

Introdução ao Aprendizado de Máquina

Prof. Milton Roman de Brito

Objetivo

Ao final da aula, vocês devem ser capazes de:

- Entender o que é Aprendizado de Máquina (Machine Learning)
- Diferenciar os principais tipos de sistemas de ML
- Reconhecer exemplos reais de aplicação em Ciência de Dados
- Reconhecer conceitos Básicos Fundamentais em Machine Learning

Conteúdo da Aula

- O que é Aprendizado de Máquina?
- Por que ML é importante?
- Tipos de sistemas de ML
- Exemplos práticos
- Conceitos Básicos Fundamentais em Machine Learning
- Discussão final

O que é Aprendizado de Máquina?

*Machine Learning **NÃO É** criar o Jarvis do Tony Stark, nem o Exterminador do Futuro!*

Aprendizado de Máquina/Machine Learning (ML) é ensinar computadores a **aprender a partir de dados**, em vez de seguir apenas instruções rígidas criadas por um programador.

De forma muito simplista

- O sistema não é programado com regras fixas;
- Ele aprende padrões a partir dos dados.

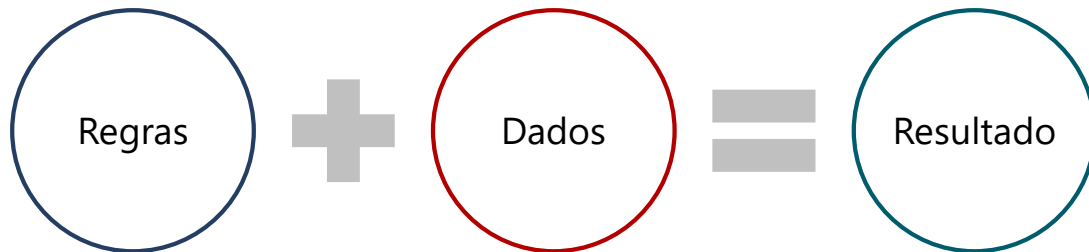
O que é Aprendizado de Máquina?

Definições de Autores

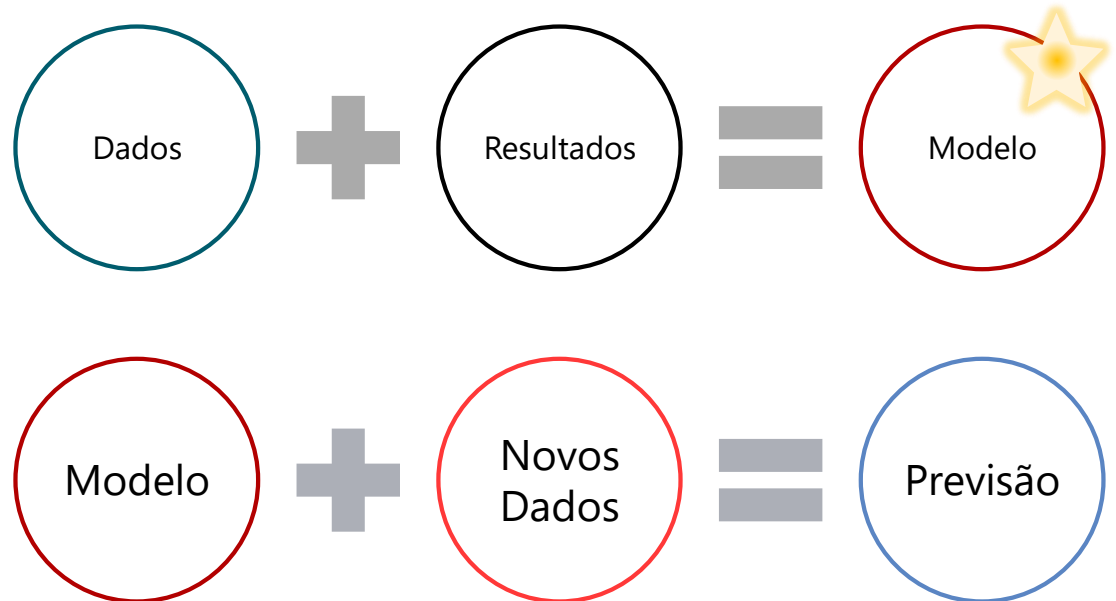
- Aprendizado de Máquina é a ciência (e a arte) da programação de computadores para que eles possam aprender com os dados (Aurélien Géron, 2017)
- Definimos aprendizado de máquina como um conjunto de métodos que podem detectar automaticamente padrões nos dados e, em seguida, usa os padrões descobertos para prever dados futuros ou para executar outros tipos de tomada de decisão (Kevin P. Murphy)

Programação Tradicional x Machine Learning

Programação Tradicional



Machine Learning



Por que usar Aprendizado de Máquina?

Problemas difíceis de resolver com regras explícitas

- Reconhecimento de imagens
- Detecção de fraude
- Recomendação de produtos
- Diagnóstico médico
- Previsões financeiras

Onde o ML aparece no dia a dia?

Exemplos conhecidos

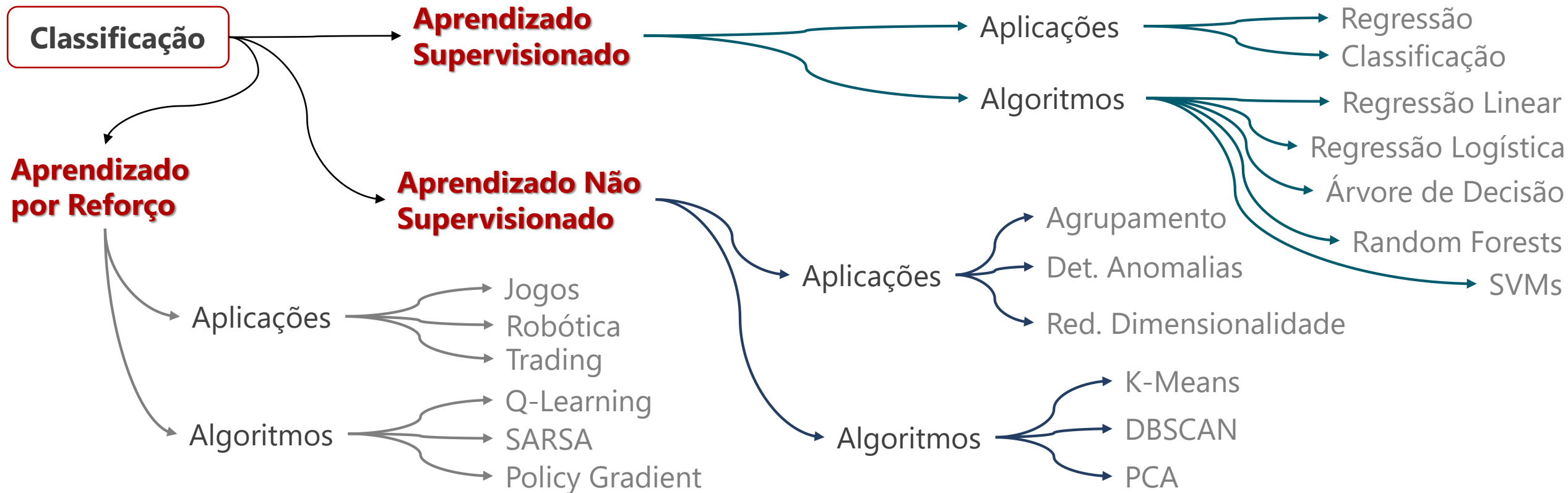
- Netflix / Spotify (recomendação)
- Google Maps (rotas e trânsito)
- Bancos (fraude)
- E-commerce (preço dinâmico)
- Redes sociais (feed)

Dúvidas e consolidação do Conteúdo

- Como vocês imaginam que seja o desenvolvimento de Modelos de ML, se comparado com a Programação Tradicional?
- Onde vocês acham que ML é usado hoje sem percebermos?
- Já conseguem ter um vislumbre de um Projeto de Aprendizado de Máquina, ou como poderia aplicar para evoluir um Projeto anterior?

Tipos de Sistemas de Aprendizizado de Máquina

Classificação Geral dos Sistemas de ML



Aprendizado Supervisionado

Ideia Central

- Os dados possuem **rótulos**
- O modelo aprende a relação: entrada → saída

Exemplos

- Classificação de e-mails (spam / não spam)
- Previsão de preços
- Diagnóstico (enfermo / saudável)

Exemplos de Tarefas Supervisionadas - Regressão

Objetivo → Prever um **valor numérico contínuo**.

Previsão de preço de imóveis

- *Entrada* → Metragem, número de quartos, localização, idade do imóvel
- *Saída* → Preço do imóvel (R\$)

Previsão de vendas futuras

- *Entrada* → Histórico de vendas, sazonalidade, promoções
- *Saída* → Quantidade de vendas esperada

Estimativa de consumo de energia

- *Entrada* → Temperatura, horário, número de equipamentos ligados
- *Saída* → Consumo em kWh

Exemplos de Tarefas Supervisionadas – Classificação

Objetivo → prever uma **Classe ou Categoria**

Detecção de spam em e-mails

- *Entrada* → Conteúdo do e-mail, remetente, palavras-chave
- *Saída* → **"Spam"** ou **"Não Spam"**

Diagnóstico médico

- *Entrada* → Exames, sintomas, histórico do paciente
- *Saída* → **"Enfermo"** ou **"Saudável"**. Ou ainda **"Saudável"** ou **"Enfermo - CID da Enfermidade"**.

Classificação de clientes (churn)

- *Entrada* → Tempo de contrato, uso do serviço, reclamações
- *Saída* → **"Vai Cancelar"** ou **"Não vai Cancelar"**

Aprendizado Não Supervisionado

Ideia Central

- Não existem rótulos
- O algoritmo busca padrões ocultos nos dados

Exemplos

- Agrupamento de clientes
- Segmentação de mercado
- Redução de dimensionalidade

Exemplos de Tarefas Não Supervisionadas (1/3)

Clustering (Agrupamento)

- **Segmentação de Clientes** → Empresas de marketing agrupam consumidores com base em comportamento de compra, idade ou interesses, sem categorias pré-definidas.
- **Agrupamento de Documentos/Notícias** → Organizar milhares de notícias em tópicos automáticos (esportes, política, finanças).

Exemplos de Tarefas Não Supervisionadas (2/3)

Detecção de anomalias

- **Detecção de Fraude** → Identificar transações de cartão de crédito que fogem ao padrão de compra habitual do cliente.
- **Ciber segurança** → Detectar tráfego de rede incomum ou tentativas de invasão (intrusão) em sistemas.
- **Manutenção Preditiva** → Analisar sensores de máquinas para detectar vibrações ou temperaturas atípicas que indicam falha iminente.

Exemplos de Tarefas Não Supervisionadas (3/3)

Redução de dimensionalidade

- **Visualização de Dados** → Reduzir um conjunto de dados com centenas de características para 2 ou 3 dimensões, permitindo gráficos visualizáveis.
- **Compressão de Imagens** → Reduzir o tamanho de arquivos de imagem retendo os detalhes essenciais.
- **Pré-processamento de Dados** → Remover "ruído" e reduzir o número de variáveis de entrada para modelos de aprendizado supervisionado, melhorando sua performance.

Aprendizado por Reforço

Ideia central

- Um **agente** aprende por tentativa e erro
- O **agente** vai interagir com um **ambiente** através de **ações**
- Recebe **recompensas** ou **punições**

Componentes

- Agente
- Ambiente
- Ações
- Recompensas

Exemplos de Aprendizado por Reforço (1/3)

Xadrez, Go ou jogos Atari

- **Agente** → Jogador controlado pelo algoritmo
- **Ambiente** → O jogo (tabuleiro, regras, estados)
- **Ações** → Movimentos possíveis
- **Recompensa** →
 - Vitória (+1)
 - Derrota (-1)
 - Empate (0)

Exemplos de Aprendizado por Reforço (2/3)

Robô aprendendo a andar ou desviar de obstáculos

- **Agente** → Robô
- **Ambiente** → Espaço físico ou simulado
- **Ações** → Mover braços, rodas ou pernas
- **Recompensa** →
 - Avanço correto (+),
 - Colisão ou queda (–)

Exemplos de Aprendizado por Reforço (3/3)

Recomendação de filmes, músicas ou produtos

- **Agente** → Sistema de recomendação
- **Ambiente** → Interação com o usuário
- **Ações** → Recomendar um item
- **Recompensa** →
 - Clique, tempo de visualização ou compra (+)
 - Rejeição ou abandono (–)

Comparando os Tipos de Aprendizado

Tipo	Usa Rótulos?	Objetivo
Supervisionado	Sim	Previsão
Não Supervisionado	Não	Descoberta de Padrões
Reforço	Não	Maximizar Recompensa

Coleta de Dados

Pré-processamento

Treinamento do Modelo

Avaliação

Deploy / Uso

Dúvidas e consolidação do Conteúdo

- Qual tipo de aprendizado parece mais intuitivo? Mais simples de compreender e aplicar.
- Onde vocês imaginam usar Aprendizado de Máquina na área profissional de vocês?
- Que tipo de dados vocês já viram no trabalho ou estudos?

Conceitos Básicos Fundamentais em Machine Learning

Conjunto de Dados (Dataset)

- É o **conjunto de informações** usado para treinar e avaliar modelos de Machine Learning, normalmente organizado em linhas e colunas

Exemplo → Cada linha representa um cliente e cada coluna uma informação sobre ele

Instância (Amostra ou Observação)

- É **uma linha do dataset**, representando um único exemplo do problema

Exemplo → Um cliente específico dentro do conjunto de dados

Conceitos Básicos Fundamentais em Machine Learning

Características (Features ou Variáveis)

- São os **atributos de entrada** usados pelo modelo para aprender padrões

Exemplo → Idade, renda e tempo de contrato de um cliente

Rótulo (Label ou Variável Alvo)

- É o **valor que o modelo deve aprender a prever** em tarefas supervisionadas

Exemplo → "Vai cancelar o serviço?" → *sim* ou *não*

Conceitos Básicos Fundamentais em Machine Learning

Divisão do Conjunto de Dados

- Processo de separar o dataset em partes para **treinar e avaliar** o modelo
 - **Treino** → Usado para aprender
 - **Teste** → Usado para avaliar o desempenho
 - (opcional) **Validação** → Usado para ajustes

Predição

- É o **resultado estimado pelo modelo** para uma nova amostra

Exemplo → O modelo prever que um cliente *vai cancelar*

Conceitos Básicos Fundamentais em Machine Learning

Modelo

- É a **representação matemática** que aprende padrões a partir dos dados

Exemplo → Uma regressão logística treinada para prever churn

Métrica / Erro

- É uma **forma de medir quão boa é a predição** do modelo

Exemplo → Acurácia, erro médio, precisão

Treinamento

- Processo em que o modelo **ajusta seus parâmetros** a partir dos dados de treino

Consolidando o conhecimento

- Em **Aprendizado de Máquina**, o modelo aprende padrões a partir de características presentes em um dataset (com rótulos ou não), é treinado com dados conhecidos e avaliado com dados que nunca viu, gerando predições para novas instâncias.



Encerramento

- Foco de Hoje → **Entender o problema antes da técnica**
- Por enquanto, nenhum código
- Próxima aula → ***Pré-processamento de Dados e Engenharia de Características***