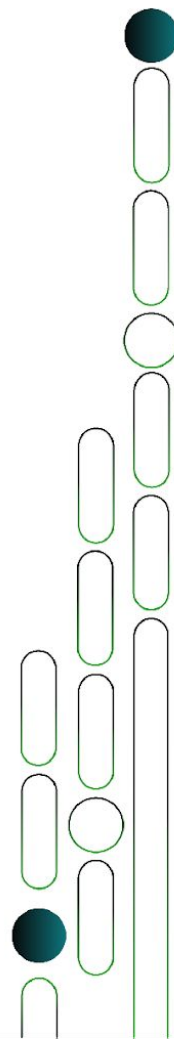


Python Avançado

Capítulo 1. *Machine Learning*

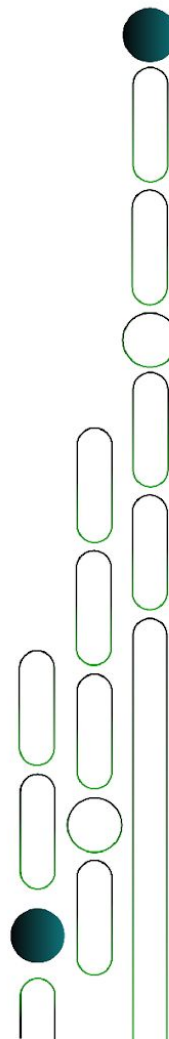
Aula 1.1. O cenário do aprendizado de máquina.

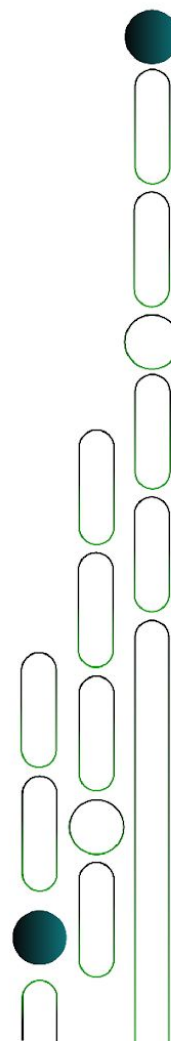
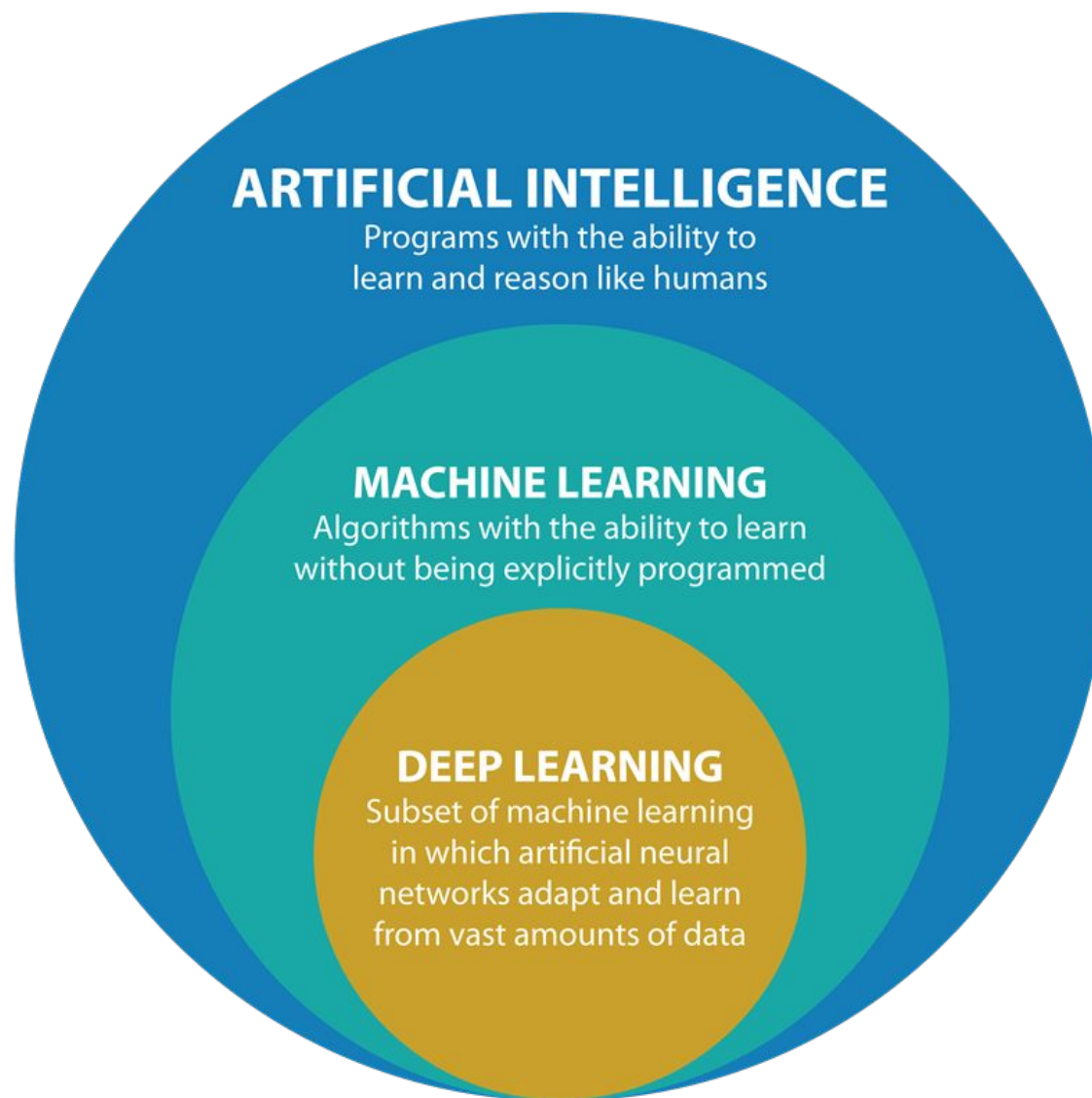
Prof. Rennan Alves



Nesta Aula

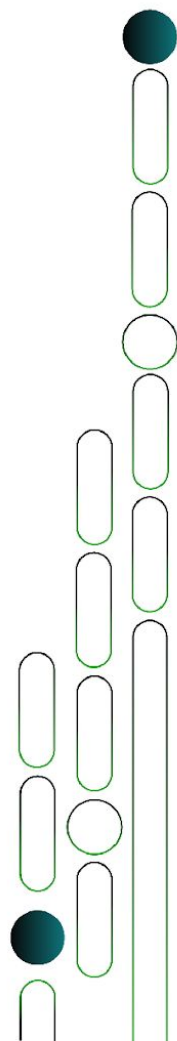
- ❑ O que é Aprendizado de Máquina?
- ❑ Por que utilizar o Aprendizado de Máquina?
- ❑ Tipos de Sistemas do Aprendizado de Máquina.
- ❑ Principais desafios do Aprendizado de Máquina.





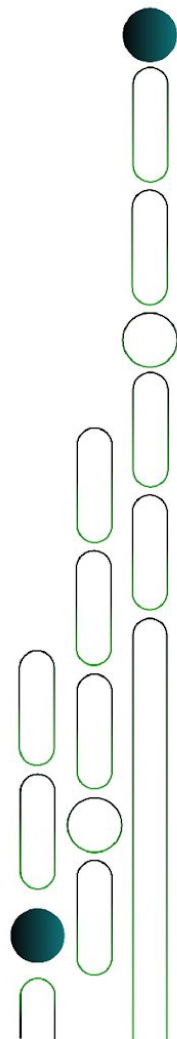
Introdução

- O Cenário do Aprendizado de Máquina:
 - Reconhecimento Ótico de Caracteres (ORC).
 - Filtro de Spam.
- Por onde começa e termina o Aprendizado de Máquina?



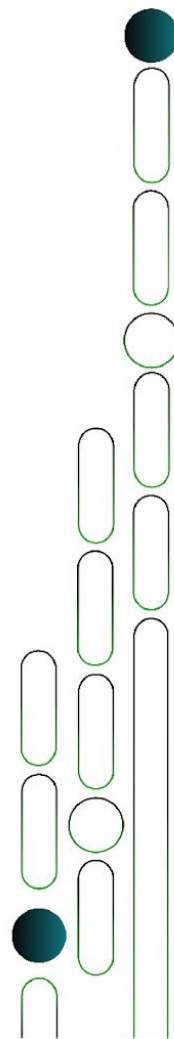
O que é o Aprendizado de Máquina?

Aprendizado de Máquina é a ciência (e a arte) da programação de computadores para que eles possam aprender com os dados.



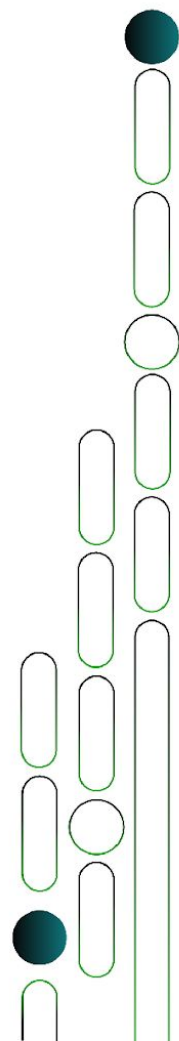
Aprendizado de maquina é o campo de estudo que dá aos computadores a habilidade de aprender sem ser explicitamente programado.

- Arthur Samuel, 1959

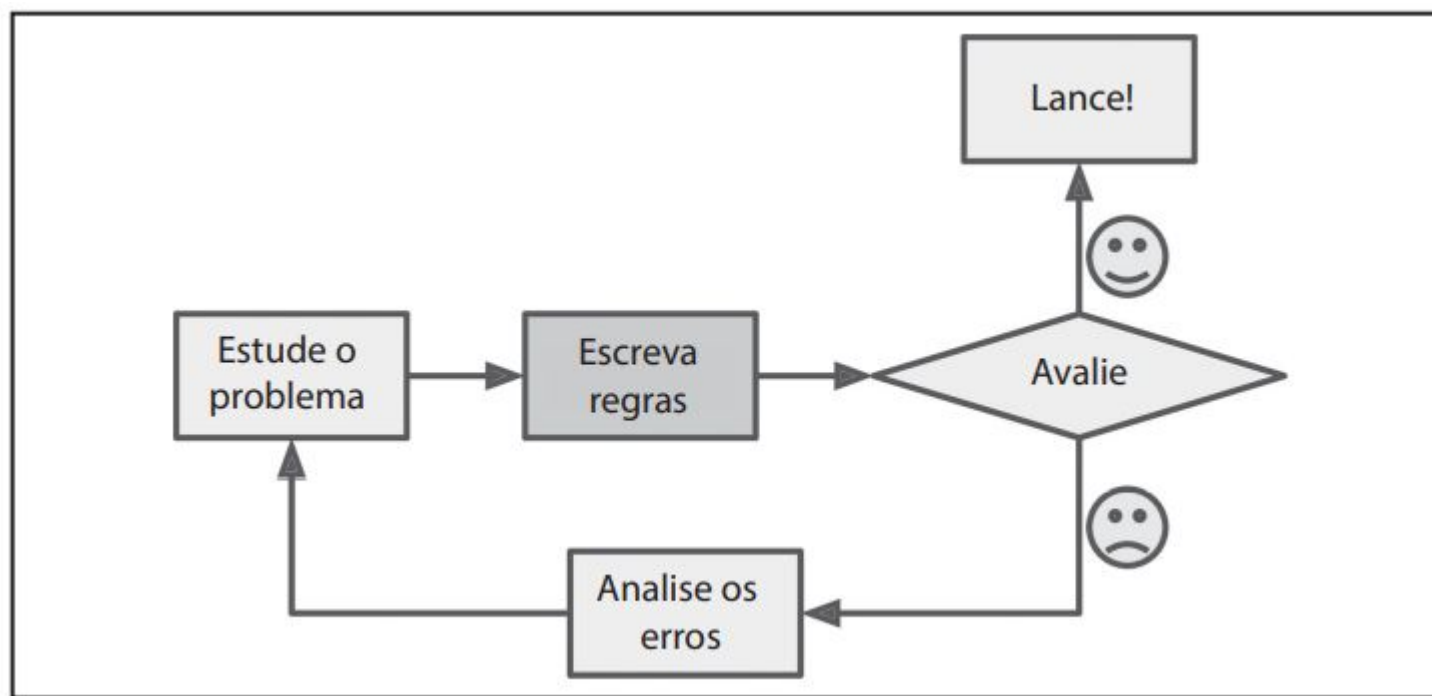


Introdução

- O Cenário do Aprendizado de Máquina:
 - Reconhecimento Ótico de Caracteres (ORC).
 - Filtro de Spam.
- Por onde começa e termina o Aprendizado de Máquina?

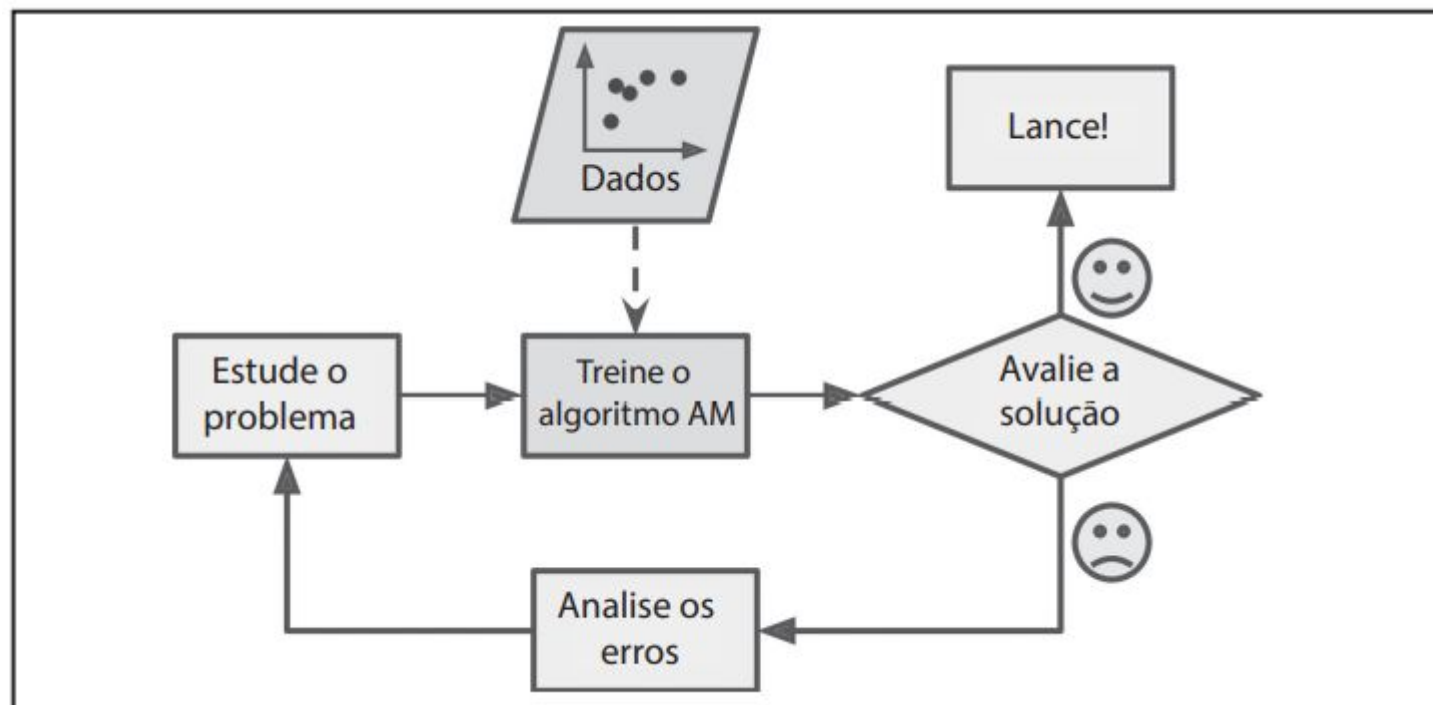


Por que utilizar Aprendizado de Máquina?



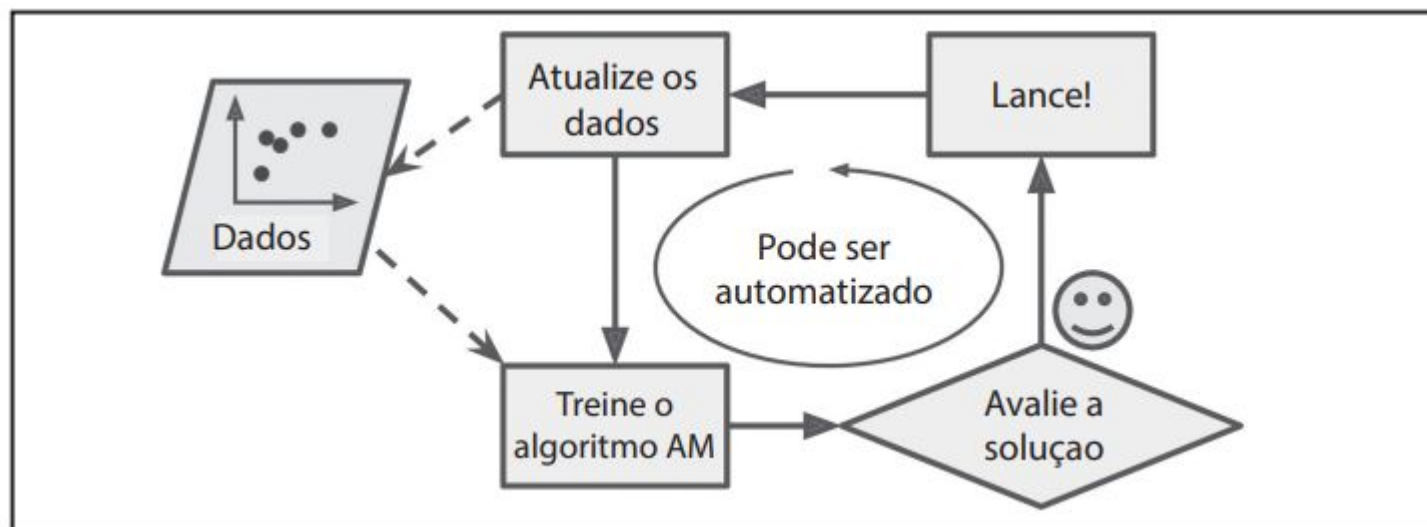
Abordagem Tradicional, Gerón (2019)

Por que utilizar Aprendizado de Máquina?



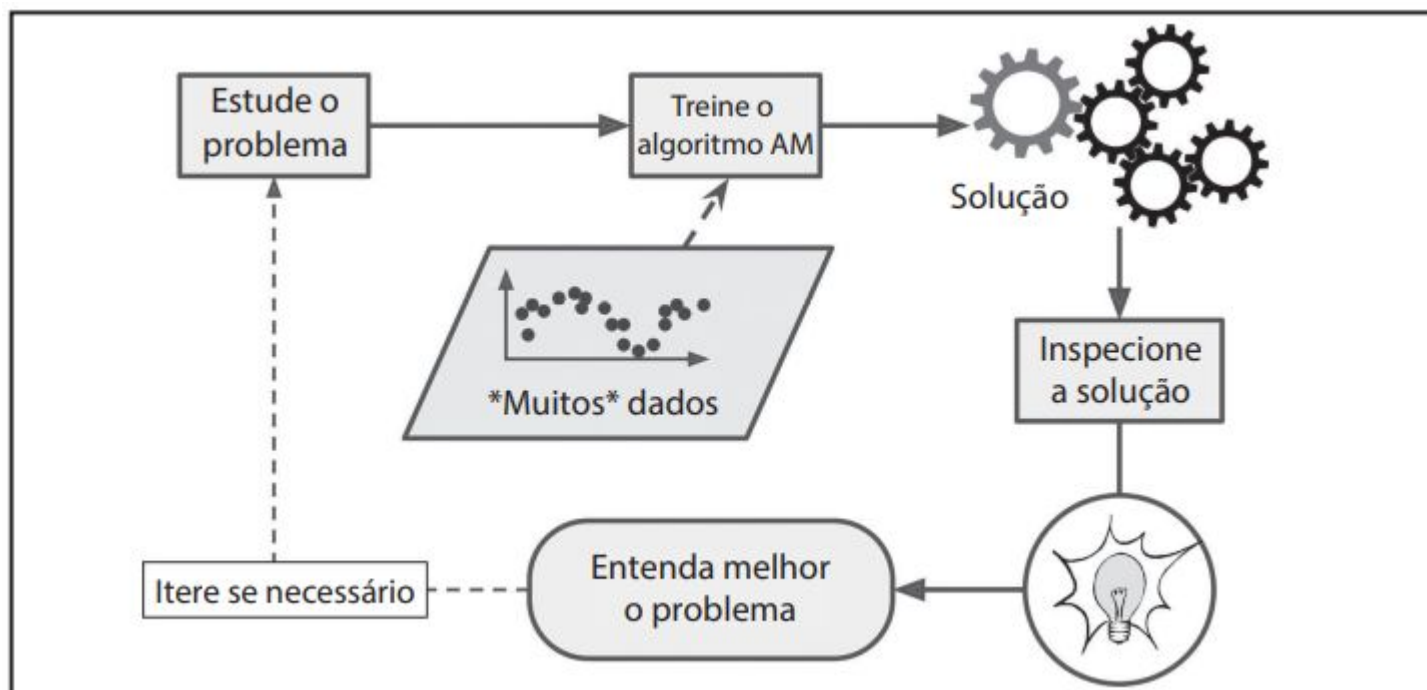
Abordagem de Aprendizado do ML, Gerón (2019)

Por que utilizar Aprendizado de Máquina?



Adaptando-se automaticamente à mudança, Gerón (2019)

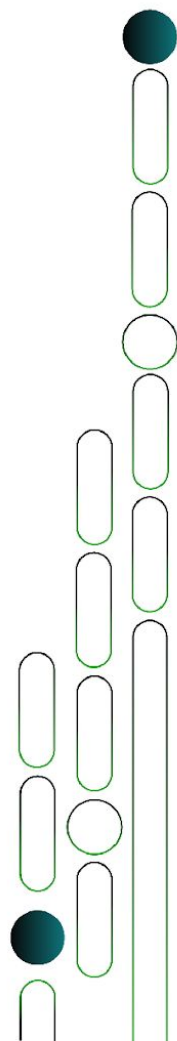
Por que utilizar Aprendizado de Máquina?



O Aprendizado de Máquina pode ajudar no ensino dos humanos, Gerón (2019)

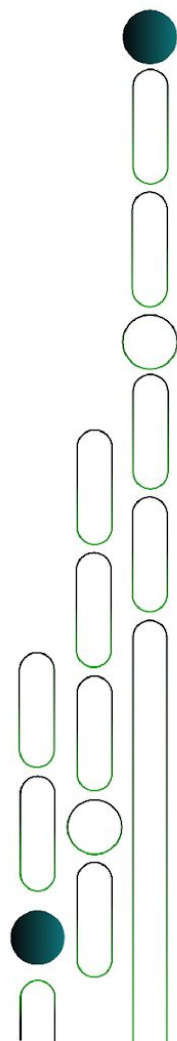
Resumindo, o Aprendizado de Máquina é ótimo para:

- Problemas para os quais as soluções existentes exigem muita configuração manual ou longas listas de regras: um algoritmo de Aprendizado de Máquina geralmente simplifica e melhora o código.
- Problemas complexos para os quais não existe uma boa solução quando utilizamos uma abordagem tradicional: as melhores técnicas de Aprendizado de Máquina podem encontrar uma solução.
- Ambientes flutuantes: um sistema de Aprendizado de Máquina pode se adaptar a novos dados.
- Compreensão de problemas complexos e grandes quantidades de dados.



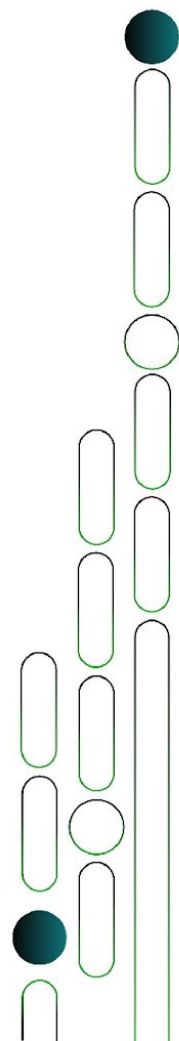
Conclusão

- O que é Aprendizado de Máquina?
- Por que utilizar o Aprendizado de Máquina?
- Tipos de Sistemas do Aprendizado de Máquina.
- Principais desafios do Aprendizado de Máquina.



Próxima aula

- ❑ Tipos de Sistemas do Aprendizado de Máquina.

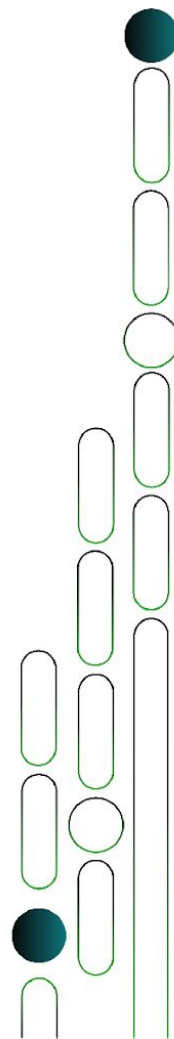


Python Avançado

Capítulo 1. *Machine Learning*

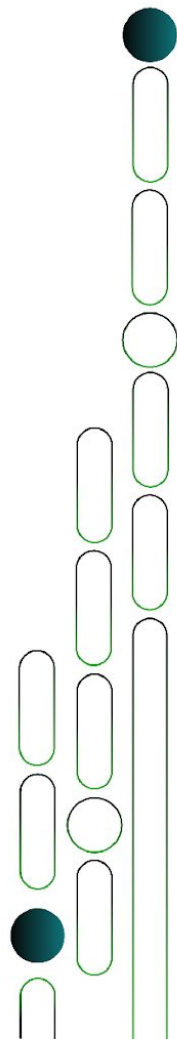
Aula 1.2. Tipos de Sistemas do Aprendizado de Máquina.

Prof. Rennan Alves



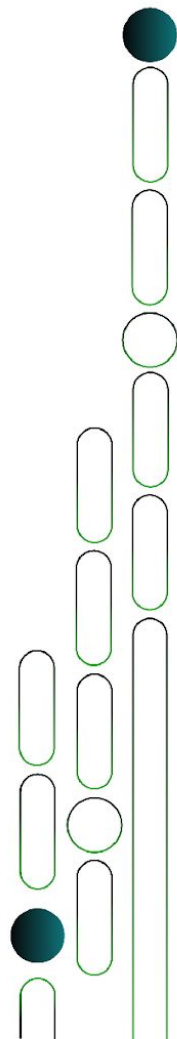
Nesta Aula

- ☐ O que é Aprendizado Supervisionado.
- ☐ O que é Não Supervisionado.
- ☐ O que é Aprendizado Semi-Supervisionado.
- ☐ O que é Aprendizado por Reforço.



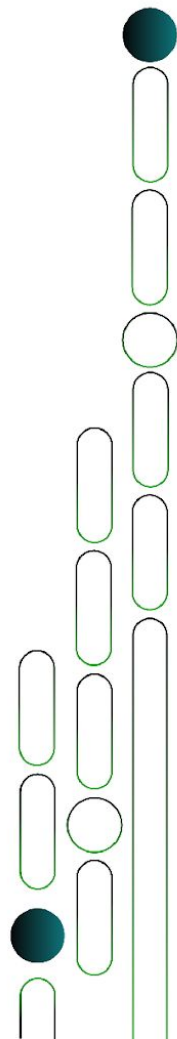
Tipos de Sistemas do Aprendizado de Máquina

- Serem ou não treinados com supervisão humana
(supervisionado, não supervisionado, semi-supervisionado e aprendizado por reforço);
- Se podem ou não aprender rapidamente, de forma incremental (aprendizado on-line versus aprendizado por lotes).



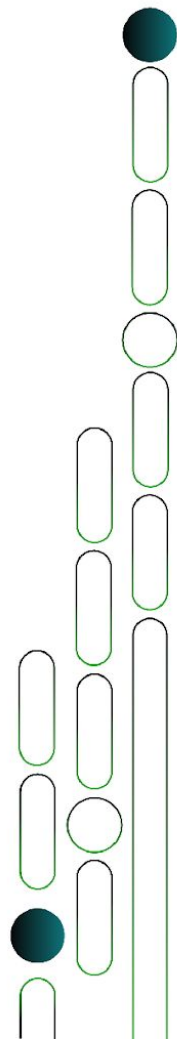
Tipos de Sistemas do Aprendizado de Máquina

- Se funcionam simplesmente comparando novos pontos de dados com pontos de dados conhecidos, ou se detectam padrões em dados de treinamento e criam um modelo preditivo, como os cientistas (aprendizado baseado em instâncias versus aprendizado baseado em modelo). Por onde começa e termina o Aprendizado de Máquina?



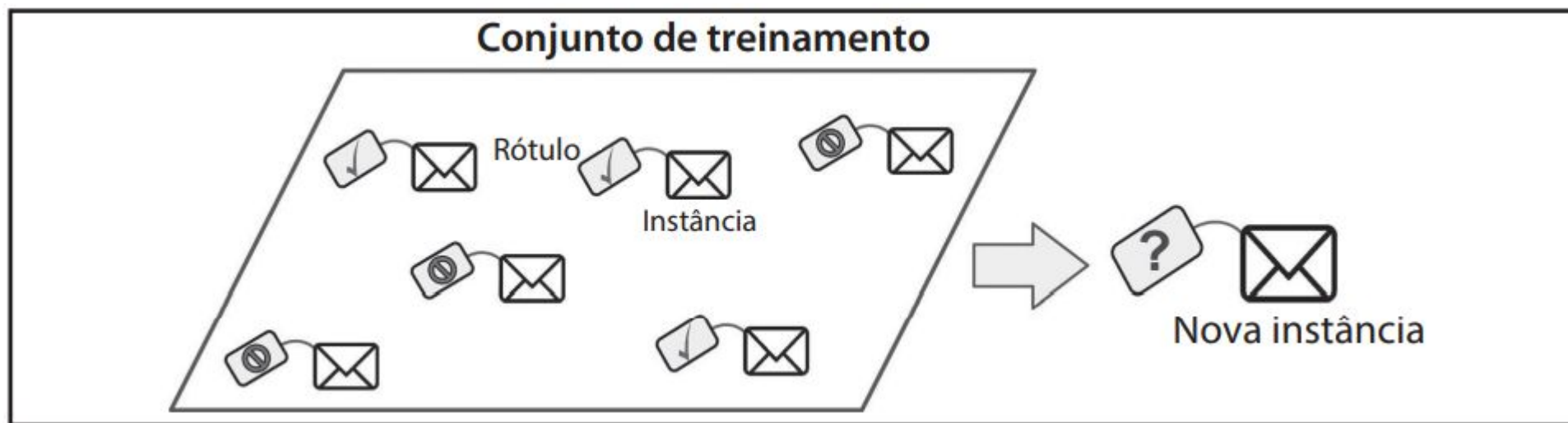
Tipos de Aprendizado de Máquina

Existem quatro categorias principais de aprendizado:
supervisionado, não supervisionado, semi-supervisionado e por
reforço.

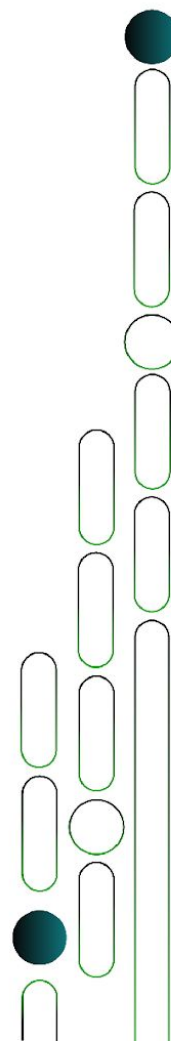


Aprendizado Supervisionado

Um conjunto de treinamento rotulado para aprendizado supervisionado (por exemplo, classificação de spam),

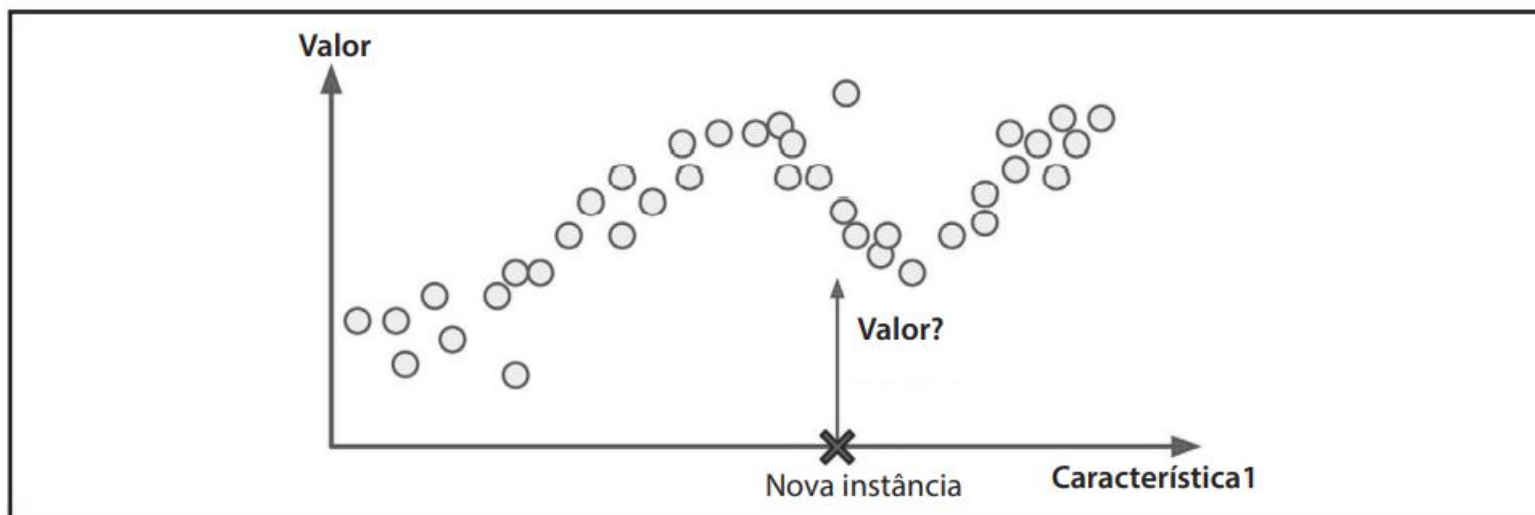


Fonte: Gerón (2019)



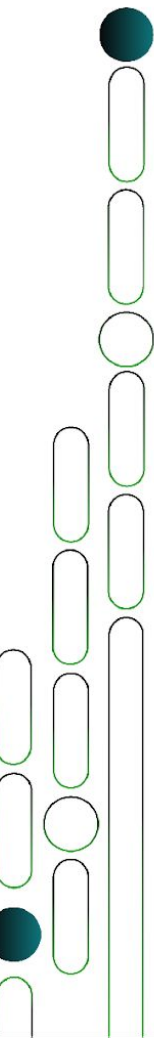
Aprendizado Supervisionado

Prever um alvo de valor numérico é outra tarefa típica, como o preço de um carro a partir de um conjunto de características (quilometragem, idade, marca etc.) denominadas previsores. Esse tipo de tarefa é chamada de **regressão**. Para treinar o sistema, você precisa fornecer muitos exemplos de carros, incluindo seus previsores e seus labels (ou seja, seus preços).



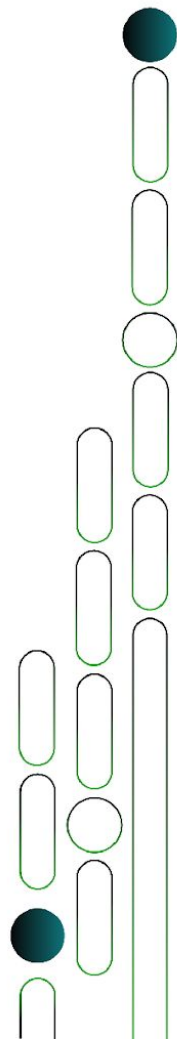
Fonte: Gerón (2019)

Curiosidade: O estranho nome REGRESSÃO é um termo de estatística introduzido por Francis Galton, enquanto estudava o fato de que os filhos de pessoas altas tendem a ser mais baixos do que os pais. Como as crianças eram mais baixas, ele chamou essa alteração de regressão à média. Este nome foi aplicado aos métodos utilizados por ele para analisar as correlações entre variáveis.

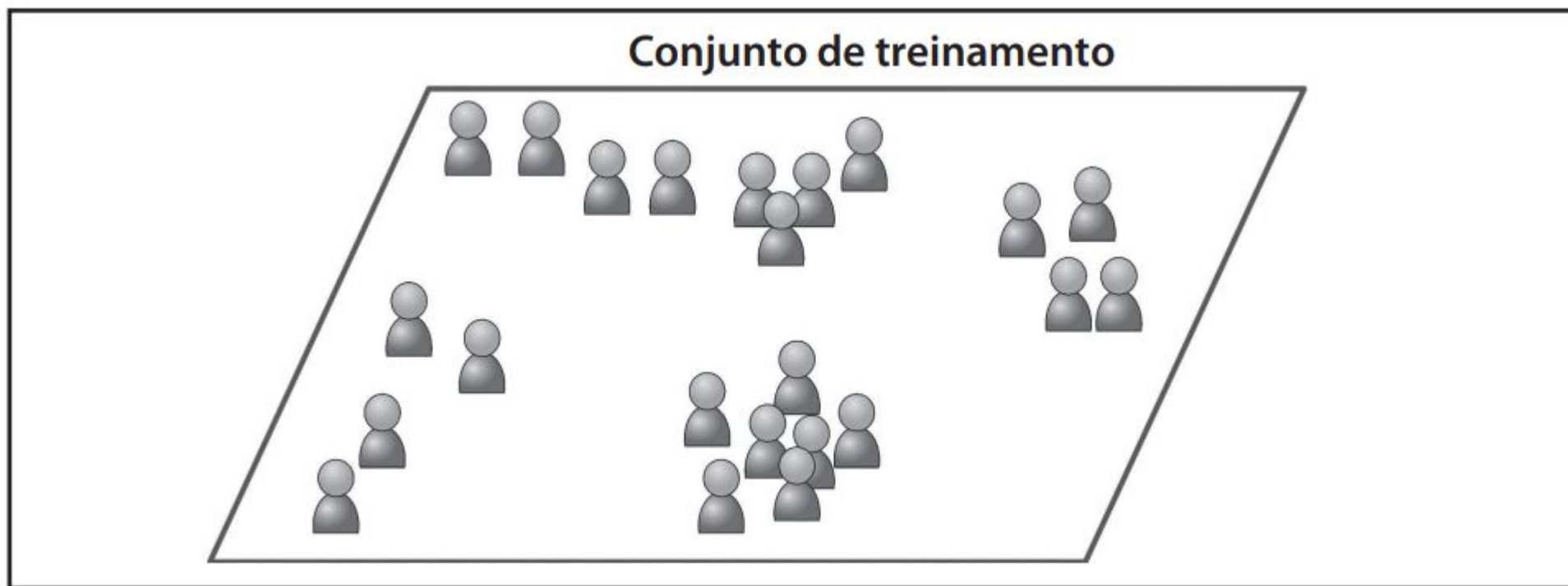


Alguns dos algoritmos mais importantes do aprendizado supervisionado

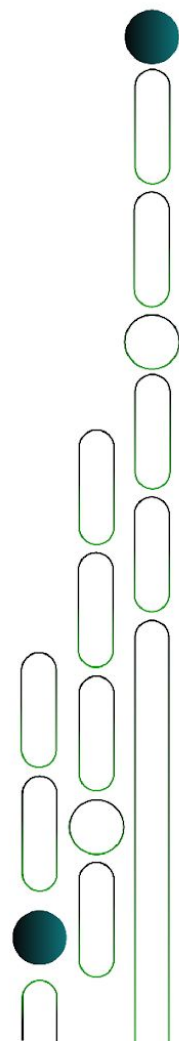
- k-Nearest Neighbours
- Regressão Linear
- Regressão Logística
- Máquinas de Vetores de Suporte (SVM)
- Árvores de Decisão e Florestas Aleatórias
- Redes Neurais*



Aprendizado Não Supervisionado

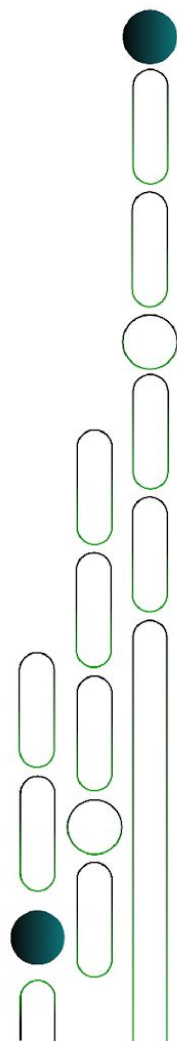


Conjunto de treinamento não rotulado para aprendizado não supervisionado, Gerón (2019)

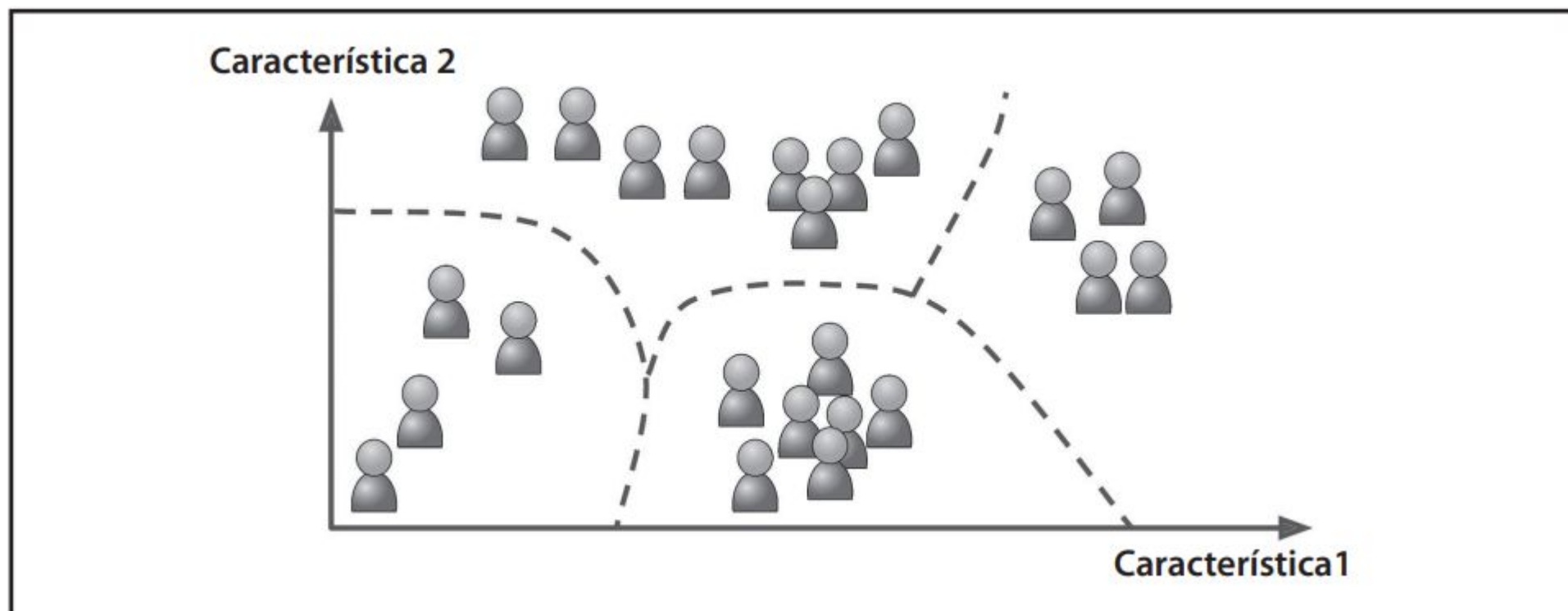


Alguns dos algoritmos mais importantes do aprendizado supervisionado

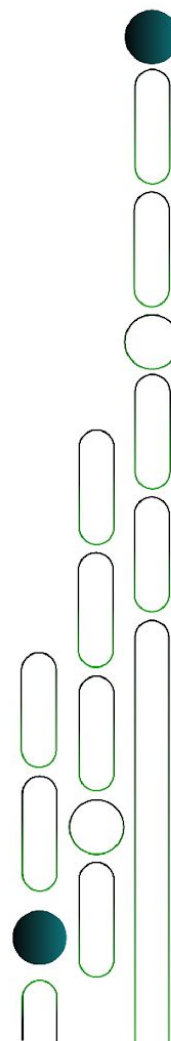
- Clustering
- Visualização e redução da dimensionalidade
- Aprendizado da regra da associação
- Redes Neurais Artificiais



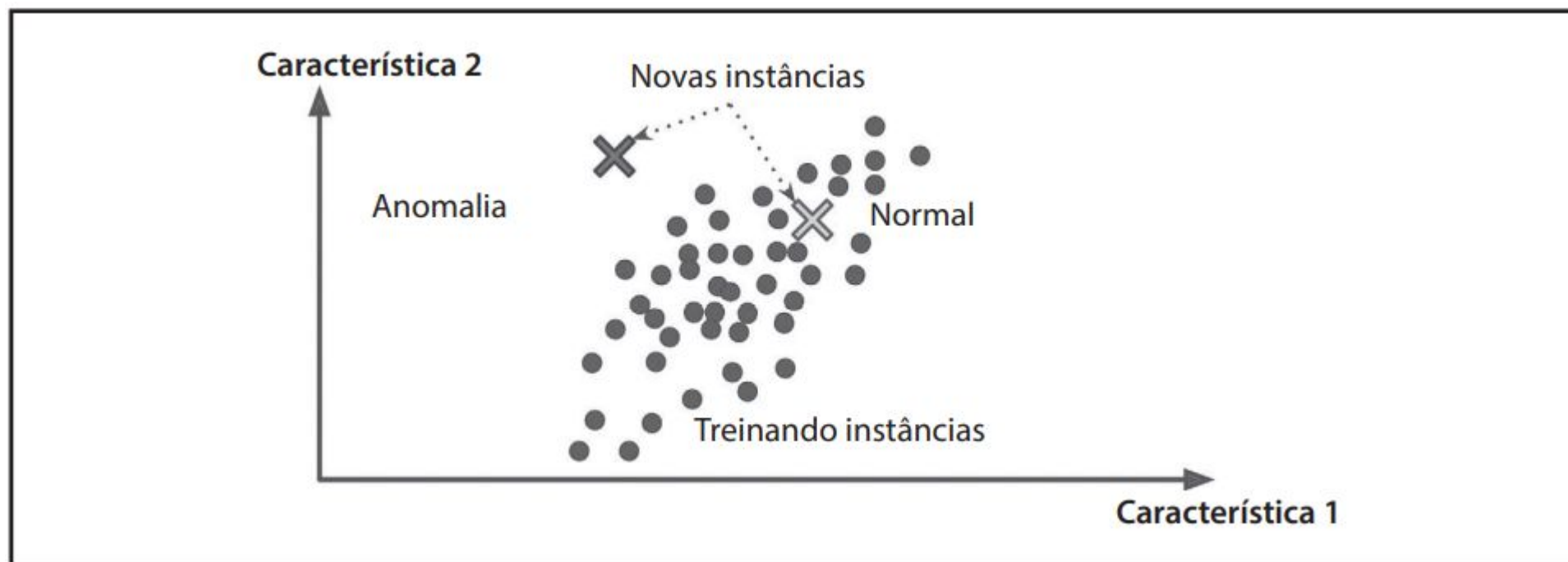
Por que utilizar Aprendizado de Máquina?



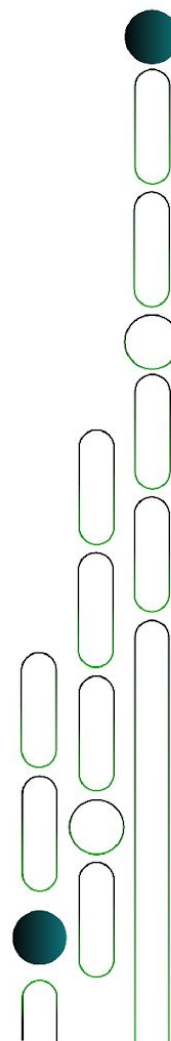
Gerón (2019)



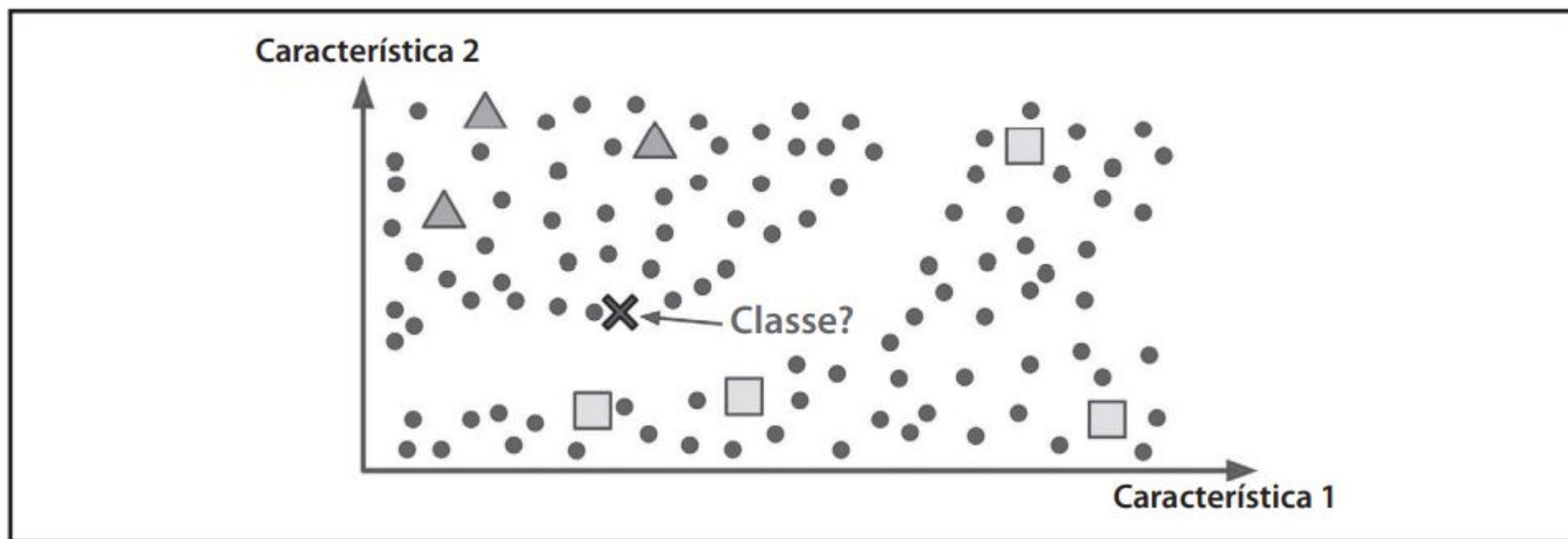
Detecção de anomalias



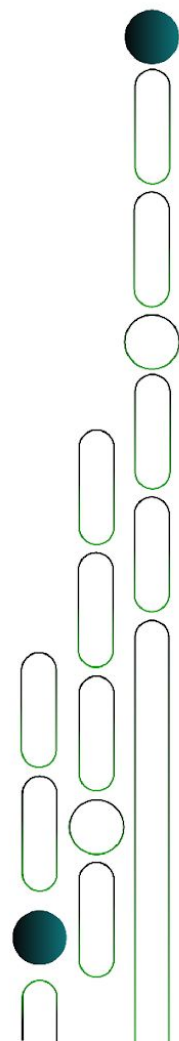
Gerón (2019)



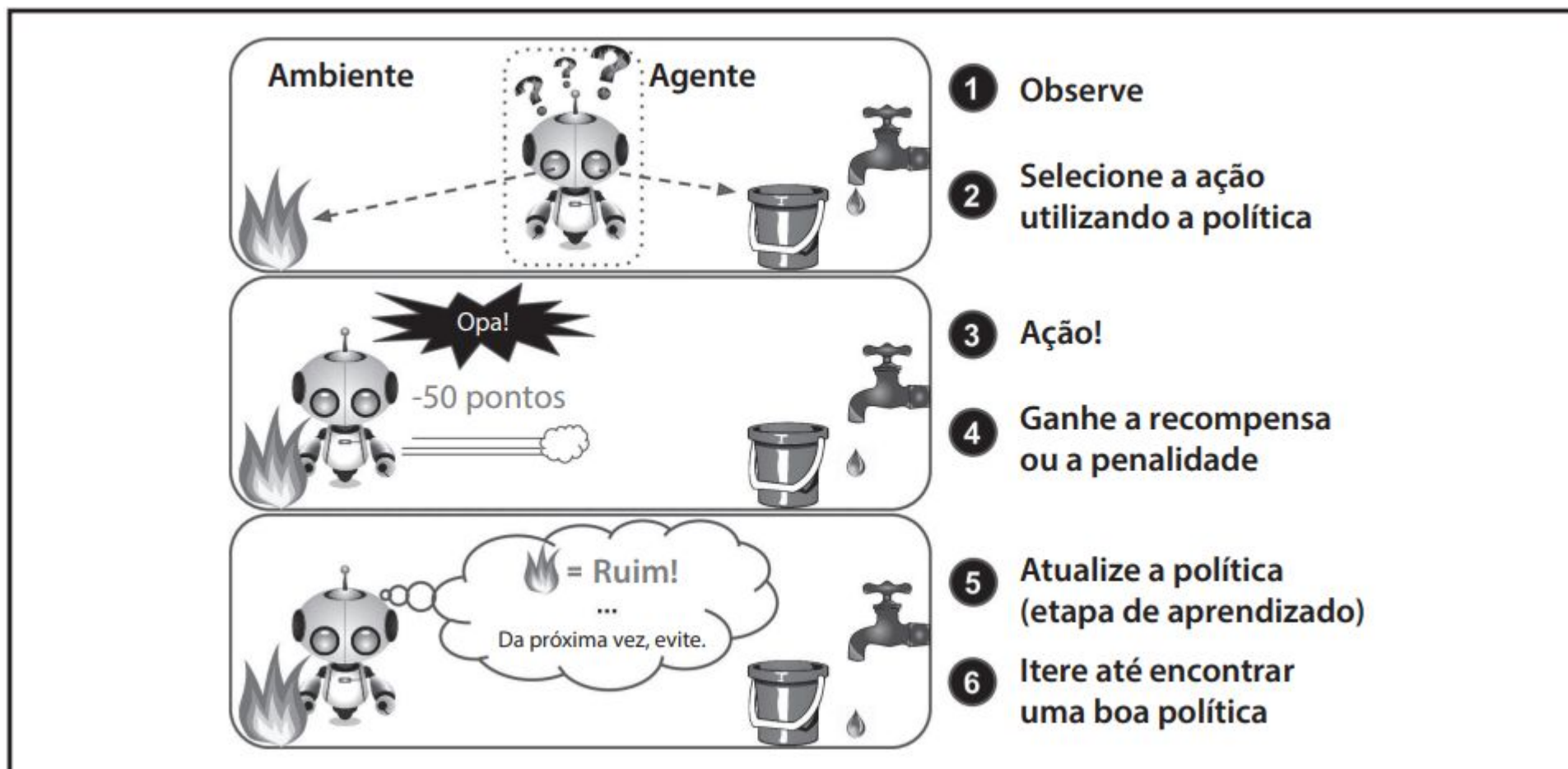
Aprendizado Semi-supervisionado



Gerón (2019)

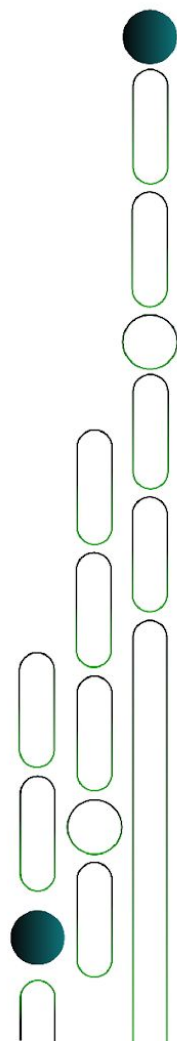


Aprendizado por Reforço



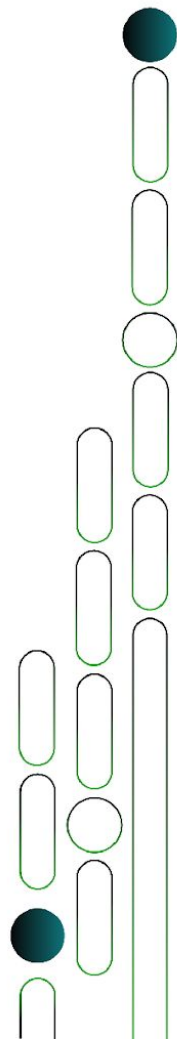
Conclusão

- ❑ O que é Aprendizado Supervisionado.
- ❑ O que é Não Supervisionado.
- ❑ O que é Aprendizado Semi-Supervisionado.
- ❑ O que é Aprendizado por Reforço.



Próxima aula

- ❑ Principais Desafios do Aprendizado de Máquina.

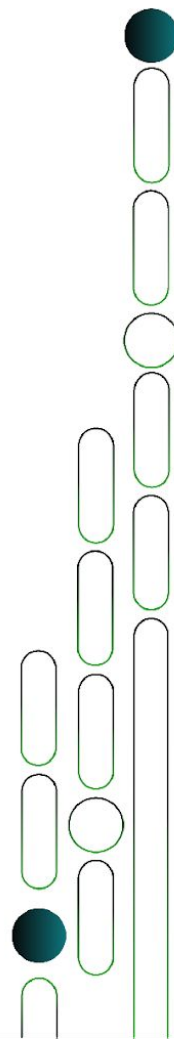


Python Avançado

Capítulo 1. *Machine Learning*

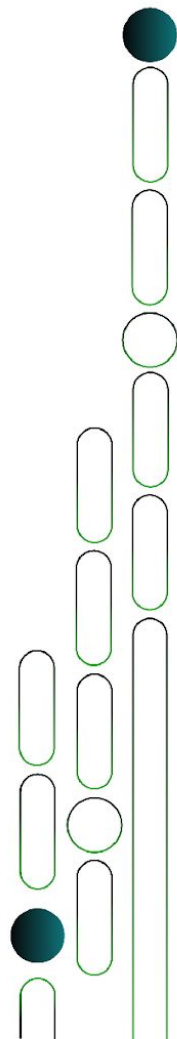
Aula 1.3. Principais Desafios do Aprendizado de Máquina.

Prof. Rennan Alves



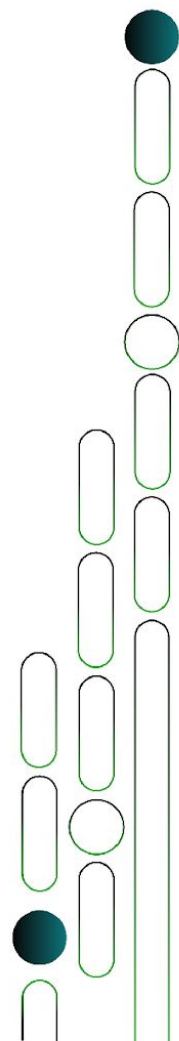
Nesta Aula

- ❑ Principais desafios sobre Aprendizado de Máquina.
- ❑ Quantidade Insuficiente de Dados de Treinamento.
- ❑ Dados de Baixa Qualidade.
- ❑ Testando e Validando.



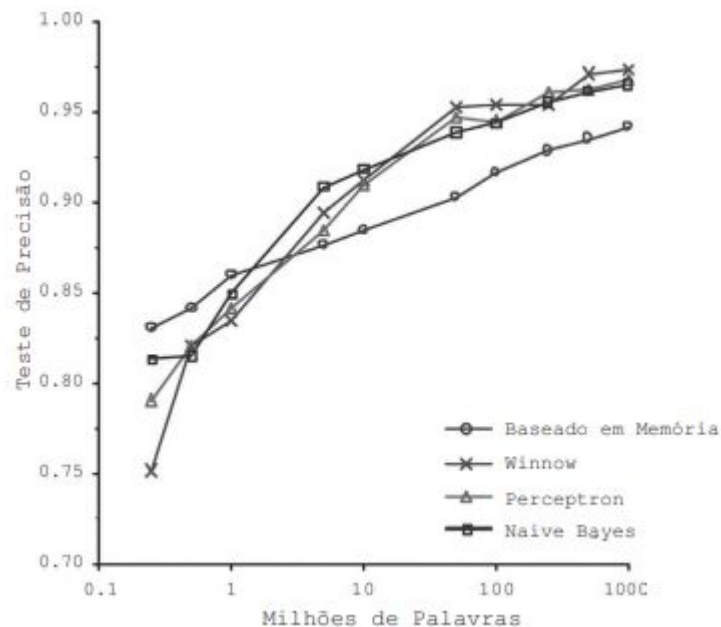
Principais Desafios do Aprendizado de Máquina

- Em suma, uma vez que a sua tarefa principal é selecionar um algoritmo de aprendizado e treiná-lo em alguns dados, as duas coisas que podem dar errado são: “**algoritmos ruins**” e “**dados ruins**”. Começemos com exemplos de dados ruins.



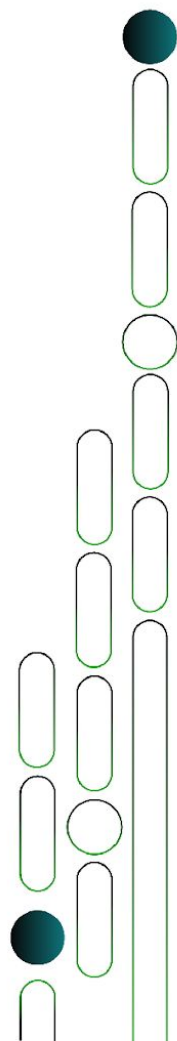
Quantidade Insuficiente de Dados de Treinamento

- Para que uma criança aprenda o que é uma maçã, é preciso que você aponte para uma maçã e diga “maçã” (possivelmente repetindo algumas vezes esse procedimento).



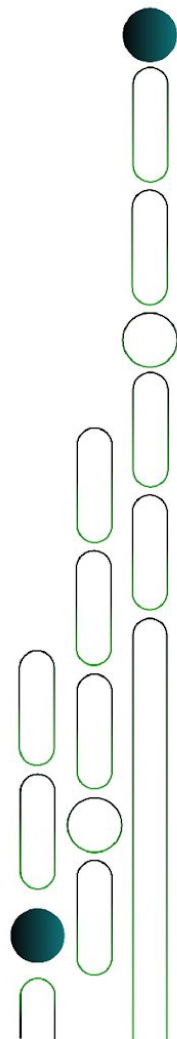
Dados de Treinamento Não Representativos

- Ao utilizar um conjunto de treinamento não representativo, treinamos um modelo que dificilmente fará previsões precisas.
- É crucial utilizar um conjunto de treinamento representativo nos casos em que desejamos generalizar.



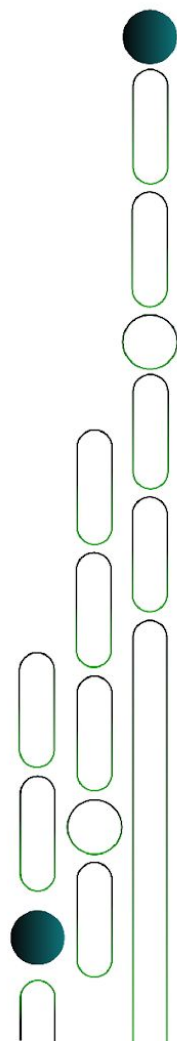
Dados de Baixa Qualidade

- Se algumas instâncias são claramente outliers, isso pode ajudar a descartá-las ou tentar manualmente a correção dos erros.
- Se faltam algumas características para algumas instâncias, você deve decidir se deseja ignorar completamente esse atributo, se deseja ignorar essas instâncias, preencher os valores ausentes, ou treinar um modelo com a característica e um modelo sem ela, e assim por diante.



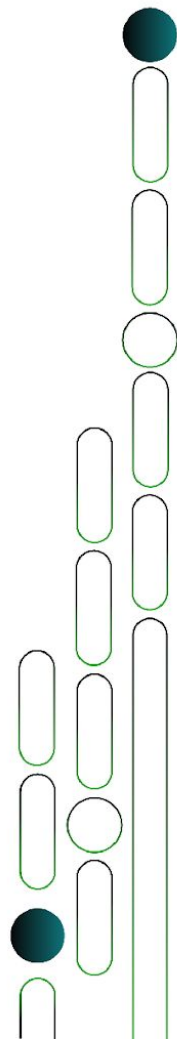
Características Irrelevantes

- Entra lixo, sai lixo.
- Uma parte crítica do sucesso de um projeto de Aprendizado de Máquina é criar um bom conjunto de características para o treinamento, processo chamado de feature engineering que envolve:
 - Seleção das características.
 - Extração das características.
 - Criação de novas características ao coletar novos dados.

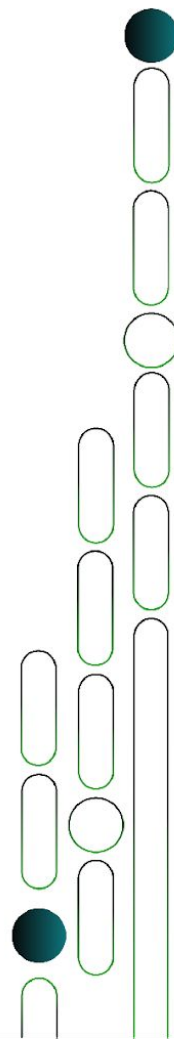


Testando e Validando

- A única maneira de saber o quão bem um modelo generalizará em novos casos é de fato testá-lo em novos casos.
- A taxa de erro em novos casos é chamada de erro de generalização (ou erro fora da amostra) e, ao avaliar seu modelo no conjunto de teste, você obtém uma estimativa desse erro.

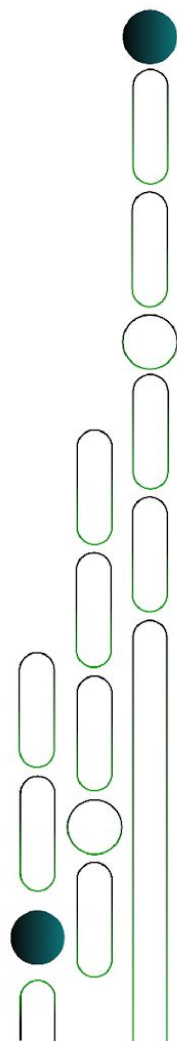


É comum utilizar 80% dos dados para treinamento e reservar 20% para o teste.



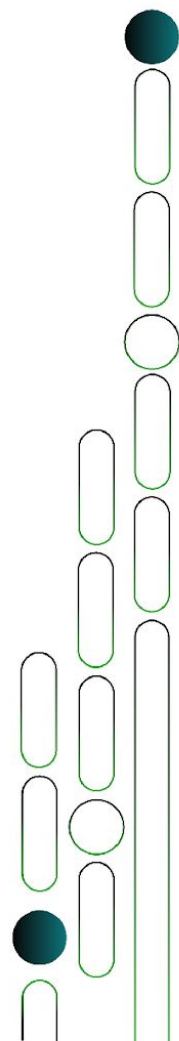
Conclusão

- ❑ Principais desafios sobre Aprendizado de Máquina.
- ❑ Quantidade Insuficiente de Dados de Treinamento.
- ❑ Dados de Baixa Qualidade.
- ❑ Testando e Validando.



Próxima Aula

- Abordaremos conceitos sobre regressão linear.

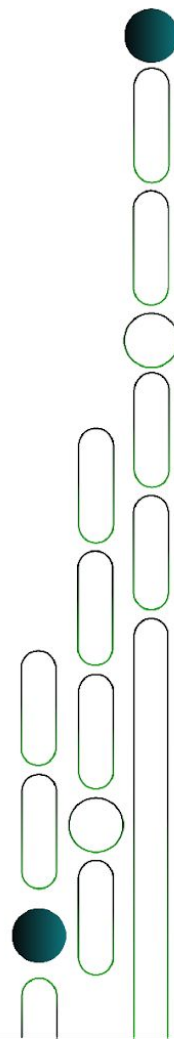


Python Avançado

Capítulo 1. *Machine Learning*

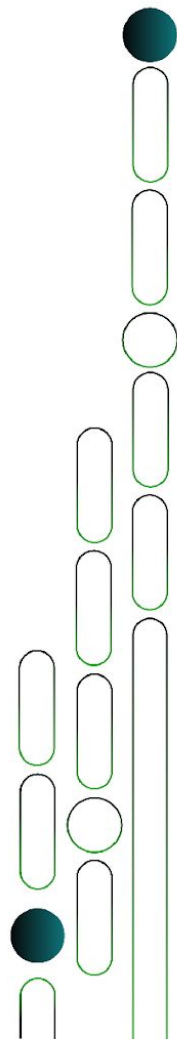
Aula 1.4. Regressão Linear

Prof. Rennan Alves



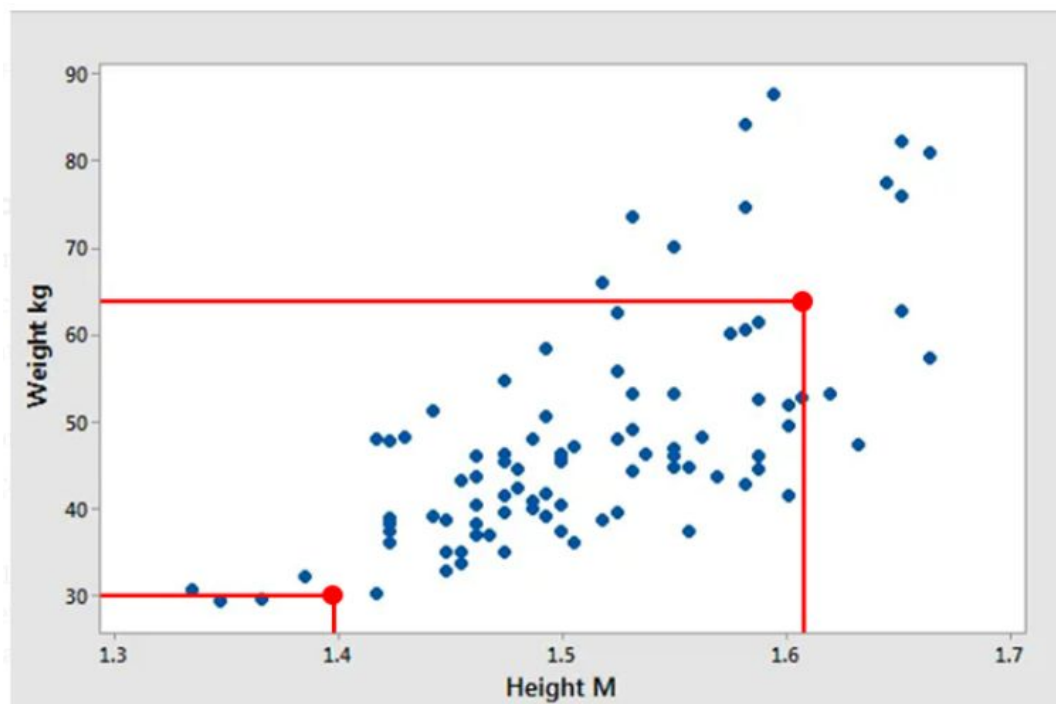
Nesta aula

- ☐ O que é Regressão ?
- ☐ Correlação.
- ☐ Interpretando a Correlação.
- ☐ Definição de Regressão.
- ☐ Regressão Linear.

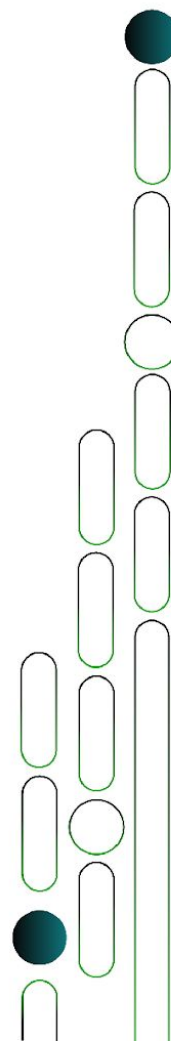


O que é regressão?

- Observação de Características.
- Variáveis Dependentes e Independente.

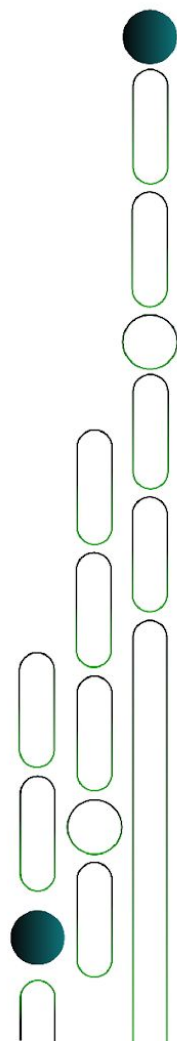


Peso x Altura - Adaptado - <https://statisticsbyjim.com/basics/correlations/>



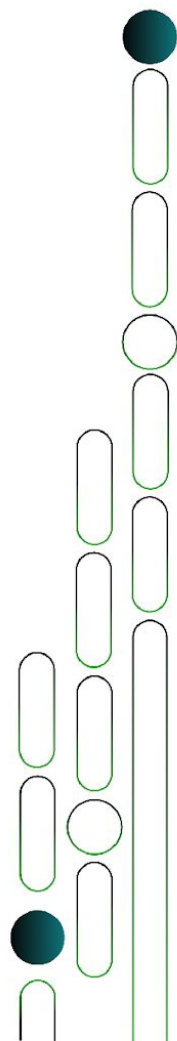
Correlação

- A correlação entre duas variáveis indica a que o valor da variável dependente y altera de acordo com a variável independente x , vamos seguir com um exemplo.
- O peso (variável dependente y) de uma pessoa, varia de acordo com a sua altura (variável independente x).



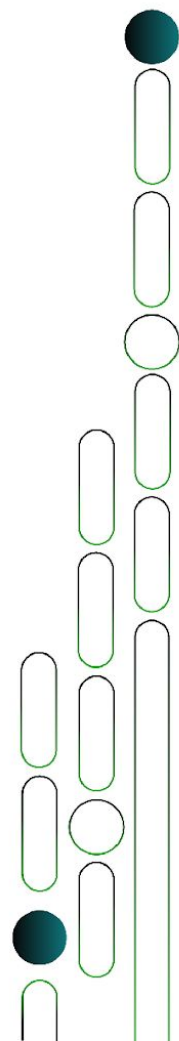
Correlação

- A correlação entre duas variáveis indica a que o valor da variável dependente y altera de acordo com a variável independente x , vamos seguir com um exemplo.
- O peso (variável dependente y) de uma pessoa, varia de acordo com a sua altura (variável independente x).

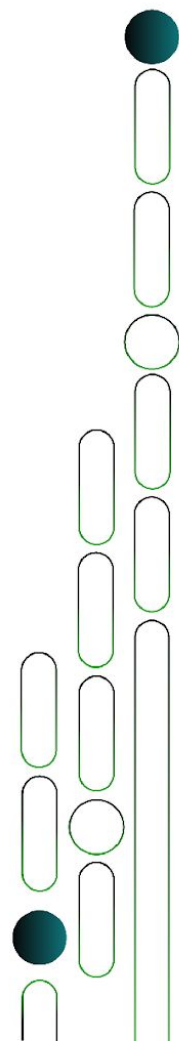
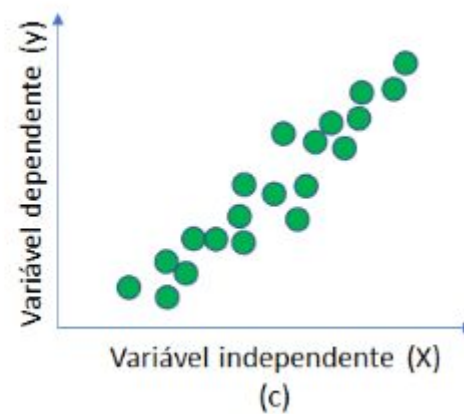
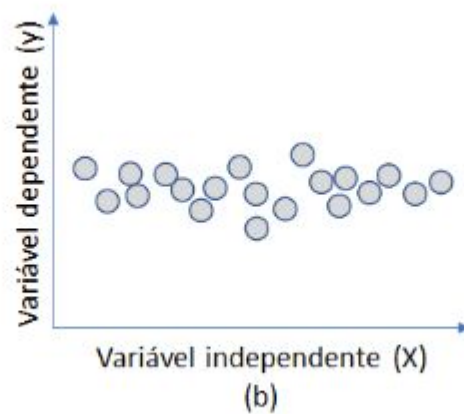
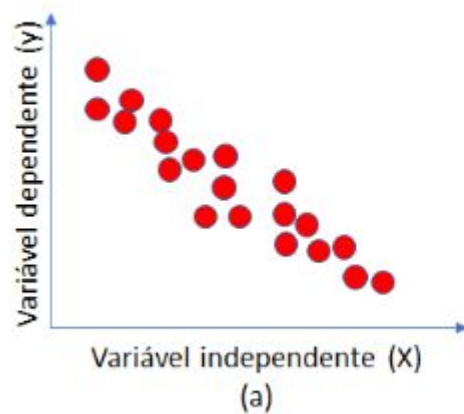


Interpretando a Correlação

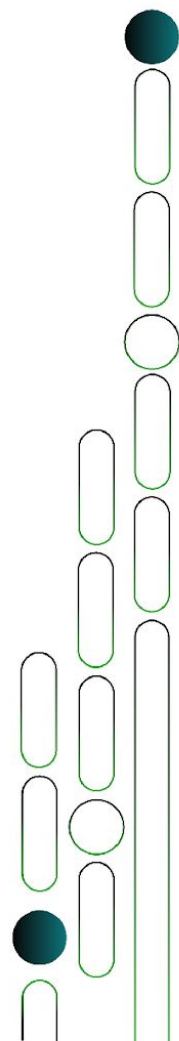
- Neste caso, falaremos do coeficiente de correlação de Pearson. O coeficiente de correlação indica a força e direção do relacionamento linear entre duas variáveis contínuas.
 - O valor de -1.
 - O valor de 0 (zero).
 - O valor de 1.



Interpretando a Correlação



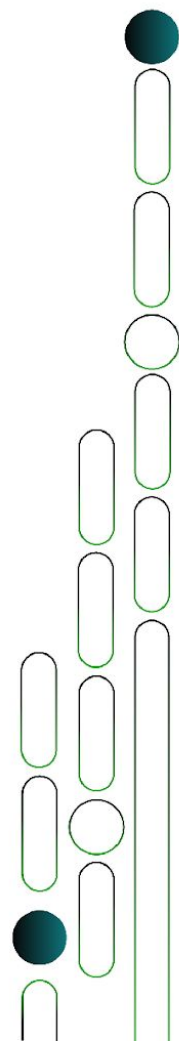
Interpretando a Correlação



Regressão Linear

- **Modo Representativo:**

- $y = b_0 + b_1.X$
- **y**: é a variável dependente, ou seja, o valor previsto.
- **X**: é a variável independente, ou seja, a variável preditora.
- **b₀**: é o coeficiente que intercepta ou que corta o eixo y.
- **b₁**: é o coeficiente que define a inclinação da reta.



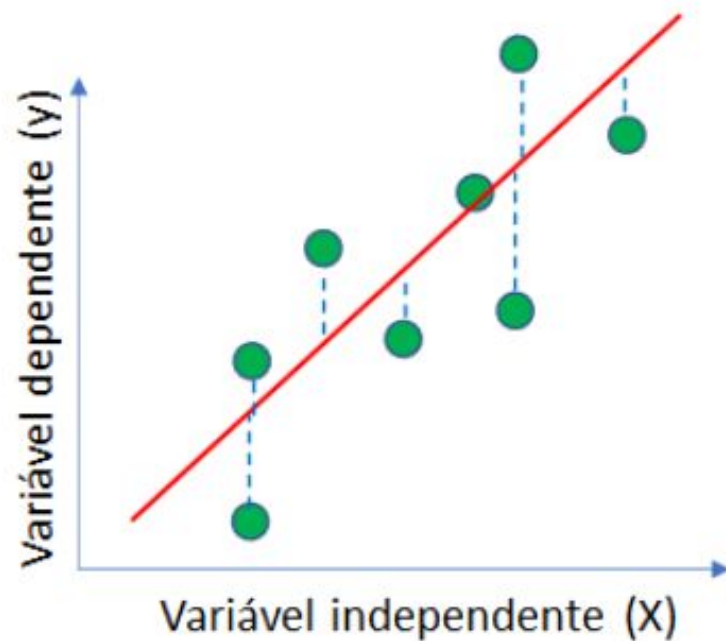
Modelo Representativo

- O objetivo é encontrar as melhores estimativas para os coeficientes, que minimizam os erros na previsão de y a partir de X .

Podemos estimar b_1 como:

$$b_1 = \frac{\text{soma}((X_i - \text{média}(x)) * (y_i - \text{média}(y)))}{\text{soma}((x_i - \text{média}(x))^2)}$$

Erro Padrão

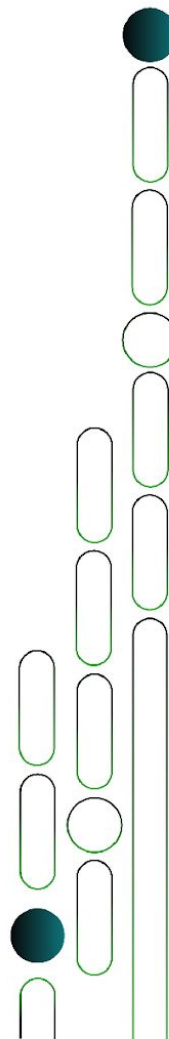


Podemos acrescentar o erro padrão à equação e, de uma forma geral, ficaria da seguinte forma:

$$y = b0 + b1.X1 + e$$

Onde:

- **e**: é o erro padrão.

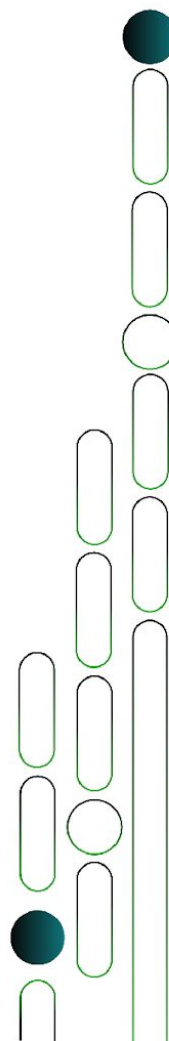


Coeficiente de Determinação R^2

- O valor do R^2 ou *R-squared* é uma medida estatística que nos mostra o quão próximos os dados estão ajustados à linha de regressão. É um valor de 0 à 1 que, quanto mais próximo de 1, melhor o ajuste e menor o erro associado.

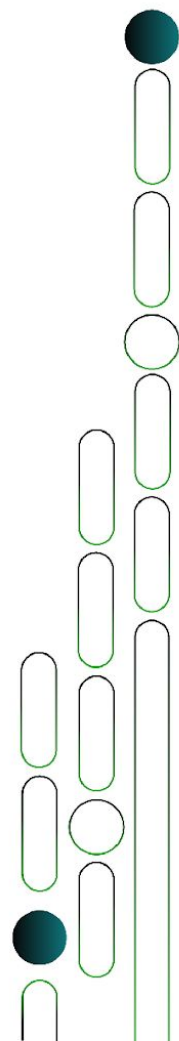
$$R^2 = \text{Variação explicada} / \text{Variação total}$$

- 0% indica que o modelo não explica nada da variabilidade dos dados de resposta ao redor de sua média.
- 100% indica que o modelo explica toda a variabilidade dos dados de resposta ao redor de sua média.



Conclusão

- Correlação.
- Interpretando a Correlação.
- Definição de Regressão.
- Regressão Linear.



Próxima aula

- Abordaremos aula prática sobre regressão linear.

