



FIAP



MACHINE LEARNING & MODELLING

Esta disciplina aborda os principais conceitos sobre aprendizado de máquina e as técnicas clássicas de modelagem



Na última aula...

- Parametrização dos modelos baseados em árvores
- Como evitar overfitting
- Criação de uma árvore de decisão via sklearn

Agenda

- Análise exploratória de dados
- Revisão de conceitos de Machine Learning
- Regressão linear
 - Intuição
 - Construção

Revendo conceitos

Aprendizado de Máquina

Supervisionada
(usa labels)

Não-supervisionada
(NÃO usa labels)

Por
reforço
(ambiente)

Regressão

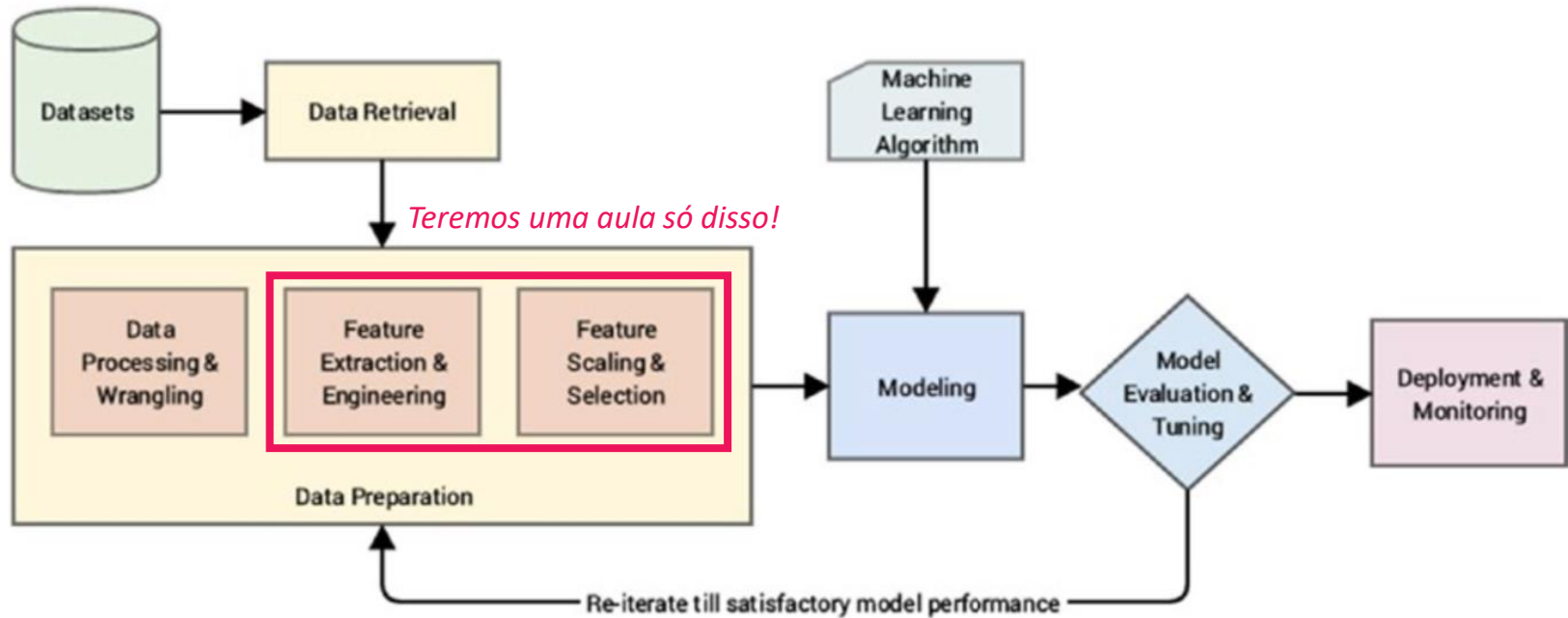
Classificação

Redução de
dimensionali
dade

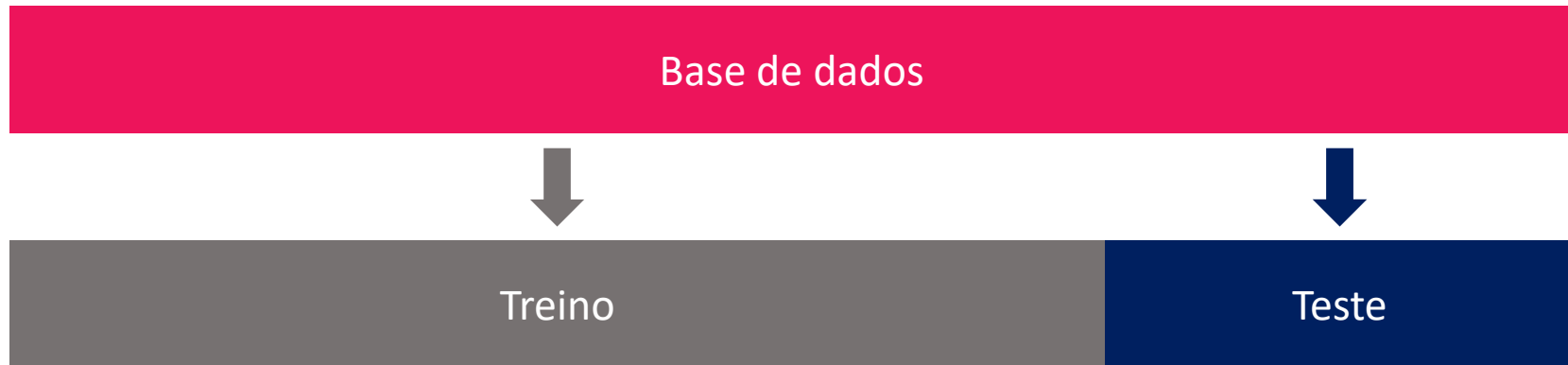
Clusterização

O pipeline de machine learning

Foco nos aspectos técnicos



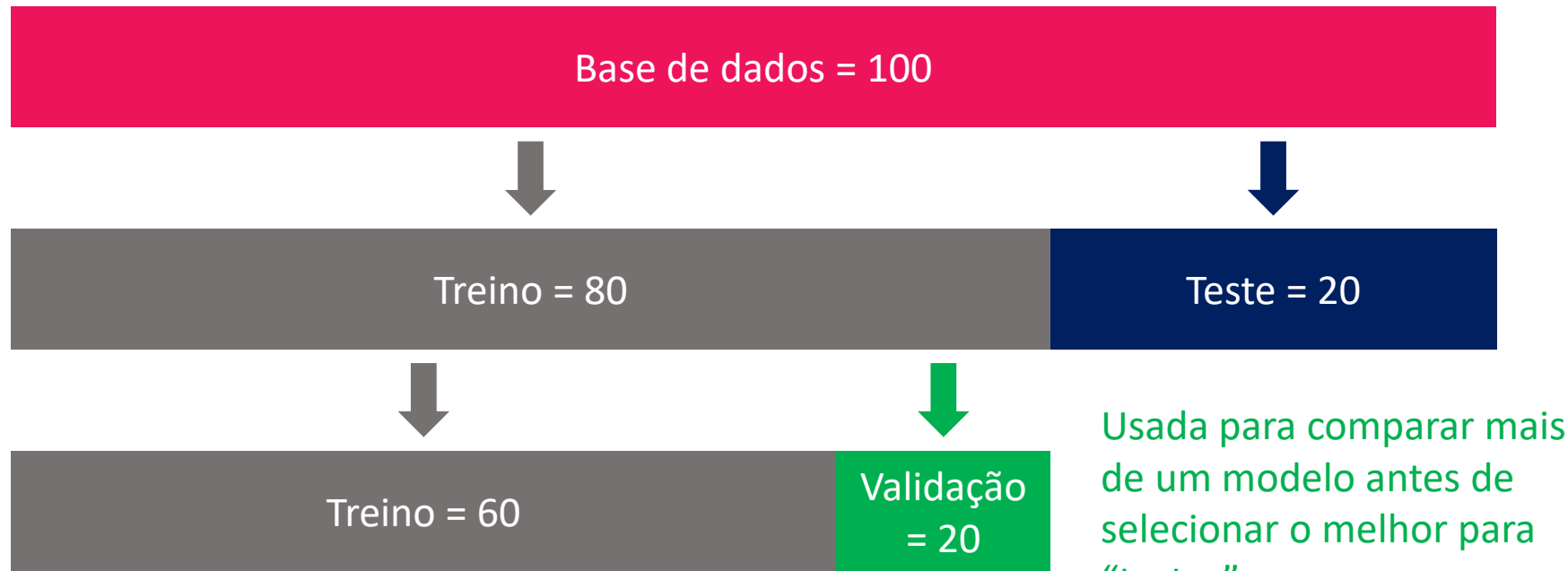
Uso da base de dados



É comum usarmos 80% dos dados para treinamento e **separar** 20% para teste.

Agora, e se precisamos escolher qual é o melhor entre 2 ou mais modelos?

Uso da base de dados



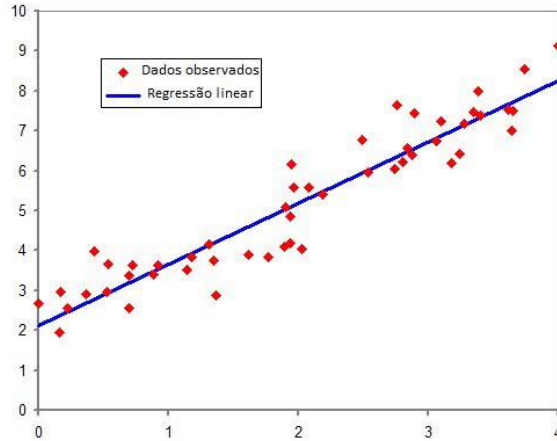
Usada para comparar mais de um modelo antes de selecionar o melhor para “testar”

Ex: árvore X floresta ; RN

Regressão linear

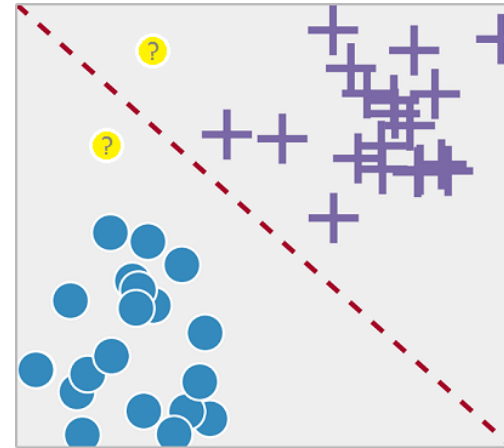
A tarefa

Regressão



A partir de um novo valor de entrada (da variável X), prever o valor da resposta (variável Y)

Classificação



A partir de um novo valor de entrada (variáveis $X1$ e $X2$), prever a qual classe (*label*) ela pertence

Problemas clássicos



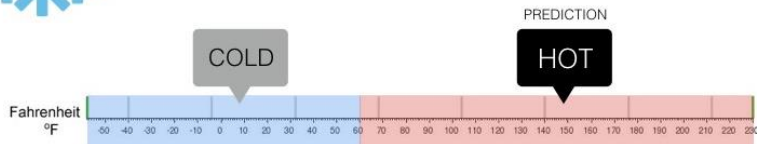
Regression

What is the temperature going to be tomorrow?



Classification

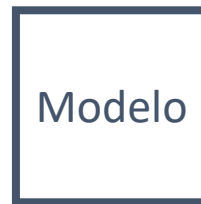
Will it be Cold or Hot tomorrow?



Regressão



Aluno

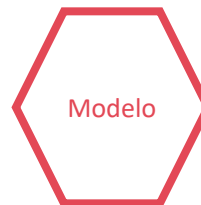


Qual nota terá?

Classificação



Aluno

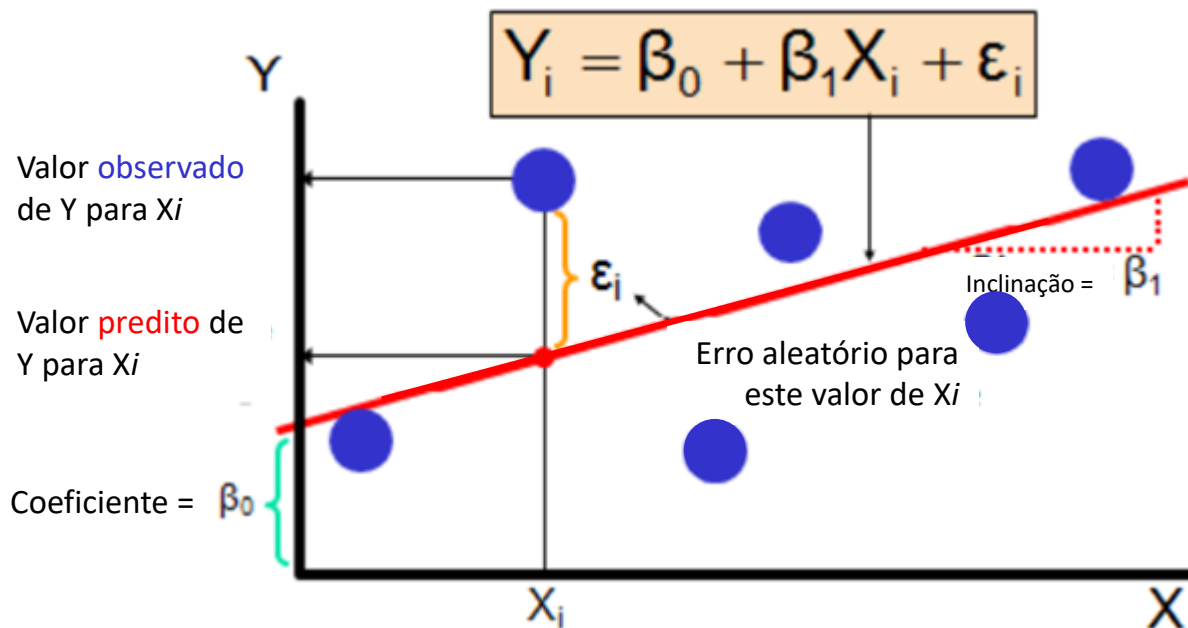


Vai passar em ML ou não?

Regressão e Classificação são intercambiáveis?

- *Sempre* conseguimos alterar os labels (rótulos) de **regressão** (valores) para rótulos de **classificação** (classes), definindo as faixas de valores de cada classe.
 - Exemplo: valor de um imóvel → faixa de valor do imóvel
- A volta não é sempre verdadeira! Nem sempre temos informação numérica que nos permita aplicar **regressão**.
 - Exemplo: imóvel acima ou abaixo do preço de mercado e sem a informação do valor em si.

Regressão linear – intuição



- Podemos descrever a reta por uma equação que mapeia os dados de entrada (X) na variável resposta (Y)
- **Objetivo** da regressão linear:
 - Qual é a “reta” que melhor se ajusta aos dados?
 - Ou seja, determinar os **coeficientes**!

Regressão linear – avaliação do modelo

- Para avaliar nossos modelos de regressão precisamos de métricas que quantifiquem os erros de forma não-binária (ao contrário da classificação)
- Principais métricas utilizadas:
 - **MSE** (*Mean Squared Error* – Erro quadrático médio)
 - **MAE** (*Mean Absolute Error* – Erro absoluto médio)
 - **R²** (“R quadrado”):

Regressão linear – avaliação do modelo

MSE – Erro Quadrático Médio

$$MSE = \frac{1}{N} \sum_{i=1}^N (y_i - \hat{y})^2$$

- Penaliza mais erros **maiores**, já que os erros (diferença entre o valor previsto e o correto) são **elevados ao quadrado**.
- Podemos ter dificuldade em interpretar no contexto real o que o erro obtido significa

MAE – Erro Absoluto Médio

$$MAE = \frac{1}{N} \sum_{i=1}^N |y_i - \hat{y}|$$

- Faz a média do **erro absoluto** (sem sinal) de cada previsão.
- Facilita a interpretação no modelo real
- Porém, temos que ter sempre mente que erros maiores (*outliers*) podem atrapalhar a ideia obtida pela média

R^2 – R quadrado

$$R^2 = 1 - \frac{\sum (y_i - \hat{y})^2}{\sum (y_i - \bar{y})^2}$$

- Esta métrica que varia entre $[-\infty, +1]$
- Indica o quão bom o nosso modelo está em comparação com um modelo “ingênuo” (previsões baseados no valor médio do *target*).
- Quanto **maior** seu valor, **melhor** nosso modelo (em relação ao ingênuo)

ONDE:

y : valor real y | \hat{y} : valor predito de y | \bar{y} : média dos valores de y

E na prática?

Construindo um pipeline de **Regressão Linear** no Sklearn



OBRIGADO!



Prof. Michel Fornaciali

<https://www.linkedin.com/in/michelfornaciali/>
profmichel.fornaciali@fiap.com.br

Copyright © 2021 | Professor Michel Fornaciali

Todos os direitos reservados. Reprodução ou divulgação total ou parcial deste documento, é expressamente proibido sem consentimento formal, por escrito, do professor/autor.