

# Aprendizagem de Máquina

## Introdução

Profa. Leticia T. M. Zoby

([leticia.zoby@udf.edu.br](mailto:leticia.zoby@udf.edu.br))

# Machine Learning - Aprendizado de Máquina

- Sendo uma área da Inteligência Artificial, Aprendizagem de Máquina -AM (*Machine Learning*) trata do desenvolvimento de algoritmos e técnicas que permitem computadores “aprender”.

Estatística



Métodos

Machine Learning



aplicados a

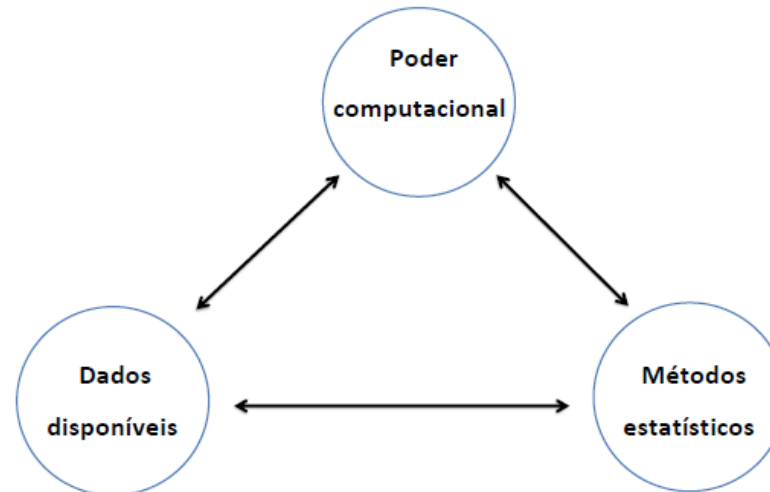
Ciência da  
computação



problemas

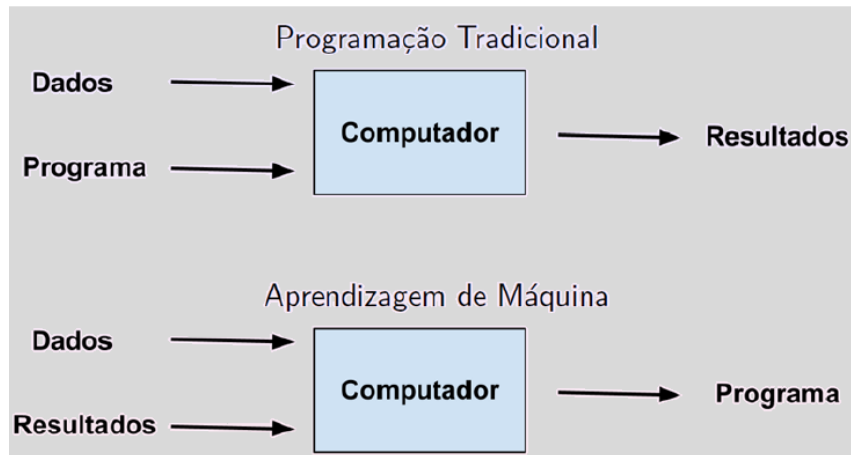
# Aprendizado de Máquina

- Definição
  - Área de estudo que dá aos computadores a habilidade de aprender sem serem explicitamente programados (Arthur Samuel, 1959).
- Interesse em capacitar os algoritmos para desenvolverem ações inteligentes a partir dos dados.



# Aprendizado de Máquina

- Definição
- AM é mais bem sucedido quando aumenta o conhecimento do especialista, ao invés de substituir
- Qualquer organização que gere ou agregue dados, provavelmente emprega ao menos um algoritmo de AM para fazer sentido aos mesmos.



# Aprendizado de Máquina

- Por que utilizar o AM?

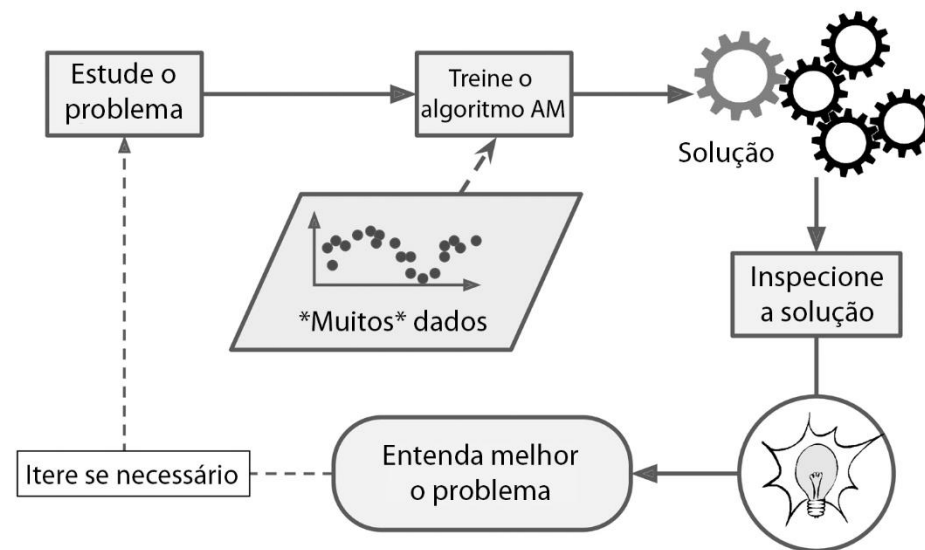


Figura 1. AM pode ajudar no ensino dos humanos (Geron, 2019)

# Aprendizado de Máquina

- Motivação: Dado um conjunto de objetos, colocar os objetos em grupos baseados na similaridade entre eles.

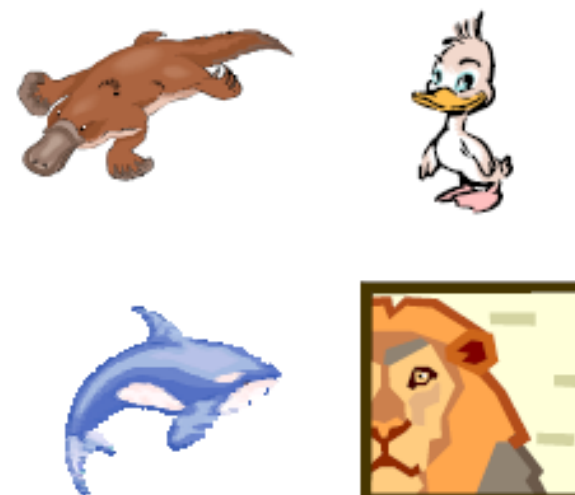


Figura 1. Grupos de Objetos (Baranauskas, 2011)

# Aprendizado de Máquina

- Motivação: Dado um conjunto de objetos, colocar os objetos em grupos baseados na similaridade entre eles.



Figura 2. Grupos de Objetos (Baranauskas, 2011)

# Aprendizado de Máquina

- Motivação: Dado um conjunto de objetos, colocar os objetos em grupos baseados na similaridade entre eles.

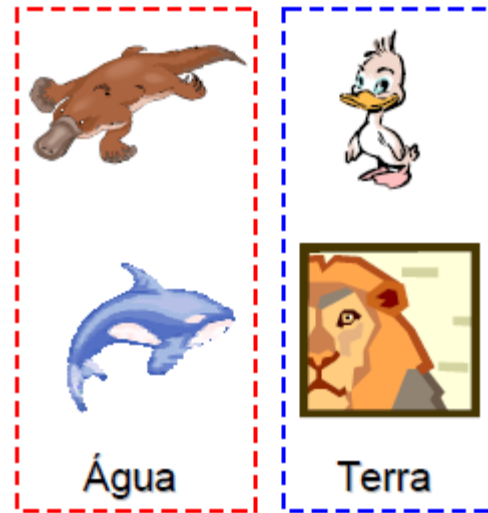


Figura 3. Grupos de Objetos (Baranauskas, 2011)



# Aprendizado de Máquina

- Motivação: Dado um conjunto de objetos, colocar os objetos em grupos baseados na similaridade entre eles.



Figura 4. Grupos de Objetos (Baranauskas, 2011)

# Aprendizado de Máquina

- Os métodos de aprendizagem de máquina tem sido empregados em problemas como:



Segmentação dos clientes para publicidade direcionada



Análise do mercado de ações  
Previsões financeiras



Deteção de fraude em cartões de crédito



Descobertas de sequências gênicas ligados a doenças

Bioinformática



Reconhecimento de objetos em visão computacional



Jogos de entretenimento



Processamento de linguagem natural  
Reconhecimento de voz e escrita



Mineração de dados em grandes



Ferramentas de busca

## Serviços

### Financeiros

- Detecção de Fraude
- Visão 360° do cliente

## Utilidades

- Smart Meter Analytics
- Gerenciamento dos Ativos

## Transporte

- otimização logística
- congestionamento do tráfego

## Mídia Digital

- Segmentação de anúncios em tempo real

# Aprendizado de Máquina

Mercado de novas possibilidades

## Saúde

- Análise de dados omni-ômicos
- Monitoramento de doenças

## Varejo

- Omni-channel Marketing
- Real-time promotions

## Telecomunicação

- Perfil de Cliente
- Análise e Otimização de redes

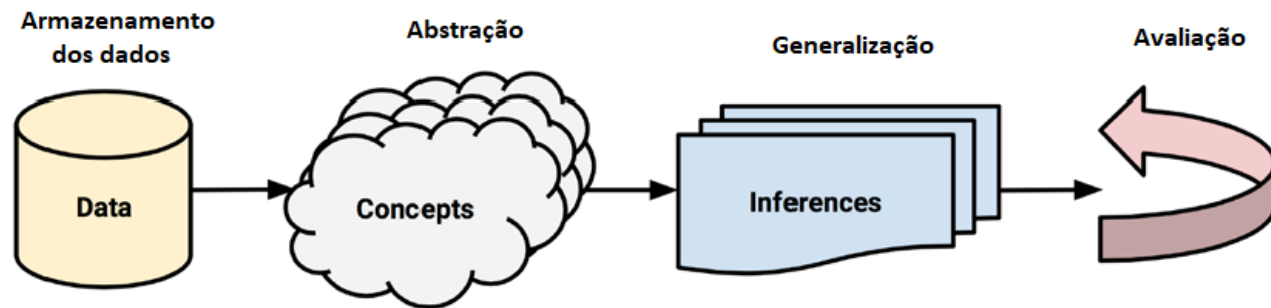
## Segurança

- Vigilância Multimodal
- Detecção de segurança cibernética

# Aprendizado de Máquina

Como as máquinas aprendem?

- ML é a ciência de descobrir estruturas e fazer previsões em conjunto de dados
- O aprendizado é efetuado a partir de raciocínio sobre exemplos fornecidos por um processo externo ao sistema de aprendizado



# Aprendizado de Máquina

## Resolução de problemas computacionais - Desafios

- Desenhar algoritmos eficientes
- Garantir que o algoritmo está correto
- Fazer implementações eficientes
- Garantir que a implementação do algoritmo está correta

# Aprendizado de Máquina

E quando não temos ou sabemos um algoritmo?

- Como descrever formalmente a relação a seguir:



⇒ 'a'



⇒ 's'



⇒ 'A'



⇒ 'L'



⇒ 'd'

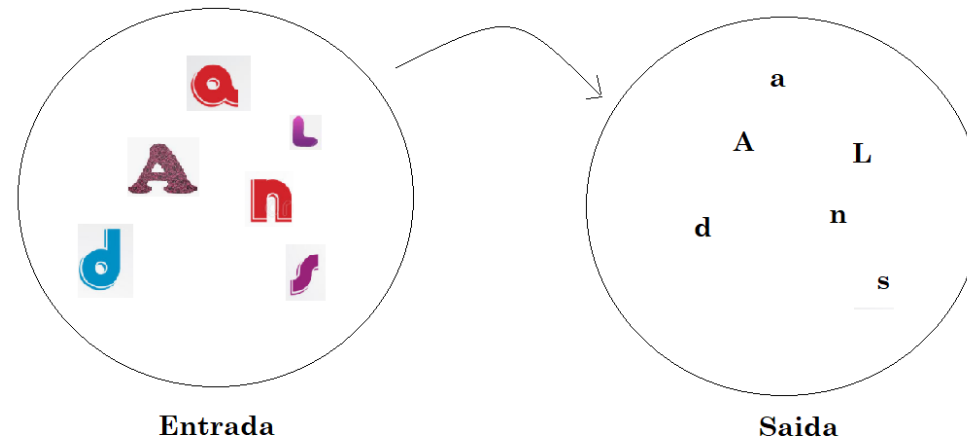


⇒ 'n'

# Aprendizado de Máquina

E quando não temos ou sabemos um algoritmo?

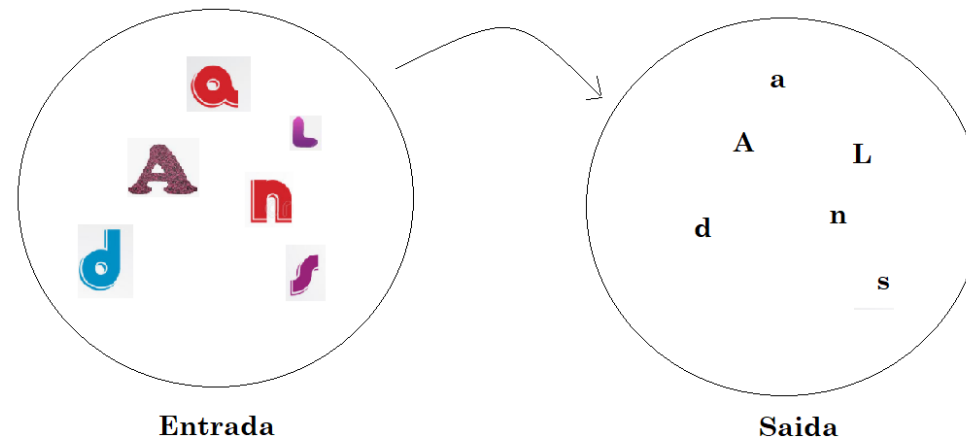
- E se já caracterizar a relação entrada-saída por meio de exemplo, ainda que a relação seja ambígua:



# Aprendizado de Máquina

## AM atuando...

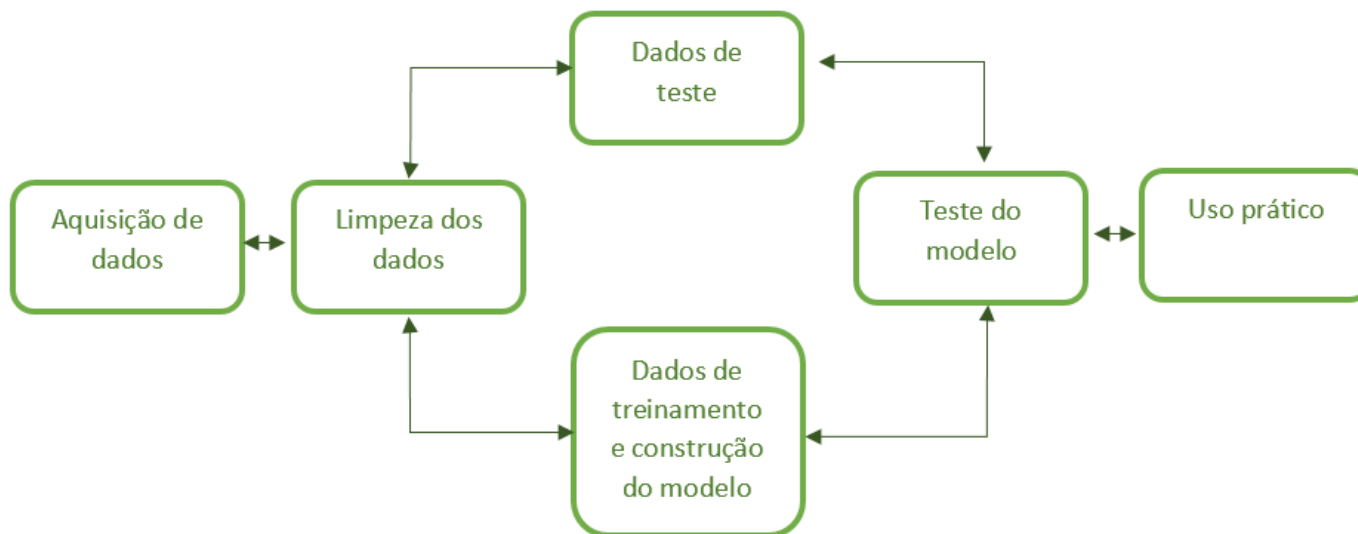
- Ideia principal do AM: “aprender” a relação a partir de exemplos
- Produto final do AM: um programa que realiza o processamento entrada-saída





# Aprendizado de Máquina

## ► Processo básico do ML



# Aprendizado de Máquina

- ▶ Atributos

  - ▶ Atributos ou Classes

  - ▶ Características

  - ▶ Tipos

    - ▶ Nominal (ou discreto ou categórico)

      - ▶ Ordenado

      - ▶ Não-ordenado

    - ▶ Contínuo (ou numérico ou real)

- ▶ Classes

  - ▶ Atributos especial que descreve o que se deseja/ interesse: Rótulo  
(somente em Aprendizado Supervisionado)

# Aprendizado de Máquina

- ▶ Conjunto de Exemplos
  - ▶ Conjunto de treinamento
  - ▶ Conjunto de teste
  - ▶ Conjunto de validação\*

Geralmente, usa-se 80% do dados para treinamento e 20% dos dados para teste.

# Aprendizado de Máquina

- ▶ Formação do Modelo:
  - ▶ Aprendizado supervisionado
  - ▶ Aprendizado não supervisionado
  - ▶ Aprendizado por reforço

# Aprendizado de Máquina

## Teste e Validação

- ▶ Teste em novos casos
  - ▶ Conjunto de treinamento
  - ▶ Conjunto de testes

Geralmente, usa-se 80% dos dados para treinamento e 20% dos dados para teste.

- ▶ Conjunto de validação

# Aprendizado de Máquina

Formação do Modelo:  
**Aprendizado supervisionado**

- ▶ Há parâmetros rotulados que são usados para construir o modelo e tentar prever os demais **rótulos**, baseados apenas nos parâmetros

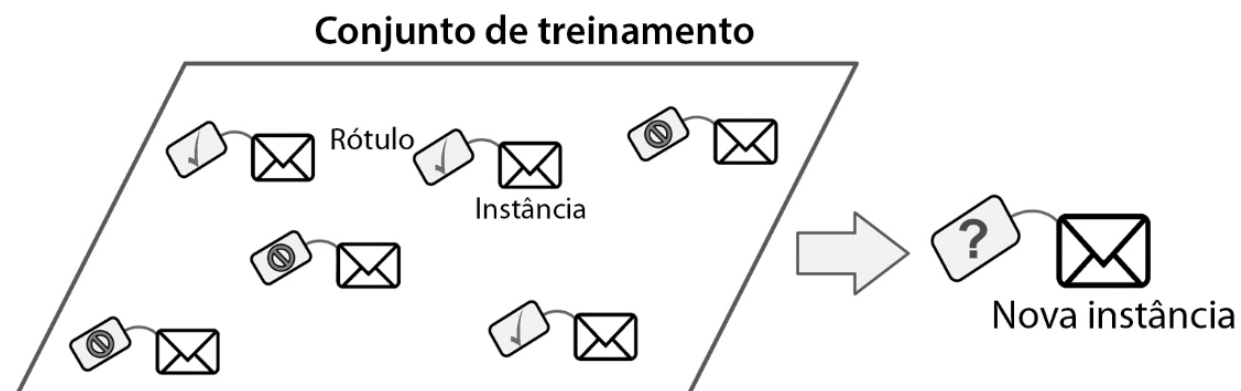


Figura 2. Conjunto de treinamento rotulado – Classificação SPAM (Geron, 2019)

# Aprendizado de Máquina

Formação do Modelo: **Aprendizado supervisionado**

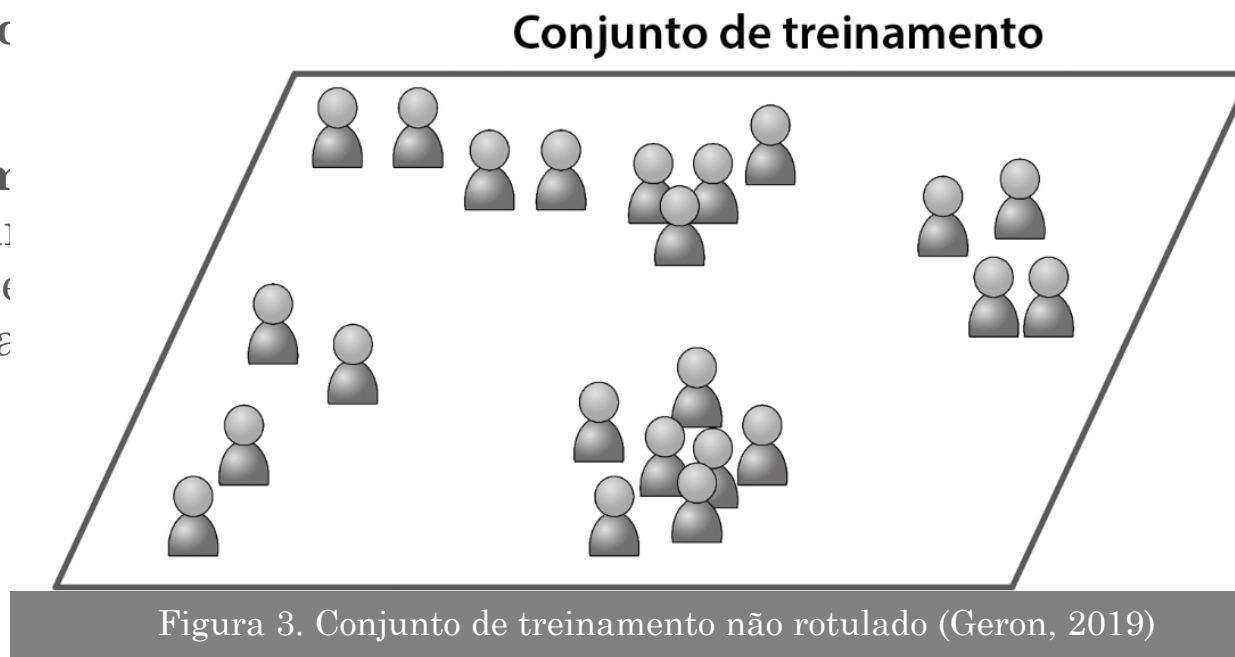
► Principais algoritmos:

- Vizinhos mais próximos (KNN – *K-Nearest Neighbours*)
- Regressão Linear
- Regressão Logística
- Máquina de Vetores de Suporte (SVM- *Support Vector Machines*)
- Árvores de Decisão e Floresta Aleatórias
- Redes Neurais

# Aprendizado de Máquina

Formação do Modelo: **Aprendizado não supervisionado**

- Há apenas os parâmetros, **sem rótulos**, e quer encontrar subgrupos dentro dos dados que possuam algum tipo de semelhança





# Aprendizado de Máquina

Formação do Modelo: **Aprendizado não supervisionado**

► Principais algoritmos:

► Clustering

► K-Means

► Visualização e redução da dimensionalidade

► Análise de Componentes Principais (PCA - *Principal Component Analysis*)

► Aprendizado da regra da associação

► Apriori

# Aprendizado de Máquina

Formação do Modelo: **Aprendizado por reforço**

- ▶ Algoritmos que aprendem a executar ações baseados em experiências do mesmo com algum meio
- ▶ Sistema de aprendizado -> agente

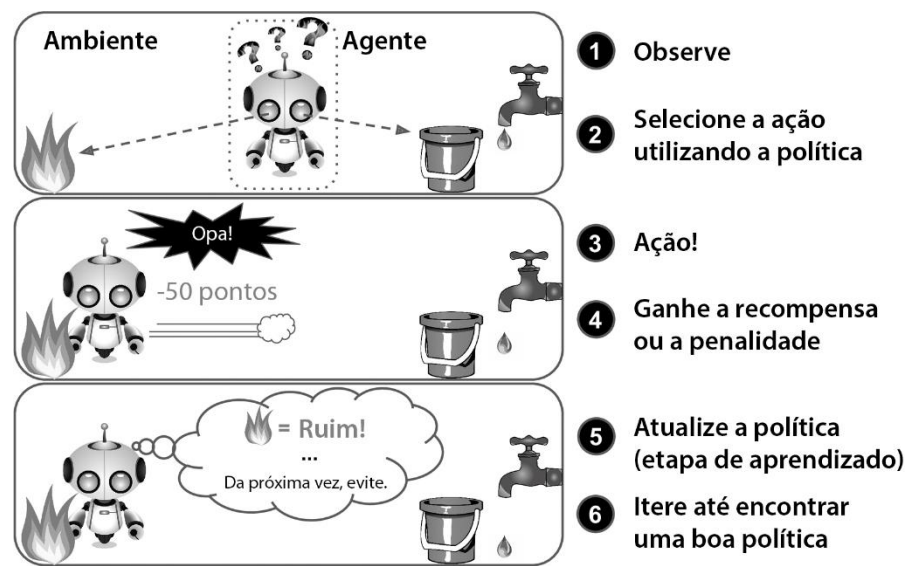


Figura 4. Aprendizado por reforço (Geron, 2019)

# Aprendizado de Máquina

## Principais desafios do AM

- ▶ Quantidade Insuficiente de Dados de Treinamento
- ▶ Dados de Treinamento não Representativos
- ▶ Dados de Baixa Qualidade
- ▶ Características Irrelevantes
- ▶ Sobreajustando os Dados de Treinamento
- ▶ Subajustando os Dados de Treinamento

# Pré Processamento

- ▶ Dados reais em geral são:
  - ▶ incompletos: faltam valores e/ou atributos (Salário = “”).
  - ▶ ruidosos: contêm erros e/ou outliers (Salário = -150).
  - ▶ inconsistentes: apresentam discrepâncias em códigos e nomes (1→A, 2→B, 3→C).
- ▶ Problemas causados por humanos, softwares, problemas de hardware, dados de diferentes fontes, entre outros.
- ▶ Preparar os dados pode consumir mais de 80% do esforço de modelagem.

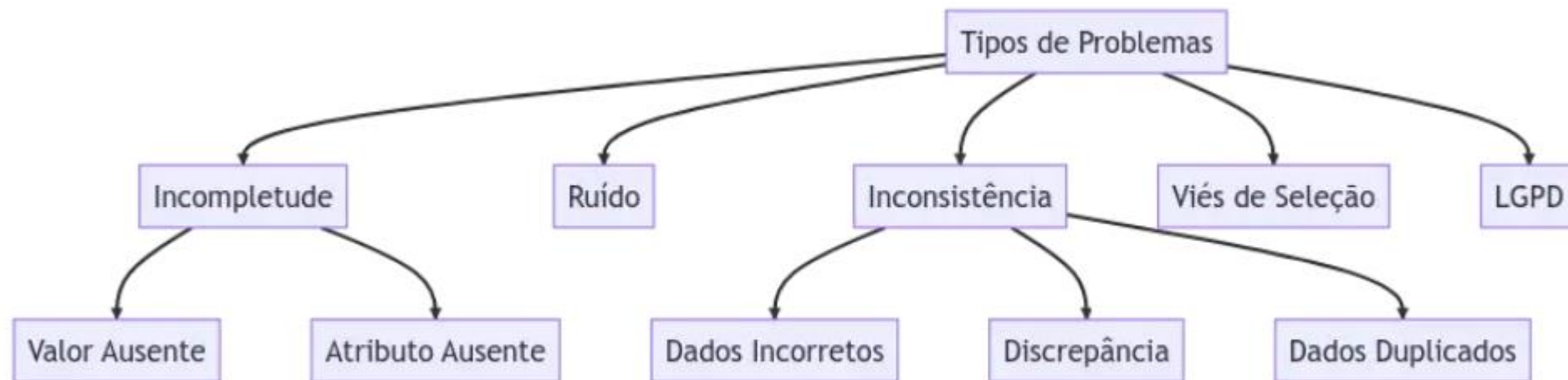
# Aprendizado de Máquina

O que há de errado nessa base?

Usuário	Idade	Cidade	Data	Salário	Npáginas	Nsessões	Nprodutos	Clique	Comprou
user1	22	Belo Horizonte	22/10/21	1200	5	2	5	no	no
user2	1	São paulo	20/09/22	5000	21	4	11	yes	yes
user 3	41	BH	23 de dezembro 2022	12000	2	1	3	NA	yes
user 4	32	Miami	22/10/01	50000	NA	5	2	yes	no
user5	20	Tokyo	23/01/25	12220	5	5	8	no	yes
user6	28	NY	March 14, 2018	250000	8	NA	NA	no	yes

# Aprendizado de Máquina

## Principais problemas



# Pré Processamento

## Atividades comuns

- ▶ Limpeza de dados: lidar com valores ausentes, dados ruidosos, outliers, inconsistências etc.
- ▶ Integração de dados de múltiplos bancos de dados e arquivos.
- ▶ Transformação de dados: normalização e agregação.
- ▶ Redução de dados: menos dados com mesmos resultados analíticos (seleção de atributos e de amostras).
- ▶ Discretização: compactar informação.
- ▶ Nunca deixar de fazer análise exploratória (médias, medianas, variâncias, min, max, gráficos etc).

# Pré Processamento

## VALORES AUSENTES

### ► Ocorrência comum:

- Mau funcionamento de dispositivos de coleta de dados;
- Dado omitido pela fonte de informação numa pesquisa;
- Falha na digitação ou na composição da base;

### ► Formas de eliminação de valores ausentes:

- Eliminar registros/atributos com valores ausentes;
  - Perda de dados pode ser considerável.

### ► Preenchimento de valores (imputação)

- Por uma constante. Ex.: Média/Moda do atributo
  - Desconsidera a relação entre atributos da base de dados
- Por valores que tentem preservar as relações entre atributos da base de dados
  - Uso de um algoritmo de aprendizado



# Referências

- ▶ Baranauskas, J. A. (2011). Notas de aula “Aprendizado de Máquina Conceitos e Definições”. USP.
- ▶ Géron, A. (2017). Hands-On Machine Learning with Scikit-Learn and TensorFlow. O'Reilly.
- ▶ Russel, S.; Norvig, P. Inteligência Artificial - Uma Abordagem Moderna. 4a. ed. GEN LTC Editora, 2022
- ▶ Silveira, G., Bullock, B. (2017). “Machine Learning- Introdução à classificação”. Casa do Código, Brasil
- ▶ Tan, P. Steinbach, M., Kumar, V. (2018). “Introduction to Data Mining”. 2 Ed. Pearson.