



GREat

Grupo de Redes de Computadores
Engenharia de Software
e Sistemas



**UNIVERSIDADE
FEDERAL DO CEARÁ**

Introdução ao Aprendizado de Máquina: Classificação e Regressão de Dados



Joyce Quintino Alves

Mestra em Ciência da Computação

Sobre mim



Joyce Quintino Alves

Formação:

- Mestra em Ciência da Computação - Universidade Federal do Ceará 2024
- Graduada em Ciência da Computação - Instituto Federal do Ceará 2019
- Técnica em Informática - Instituto Federal do Ceará 2015

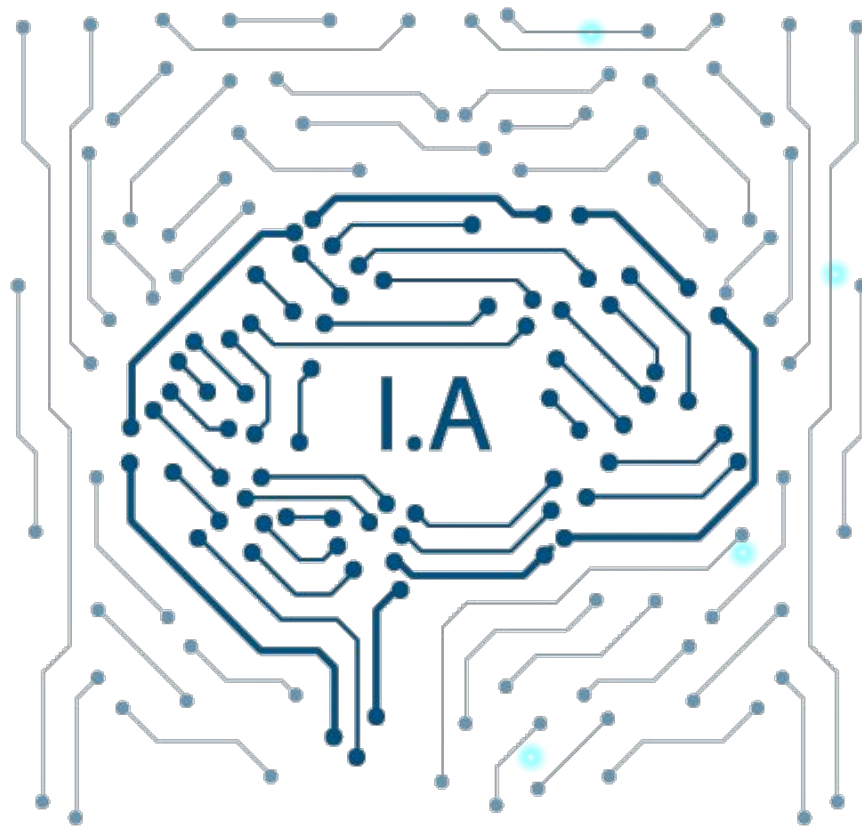
Experiência:

- Full-Stack Dev - Instituto Iracema de Pesquisa e Inovação 2024
- Cientista de Dados - Instituto Iracema de Pesquisa e Inovação 2023
- Analista de Sistemas - GREat 2019 a 2022

Roteiro

- ⬡ Inteligência Artificial
- ⬡ Aprendizado de Máquina
- ⬡ Pré-processamento de Dados
- ⬡ Algoritmos do Aprendizado Supervisionado
- ⬡ Métricas de Avaliação
- ⬡ Linguagens de Programação
- ⬡ Ferramentas
- ⬡ Vamos ao código?!

Inteligência Artificial

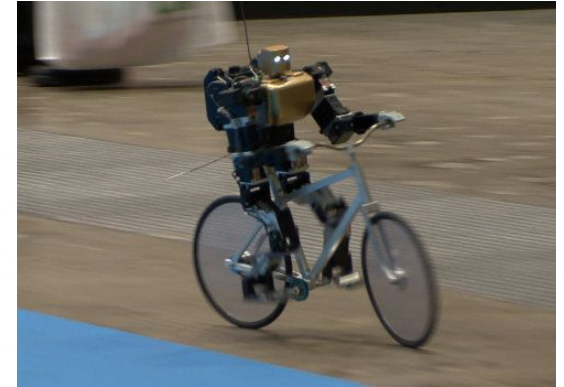
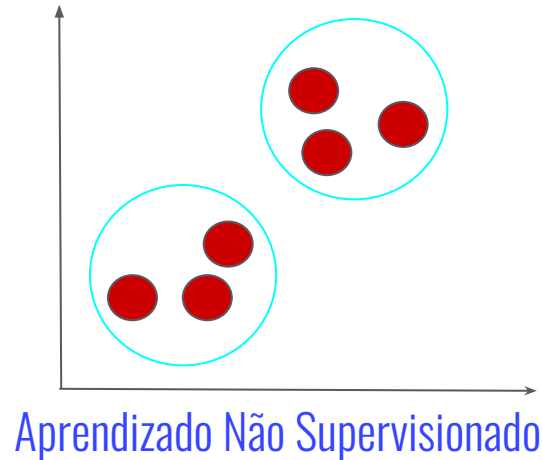
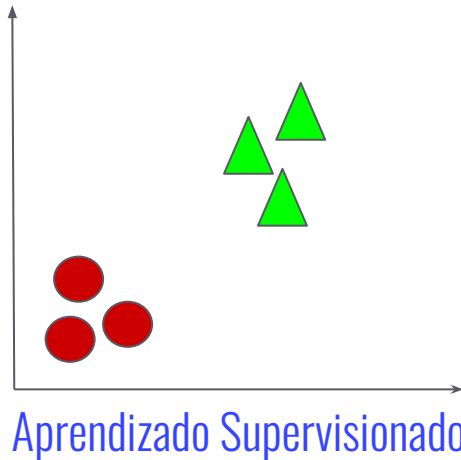


Aprendizado de Máquina

- Capacidade de uma máquina aprender
- A máquina aprende interagindo com o ambiente
- O ambiente é representado por dados
- A máquina aprende com os dados e cria um modelo que representa sua inteligência

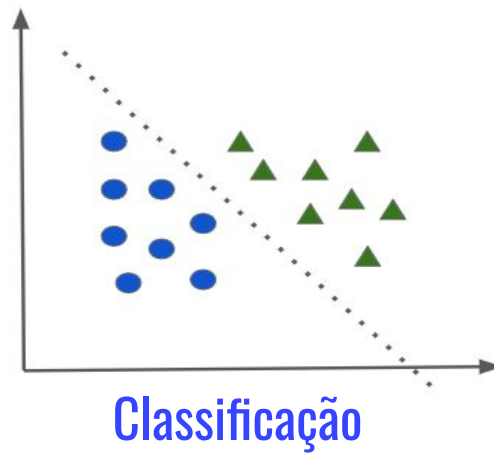
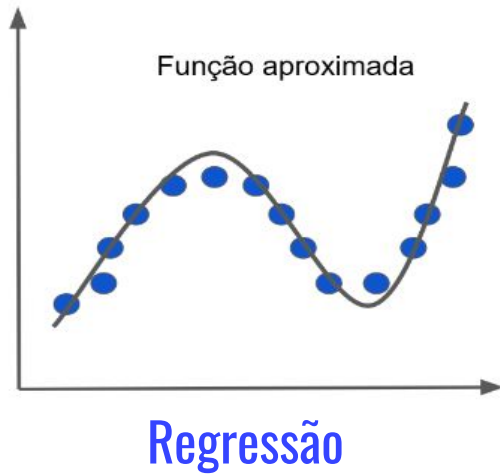


Tipos de Aprendizado de Máquina

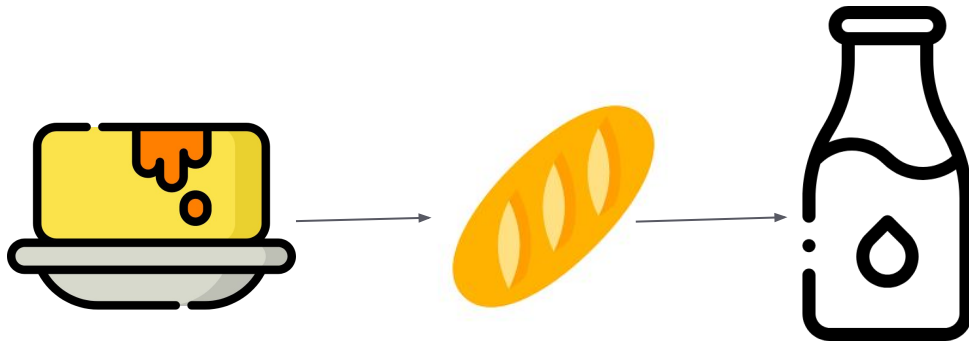


Aprendizado Por Reforço

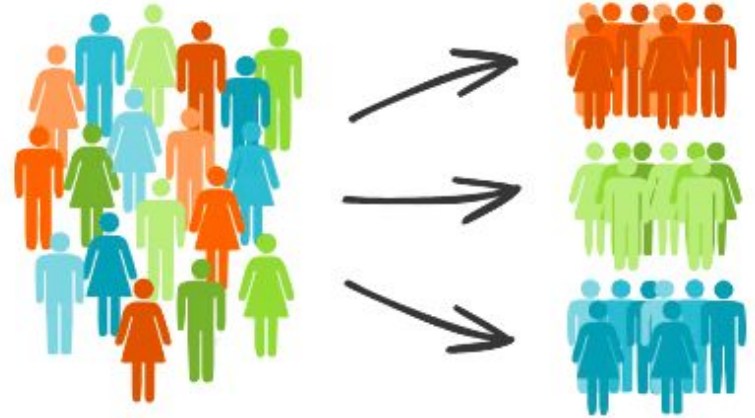
Aprendizado Supervisionado



Aprendizado Não Supervisionado

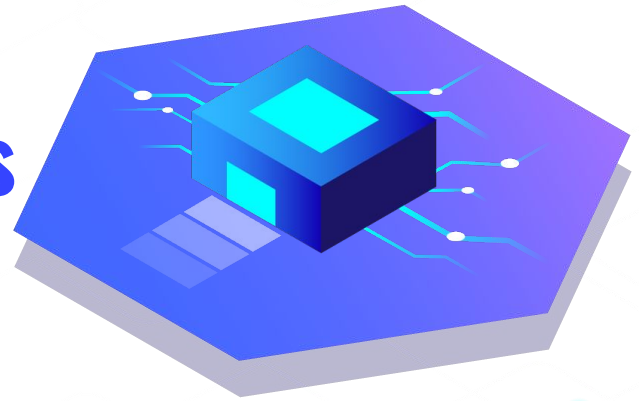


Regras de Associação



Agrupamento

Pré-processamento de Dados

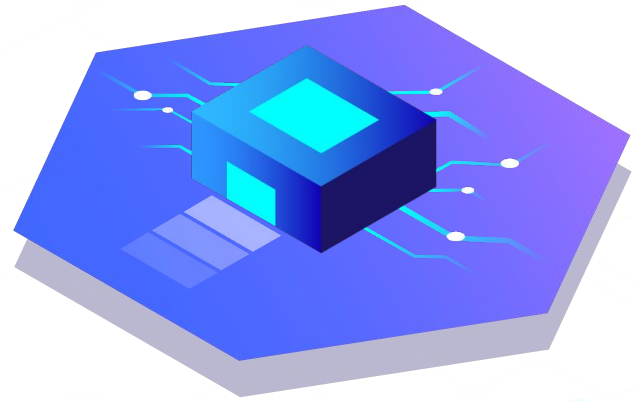


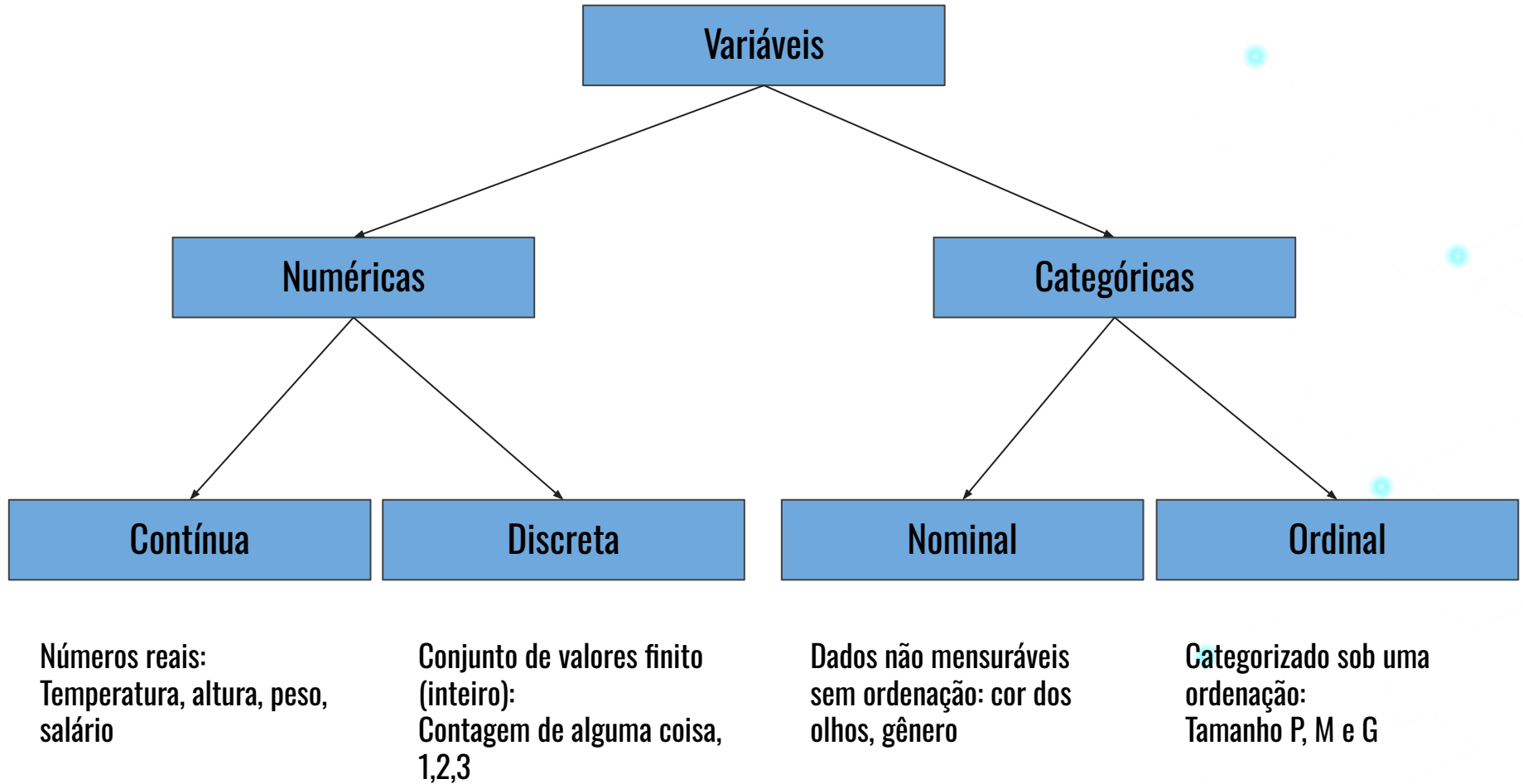
Pré-processamento de Dados

- É um conjunto de atividades que envolvem converter dados brutos em dados preparados, ou seja, em formatos úteis para o processamento dos algoritmos
- É um processo que compreende a preparação, organização e estruturação de dados
- É uma etapa para escolhermos quais dados fazem sentido fazerem parte do conjunto de dados

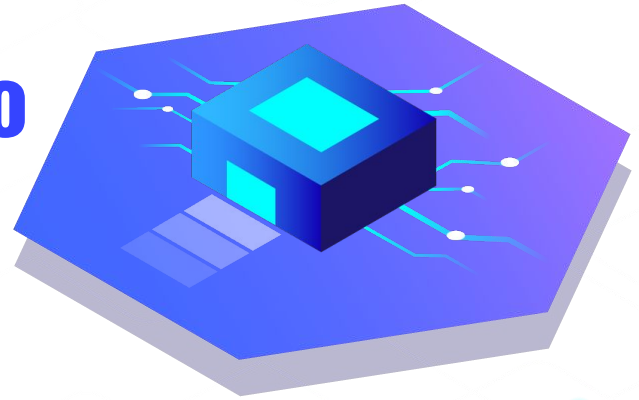


Tipos de Variáveis

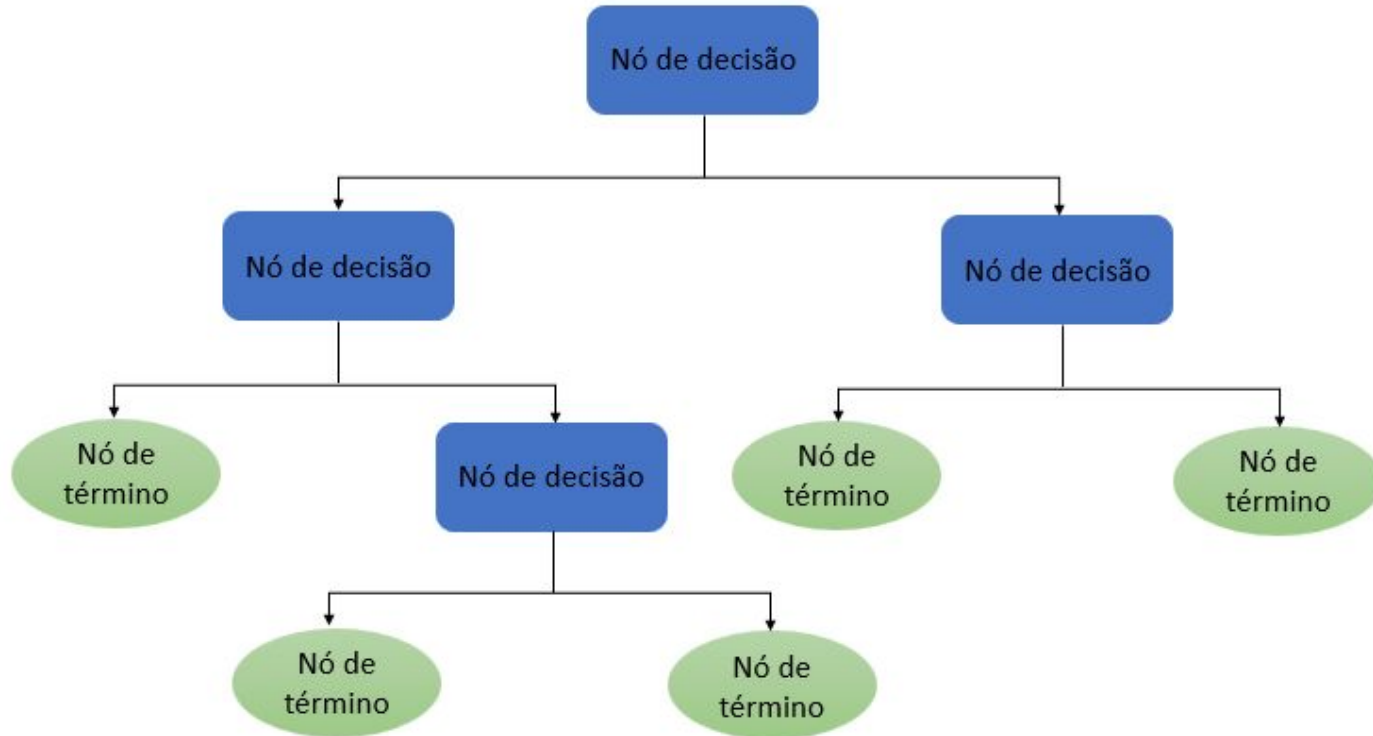




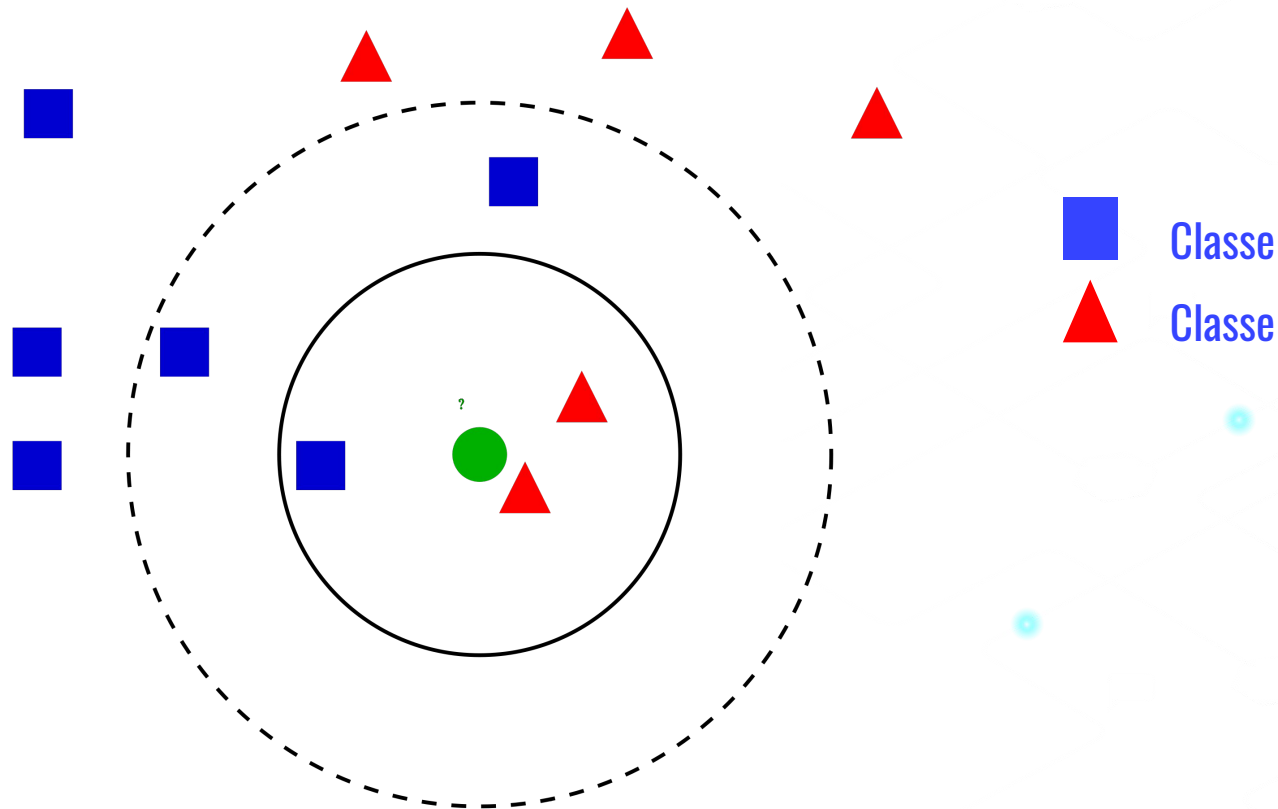
Algoritmos do Aprendizado Supervisionado



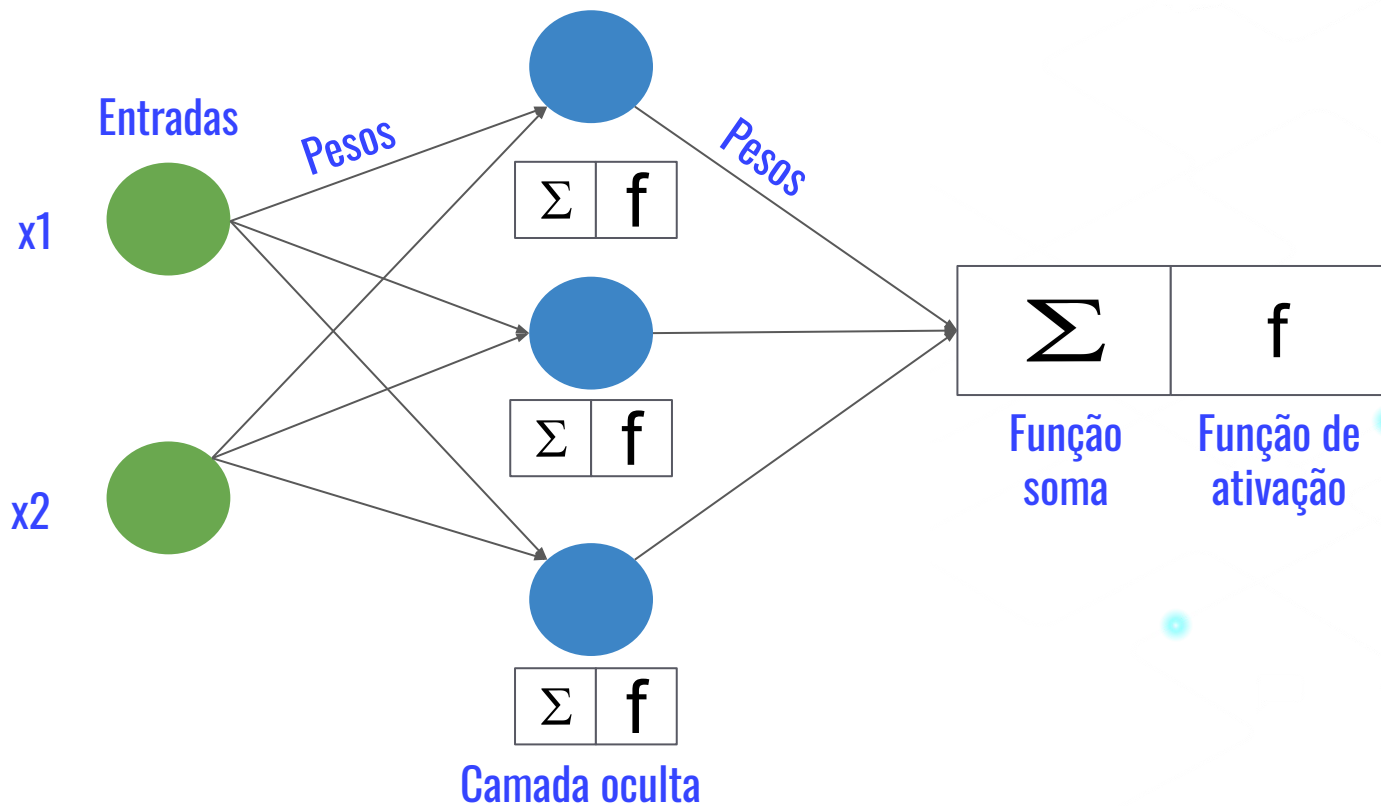
Árvore de Decisão



K-Vizinhos Mais Próximos (KNN)

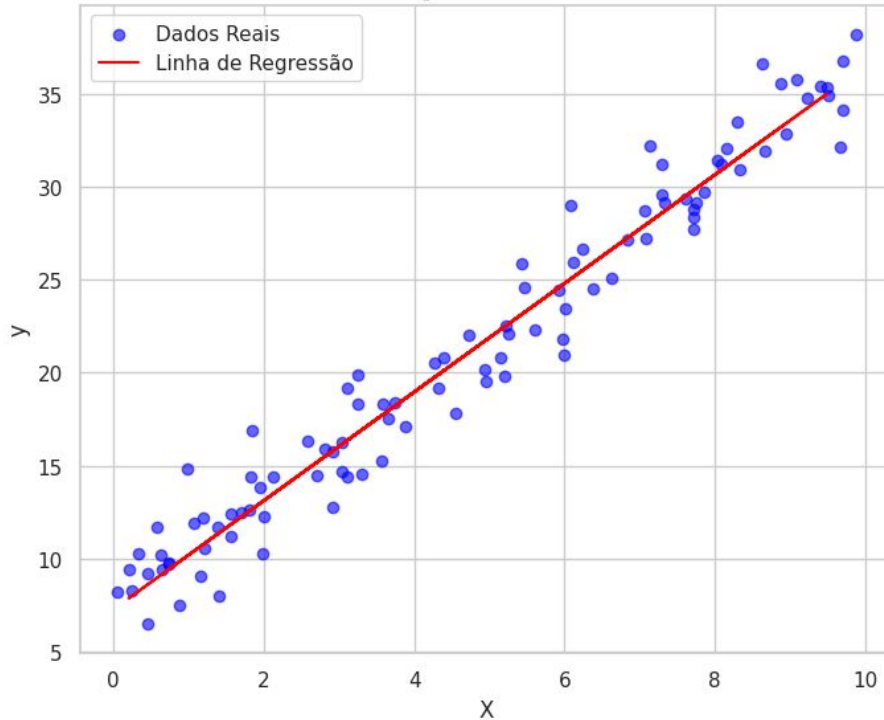


Rede Neural Perceptron

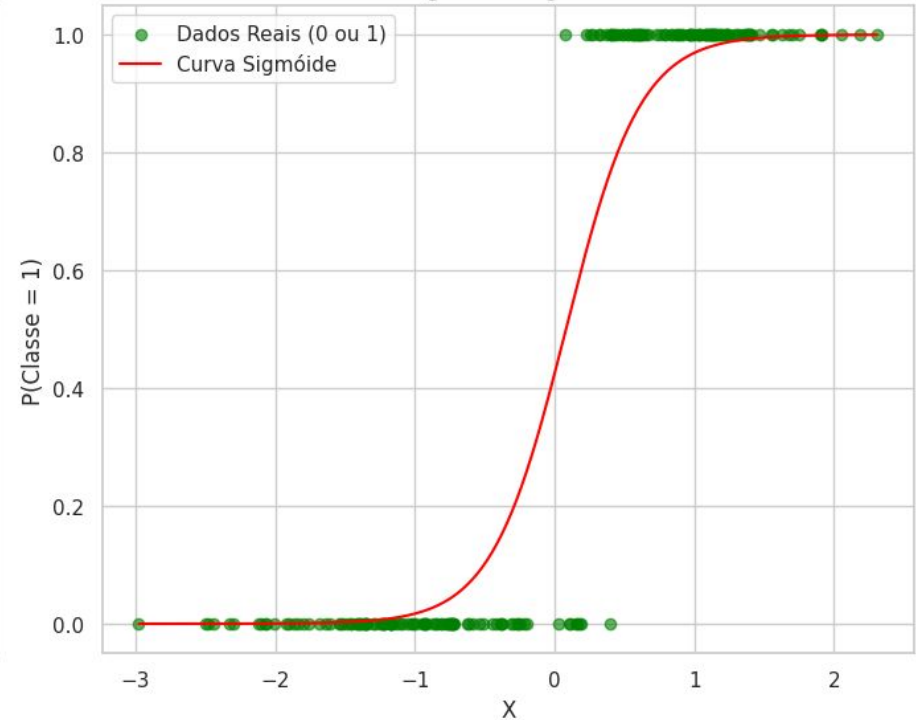


Regressão Linear e Logística

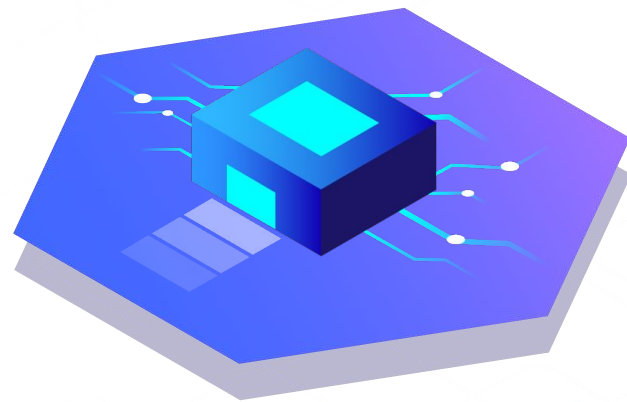
Regressão Linear



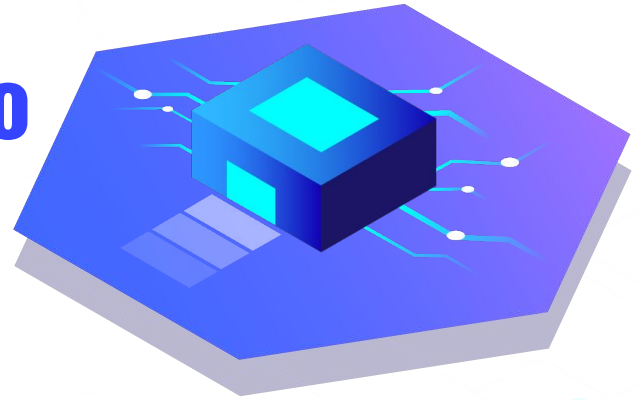
Regressão Logística



Métricas de Avaliação



Métricas para Classificação



Matriz de Confusão

		Classe Prevista	
		Sim	Não
Classe Real	Sim	Verdadeiro Positivo (VP)	Falso Positivo (FP)
	Não	Falso Negativo (FN)	Verdadeiro Negativo (VN)

Verdadeiro Positivo (VP)

- Predição **correta** da classe **Sim**
- (previu que é, e de fato é)

Verdadeiro Negativo (VN)

- Predição **correta** da classe **Não**
- (previu que não é, e de fato não é)

Falso Positivo (FP)

- Predição **incorreta** da classe **Não**
- (previu que é, mas na verdade não é)

Falso Negativo (FN)

- Predição **incorreta** da classe **Sim**
- (previu que não é, mas na verdade é)

Matriz de Confusão: Acurácia

- ⬡ **Acurácia:** a fração de predições corretas:
“Quanto que o modelo está acertando”

$$\text{Acurácia} = \frac{VP + VN}{VP + FP + FN + VN}$$

Matriz de Confusão: Precisão e *Recall*

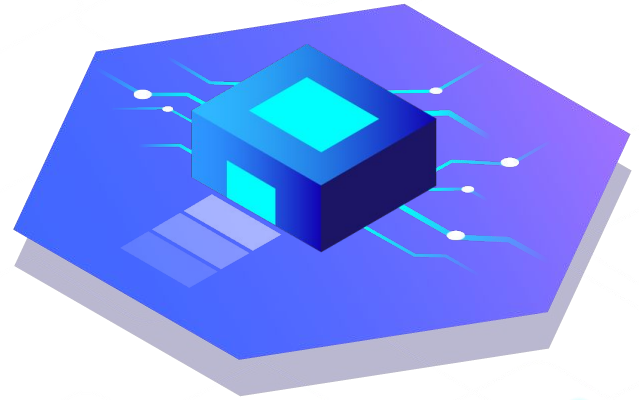
- **Precisão:** Dentre os previstos como positivo, qual a proporção de acerto?

$$\text{Precisão} = \frac{\text{VP}}{\text{VP} + \text{FP}}$$

- ***Recall* (Sensibilidade):** Dentre os positivos reais, qual a proporção de acerto?

$$\text{Recall} = \frac{\text{VP}}{\text{VP} + \text{FN}}$$

Métricas para Regressão



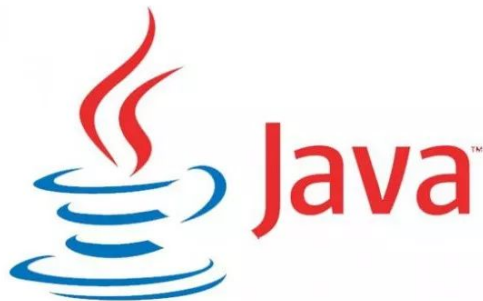
Métricas para Regressão

- ⬡ **RMSE (*Root Mean Squared Error*):** É a medida da diferença média entre os valores previstos pelo modelo e os valores reais. Quanto menor o RMSE, melhor o modelo está ajustado aos dados

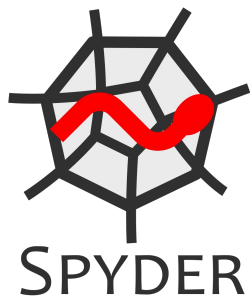
$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - p_i)^2}$$

- ⬡ **R² (Coeficiente de Determinação):** Mede a proporção da variância dos dados que é explicada pelo modelo. Vai de 0 a 1 (ou negativo, se o modelo for muito ruim)

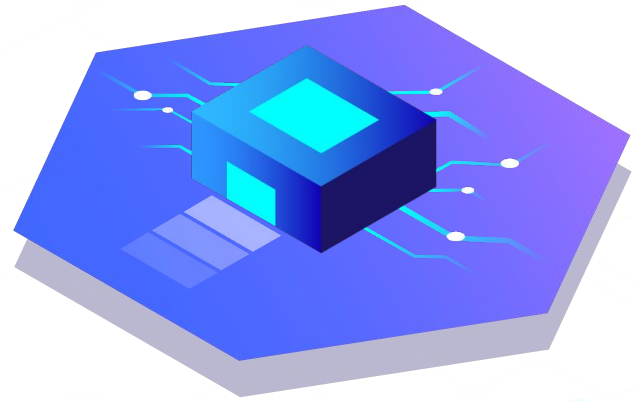
Linguagens de Programação



Ferramentas



Vamos ao código?!



Prática

⬡ Classificação de Dados:

- Conjunto de Dados de Análise de Crédito

⬡ Regressão de Dados:

- Conjunto de Dados de Diabetes da sklearn

