

# Aprendizado Supervisionado e Não Supervisionado

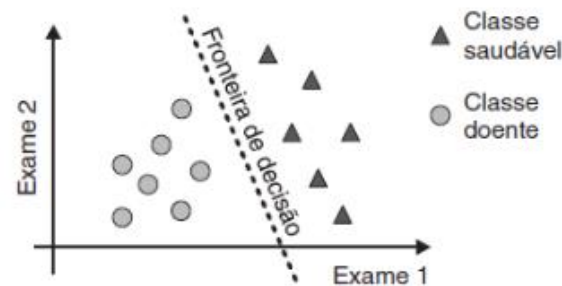
**Professor**  
Jefferson Moraes

# Aprendizado Supervisionado e não Supervisionado

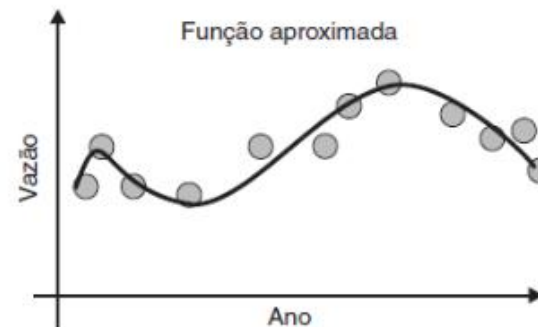
- Supervisionado (Preditivo)
  - Objetivo: encontrar uma função (modelo ou hipótese) que possa ser utilizado para prever um rótulo ou valor para novos dados
  - Objetos na forma de (entrada, saída)
- Não-supervisionado (Descritivo)
  - Objetivo: explorar ou descrever um conjunto de dados. Encontrar grupos de objetos semelhantes
  - Objetos não possuem saída associada

# Aprendizado Supervisionado

- O algoritmo de aprendizado supervisionado (preditivo) recebe um conjunto de exemplos de treinamento para os quais os rótulos da **classe** associada são **conhecidos**
  - Representado por um conjunto de pares  $(\mathbf{x}, y)$ 
    - Ex:  $\mathbf{x}$ =sintomas e  $y$ =diagnóstico
- Métodos supervisionados distinguem pelo tipo dos rótulos dos dados
  - Rótulos discretos (**classificação**)
    - Ex: diagnóstico, bom/mau pagador, etc.
  - Rótulos contínuos (**regressão**)
    - Previsão de séries temporais



(a) [Classificação]



(b) [Regressão]

# Aprendizado supervisionado

- Classificação

- Problema do filtro de spam

- Treinado com muitos exemplos de e-mails rotulados como spam e não spam
    - Deve aprender a classificar novos e-mails

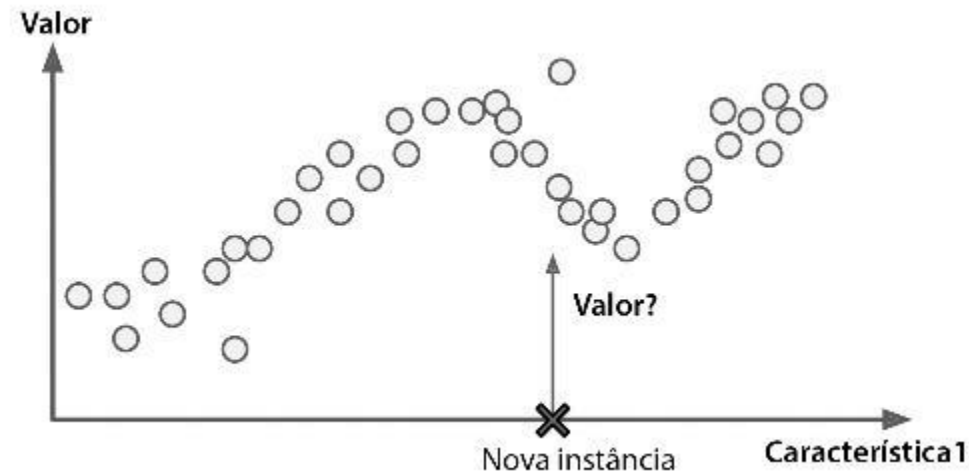


# Aprendizado supervisionado

- Regressão

- Problema do preço do veículo

- Dado um conjunto de características (*features*)
      - Ex: quilometragem, tempo de uso, marca, etc
    - Tarefa é prever um valor numérico alvo (*target*), neste o preço do carro



# Aprendizado Supervisionado

- Principais algoritmos de aprendizado supervisionado
  - K-ésimo vizinho mais próximo (KNN)
  - Regressão Linear
  - Regressão logística
  - Máquinas de vetores de suporte (SVMs)
  - Árvores de decisão e florestas aleatórias (Random Forest)
- Obs: Alguns algoritmos de regressão também pode ser utilizados para classificação e vice-versa
  - Ex: Regressão logística é comumente utilizada para classificação

# Aprendizado não supervisionado

- No aprendizado não supervisionado os dados de treinamento **não são rotulados**
- O sistema tentar aprender “sem um professor”



# Aprendizado não supervisionado

- Principais algoritmos de aprendizado não supervisionado divididos por tarefa
  - Agrupamento (Clustering)
    - K-Means (Clusterização K-Média)
    - DBSCAN (Clusterização baseada em densidade espacial)
    - Análise de cluster hierárquica (HCA)
  - Detecção de *outliers*
    - One-class SVM
    - Floresta de isolamento (*Isolation Forest*)
  - Redução de dimensionalidade
    - Análise de componentes principais (PCA)
  - Aprendizado por regras de associação
    - Apriori
    - Eclat



# Aprendizado não supervisionado

- Agrupamento (*clustering*)
  - Dados de visitantes de um *blog*
  - Objetivo do algoritmo de agrupamento: detectar grupos (*clusters*) com características semelhantes
  - Em nenhum momento se informa ao algoritmo a qual grupo o visitante pertence
    - Ele encontrará essas relações por si só



# Aprendizado não supervisionado

- Redução da dimensionalidade
  - Tarefa cujo objetivo é **simplificar** os dados sem perder muita informação
  - Para isso, pode-se fazer o **merge** de diversas características (atributos) correlacionadas em uma
  - Diferentes técnicas originárias de áreas de pesquisa como
    - Reconhecimento de Padrões, Estatística e Teoria da Informação podem ser utilizadas para a redução do número de atributos.
  - Essas técnicas podem ser divididas em duas grandes abordagens:
    - **Extração de características** (Feature Extraction)
      - Substituem as características originais por novas características formadas pela combinação de grupos de características
      - Ex: Algoritmo PCA
    - **Seleção de características** (Feature Selection)
      - Mantêm uma parte das características originais e descartam as demais
      - Ex: CFS (*Correlation Feature Selection*)

# Aprendizado não supervisionado

- Detecção de outliers
  - Detecção de transações incomuns em cartões de crédito para evitar fraudes
  - Identificar defeitos de fabricação
  - Remover automaticamente outliers de um conjunto de dados, antes de fornecê-lo a outro algoritmo de AM



# Aprendizado não supervisionado

- Aprendizado de regras de associação
  - Objetivo: investigar grandes quantidades de dados e descobrir relações interessantes entre as características (atributos)



TID	Lista de Itens
T1	Pão, Leite
T2	Pão, Fralda, Cerveja, Ovos
T3	Leite, Fralda, Cerveja, Coca
T4	Pão, Leite, Fralda, Cerveja
T5	Pão, Leite, Fralda, Coca

- Uma regra de associação é uma implicação da forma  $(X \rightarrow Y)$ , onde  $X$  e  $Y$  são conjunto de itens e  $X \cap Y = \emptyset$ .
- R1: {Cerveja}  $\rightarrow$  {Fralda}.
  - R2: {Cerveja, Pão}  $\rightarrow$  {Leite}.
  - R3: {Leite, Pão}  $\rightarrow$  {Fralda, Coca}.

## **Aprendizado semi-supervisionado e Aprendizado por Reforço**