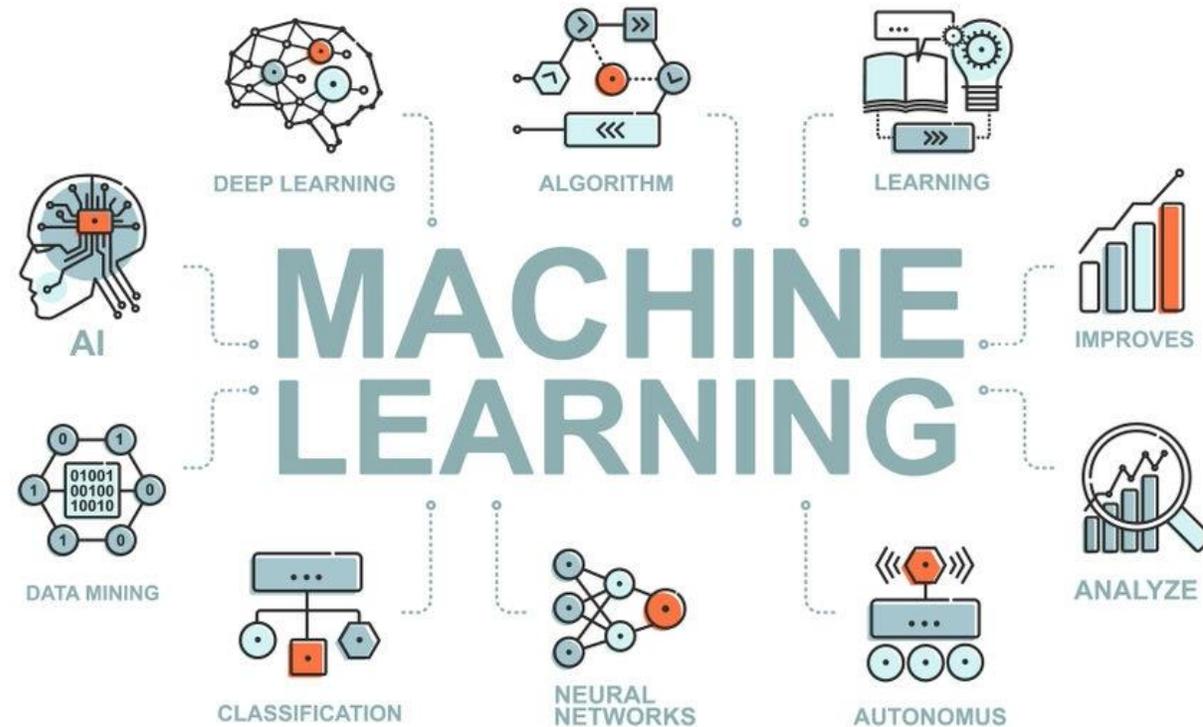


Introdução ao Aprendizado de Máquina



O que é Machine Learning?

*IBM: "O aprendizado de máquina é um ramo da Inteligência Artificial e da ciência da computação que se concentra no uso de **dados e algoritmos para imitar a maneira como os humanos aprendem**, melhorando gradualmente sua precisão"*

"Programar computadores para aprender com a experiência deve eventualmente eliminar a necessidade de grande parte desse esforço de programação detalhada"

–Arthur L. Samuel (1959)

*UC Berkeley: "O conceito básico do aprendizado de máquina envolve o uso de **aprendizado estatístico** (ou seja, encontrar uma **função preditiva** com base em dados) e métodos de otimização que permitem que os computadores analisem conjuntos de dados e identifiquem padrões"*

*O aprendizado de máquina é sobre fazer **previsões e classificações***

O que é Machine Learning?

Springer Series in Statistics

Trevor Hastie
Robert Tibshirani
Jerome Friedman

The Elements of Statistical Learning

Data Mining, Inference, and Prediction

*"Grandes quantidades de dados estão sendo geradas em muitos campos, e o trabalho do estatístico é dar sentido a tudo isso: extrair padrões e tendências importantes e entender 'o que os dados dizem'. Chamamos isso de **aprendizado estatístico**"*

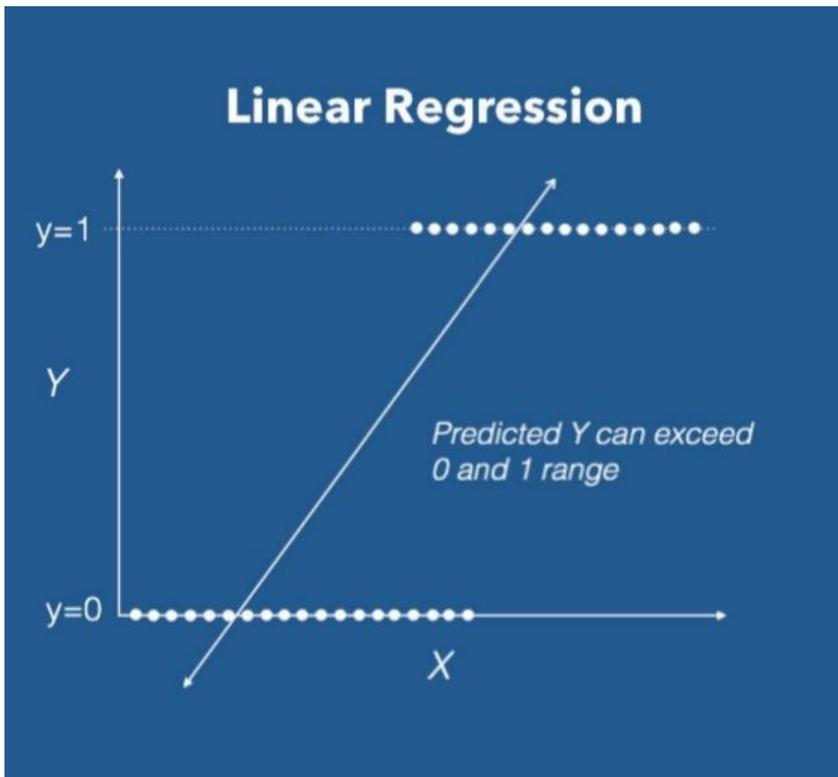
O que envolve Machine Learning?

Um processo de aprendizado baseado em dados: *A algoritmo analisa dados de entrada para identificar padrões que possam ser generalizados. O objetivo é construir um modelo capaz de realizar tarefas como previsão ou classificação utilizando novos (“unseen”) dados*

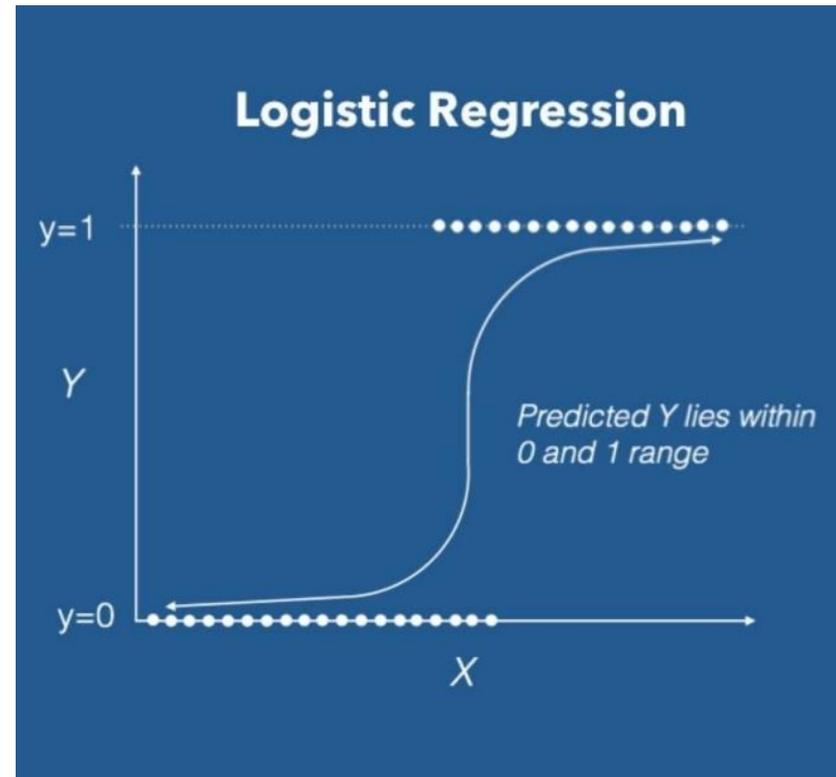
Uma função de erro: *Quantifica a diferença entre as previsões do modelo e os valores reais observados. Essa função orienta o processo de aprendizado, indicando o quão bem o modelo está performando*

Um processo de otimização: *Ajusta iterativamente os parâmetros do modelo (como os pesos em redes neurais), com o objetivo de minimizar a função de erro. Esse processo continua até que um critério de parada seja atingido (como um número máximo de iterações ou um nível aceitável de erro)*

Algoritmos comuns de Machine Learning

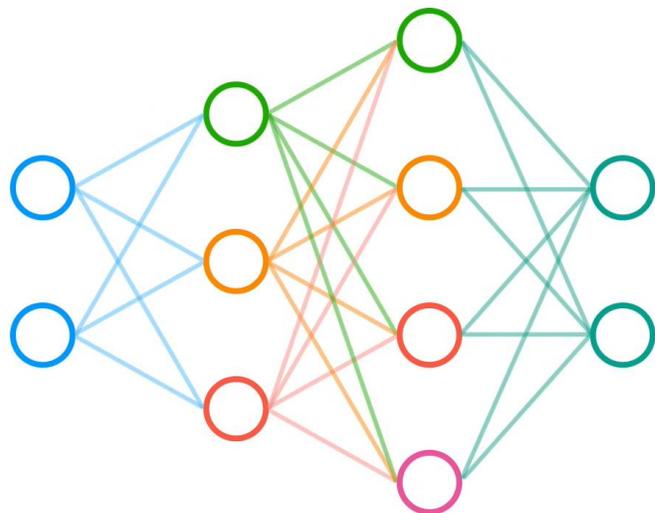


As regressões lineares preveem valores numéricos, com base em uma relação linear entre diferentes valores/variáveis

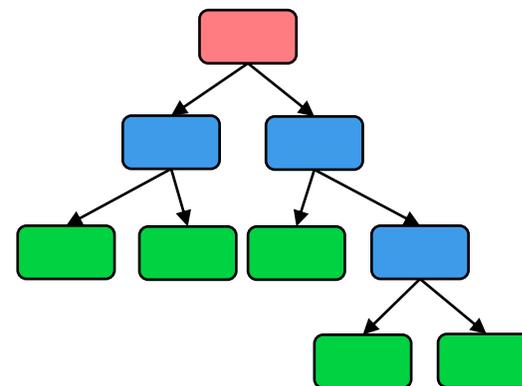


As regressões logísticas fazem previsões para variáveis de resposta categóricas (por exemplo, dados sim/não ou dados 0/1)

Algoritmos comuns de Machine Learning

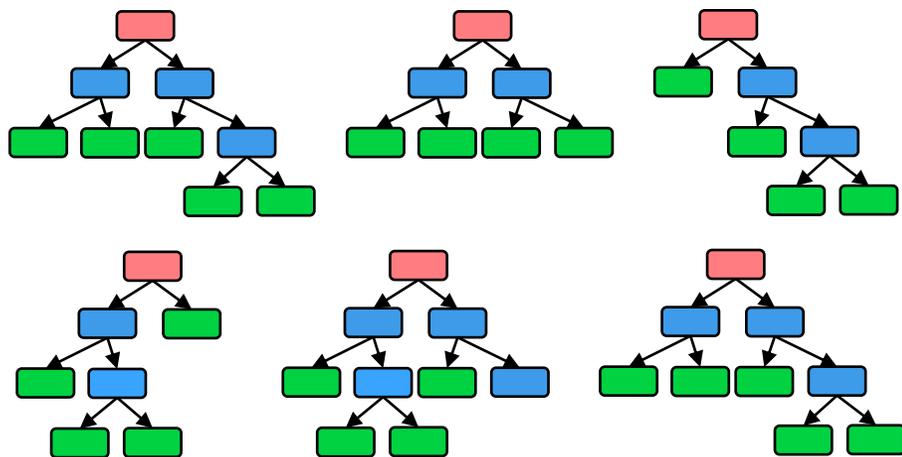


Redes Neurais Artificiais são algoritmos de inteligência artificial que tentam replicar a maneira como o cérebro humano processa informações

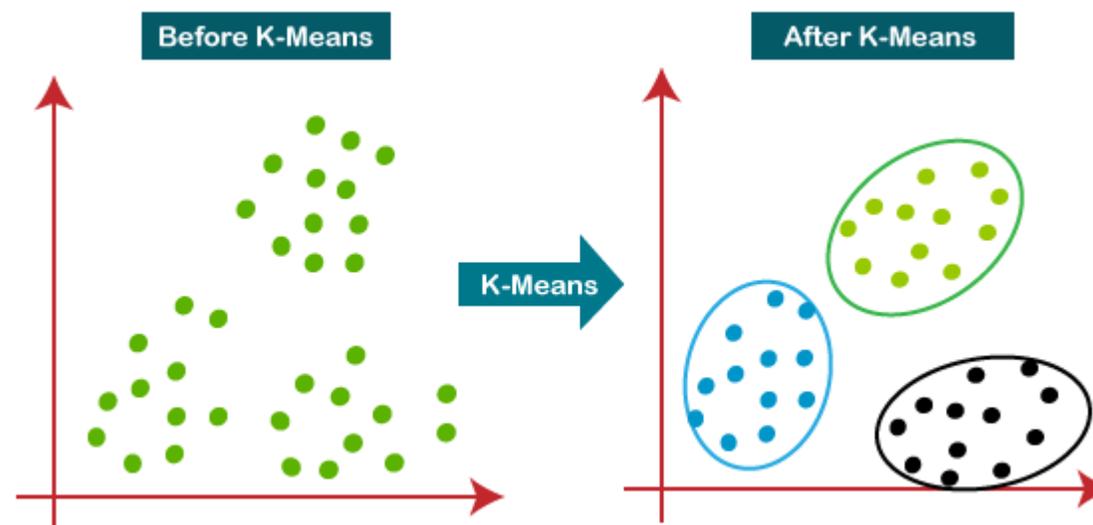


As Árvores de Decisão usam uma sequência ramificada de decisões vinculadas que podem ser representadas com um “diagrama de árvore”

Algoritmos comuns de Machine Learning

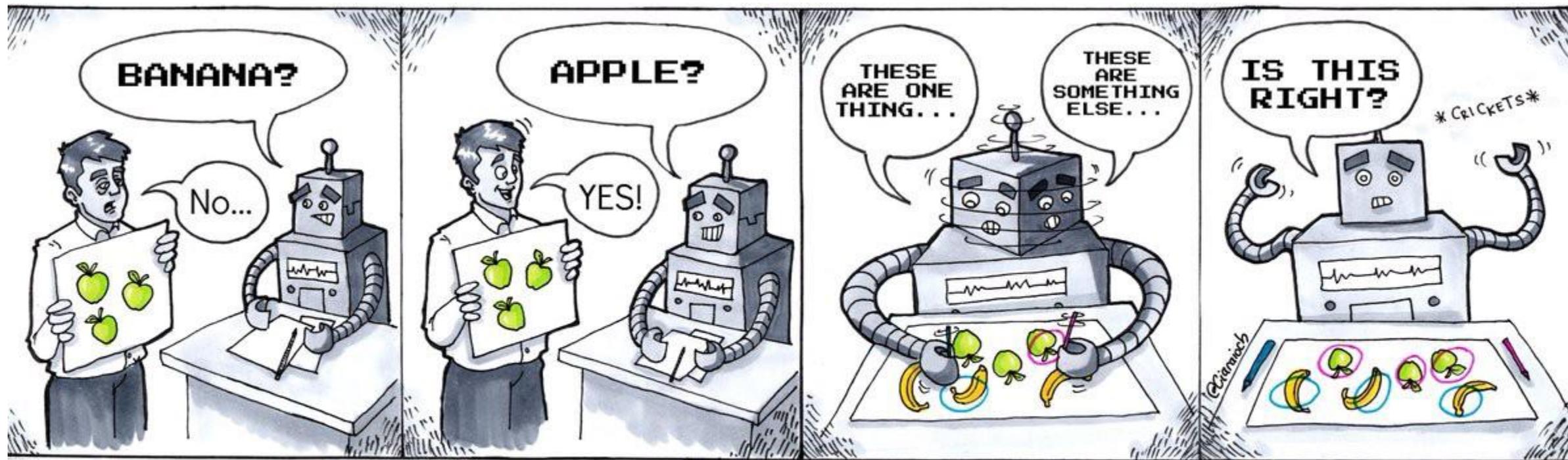


Florestas aleatórias preveem valores ou categorias combinando os resultados de muitas árvores de decisão



K-means divide seus dados em clusters (grupos)

Tipos de "aprendizado" em Machine Learning



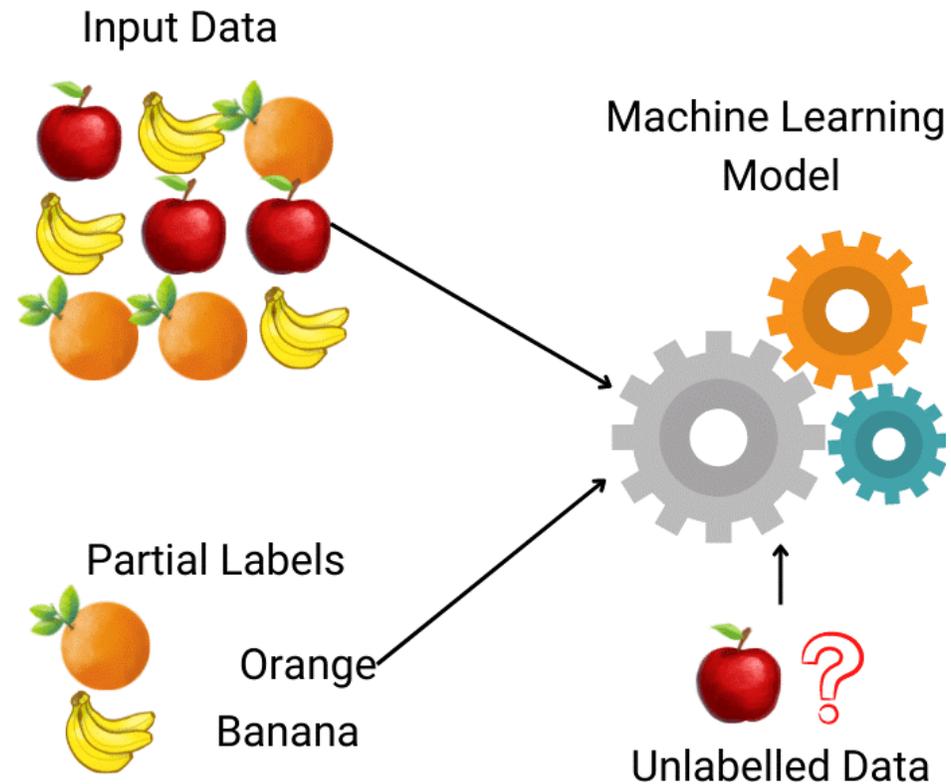
Supervised Learning

Os dados foram pré-rotulados e classificados

Unsupervised Learning

Os dados não são rotulados e o algoritmo identifica padrões e relacionamentos dentro dos dados sem ajuda

Tipos de "aprendizado" em Machine Learning



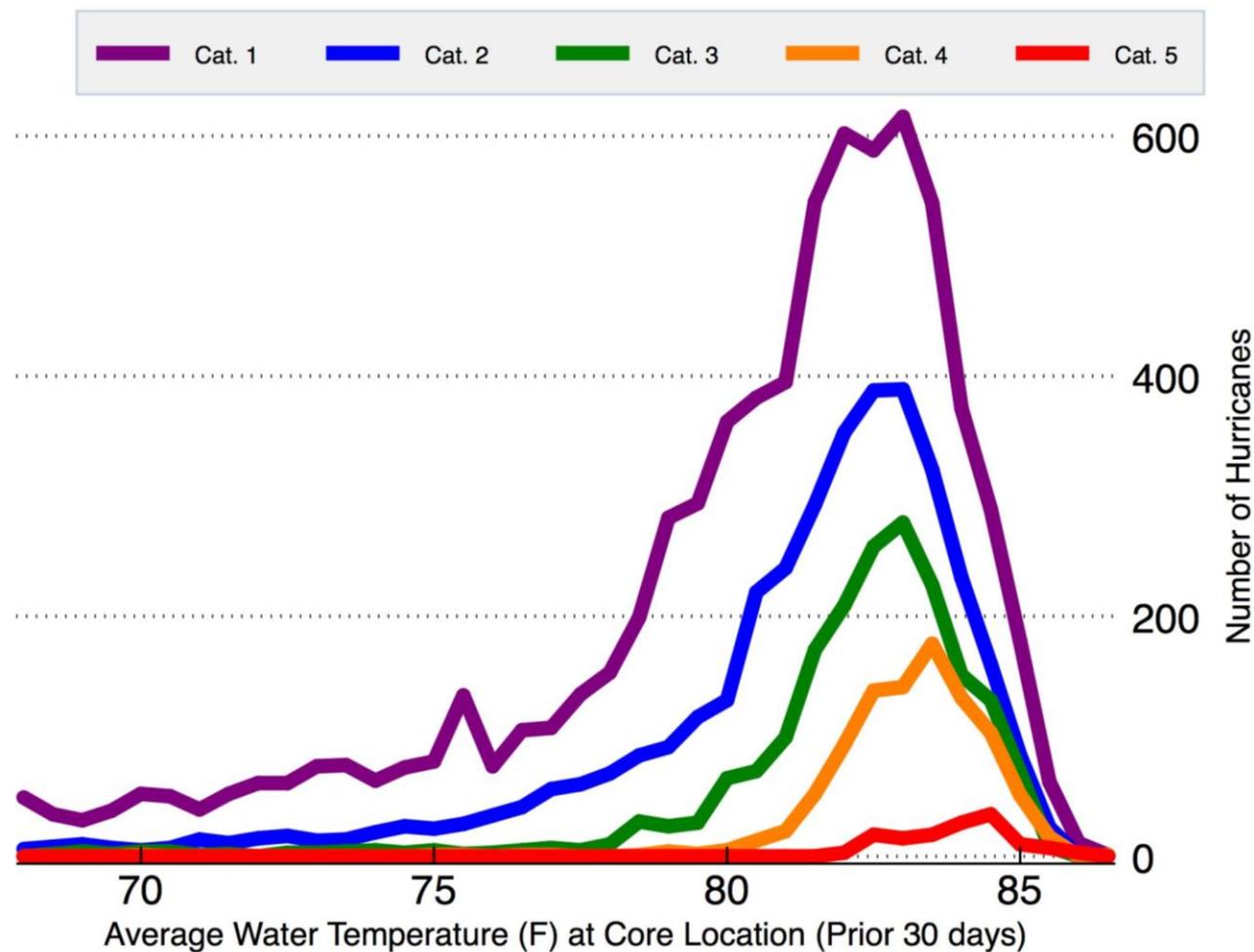
Semi-supervised Learning

*Apenas algumas características dos dados usados foram pré-rotuladas e classificadas
O algoritmo aprende a rotular os dados não rotulados*



O que fazemos com Machine Learning?

Hurricane Strength and Ocean Temperatures



Kernal density functions of SSTs by hurricane category. Area under each curve represents 100% of hurricanes of that type. Hurricane wind speeds via HURDAT.

RECORD OCEAN HEAT

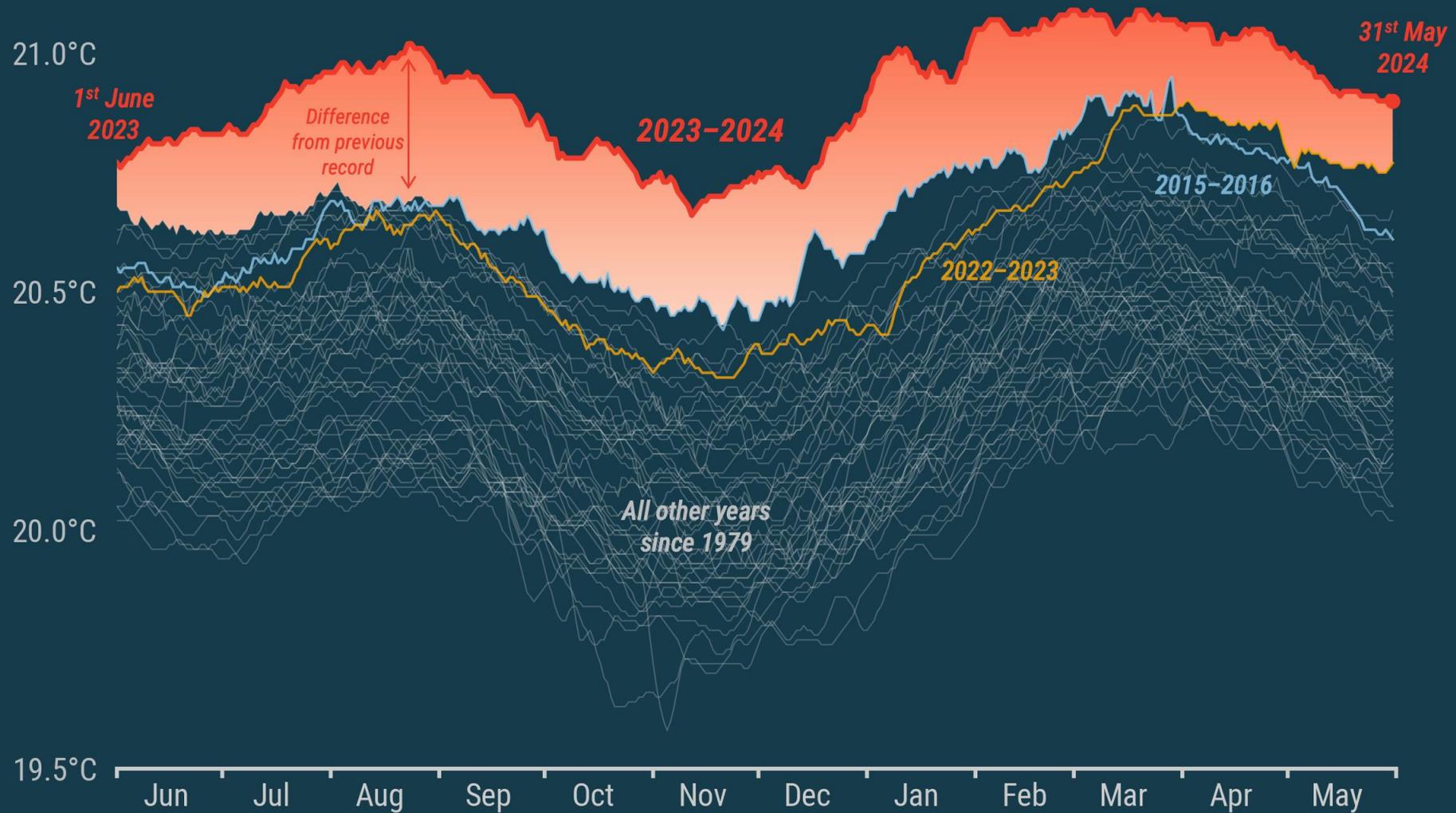
Daily global sea surface temperature (°F)



Source: NOAA OISST

Daily sea surface temperature for 60°S-60°N

Data: ERA5 1979-2024 • Credit: C3S/ECMWF



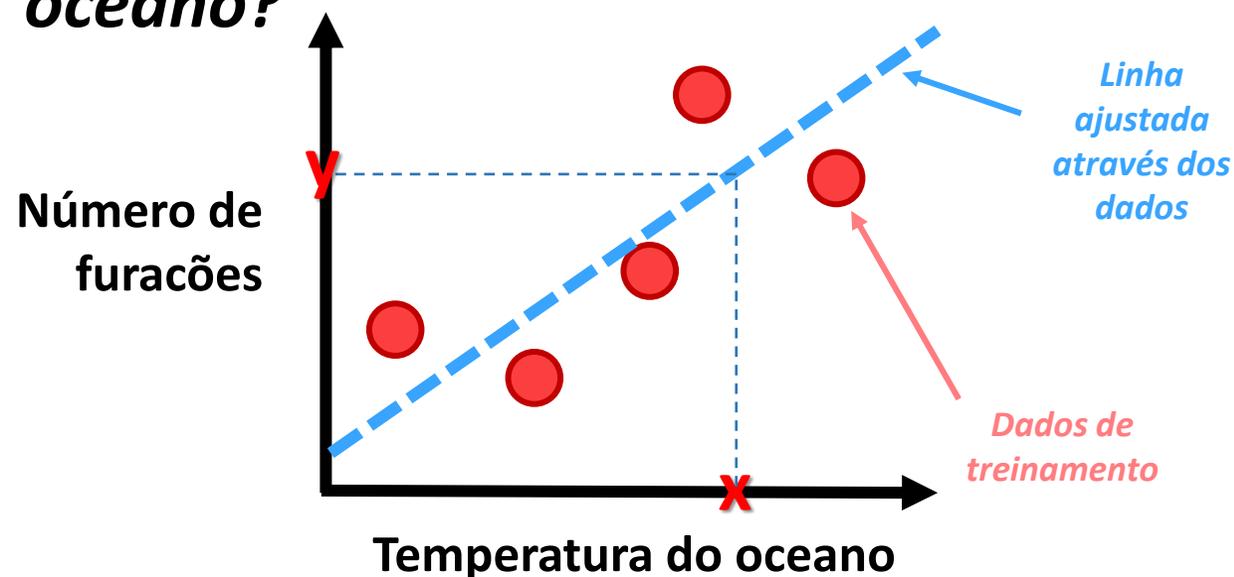
PROGRAMME OF THE
EUROPEAN UNION



O que fazemos com Machine Learning?

O aprendizado de máquina é sobre fazer *previsões* e *classificações*

Podemos prever furacões com base na temperatura do oceano?

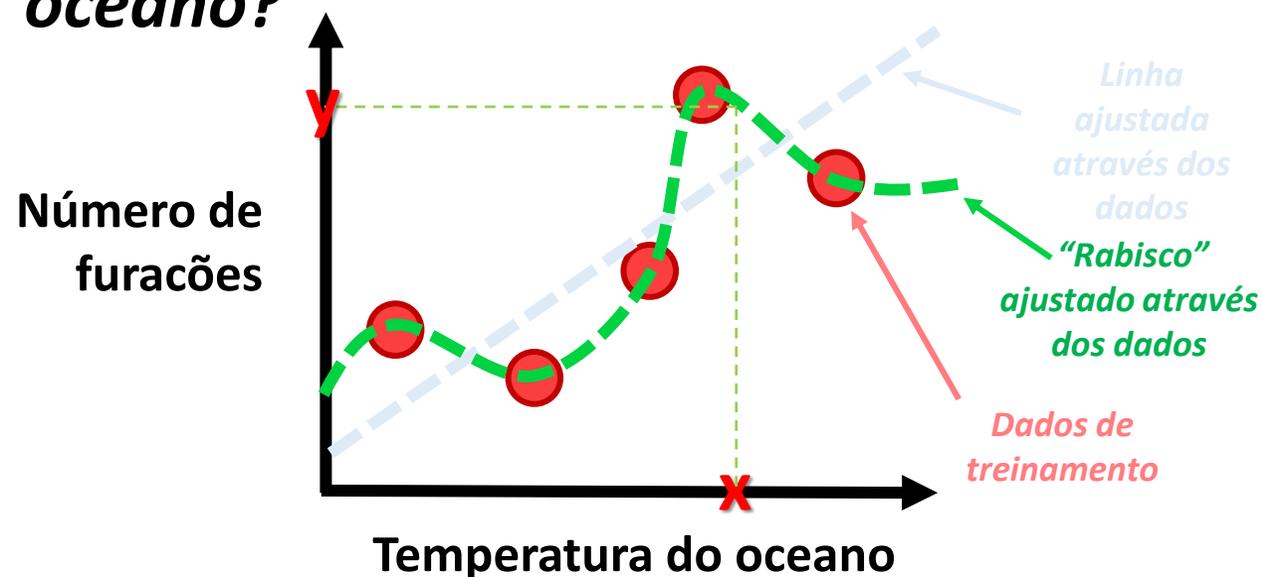


Podemos usar a *linha ajustada* para prever o número de furacões (y) com base na temperatura do oceano (x)

O que fazemos com Machine Learning?

O aprendizado de máquina é sobre fazer *previsões* e *classificações*

Podemos prever furacões com base na temperatura do oceano?

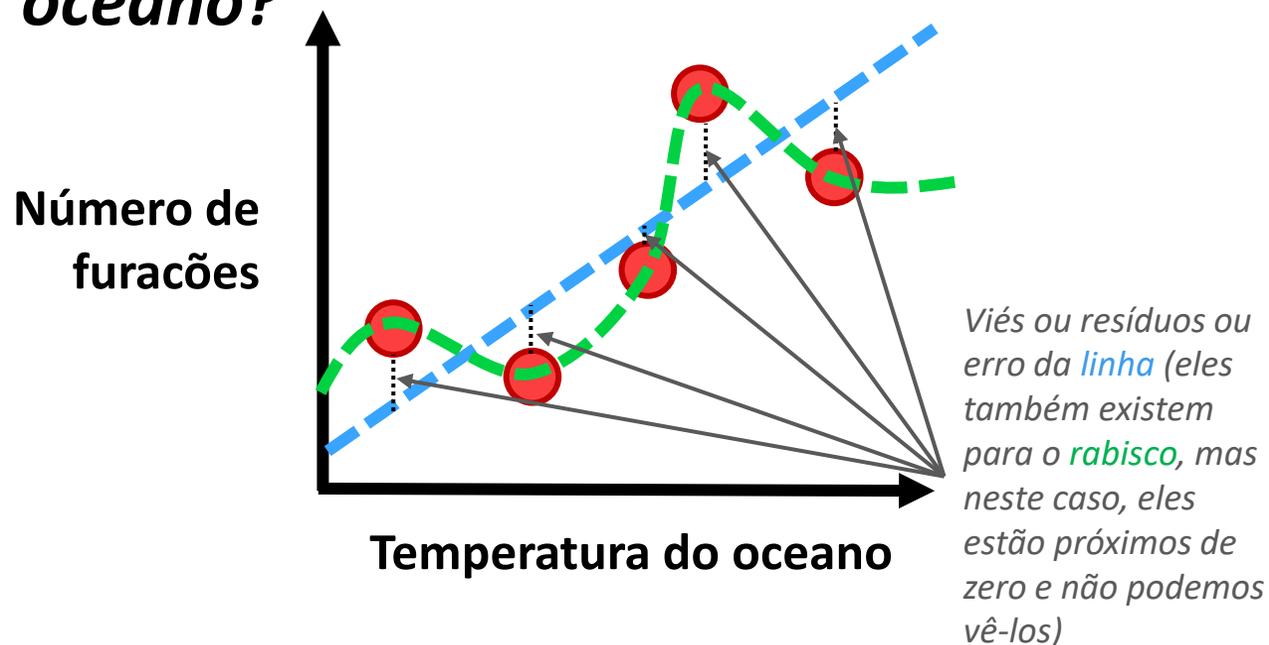


Podemos usar o “**rabisco**” **ajustado** para prever o número de furacões (y) com base na temperatura do oceano (x)

O que fazemos com Machine Learning?

O aprendizado de máquina é sobre fazer **previsões e classificações**

Podemos prever furacões com base na temperatura do oceano?



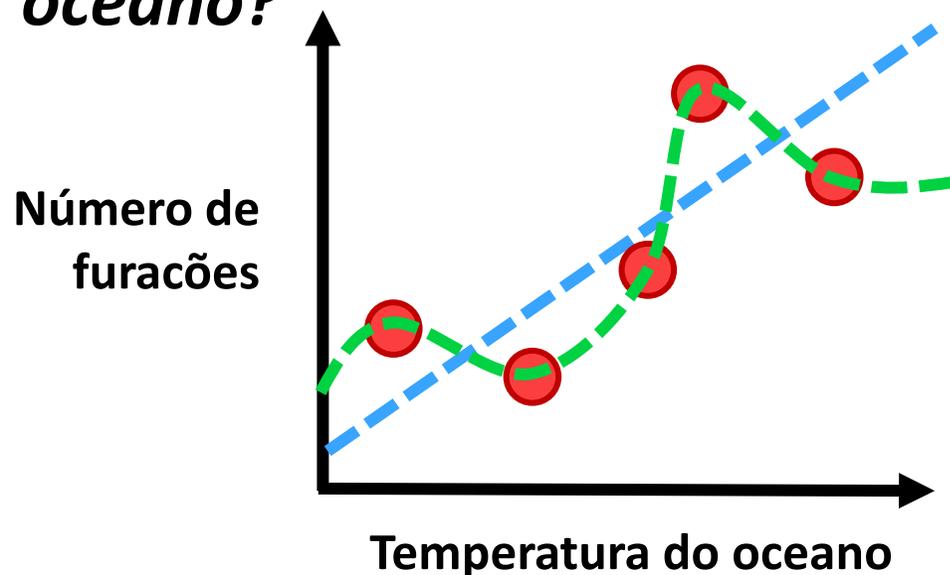
Podemos dizer que o **rabisco** se encaixa muito melhor nos **dados de treinamento** = **menor viés (bias) ou resíduos ou erro**

Nota: Na Análise de Regressão, "resíduos" é o termo mais popular, no Machine Learning, "viés" ou simplesmente "erro" tendem a ser os termos mais populares

O que fazemos com Machine Learning?

O aprendizado de máquina é sobre fazer *previsões* e *classificações*

Podemos prever furacões com base na temperatura do oceano?



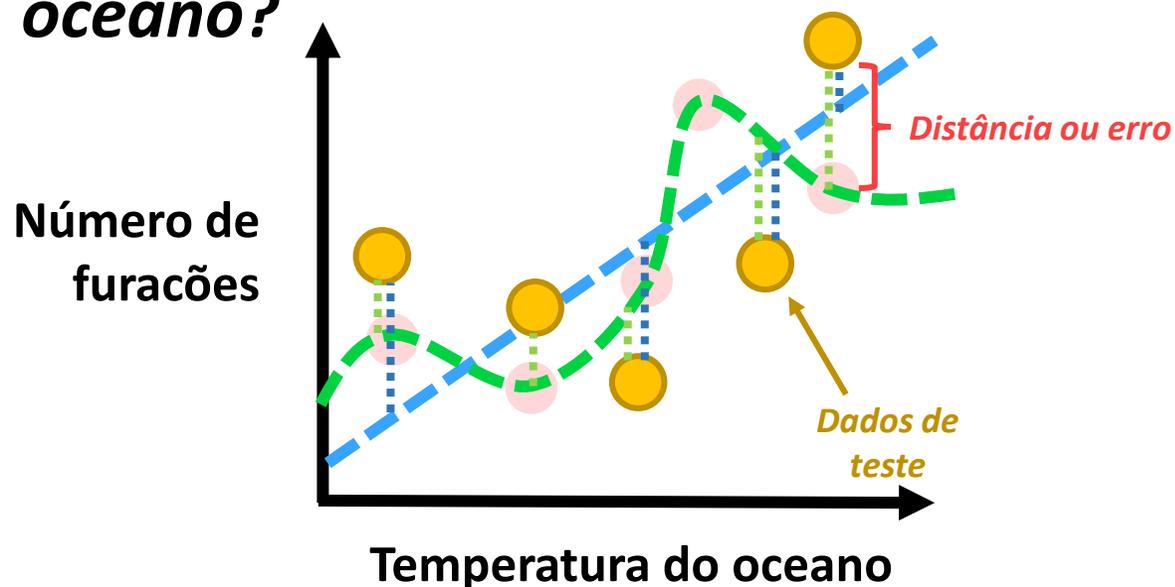
... mas se queremos usar o melhor modelo para prever o número de furacões, como saber qual modelo é melhor?

→ **Dados de teste**

O que fazemos com Machine Learning?

O aprendizado de máquina é sobre fazer **previsões** e **classificações**

Podemos prever furacões com base na temperatura do oceano?



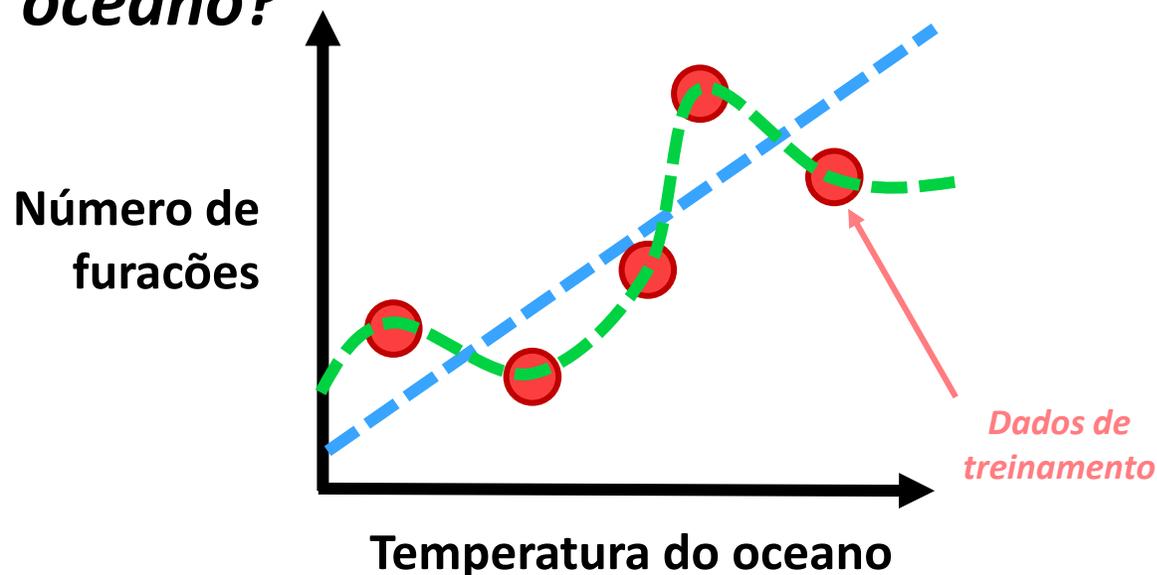
Normalmente, somamos as distâncias quadradas associadas à **linha ajustada** e ao **rabisco ajustado** e verificamos qual delas tem o **menor erro** → **“maior precisão”**

Obs.: Existem diferentes métricas para medir o erro ou a precisão do modelo, não apenas a soma das distâncias quadradas

O que fazemos com Machine Learning?

O aprendizado de máquina é sobre fazer *previsões* e *classificações*

Podemos prever furacões com base na temperatura do oceano?

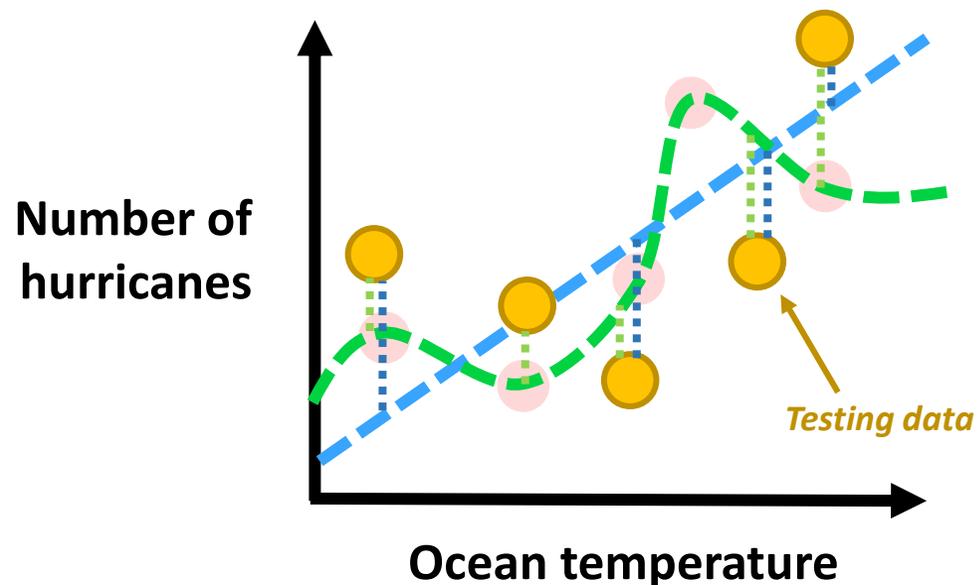


Mesmo que o *rabisco* se ajuste melhor aos *dados de treinamento* do que a *linha* = “viés” mais baixo ...

O que fazemos com Machine Learning?

O aprendizado de máquina é sobre fazer *previsões* e *classificações*

Can we predict hurricanes based on ocean temperature?

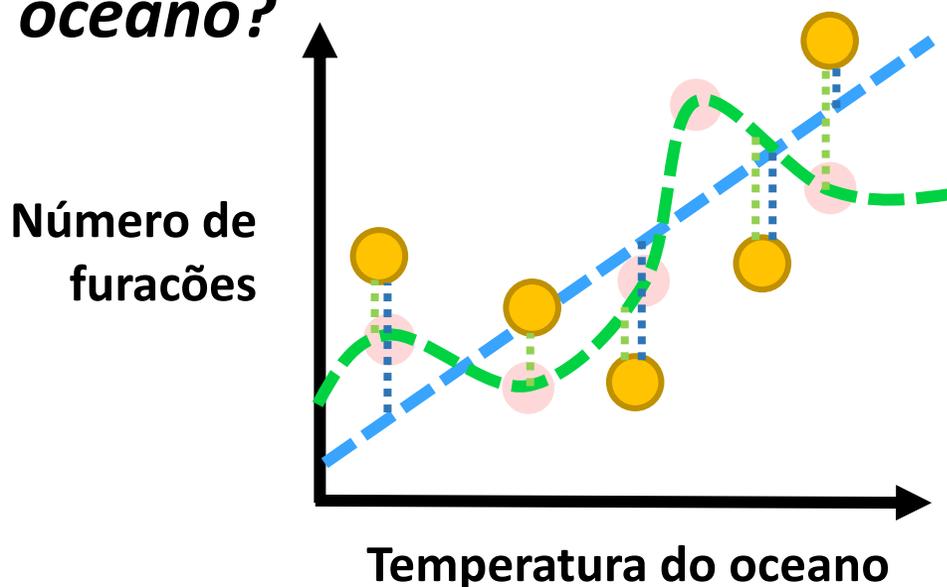


A *linha* pode se ajustar melhor aos *dados de teste* do que o *rabisco* = menor “variância”

O que fazemos com Machine Learning?

O aprendizado de máquina é sobre fazer **previsões e classificações**

Podemos prever furacões com base na temperatura do oceano?



O **rabisco** pode fazer um péssimo trabalho em fazer previsões às vezes (por causa da variância alta) = **previsões inconsistentes, algumas ótimas/péssimas...**

A linha pode nunca fazer previsões super boas (por causa do viés alto), mas também não fará previsões terríveis (por causa da variância baixa) = **previsões consistentemente boas, não ótimas...**

Isso é chamado de **trade-off de viés-variância** em Machine Learning (**bias-variance trade-off**)

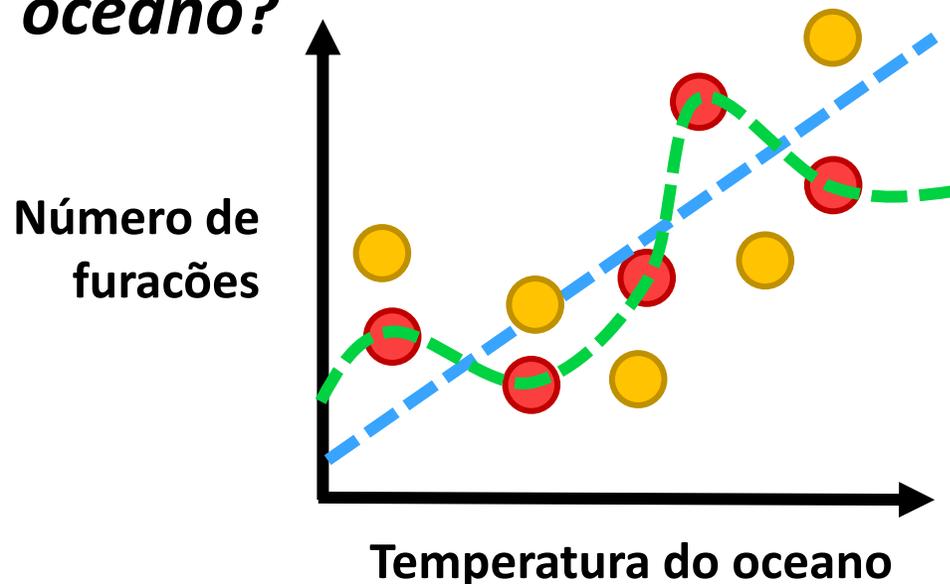
Nota:

Modelos simples tendem a ter um viés alto, mas uma baixa variância
Modelos complexos tendem a ter um viés baixo, mas uma alta variância

O que fazemos com Machine Learning?

O aprendizado de máquina é sobre fazer **previsões** e **classificações**

Podemos prever furacões com base na temperatura do oceano?



Podemos dizer que um modelo está sobreajustado (overfitted) aos **dados de treinamento** se suas previsões estiverem muito erradas quando comparadas aos **dados de teste**

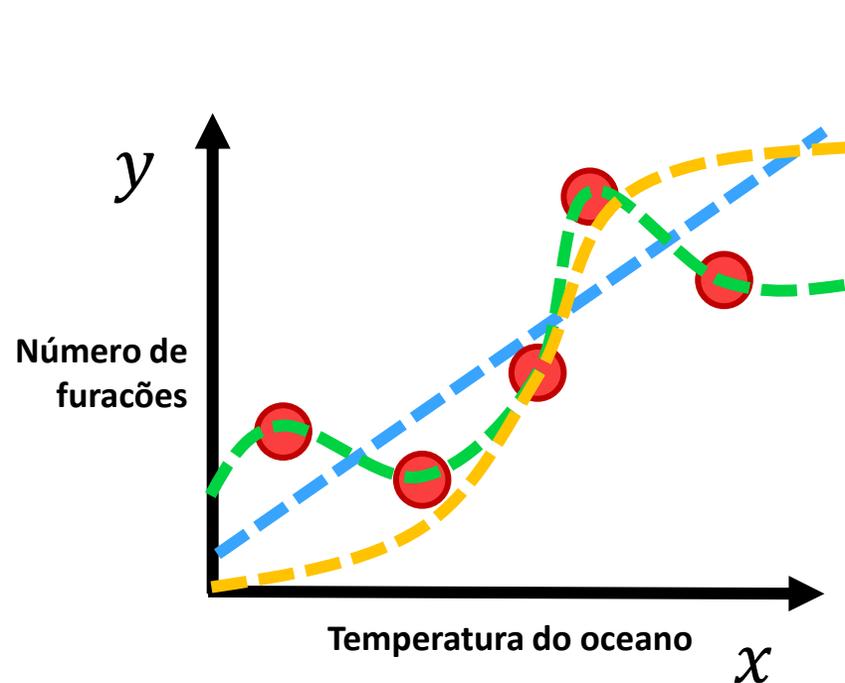
O objetivo é encontrar um modelo com o menor viés e variância possíveis
→ **Trade-off ótimo de viés-variância**

O que fazemos com Machine Learning?

Exemplo em Excel...

O que fazemos com Machine Learning?

Algoritmos diferentes para propósitos diferentes...



$$y = \alpha + \beta x + \varepsilon$$

Intercepto

Parâmetro

Erro
(resíduo)

$$y = \alpha + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_n x_n + \varepsilon$$

Interceptos e parâmetros estimados (e sua significância) podem ser facilmente interpretados e de grande relevância

Algoritmos diferentes para propósitos diferentes

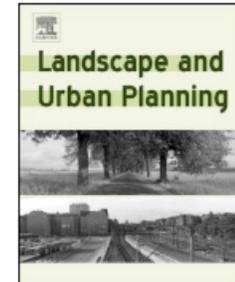
Landscape and Urban Planning 198 (2020) 103780



Contents lists available at [ScienceDirect](#)

Landscape and Urban Planning

journal homepage: www.elsevier.com/locate/landurbplan



Comparison of spatial modelling frameworks for the identification of future afforestation in New Zealand

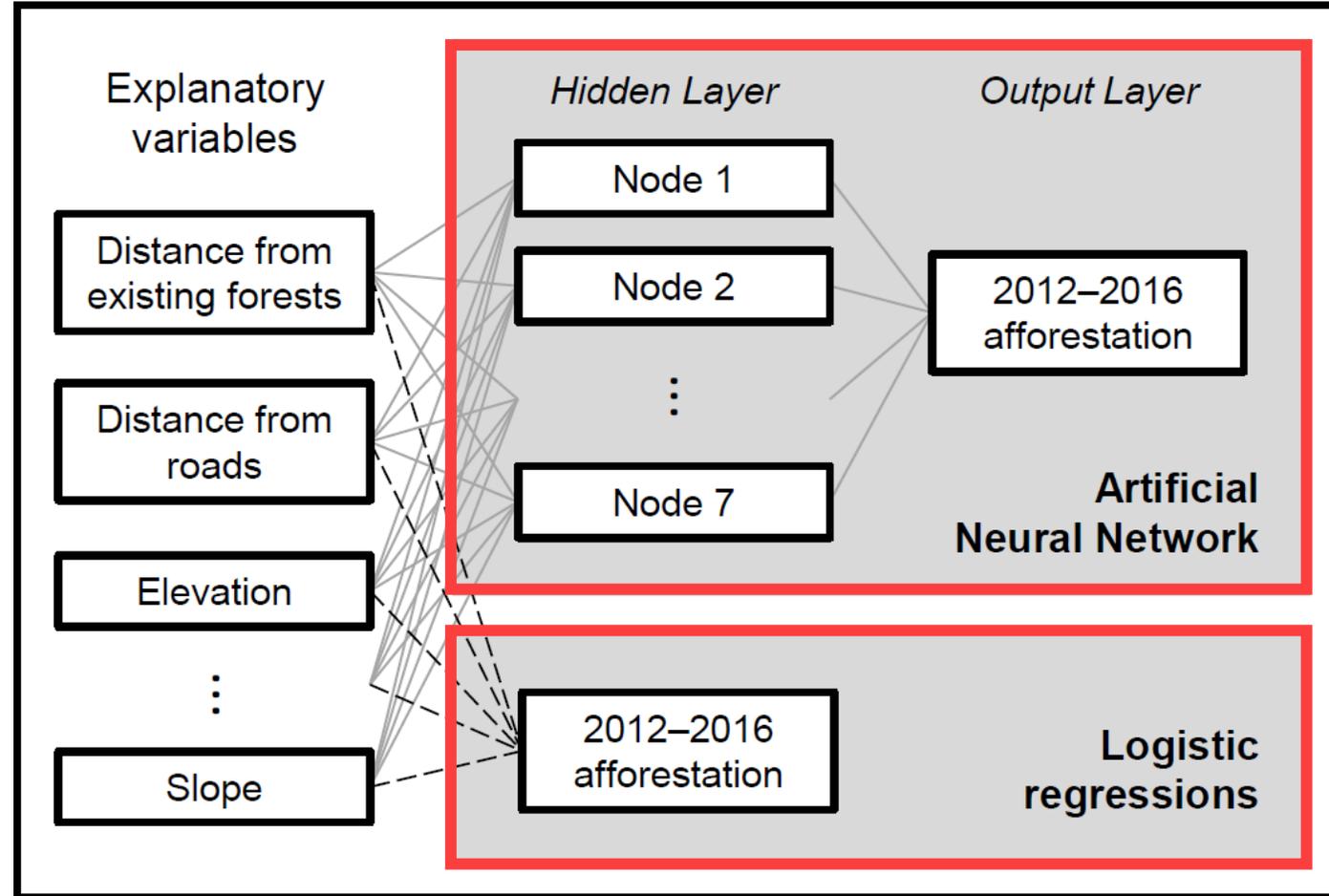


Thales A.P. West^{a,*}, Juan J. Monge^a, Les J. Dowling^a, Steve J. Wakelin^a, Richard T. Yao^a, Andrew G. Dunningham^a, Tim Payn^{a,b}

^a Scion—New Zealand Forest Research Institute, 3010 Rotorua, New Zealand

^b Toi Ohomai Institute of Technology, 3015 Rotorua, New Zealand

Algoritmos diferentes para propósitos diferentes



Algoritmos diferentes para propósitos diferentes

Regressão Logística

Table 1

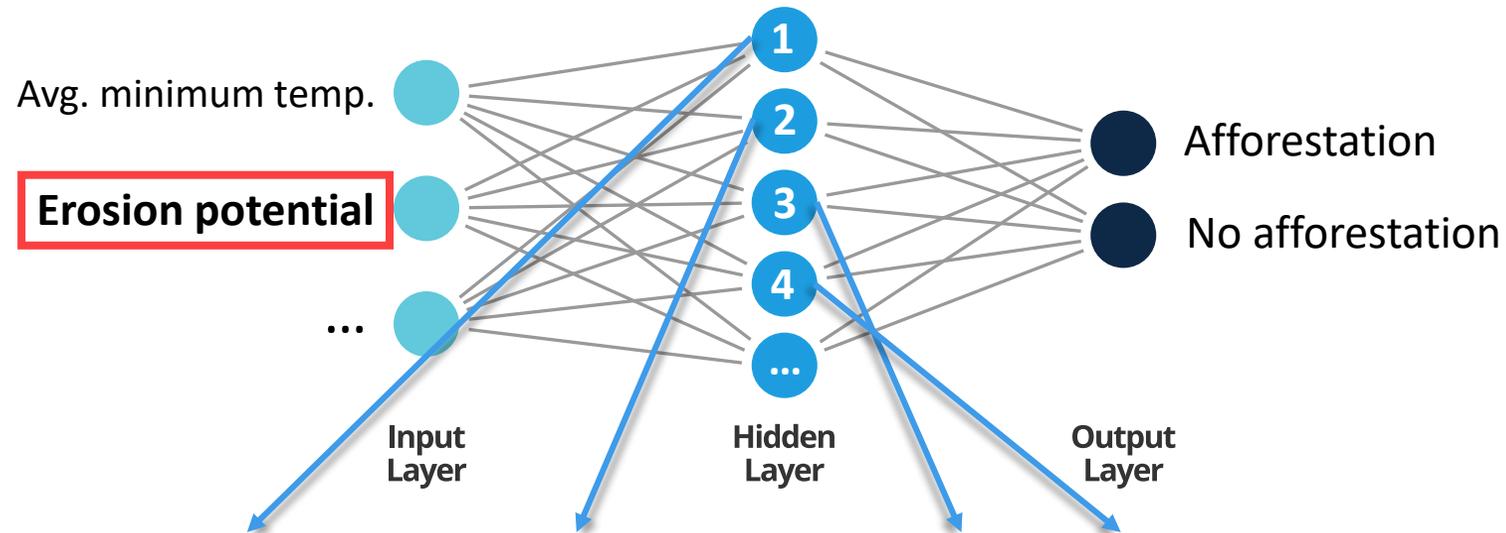
Results from the logistic regressions and the artificial neural network (ANN) model.

Variable	Logistic regression	
	Estimate	P-value
Mean annual minimum temperature	-0.05 (0.091)	0.58
Erosion potential	0.37 (0.048)	< 0.00***
Distance from 2012 exotic forests	-1.167 (0.055)	< 0.00***
Plant-rooting potential	0.106 (0.035)	< 0.00***
<i>Pinus radiata</i> Site Index	0.176 (0.084)	0.04**
Low-productivity grasslands (dummy)	1.697 (0.094)	< 0.00***
Soil class	-	-
Woody grasslands (dummy)	1.875 (0.12)	< 0.00***
Elevation	0.432 (0.056)	< 0.00***
Slope	0.192 (0.043)	< 0.00***

[West et al. \(2020\)](#)

Algoritmos diferentes para propósitos diferentes

Rede Neural Artificial

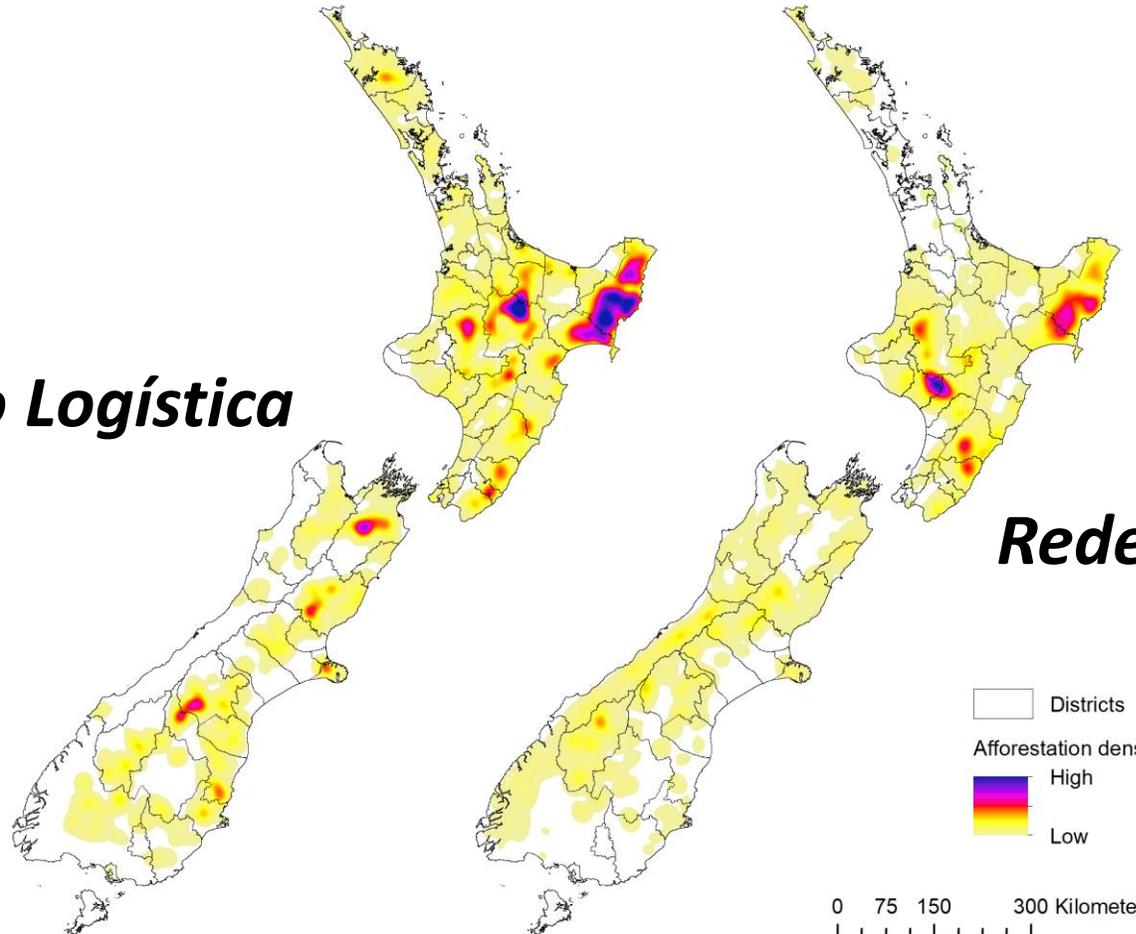


	Hidden Node 1	Hidden Node 2	Hidden Node 3	Hidden Node 4	...
Avg. minimum temp.	-4.7846	0.6514	-0.0377	3.2254	...
Erosion potential	4.8254	0.1605	0.0891	-0.2653	...
Distance to exotics	0.5620	0.2345	0.0892	-0.4920	...
Plant-rooting potent.	0.5114	0.2038	0.1367	-0.3558	...
Site index	0.3542	0.3040	0.1179	-0.1726	...
Low prod. grassland	-2.3447	0.6591	0.1593	2.2832	...
...

Algoritmos diferentes para propósitos diferentes

Previsão de futuros focos de reflorestamento na Nova Zelândia

Regressão Logística



***Qual modelo/mapa
é o "melhor"?
Validação de
modelo***

Rede Neural Artificial

[West et al. \(2020\)](#)

Resumo

- *O aprendizado de máquina é sobre fazer **previsões e classificações***
- *Não existe "o melhor" algoritmo de aprendizado de máquina*
- *Alguns algoritmos são fáceis de interpretar, outros não*
- *Otimizar o **trade-off de viés-variância** é fundamental*
- *Evitar o **sobreajuste (overfitting)** também é fundamental*
- ***Validação de modelo** também é fundamental*