

Mais sobre Seleção de Modelos

Lasso



INFORMAÇÃO,
TECNOLOGIA
& INOVAÇÃO

Revisão Atividade 1



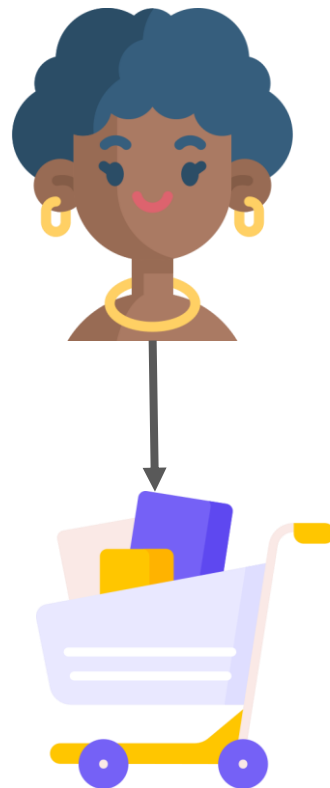
Atividade 1

Objetivo

Colocar em prática os conceitos de aprendizado supervisionado vistos em aula, à fim de ajustar corretamente uma regressão linear para um problema real

O que era esperado

- Análise e tratamento dos dados
- Utilização de técnicas de ajuste de modelos
- Comparação de modelos
- Entendimento total do código
- Impactos e conclusões de negócio



Atividade 1 - Principais erros

Análise descritiva

- A análise descritiva faz parte do processo de modelagem
- Entenda os dados
- Identifiquem o que é variável resposta e o que é variável explicativa
- Identifique missing, outliers e outros problemas de dados

Modelagem

- Entendam qual técnica de aprendizado faz mais sentido para o dados
- Testem diferentes modelos
- Avaliem os modelos por alguma medida de comparação
- Seleccionem variáveis de forma inteligente
- Conclua qual modelo fornece o melhor ajuste

Entendimento do código

- Faça comentários explicando seu raciocínio (o porquê de utilizar determinada função, alguma característica importante dos dados, motivação para próximos passos)
- Tenha uma solução de negócio: são dados reais, a análise vai guiar alguma tomada de decisão, deixe isso claro!

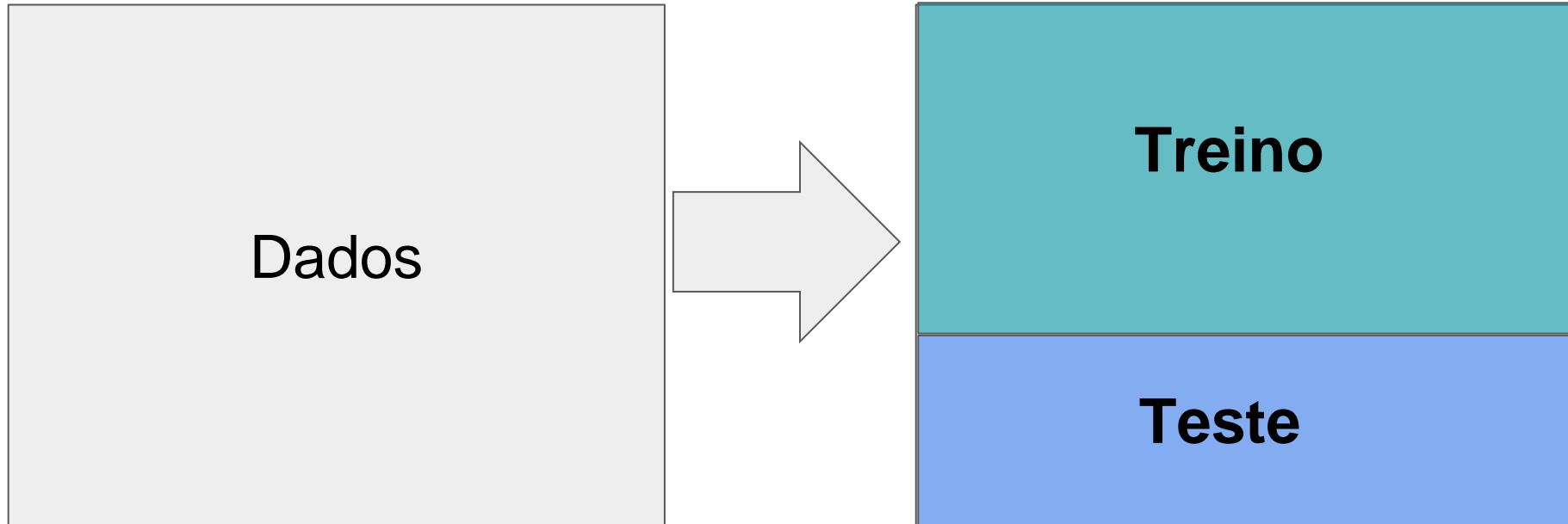
Estrutura

- Siga uma ordem lógica no desenvolvimento da sua análise
- Explícite a interpretação do modelo
- Deixe claro os membros do grupo
- Retire comentários desnecessários ou que não impactem diretamente o código ou a interpretação
- Façam a atividade utilizando o jupyter notebook

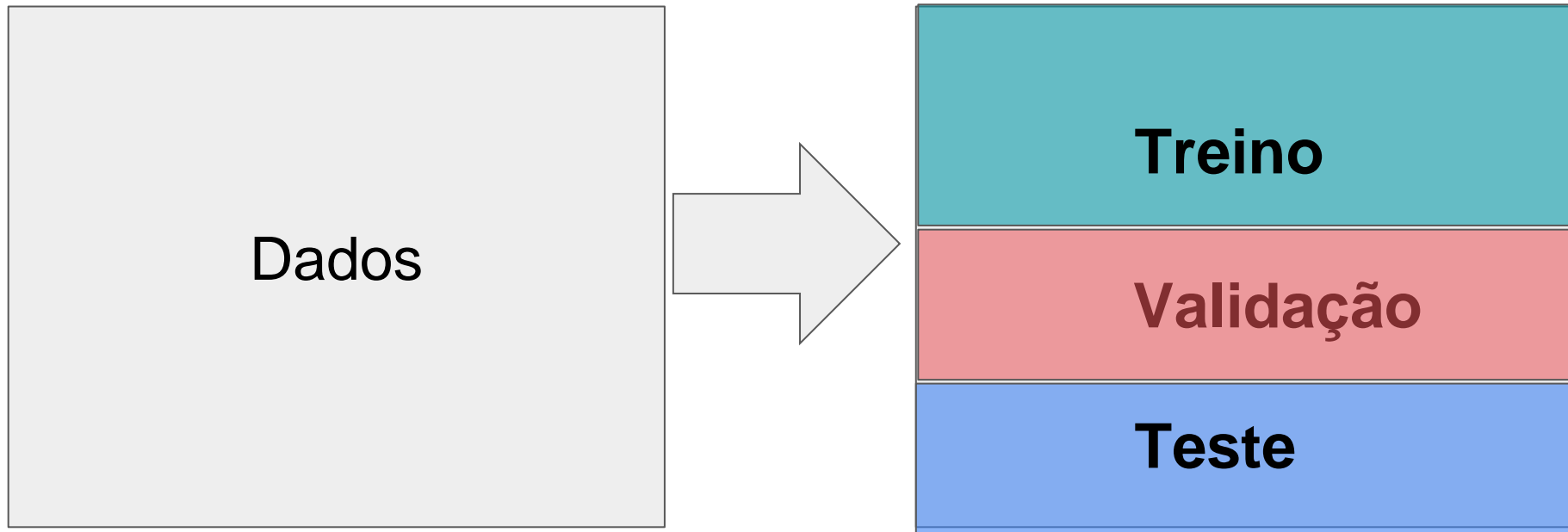
MAIS SOBRE DATA-SPLITTING



Data Splitting



Data Splitting



Nested CV

Dados



Tamanho dos bancos



LASSO



Esparsidade

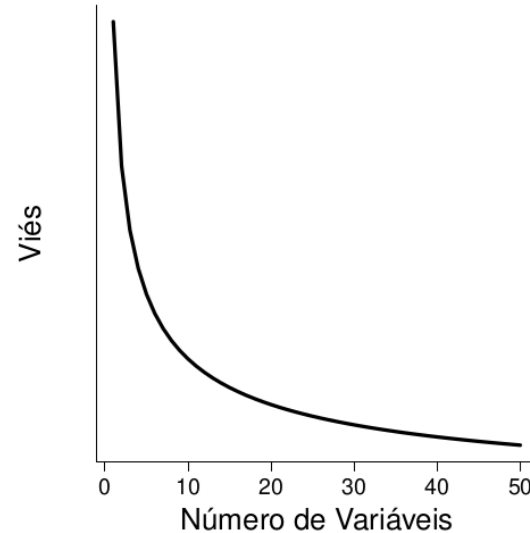
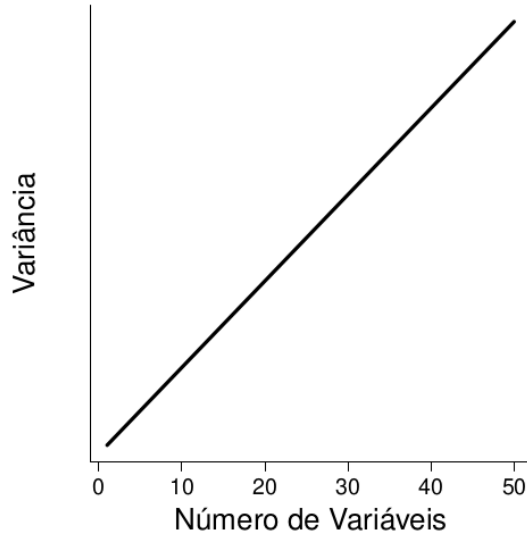
$$\text{Risco} = \text{Variabilidade Intrínseca} + \text{Variância} + \text{Viés}$$

- Redução na variância para melhorar o risco
- Usual que muitos features não sejam relevantes



Esparsidade

$$\text{Risco} = \text{Variabilidade Intrínseca} + \text{Variância} + \text{Viés}$$



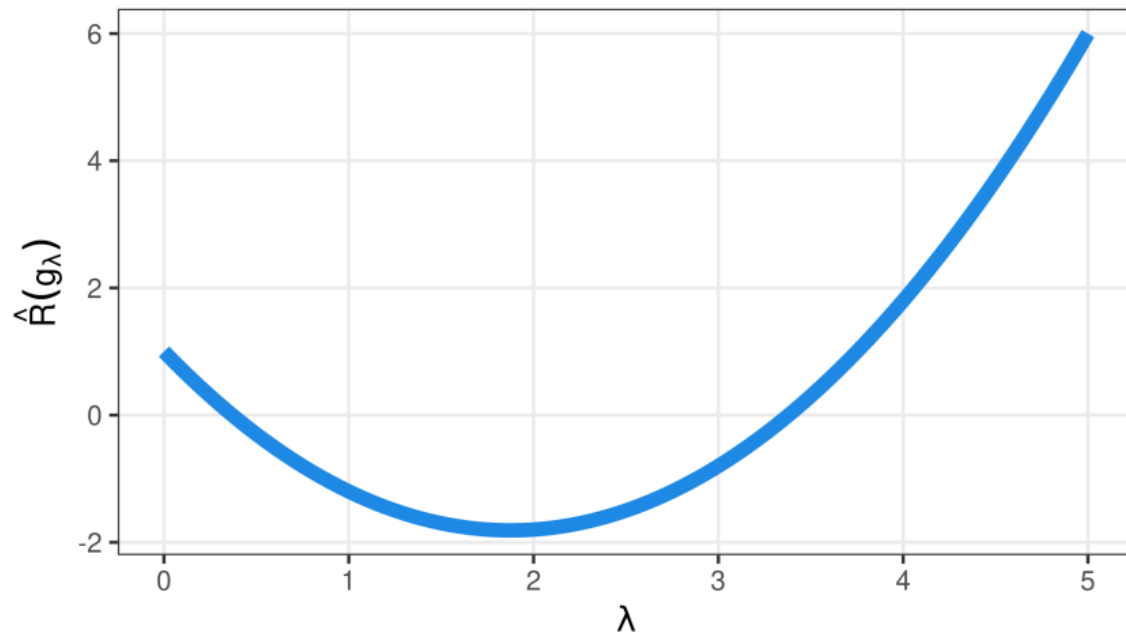
Lasso

$$\arg \min_{\beta} \text{EQM}(g_{\beta}) + \lambda \sum_{j=1}^d |\beta_j|$$

Shiny



Como escolher λ ?



Amazon



Outros Métodos (overview)

