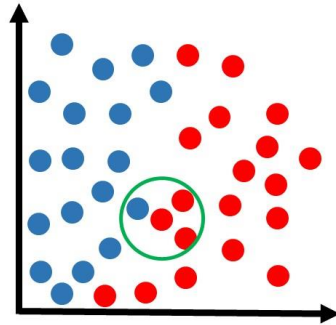


# Algoritmos de Classificação

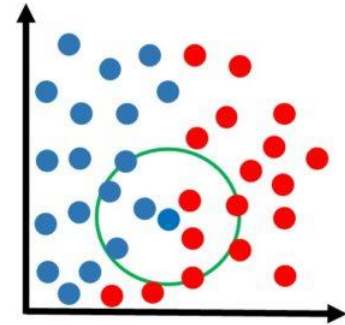
- **KNN (K-Nearest Neighbors):** Algoritmo de classificação que se baseia na proximidade dos dados. Ele classifica uma nova amostra com base nas classes das amostras vizinhas mais próximas.



$K = 3$



$K = 5$



# Avaliação dos Classificadores

---

- **Matriz de Confusão:** Resume o desempenho de um modelo de classificação em relação aos dados de testes.

		Valor Predito	
		Sim	Não
Real	Sim	Verdadeiro Positivo (TP)	Falso Negativo (FN)
	Não	Falso Positivo (FP)	Verdadeiro Negativo (TN)

# Avaliação dos Classificadores

---

- **Acurácia (acc):** Número de previsões corretas de um modelo para o total de amostras:

$$\frac{VP + VN}{N}$$

- **Precisão:** Observações positivas classificadas corretamente entre todos aqueles preditos como positivos:

$$\frac{VP}{VP + FP}$$

# Avaliação dos Classificadores

---

- **Recall:** Taxa de verdadeiros positivos:

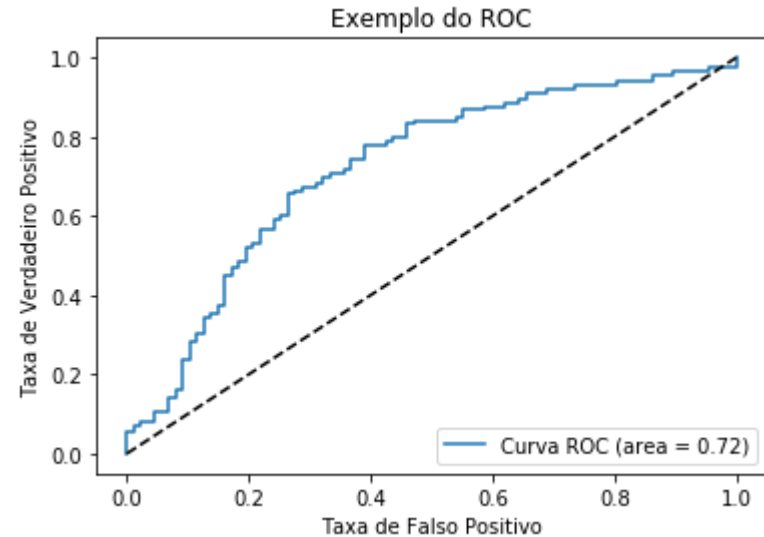
$$\frac{VP}{VP + FN}$$

- **F1-Score:** Define a média harmônica da precisão e recall:

$$\frac{2 \times precisao \times recall}{precisao + recall}$$

# Avaliação dos Classificadores

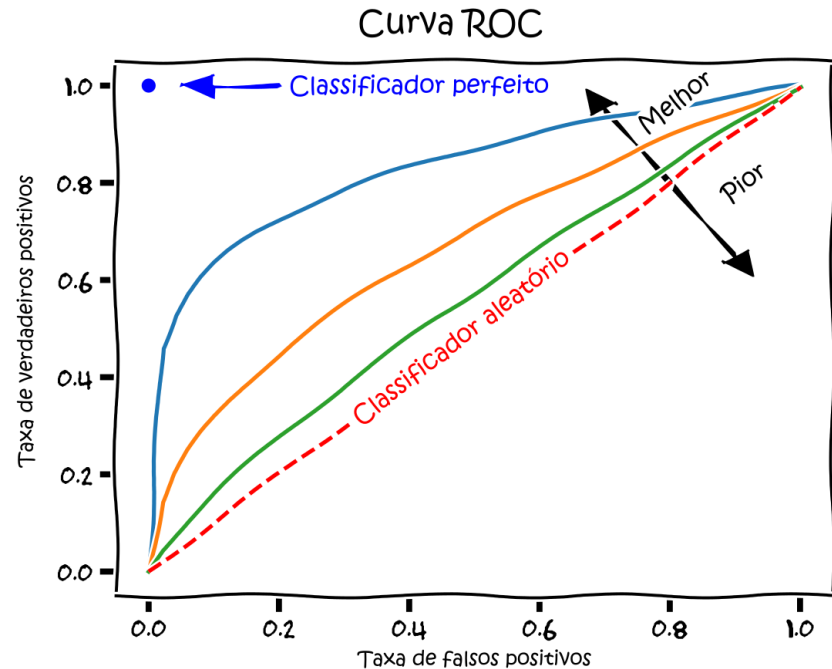
- **Curva ROC:** Quanto mais próximo o ponto da curva ROC estiver do canto superior esquerdo ( $TPR=1$  e  $FPR=0$ ), melhor é a performance do classificador.





# Avaliação dos Classificadores

- **AUC:** Representa a área sob a curva, variando de 0 a 1. Quanto maior a AUC, melhor é a capacidade do classificador em distinguir corretamente entre as classes.



## Referências

---

- [1] Grus, Joel Data Science do Zero / Joel Grus; traduzido por Welington Nascimento. - Rio de Janeiro : Alta Books, 2016. 336 p. : il. ; 3,8 MB.
- [2] Scikit-Learn - <https://scikit-learn.org/stable/>

# Obrigada!

---

- *Alguma dúvida?*  
*andressagomes@alu.ufc.br*