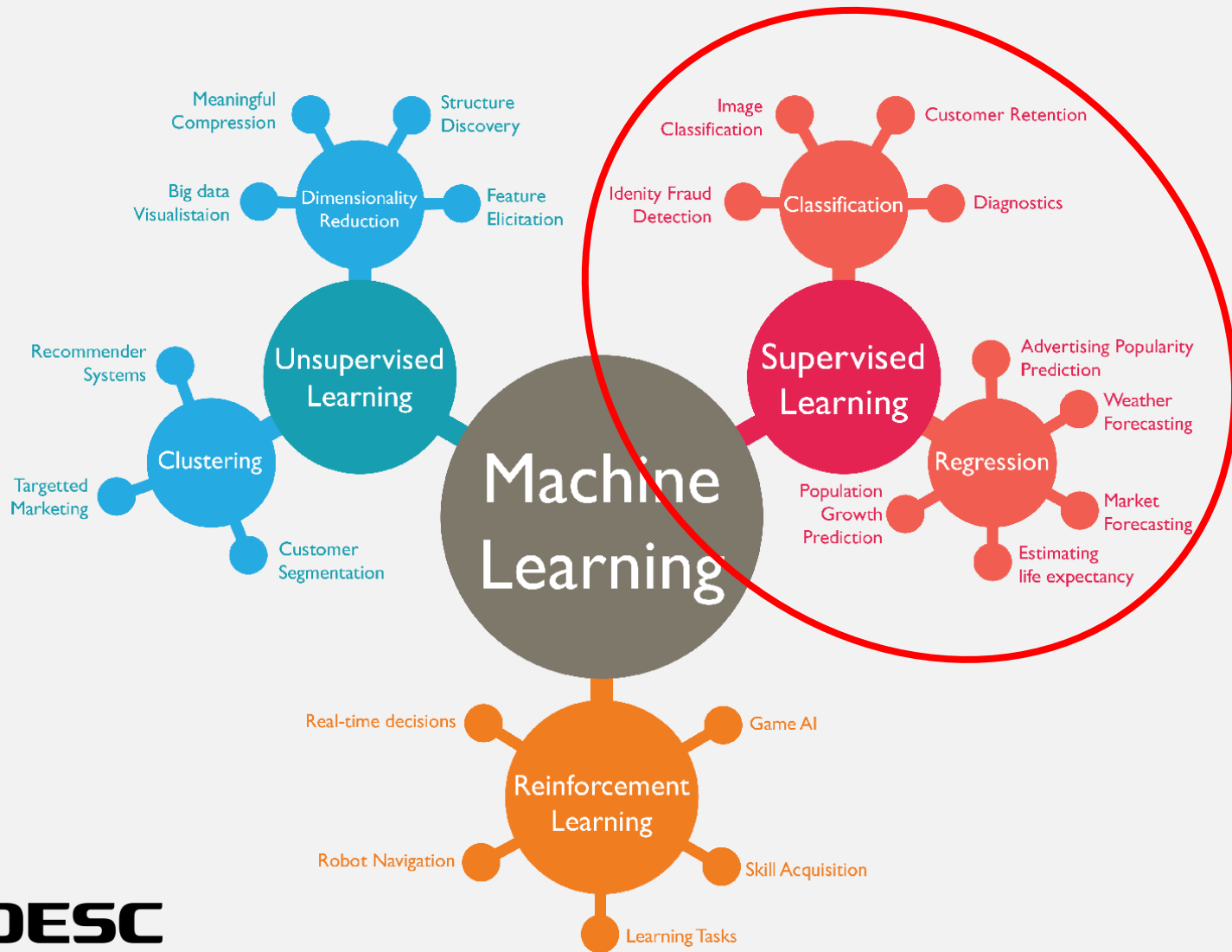


Modelos de Aprendizagem Supervisionada

Bernard da Silva
Orientador: Rafael Parpinelli
16/11/2022

Aprendizagem Supervisionada

- Uma sub-área da Aprendizagem de Máquina (Machine Learning)



■ Aprendizagem Supervisionada

- O que é aprendizagem?
- E por quê essa pergunta é importante?



Aprendizagem Supervisionada

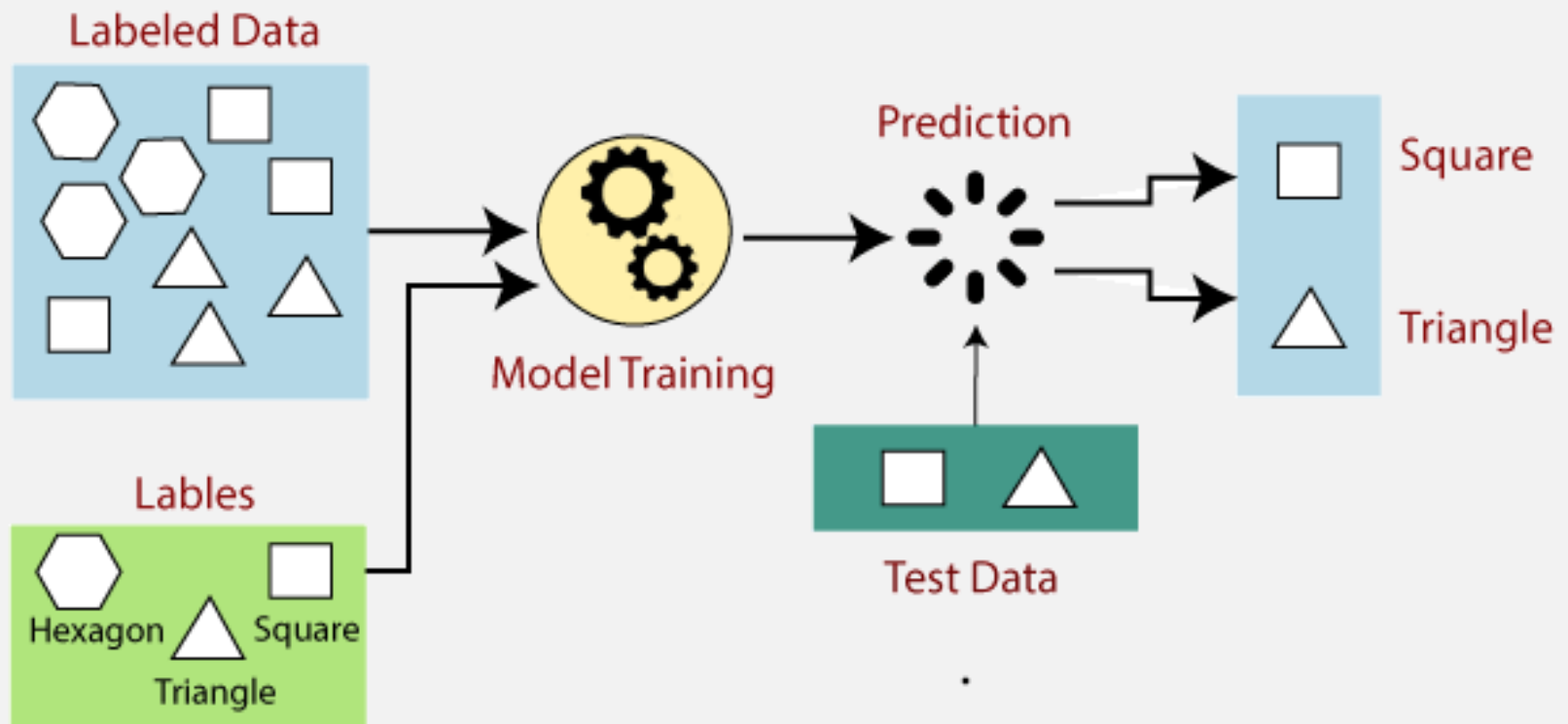
- O que que é aprendizagem?
 - Existem várias definições, inclusive:
 - O ato de obter conhecimento.
 - "Aprendizado é qualquer processo pelo qual um sistema melhora sua performance pela experiência." (Herbert Alexander Simon).

Aprendizagem Supervisionada

- E por quê essa pergunta é importante?
 - Amadurecer como funcionam programas com aprendizagem supervisionada.
 - Habilidade de imitar humanos e substituí-los em certas tarefas monótonas, que exigem alguma inteligência.

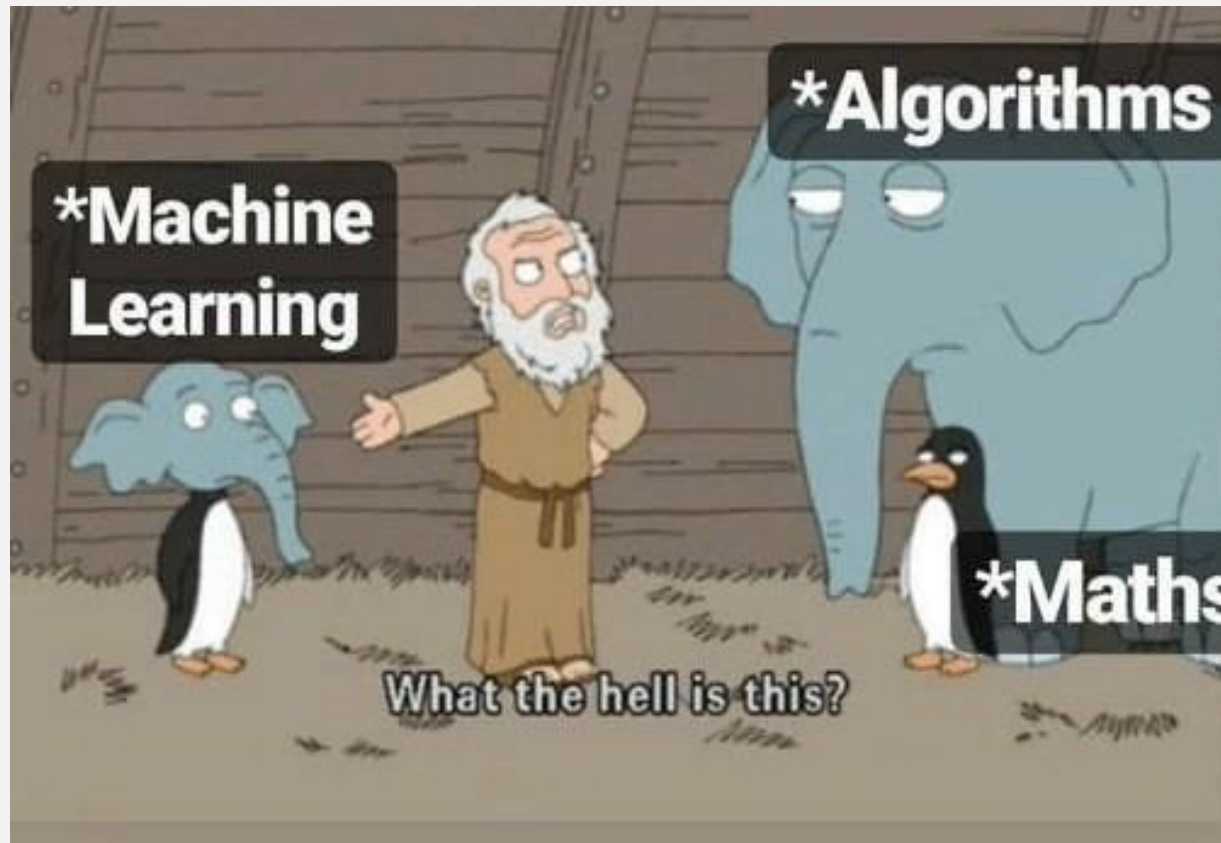
Aprendizagem Supervisionada

- Consiste em “aprender” uma função que mapeia dados de entrada em relação a saída através de **exemplos (labels ou rótulos bem estabelecidos)**



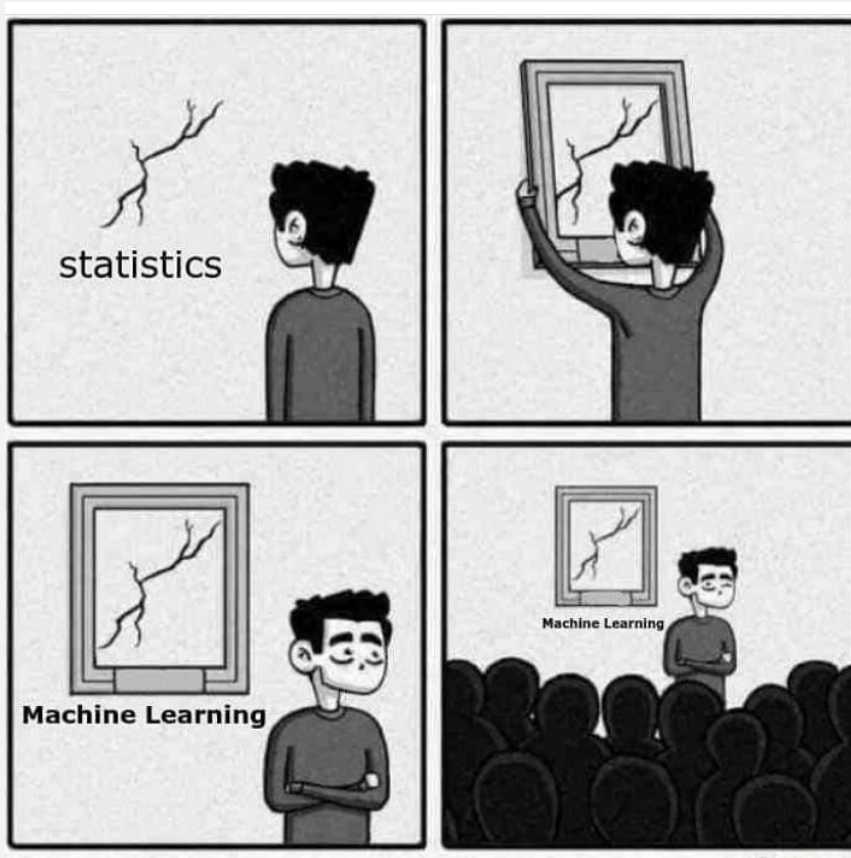
Aprendizagem Supervisionada

- Machine Learning é fruto do avanço de várias outras áreas.
- O amadurecimento de ML vem de uma demanda por algoritmos de modelagem mais sofisticados.



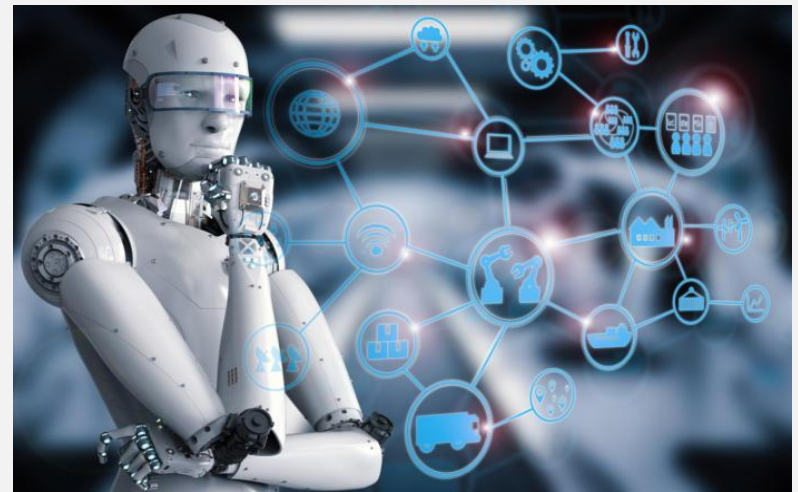
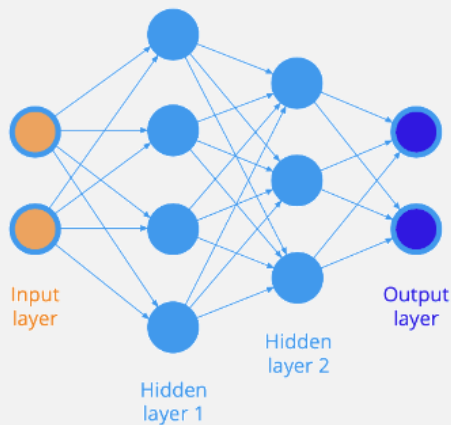
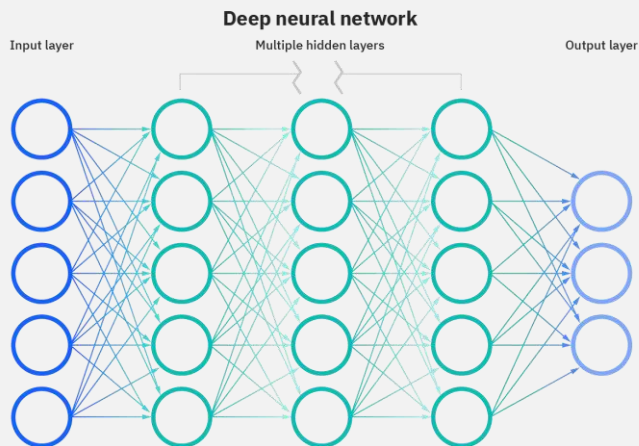
Aprendizagem Supervisionada

- O que realmente é Machine Learning?



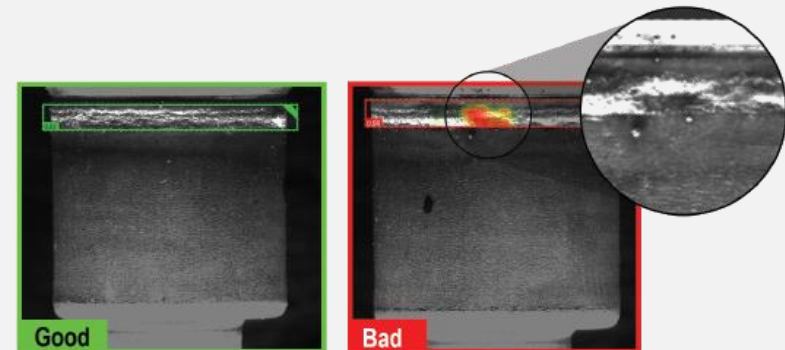
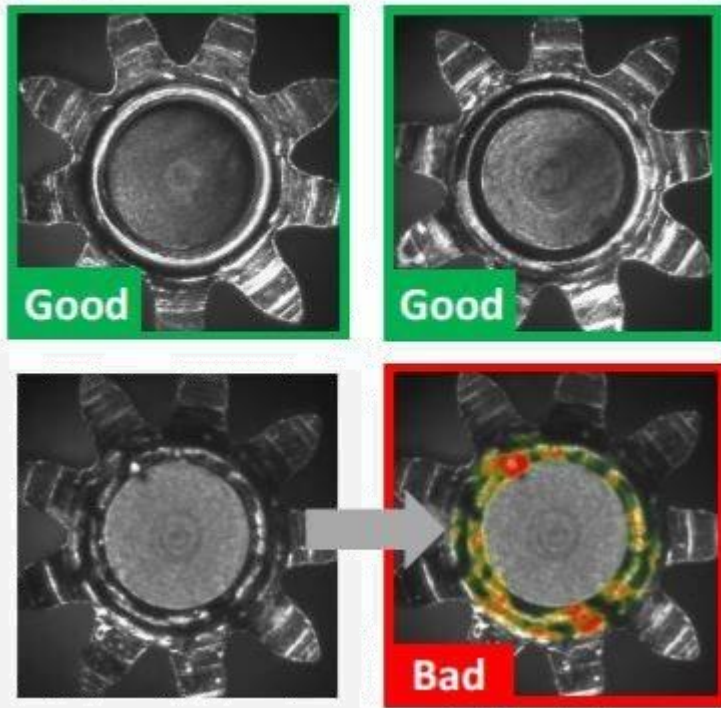
Alguns esclarecimentos...

- Então por quê se fala tanto de Machine Learning hoje em dia? **Marketing**
- Resultados de uma pesquisa por “Neural Network” e “Machine Learning” no Google



Alguns esclarecimentos...

- Então por quê se fala tanto de Machine Learning hoje em dia? **Demanda**



Alguns esclarecimentos...

- Machine Learning é a solução de todos os problemas?



Alguns esclarecimentos...

- Machine Learning é a solução de todos os problemas?



Aplicações



Aprendizagem Supervisionada

- Qual tipo de modelo treinar?
 - DEPENDE!
- Qual é o tipo de dado?
 - Linear
 - Não linear
- Qual a prioridade?
 - Acurácia
 - Precisão
 - Velocidade de treino
 - Velocidade de inferência
 - Simplicidade
 - **Interpretabilidade**
- Qual a finalidade?
 - Regressão
 - Classificação



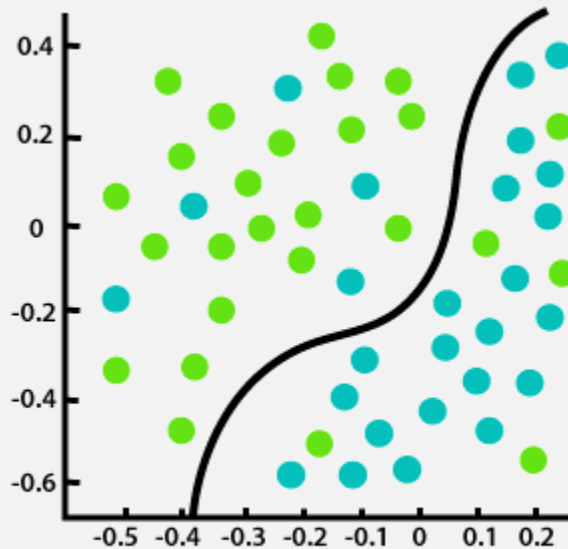
Aprendizagem Supervisionada

- Conheça seu PROBLEMA!

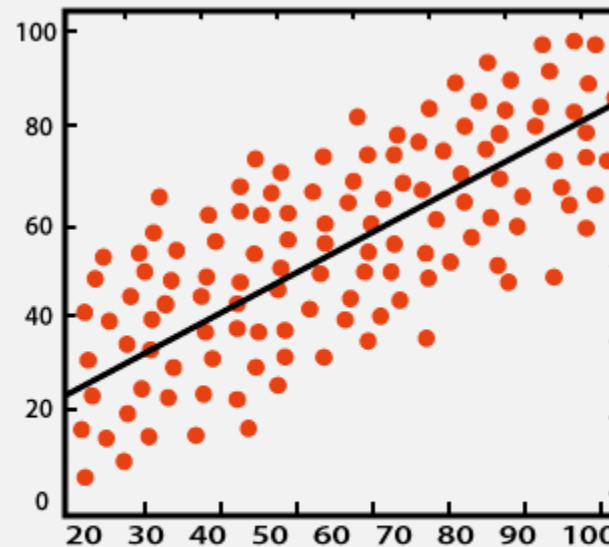


Aprendizagem Supervisionada: Tipos

- Se divide principalmente em 2 tipos de aplicações:
 - Regressão
 - Classificação



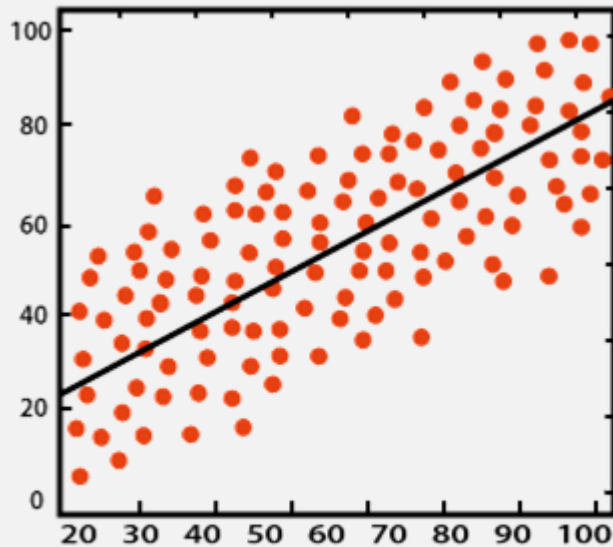
Classification



Regression

Regressão

- Objetivo: Gerar um modelo (analítico ou não) que melhor represente a distribuição de dados contínuos.
- Realizar previsões numéricas sobre algum fenômeno.

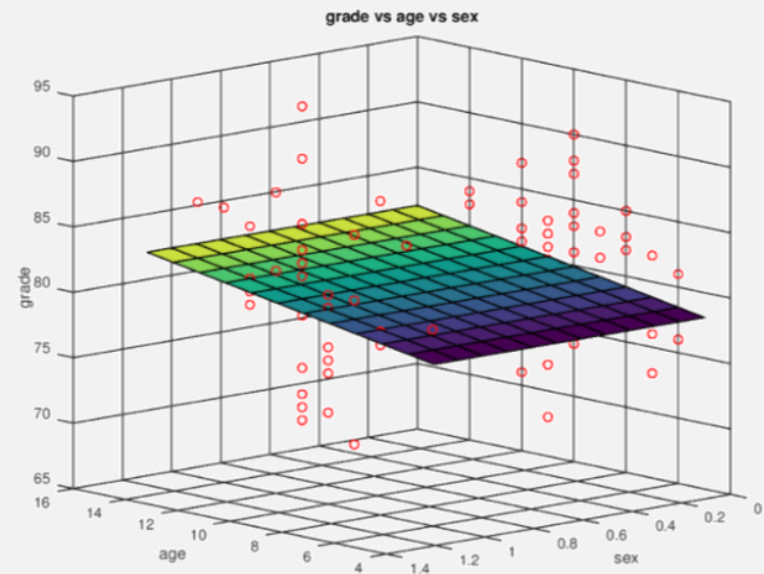
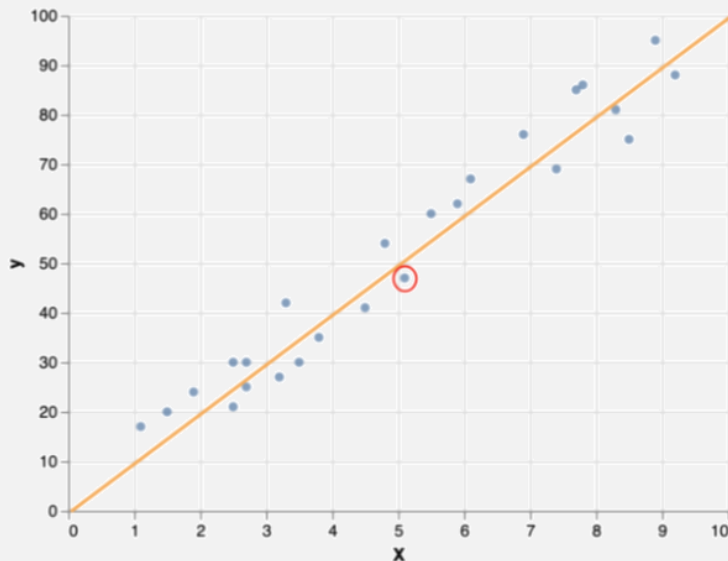


Regression



Regressão Linear

- Existem dois tipos de regressões lineares:
- Simples: refere-se quando temos somente uma variável independente (X) para fazermos a predição.
- Múltipla: refere-se a várias variáveis independentes (X) usadas para fazer a predição.



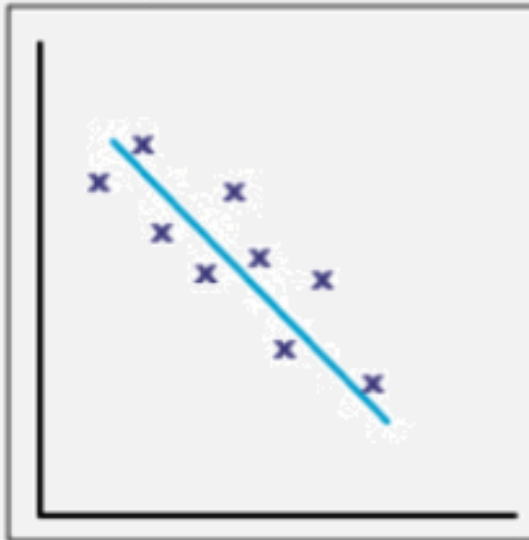
Regressão Linear

- Quando aplicar regressão linear?
- Quando há uma boa correlação linear (positiva ou negativa) entre os dados.

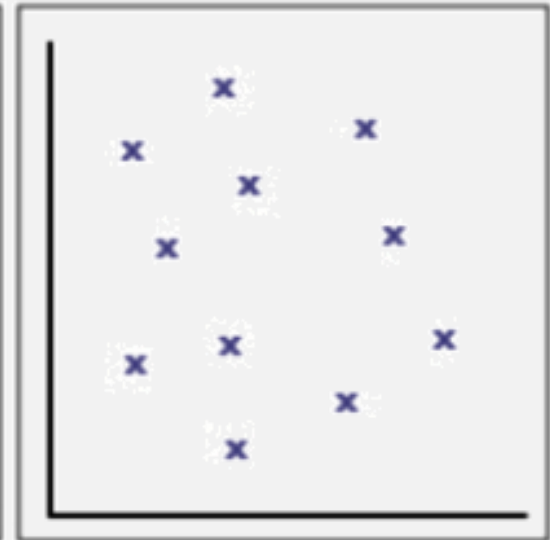
Positive correlation



Negative correlation



No correlation



Regressão Linear

O objetivo da regressão linear é encontrar uma reta que consiga definir bem os dados e minimizar a diferença entre o valor real e a saída calculada pelo modelo.

$$\textit{Simples} \Rightarrow y_i = \alpha + \beta x_i + \varepsilon$$

$$\textit{Múltipla} \Rightarrow y_i = \alpha + \beta x_{i1} + \beta x_{i2} + \dots + \beta x_{in} + \varepsilon$$

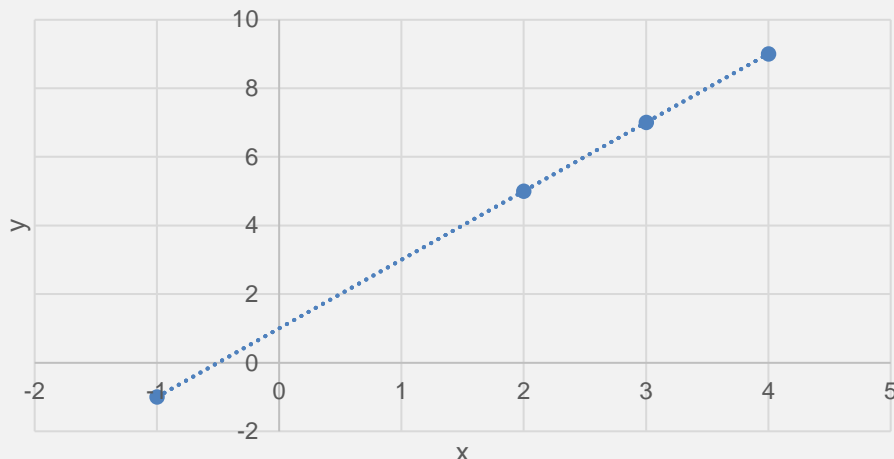
y_i = Variável alvo
 α = Intercepto (valor de y quando x for = 0)
 β = Coeficiente angular (inclinação da reta)
 x_i = Variável desejada
 ε = Erro

Existem diversas abordagens para se calcular os coeficientes α e β da equação da regressão linear, as técnicas baseadas em mínimos quadrados ordinários e gradiente descendente são as mais comuns.

Regressão Linear

Método dos mínimos quadrados – método usado para definir a reta e obter α e β

Exemplo regressão linear



	x	y	x ²	y ²	xy
	3	7	9	49	21
	2	5	4	25	10
	-1	-1	1	1	1
	4	9	16	81	36
Soma	8	20	30	156	68

$$\beta = \frac{n \sum x_i \cdot y_i - \sum x_i \cdot \sum y_i}{n \sum x_i^2 - (\sum x_i)^2} \longrightarrow \beta = \frac{4 \cdot 68 - 8 \cdot 20}{4 \cdot 30 - (8)^2} \longrightarrow \beta = 2$$

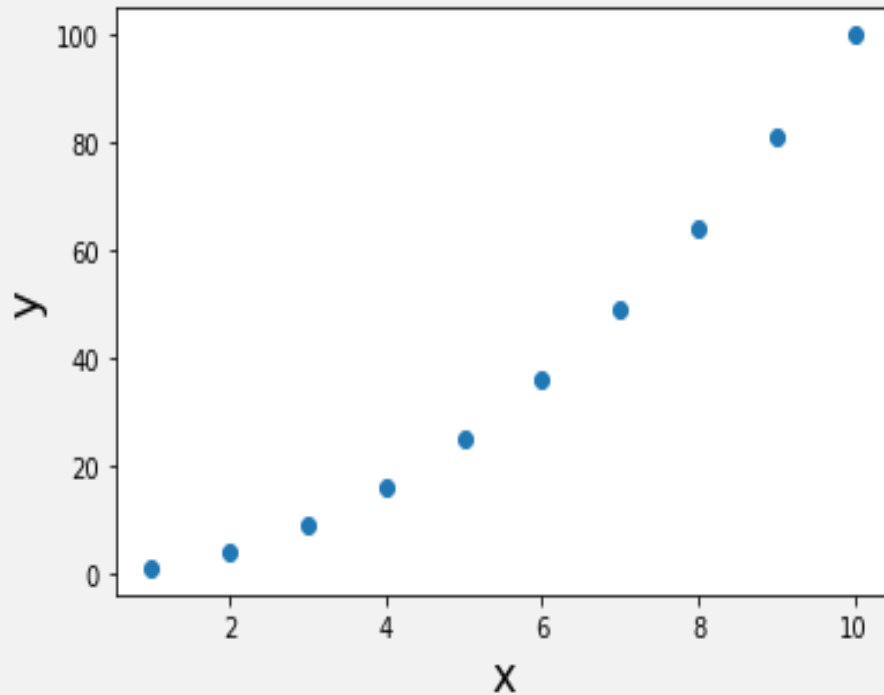
$$\alpha = \frac{\sum y - \beta \cdot \sum x}{n} \longrightarrow \alpha = \frac{20 - 2 \cdot 8}{4} \longrightarrow \alpha = 1$$

$$y_i = \alpha + \beta x_i$$

$$y_i = 1 + 2x_i$$

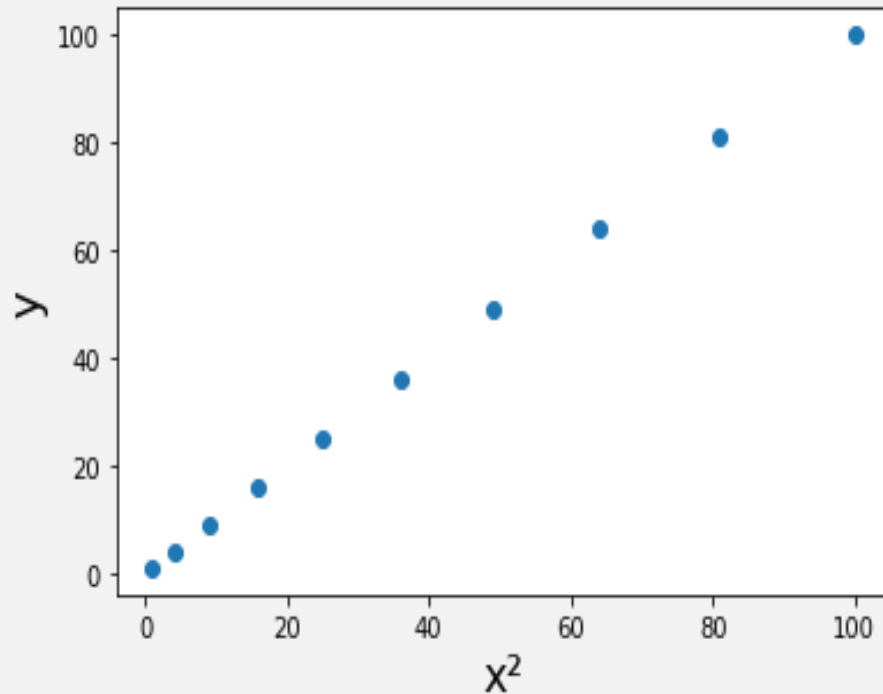
Regressão Linear

- Estes dados são lineares?



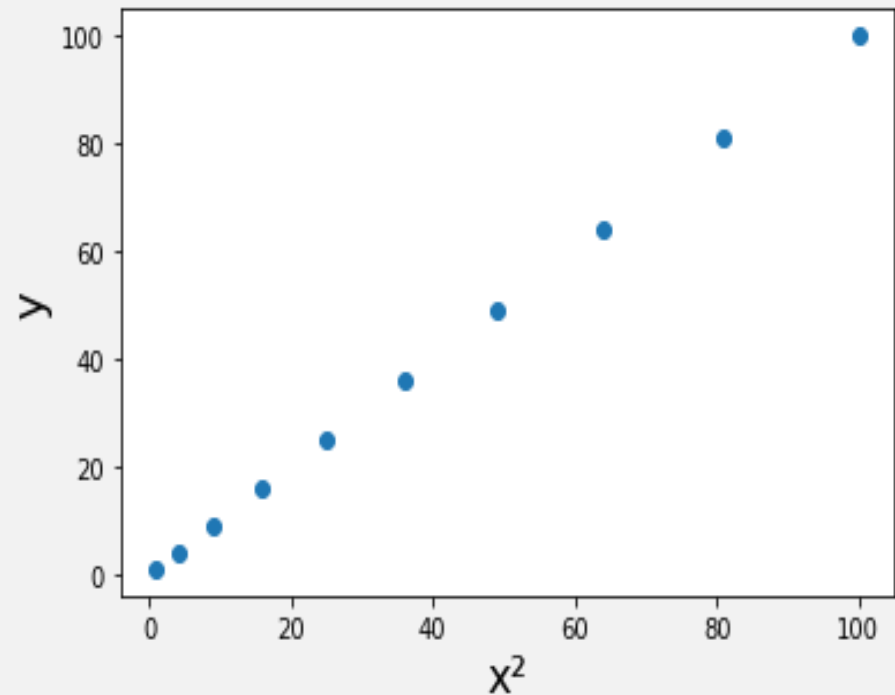
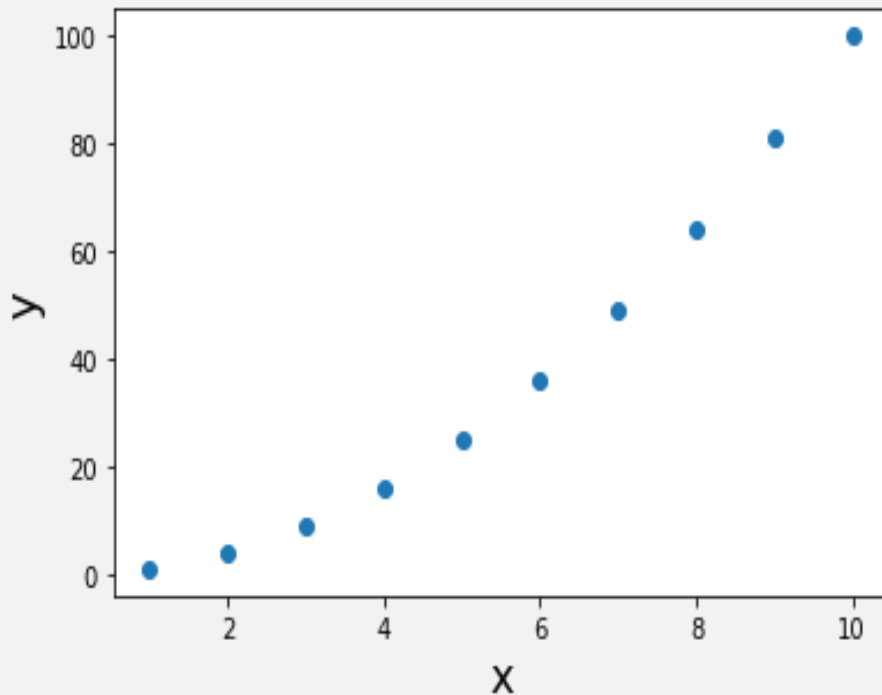
Regressão Linear

- Estes dados são lineares?



Regressão Linear

- Estes dados são lineares?

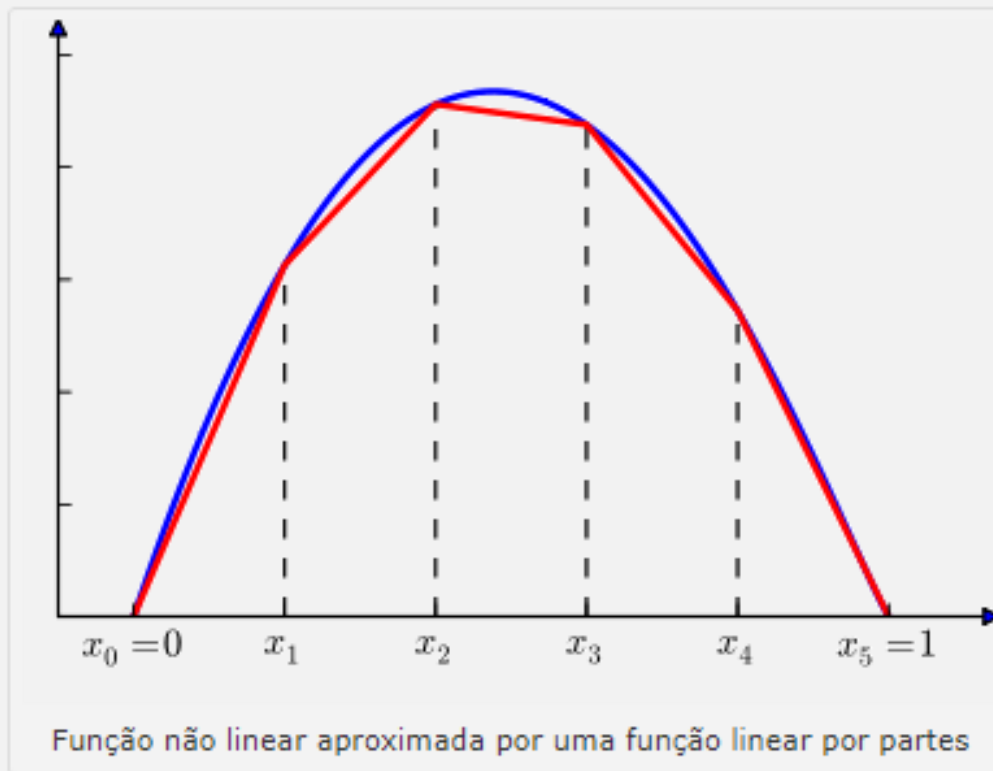


Y é não linear em relação a X

Y é linear em relação a X^2

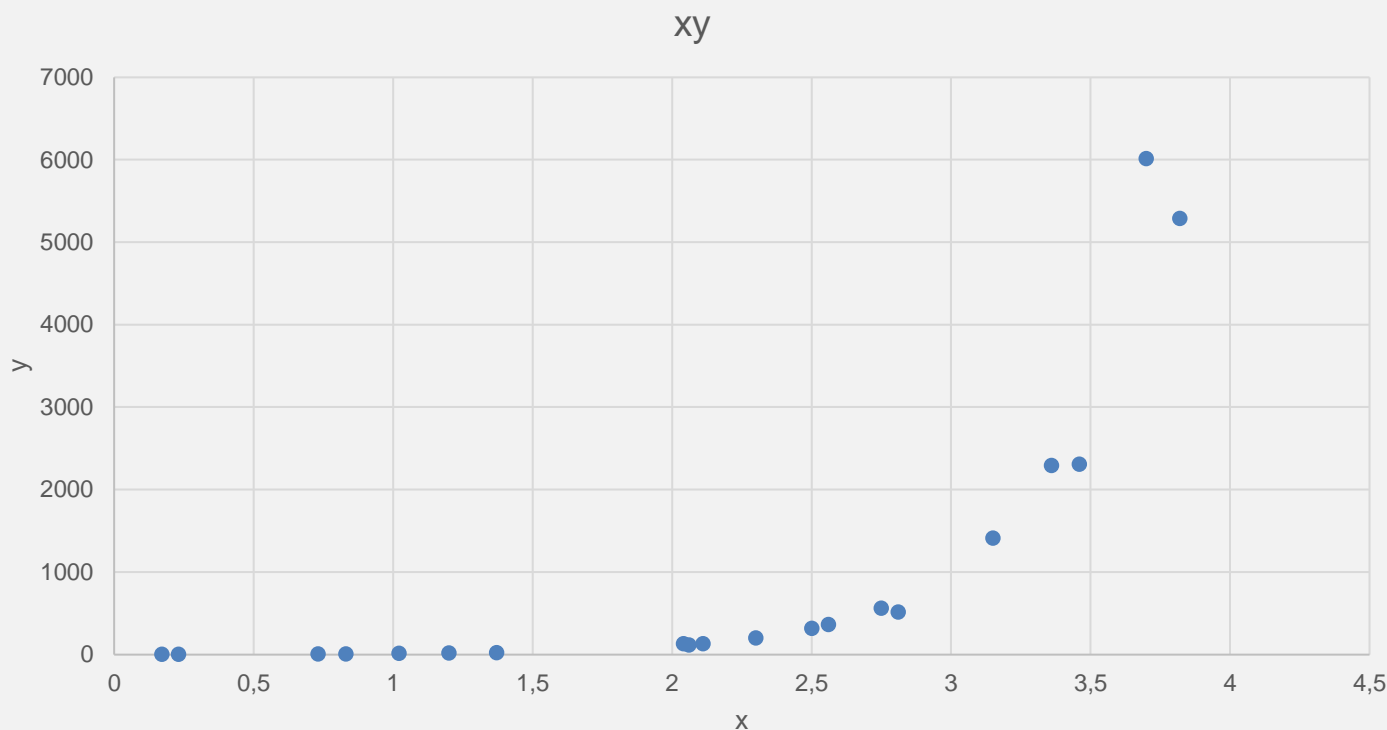
Regressão polinomial

- Funções não lineares podem ser LINEARIZADAS!
- Função linear por partes (Piecewise linear function)



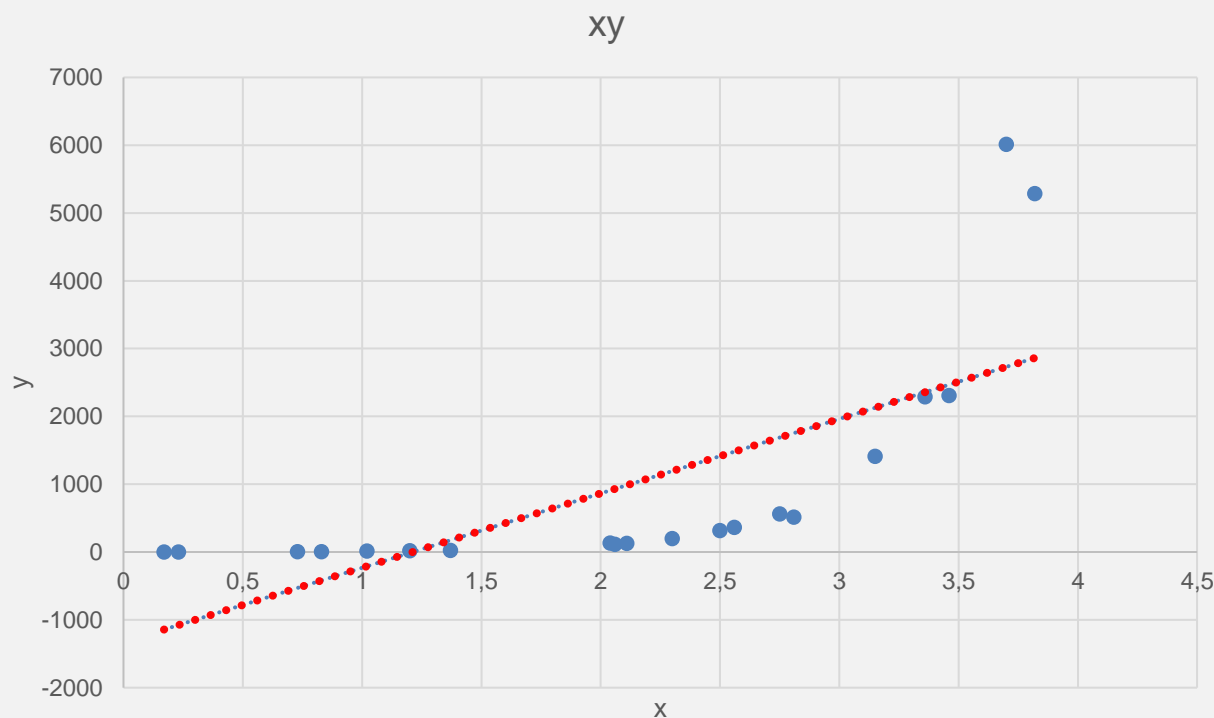
Regressão polinomial

- Funções não lineares podem ser LINEARIZADAS!
- Transformação de dados



Regressão polinomial

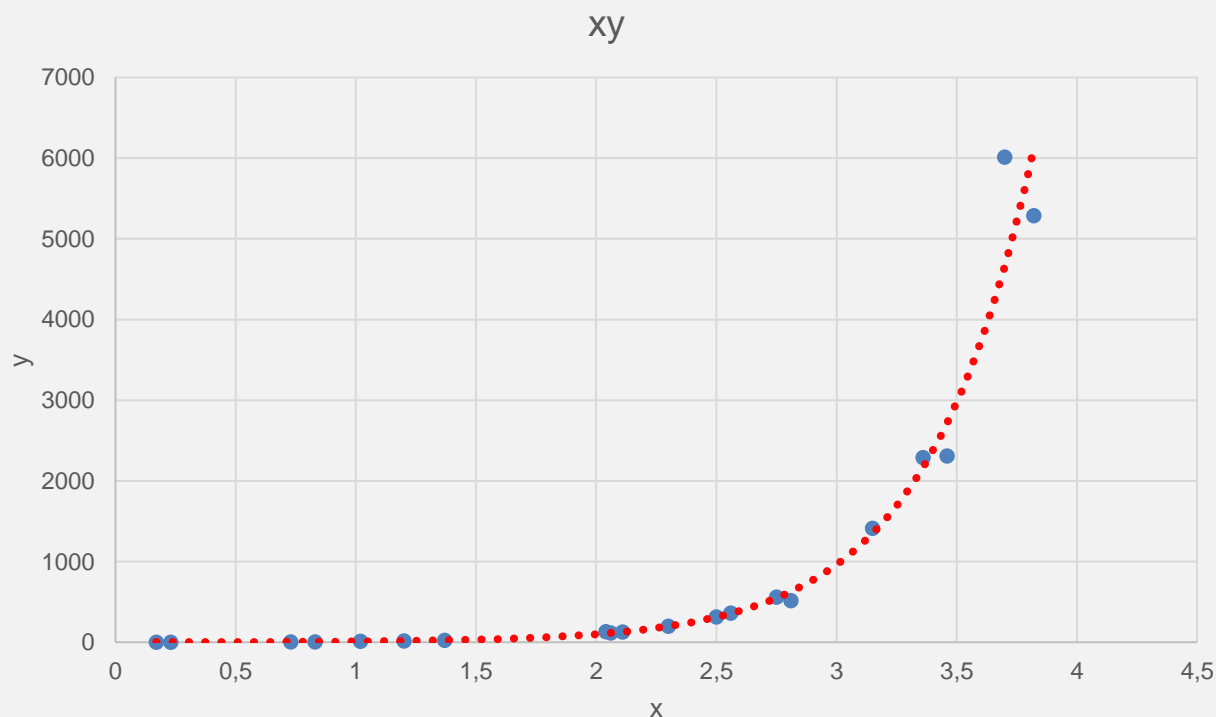
- Funções não lineares podem ser LINEARIZADAS!
- Transformação de dados



x	y
3,46	2307,23
3,36	2290,87
1,37	23,44
3,82	5285,55
2,75	562,34
0,73	6,44
0,17	1,77
1,2	19,02
2,3	199,53
2,06	114,82
2,56	363,08
2,04	131,58
3,7	6014,25
2,5	316,23
1,02	12,57
2,11	128,82
3,15	1412,54
0,83	5,41
2,81	516,52
0,23	1,7

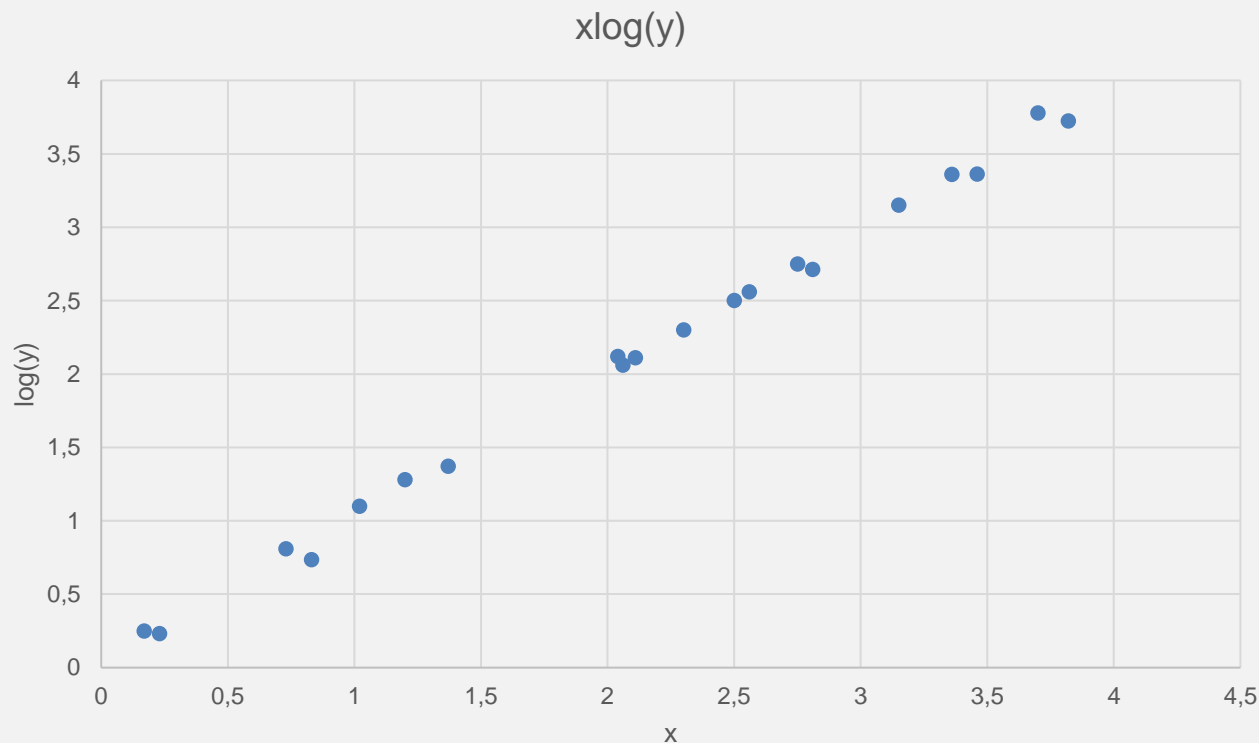
Regressão polinomial

- Funções não lineares podem ser LINEARIZADAS!
- Transformação de dados



Regressão polinomial

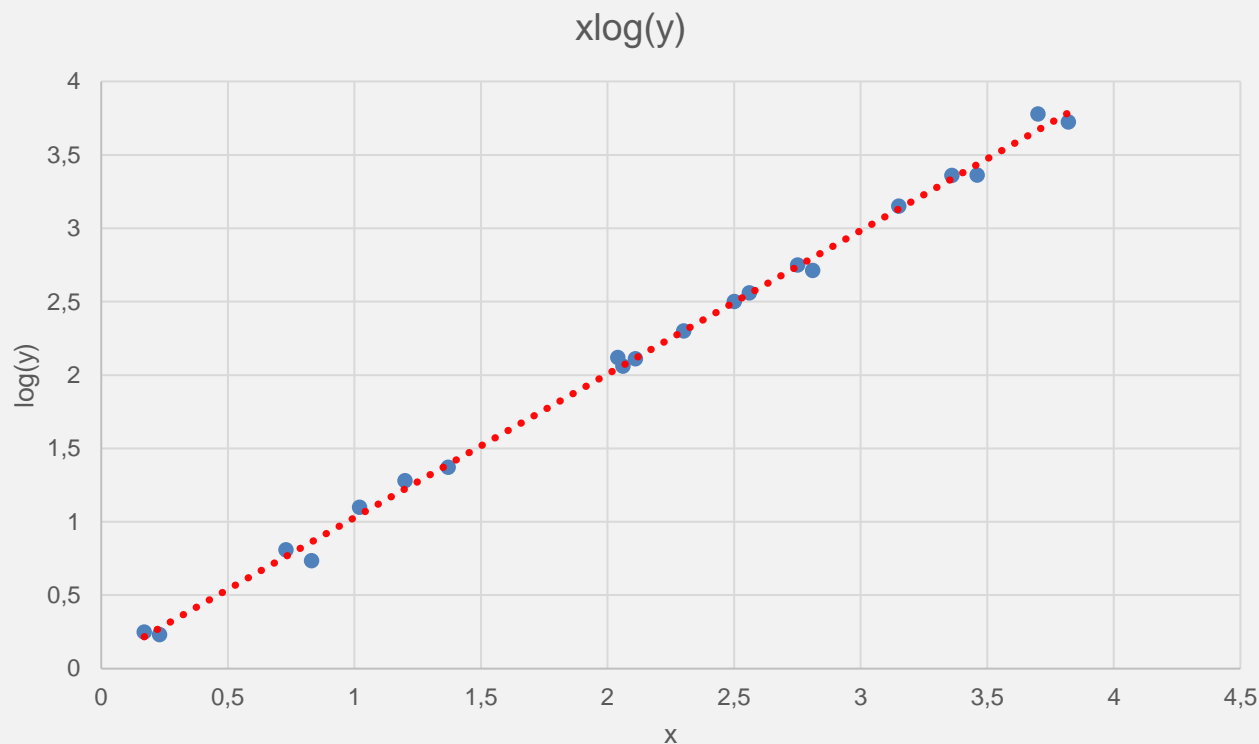
- Funções não lineares podem ser LINEARIZADAS!
- Transformação de dados



x	log(y)
3,46	3,36309
3,36	3,36
1,37	1,36996
3,82	3,72309
2,75	2,75
0,73	0,80889
0,17	0,24797
1,2	1,27921
2,3	2,30001
2,06	2,06002
2,56	2,56
2,04	2,11919
3,7	3,77918
2,5	2,5
1,02	1,09934
2,11	2,10998
3,15	3,15
0,83	0,7332
2,81	2,71309
0,23	0,23045

Regressão polinomial

- Funções não lineares podem ser LINEARIZADAS!
- Transformação de dados



x	log(y)
3,46	3,36309
3,36	3,36
1,37	1,36996
3,82	3,72309
2,75	2,75
0,73	0,80889
0,17	0,24797
1,2	1,27921
2,3	2,30001
2,06	2,06002
2,56	2,56
2,04	2,11919
3,7	3,77918
2,5	2,5
1,02	1,09934
2,11	2,10998
3,15	3,15
0,83	0,7332
2,81	2,71309
0,23	0,23045

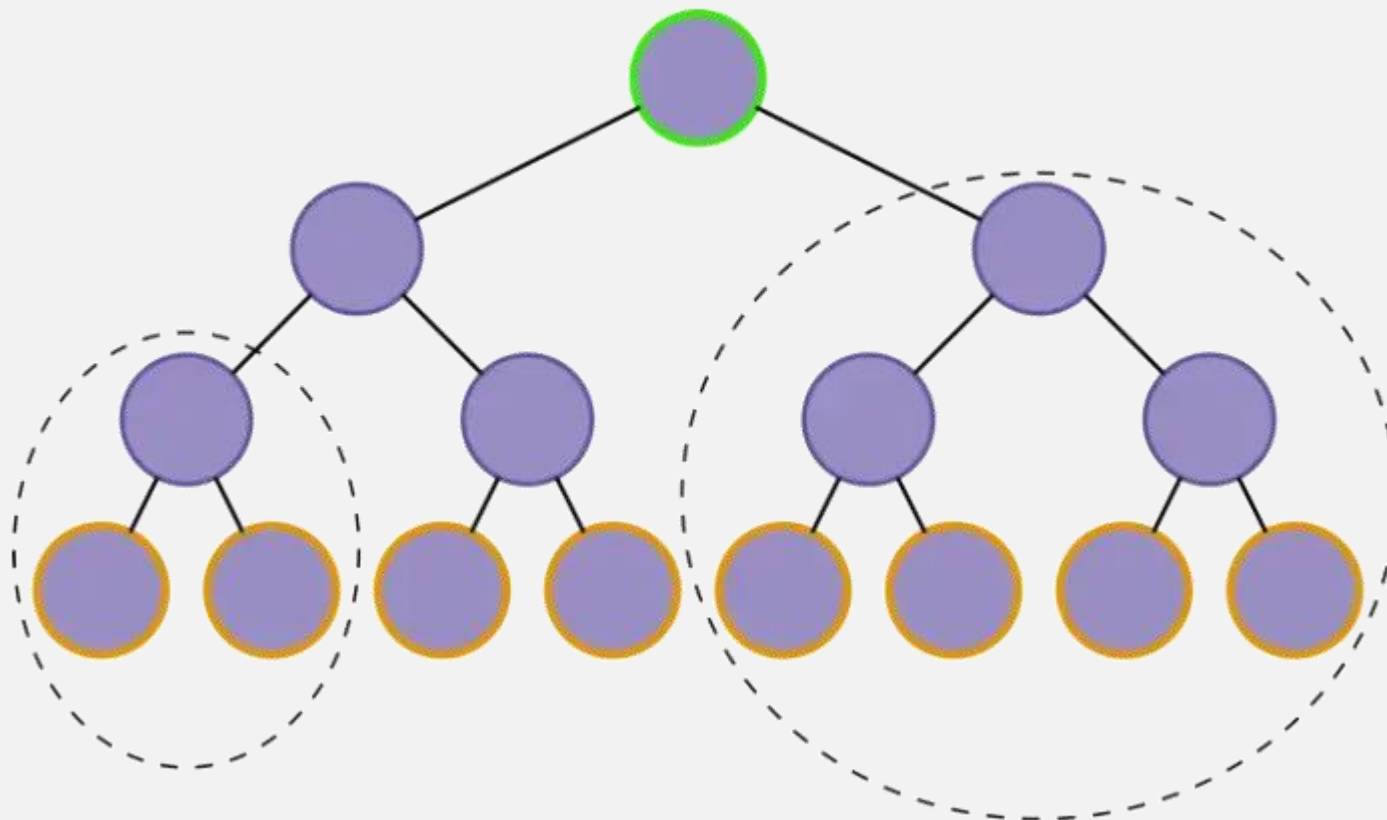
Regressão - Decision Trees



- Aprende simples **decisões** a partir de um conjunto de dados
- Você já jogou cara a cara?



Regressão - Decision Trees



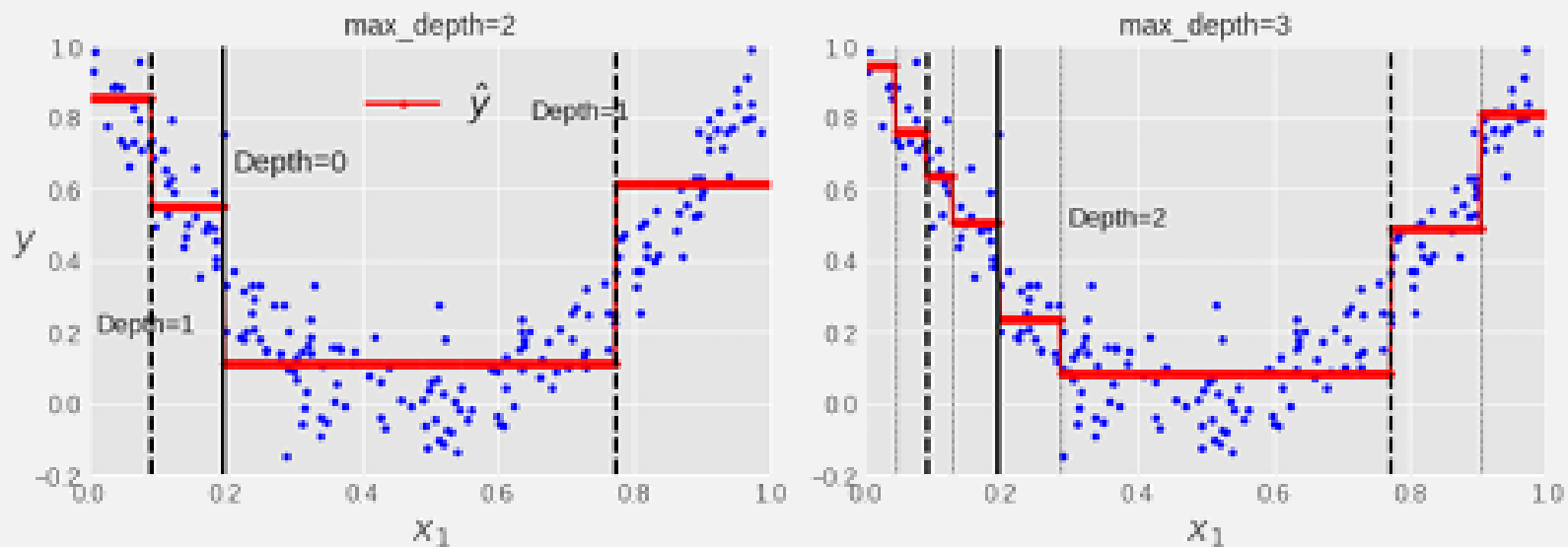
- Círculos roxos - Nós
- Linhas pretas – Ramos
- Nós sem descendentes (com borda laranja) - Folhas
- Primeiro nó (com borda verde) - Raiz

Regressão - Decision Trees



Utiliza métodos como Entropia e Impureza de Gini para classificação.

Erro Quadrático Médio (MSE) e o Erro Médio Absoluto (MAE) para regressão.



- Exemplo de árvores de regressão com diferentes profundidades

Regressão - Decision Trees



Vantagens	Desvantagens
Fácil de entender e principalmente de interpretar.	Muito suscetível a overfit.
Precisa de pouco pré processamento.	Bastante sensível aos dados ou parâmetros do modelo.
Boa complexidade computacional de inferência	Predição não é contínua, e sim linear em trechos. Péssimo para extrapolação.
Capaz de lidar com múltiplas saídas (múltiplos regressores)	Existem conceitos que não funcionam bem na forma lógica de árvore de decisão.
Consegue lidar tanto com variáveis numéricas como categóricas	Dados desbalanceados geram árvores tendenciosas.

Regressão - Random Forest

- É um tipo de modelo que aprende simples decisões a partir de um conjunto de dados utilizando várias árvores de decisão diferentes
- Random Forest = um Ensemble (agrupamento) de Decision Trees



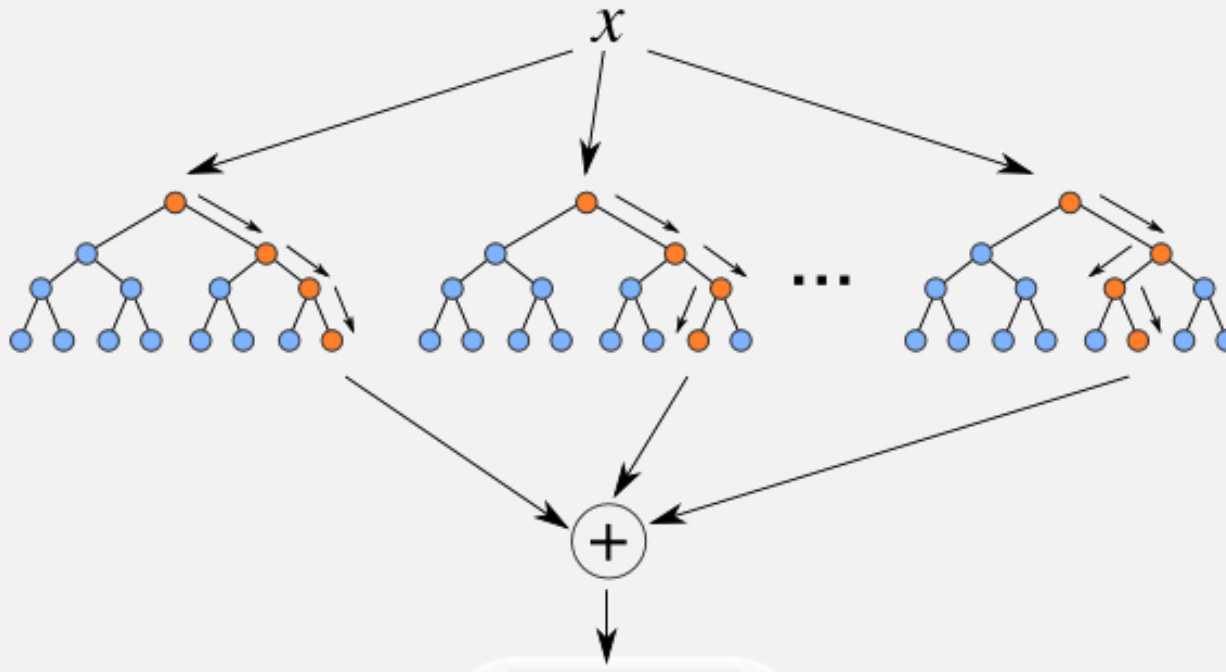
Random Forest



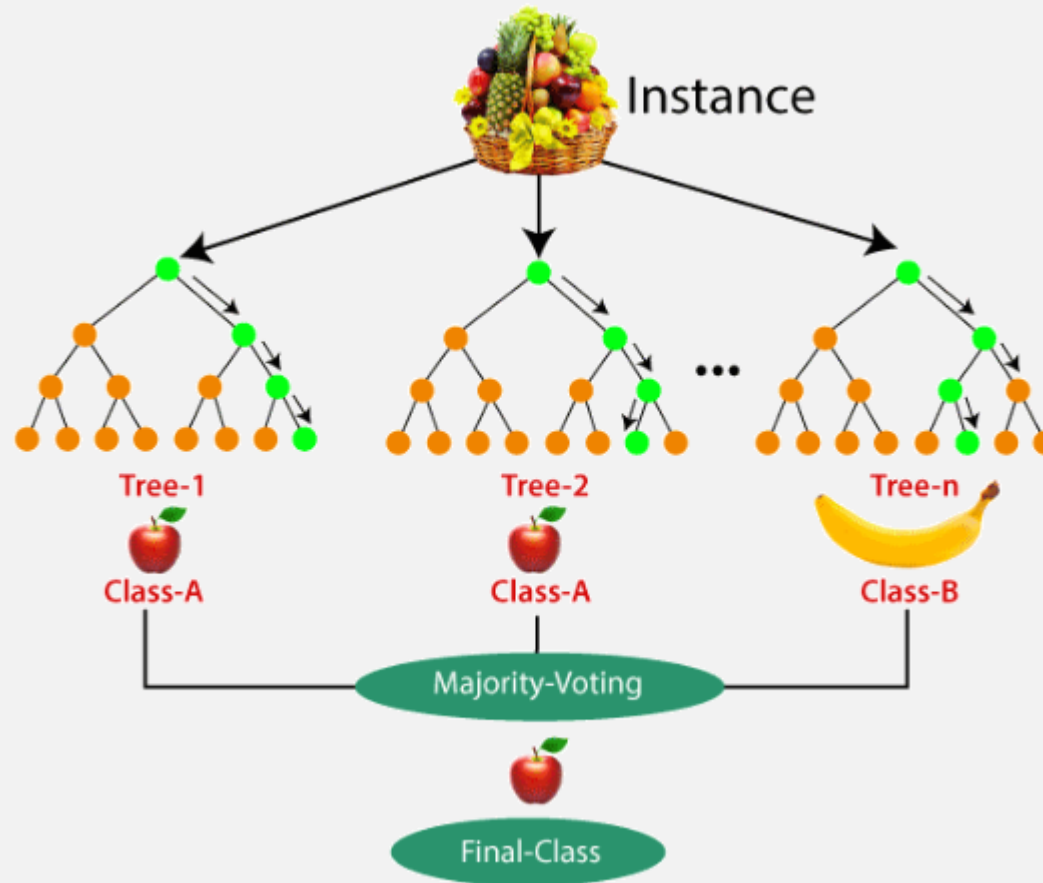
Decision Trees

Regressão - Random Forest

Utiliza métodos como Bagging e Feature Randomness



Regressão - Random Forest

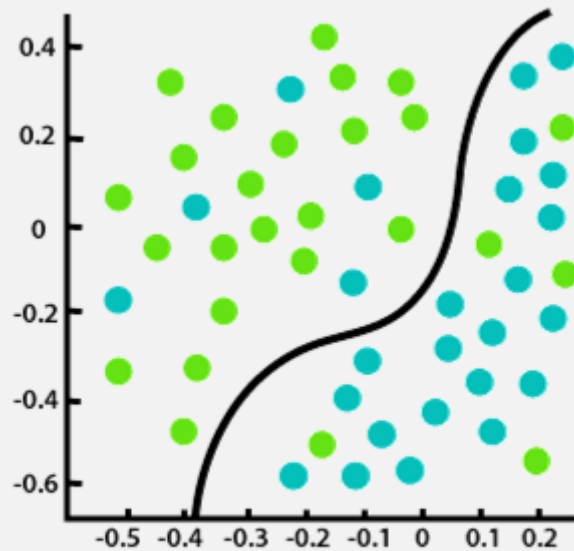


Regressão - Random Forest

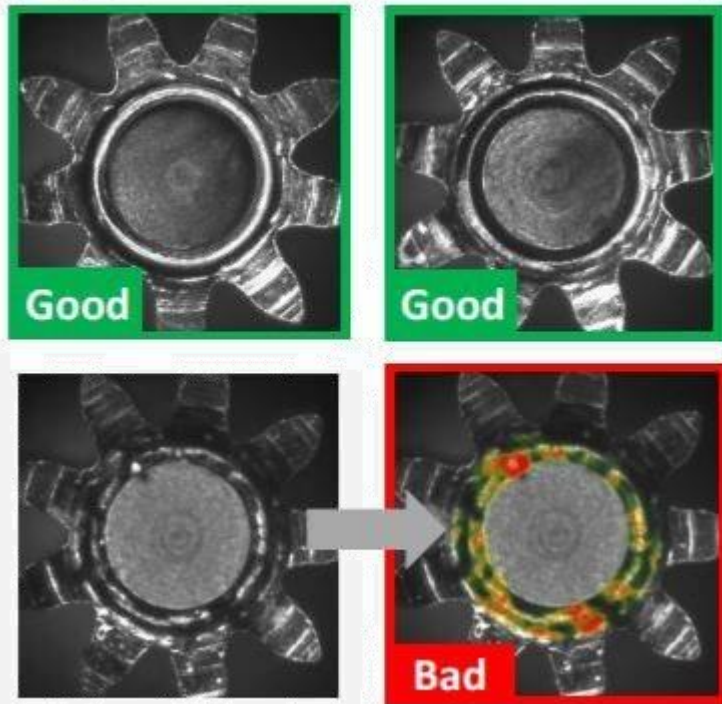
Vantagens	Desvantagens
Fácil de entender.	Predição lenta dependendo do tamanho da floresta
Precisa de pouco pré processamento.	
Boa complexidade computacional de inferência	
Capaz de lidar com múltiplas saídas (múltiplos regressores)	
Consegue lidar tanto com variáveis numéricas como categóricas	

Classificação

- Objetivo: Prever a categoria de uma observação dada a partir dos dados de entrada.

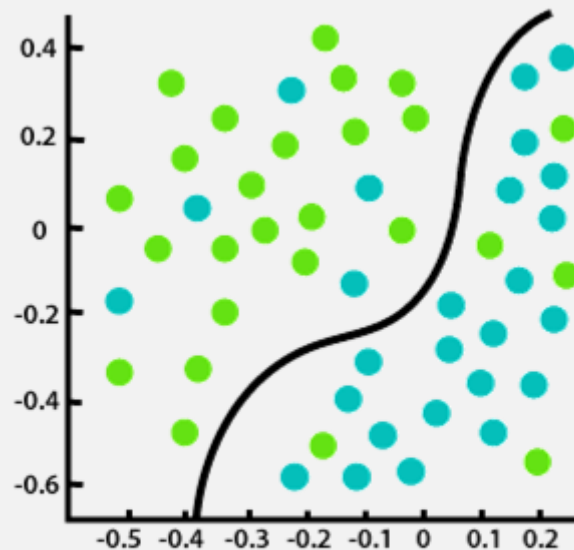


Classification



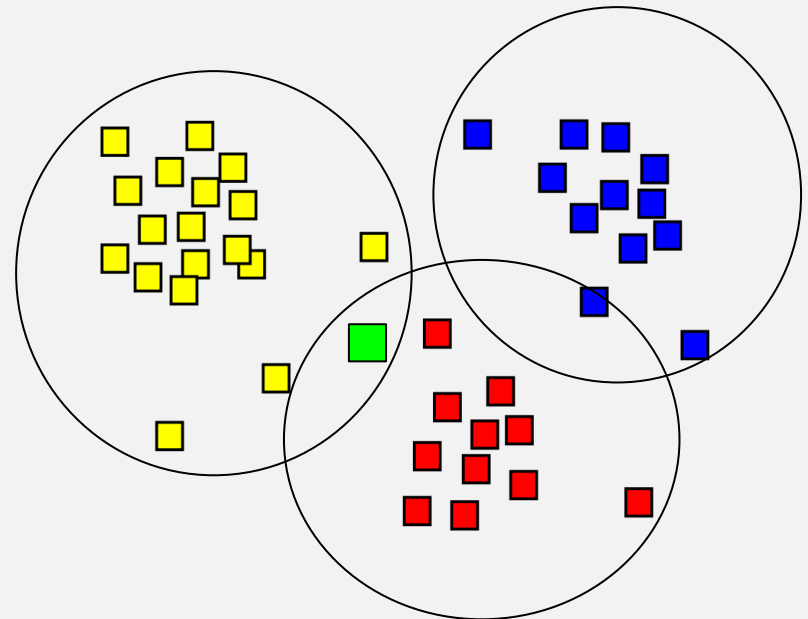
Classificação

- Objetivo: Prever a categoria de uma observação dada a partir dos dados de entrada.
- Modelos de classificação podem ser analíticos ou não.



Classification

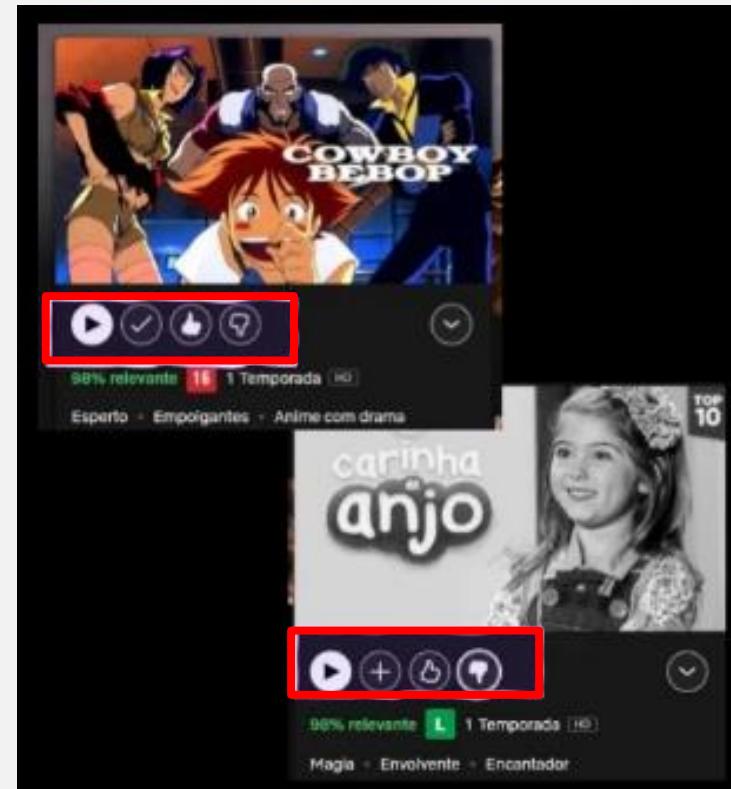
- Classificações podem ser:
 - Binárias ou
 - Multi-classe
 - Multi-label / Multi-rótulo



Classificação

- Classificação binária

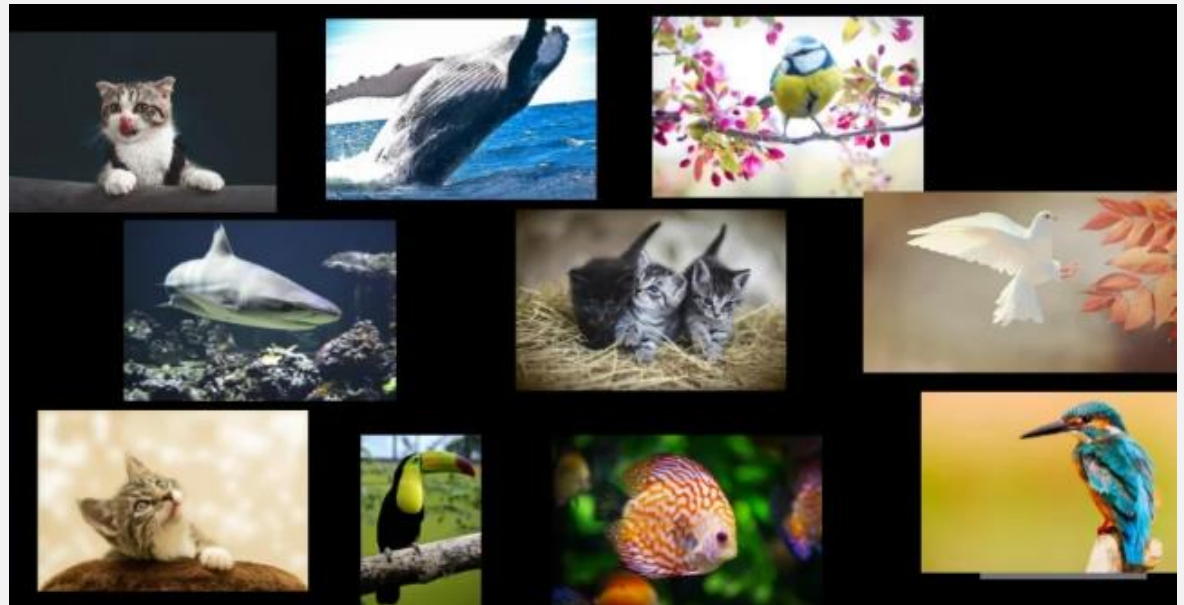
É ou não é
Pertence não pertence
Um ou zero
Gostou não gostou
Etc...



Classificação

- Classificação multiclasse
- Pertence ou não a uma classe específica

- Animal marinho
- Animal voador
- Animal doméstico



Classificação

- Classificação multilabel
- Pode pertencer um ou mais classes do problema



Resumo

- Machine Learning nada mais é que matemática e estatística;
- Entender o SEU problema é o primeiro passo;
- Machine learning não resolve TODOS os problemas;
- Existem 2 principais tipos de modelagem de aprendizagem supervisionada: Regressão e Classificação;
- Existem diversos métodos e ferramentas disponíveis para auxiliar na resolução do seu problema.

Trabalho 1

Lista de exercícios de regressão linear
Data de entrega: 21/11

Dúvidas:

e-mail: bernarddss62gmail.com