

A 3D rendering of a warehouse conveyor belt system. Several cardboard boxes are positioned on the belt, which is flanked by blue guides. Red laser lines are projected across the scene, creating a grid-like pattern on the floor and highlighting the boxes. The perspective is from a low angle, looking down the length of the conveyor.

Aprendizado de Máquina

Conceitos Iniciais

Daniel Sabino A. de Araújo

Volume de Dados

- 90% de todos os bytes existentes hoje foram produzidos nos últimos dois anos.
 - Mídias sociais,
 - vídeos,
 - fotos,
 - dados de geolocalização via GPS,
 - transações financeiras,
 - dados biométricos,
 - dados de sensores climáticos,
 - entre muitos outros.

Volume de Dados

CONTEXT: WHAT'S BIG DATA?

7

BIG IN GROWTH, TOO.

1 exabyte (EB) = 1,000,000,000,000,000 bytes



Onde estão/conseguimos esses dados?

- Bases de dados da própria organização: registros diários, cadastros, etc.
 - Primeira opção, mas não única!
- Segunda opção: encontrar outras fontes de dados adicionais na própria organização ou extrair informação da web.
 - Desafio!!



Desafio para trabalhar com outras fontes

Agregar dados.

Relacionar e rotular dados para conseguir trabalhar com eles.

Construir conhecimento, discernimento, noção de perspectiva.

Implementar e agregar ao processos do negócio.

Motivação para AM

- Escrever um programa de computador que:
 - Reconheça pessoas pelo rosto
 - Problemas:
 - Diferentes expressões faciais
 - Alterações na face (ex. óculos, bigode)
 - Cortes de cabelo
 - Etc.
 - Que características considerar???
 - Seres humanos: reconhecimento de padrões, aprendizado do que deve ser observado após vários exemplos

Motivação

- Escrever um programa de computador que:
 - Faça diagnóstico de pacientes por sintomas e exames
 - Médico: formação e experiência
 - Responda a questões sobre vendas como:
 - Quais produtos são vendidos em conjunto?
 - Que produto recomendar a um cliente?
 - Como agrupar clientes para melhor marketing?

Inteligência Artificial e Aprendizado de Máquina

- Necessidade de ferramentas mais autônomas
 - Reduzindo necessidade de intervenção humana e dependência de especialistas
- Aprendizado de Máquina: técnicas capazes de criar, a partir de experiência passada, uma hipótese (função) capaz de resolver o problema

Aprendizado?

Essencial para comportamento inteligente

Algumas atividades:

- Memorizar algo
- Observar e explorar situações para aprender fatos
- Melhorar habilidades motoras/cognitivas por prática
- Organizar conhecimento novo em representações apropriadas

Definição

Diz-se que um programa de computador aprende pela experiência E , com respeito a algum tipo de tarefa T e performance P , se sua performance P nas tarefas em T , na forma medida por P , melhoram com a experiência E .

A close-up photograph of two glass chess pieces on a reflective surface. In the foreground, a king piece with a cross-shaped crown is in sharp focus. Behind it and to the left, a pawn piece is slightly out of focus. The pieces are made of clear, faceted glass and are reflected on the surface below them.

Exemplos

- Xadrez
- Filtro de SPAMS
- Reconhecimento de escrita
- Conduzir veículos
- Recomendação de itens
- Autenticação biométrica
- Identificação de doenças
- muitos outros...

Inferência Indutiva

- A Inferência Indutiva é um dos principais meios para a aquisição de novos conhecimentos
- Indução: raciocínio para obter conclusões sobre todos os membros de uma classe pelo exame de alguns membros da classe
 - Raciocínio do particular para o geral

AM é interdisciplinar

Teoria da
Informação

Probabilidade e
Estatística

...

Aprendizado de Máquina

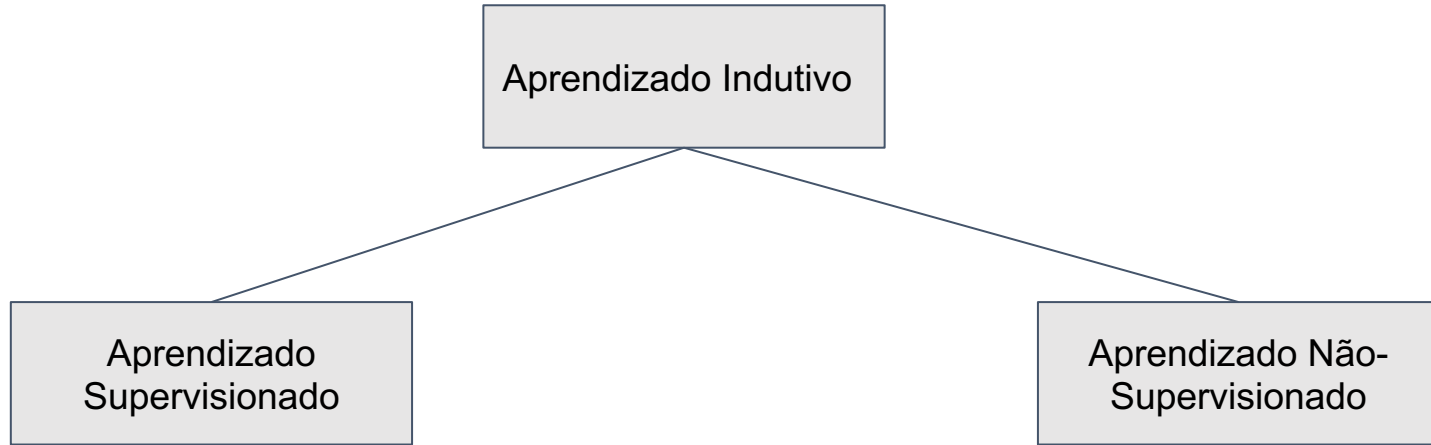
Teoria da
Computação

Neurociência

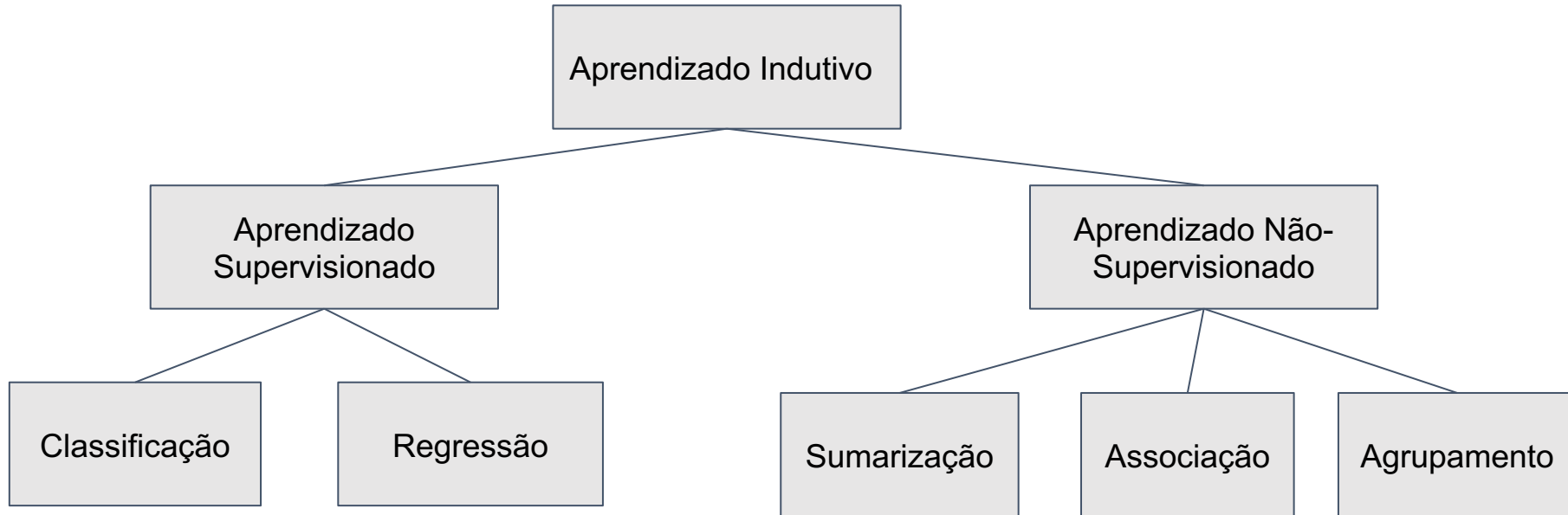
Tarefas de Aprendizado

- Previsão
 - Encontrar função (modelo ou hipótese) que possa ser utilizada para prever um rótulo ou valor para novos dados
 - Objetos na forma (entrada, saída)
- Descrição
 - Explorar ou descrever um conjunto de dados
 - Objetos não possuem saída associada

Hierarquia de Aprendizado



Hierarquia de Aprendizado



Outros aprendizados

- Aprendizado por reforço
- Aprendizado semissupervisionado
- Deep Learning
- AutoML
- ...

Viés Indutivo

- Aprendizado: busca de hipótese em espaço de possíveis hipóteses
 - que descreva relacionamentos entre os dados
 - e se ajuste aos dados de treinamento
- Todo algoritmo de AM indutivo tem um viés na escolha de uma hipótese (ou conjunto)

Sem viés, não haveria generalização (modelos seriam especializados para os exemplos individuais)

Viés Indutivo

- Viés de preferência ou busca
 - Como hipóteses são pesquisadas
 - Preferência de algumas hipóteses sobre outras
 - Ex.: preferência por hipóteses curtas
 - Navalha de Occam

"Se em tudo o mais forem idênticas as várias explicações de um fenómeno, a mais simples é a melhor"

AM: outras motivações

- Quantidade de conhecimento disponível pode ser muito grande para ser descrito (e programado) por humanos
 - Ex.: diagnóstico médico
- Algumas tarefas exigem cálculos complexos, possíveis apenas com computador
 - Ex.: interrelacionar/correlacionar grandes quantidades de dados
- Modelos podem se adaptar a novas situações

Repositórios

- UCI - <http://archive.ics.uci.edu/ml/index.html>
- Edinburgh - <http://www.inf.ed.ac.uk/teaching/courses/dme/html/datasets0405.html>
- StatLib - <http://lib.stat.cmu.edu/datasets/>
- Kdnuggets - <http://www.kdnuggets.com/competitions/index.html>