UFRN
UNIVERSIDADE FEDERAL DO RIO GRANDE DO NORTE

CAPÍTULO 05

DESIGNING MACHINE LEARNING SYSTEMS

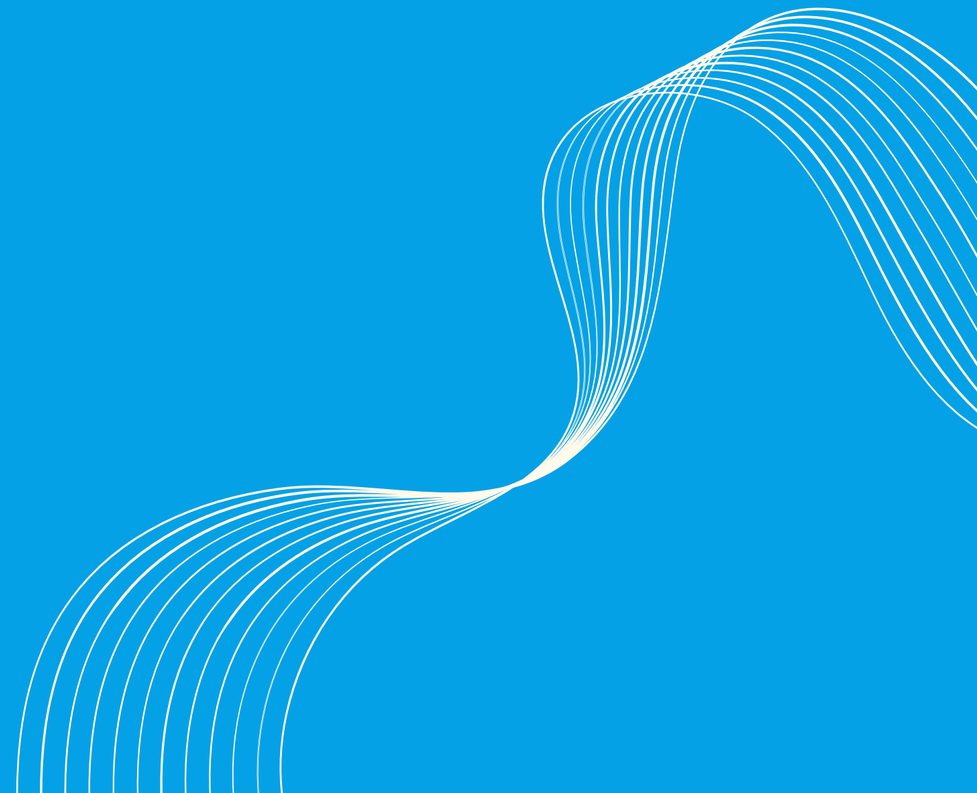
Features aprendidas vs projetadas

Operações comuns de engenharia de features

- Lidando com valores ausentes
  - Classificação
    - Missing not at random
    - Missing at random
    - Missing completely at random
  - Como lidar: deleção ou imputação

Operações comuns de engenharia de features

- Escalonamento
- Discretização
- Codificação de features categóricas
- Feature crossing
- Embeddings posicionais discretos e contínuos



Data leakage

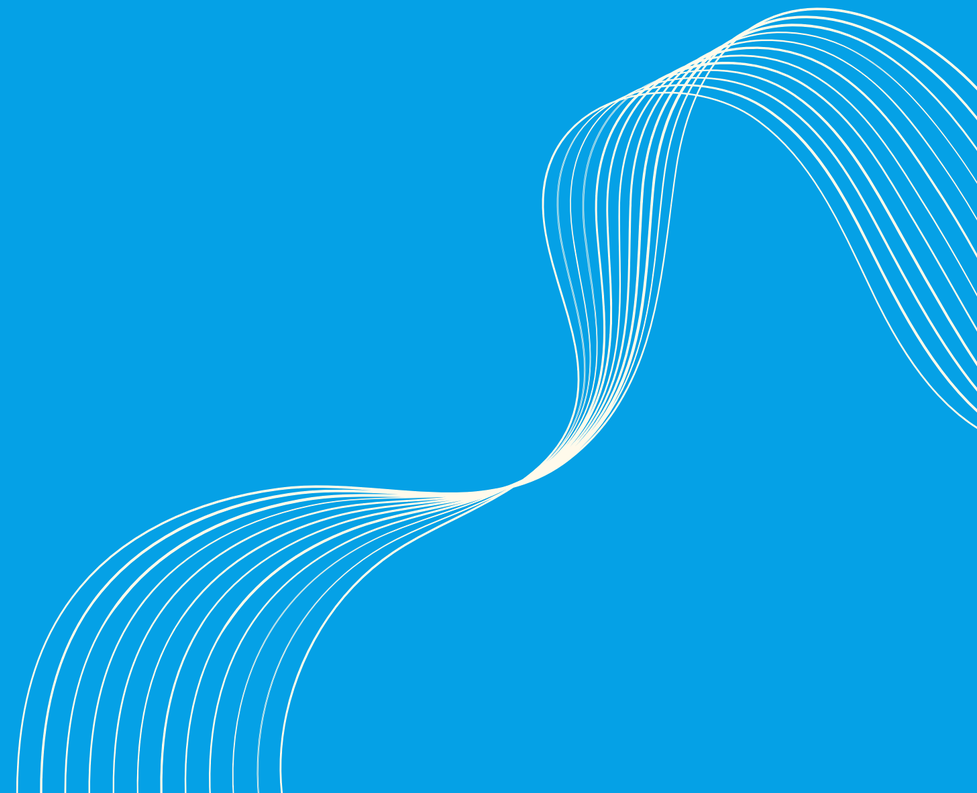
- Causas comuns
  - Divida os dados correlacionados ao tempo aleatoriamente em vez de por tempo
  - Escalone antes de dividir
  - Preencha os dados ausentes com estatísticas da divisão de teste
  - Manipulação inadequada da duplicação de dados antes da divisão
  - Grupo de leakage
  - Vazamento a partir do processo de geração de dados
- Detectando data leakage

Engenharia de boas features

- Muitas features podem ser ruins tanto durante o treinamento quanto na disponibilização do seu modelo, pelos seguintes motivos:
  - data leakage
  - sobreajuste
  - memória exigida
  - latência da inferência
  - dívidas técnicas.

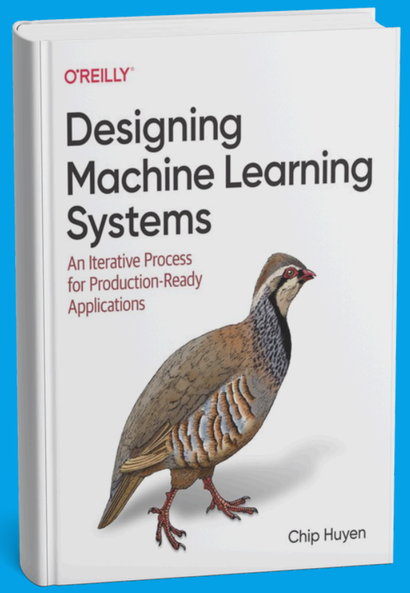
Engenharia de boas features

- Importância
- Generalização
  - Cobertura
  - Distribuições



Bibliografia

Designing Machine Learning Systems - Chip Huyen (O'Reilly, 2022)




That's all Folks!