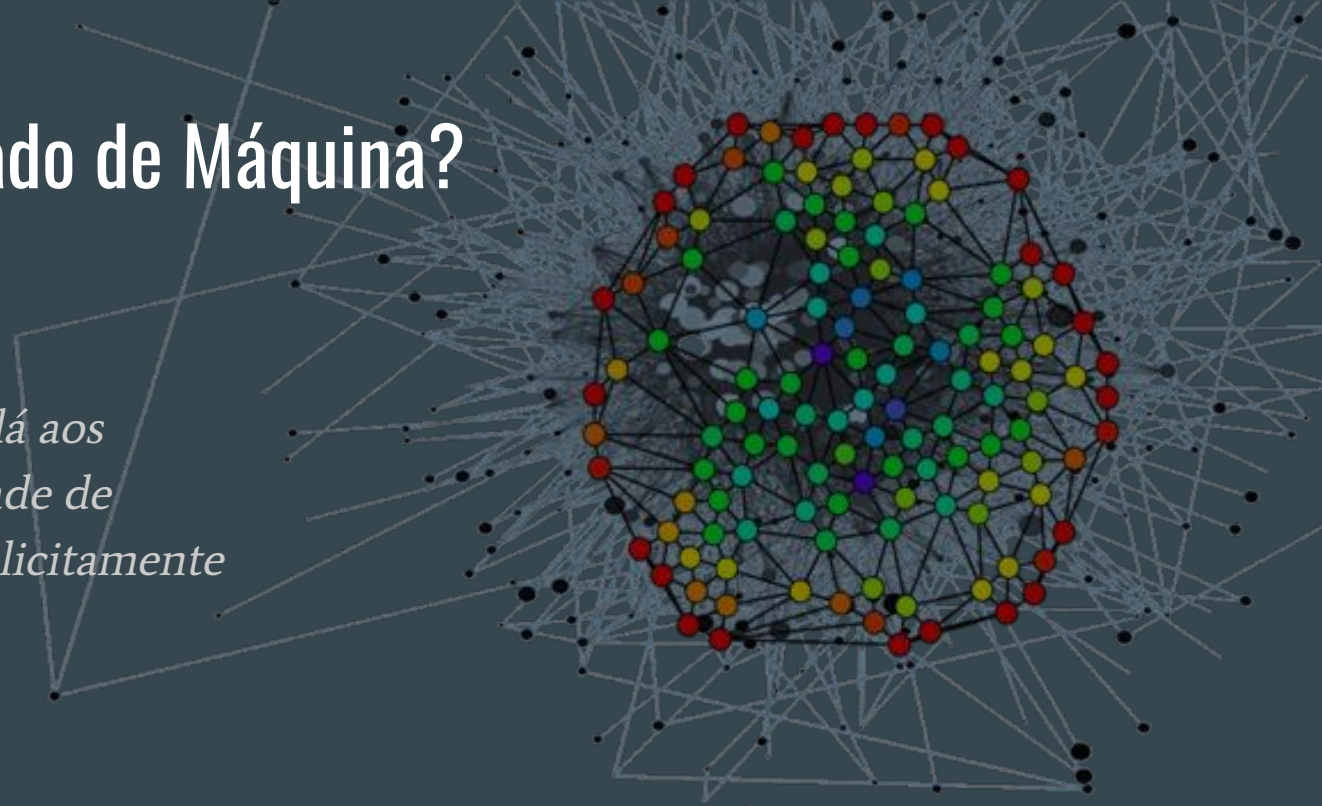


Introdução à Aprendizado de Máquina em Python

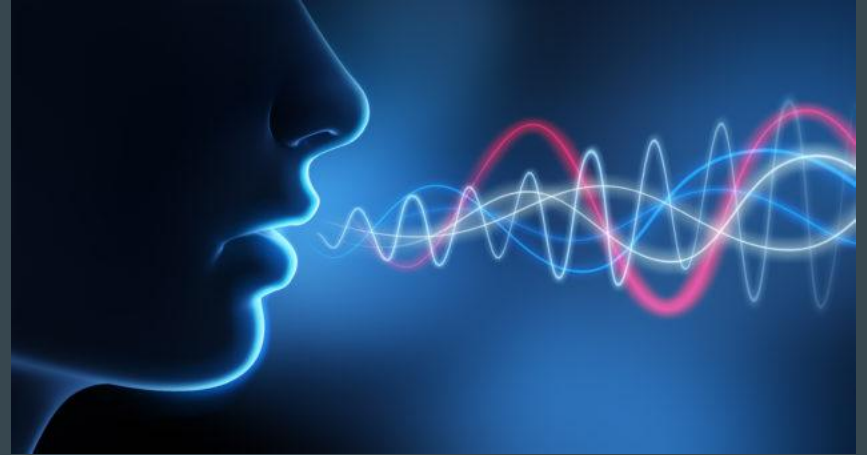


O que é Aprendizado de Máquina?

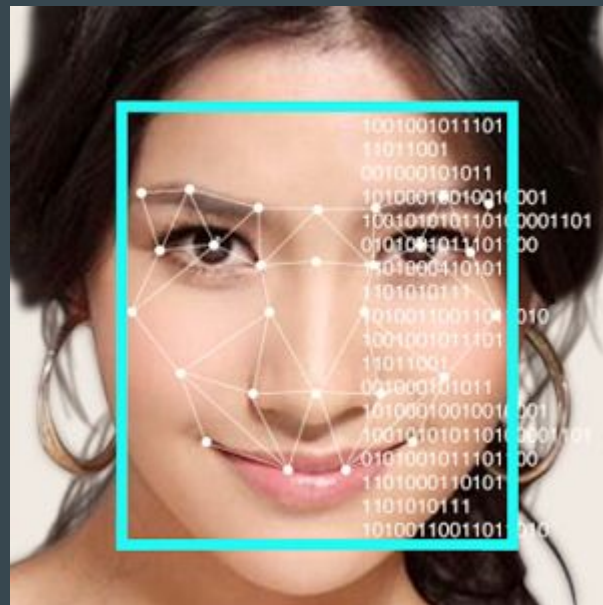
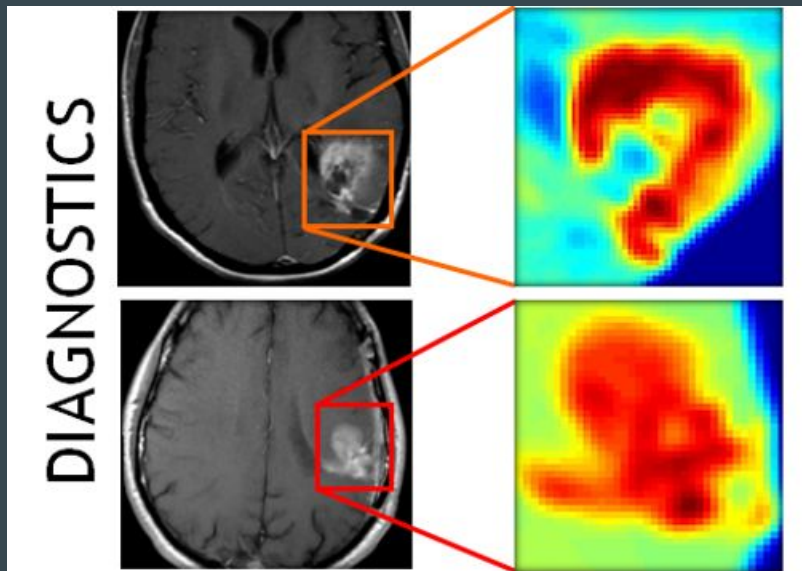
“Campo de estudo que dá aos computadores a habilidade de aprender sem serem explicitamente programados.”



Aplicações de Aprendizado de Máquina



Aplicações de Aprendizado de Máquina



Tipos de Aprendizado

- Supervisionado
 - Recebe pares de entradas e saídas
 - Objetivo: definir a relação entre entradas e saídas
 - Problema: casos onde há muitas dados
- Não Supervisionado
 - Recebe apenas entradas
 - Objetivo: definir padrões para as entradas

Aprendizado Supervisionado

- Regressão
 - Baseado no processo estatístico para estimar relações entre variáveis
 - Uma variável dependente influenciada por muitas variáveis independentes
- Classificação
 - Entradas classificadas
 - Aprender a classificar novas entradas
- Alguns termos:
 - Target/label: alvo, resposta
 - Feature: Descritor, atributo

Medidas de Desempenho

Por que existem tantos diferentes métodos de aprendizado?

Existe um método melhor de todos?

Como decidir qual método produz melhores resultados?

Medidas de Desempenho

Para modelos de classificação, podemos medir sua ***acurácia***:

$$\text{acurácia} = \text{número de predições corretas} / \text{número de predições feitas}$$

Quando fazemos isso com o mesmo conjunto que usamos para treinar o modelo, falaremos de ***acurácia de treino***.

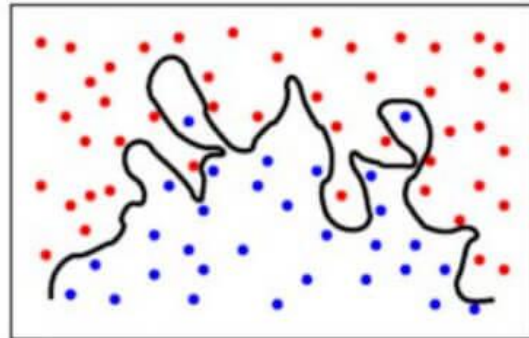
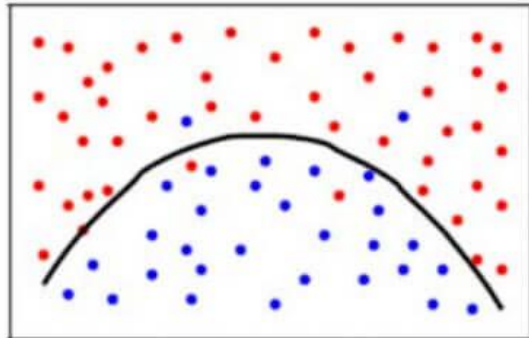
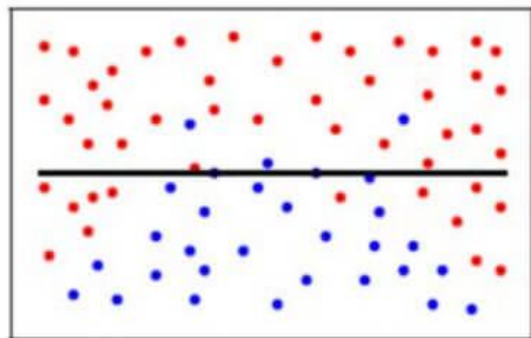
Desvantagem: não diz como o modelo se comporta com novos dados, mas simplesmente "recompensa" *overfitting* (sobreajuste).

Generalization Problem in Classification

Underfitting



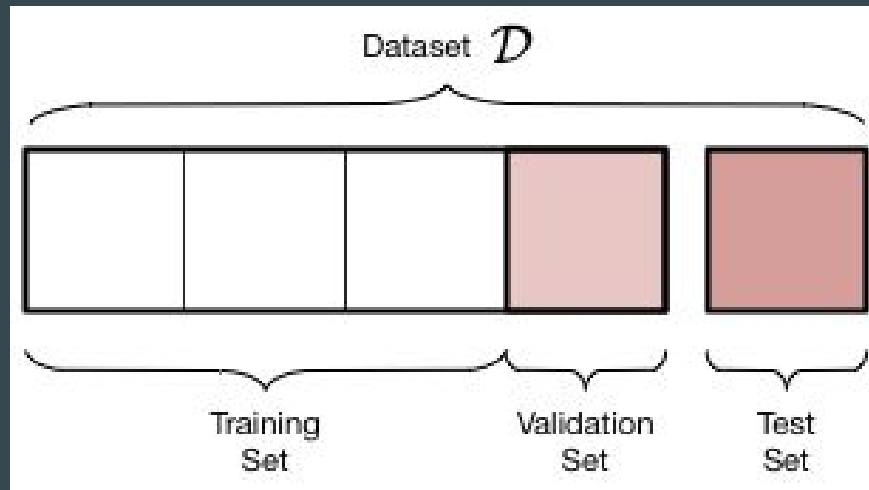
Overfitting



Medidas de Desempenho

Divisão do conjunto de dados:

- Conjunto de Treino
- Conjunto de Validação
- Conjunto de Teste (Generalização)



Assim, dividindo a base em treino e teste, podemos calcular a ***acurácia de teste***.

Medidas de Desempenho

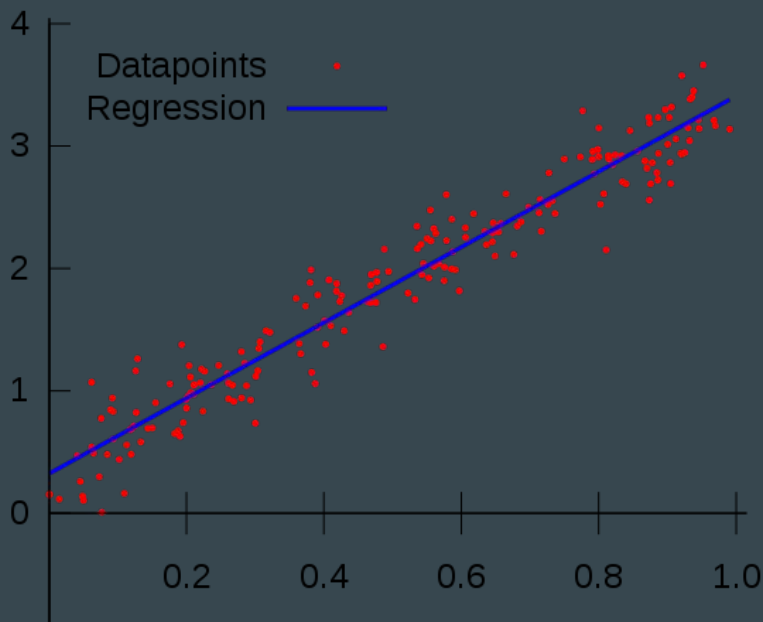
E para modelos de regressão?

A métrica mais usada nesse caso é ***Mean Squared Error (MSE)*** ou erro quadrático médio (EQM): a média dos quadrados dos erros nas previsões feitas.

$$\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2$$

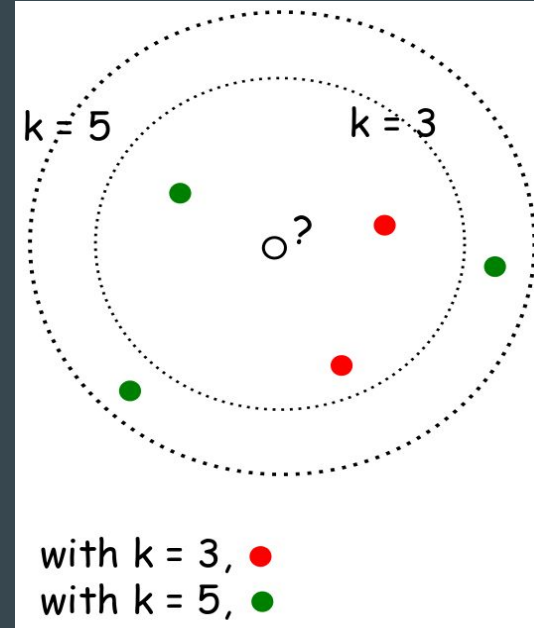
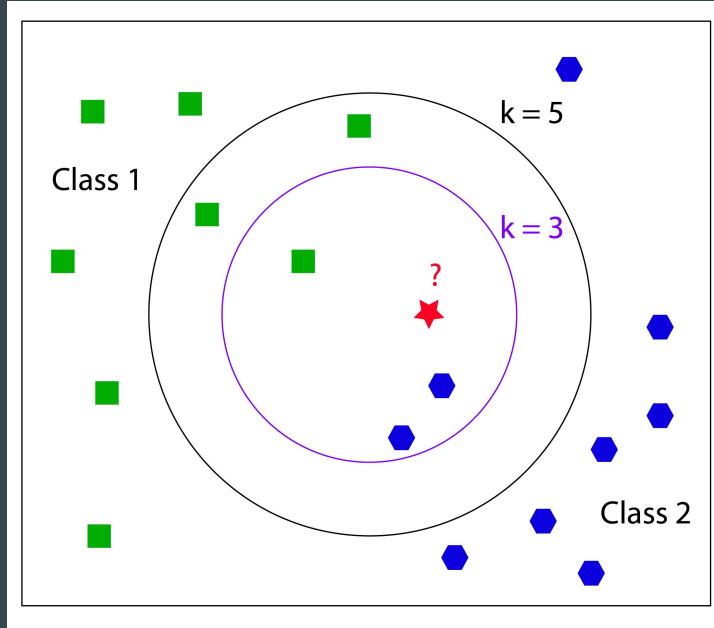
Aprendizado Supervisionado: Regressão Linear

Estimar o valor da variável dependente com base em uma combinação linear das variáveis independentes.



Aprendizado Supervisionado: KNN (K-Nearest Neighbors)

Classificação baseado nos k elementos mais próximos.



Aprendizado Supervisionado: Árvore de decisão (CART)

Mapeamento do processo de raciocínio.

Tabela de decisão: representação em forma de tabela do conjunto de condições para que determinada ação seja escolhida/executada, mapeando a árvore de decisão.



Aprendizado Supervisionado: SVM

Linearmente inseparável? Go to a higher space.

