

# Machine learning para predições em saúde

Prof. Dr. Alexandre Chiavegatto Filho





**Repositório de Arquivos**  
**Machine Learning para Predições em Saúde**

- Desenvolver algoritmos que façam boas previsões em saúde.
- Principais razões técnicas pelas quais algoritmos às vezes não apresentam boa performance preditiva:
  - Extrapolação inadequada dos resultados.
  - Pré-processamento inadequado dos dados.
  - Sobreajuste (mais importante).
  - Validação inadequada da qualidade dos algoritmos.
- Se fizer a parte técnica correta, o motivo para a baixa performance preditiva é que não foram usadas variáveis preditoras fortes.



## Extrapolação inadequada

- Desenvolver os algoritmos para uma população e esperar que funcionam corretamente para outra diferente.
  - Importar algoritmos dos EUA/Europa: nossas características genéticas e socioeconômicas são muito diferentes.
  - Extrapolação para períodos diferentes (cuidado com doenças sazonais).



# Extrapolação inadequada

- Editorial da Lancet, 12 agosto 2023.
- A maioria dos dados de saúde vem de países de alta renda, o que pode influenciar os modelos, exacerbando a injustiça histórica e a discriminação quando usados em outro lugar.
- Sem investimento em infraestrutura e pesquisa local, os países de baixa e média renda continuarão dependendo da IA desenvolvida nos EUA e na Europa, e os custos podem ser proibitivos.

## AI in medicine: creating a safe and equitable future

The meteoric progress of generative artificial intelligence (AI)—such as Open AI’s ChatGPT, capable of holding realistic conversations, or others of creating realistic images and video from simple prompts—has renewed interest in the transformative potential of AI, including for health. It has also sparked sobering warnings. Addressing the UN Security Council in July, Secretary General António Guterres spoke of the “horrific levels of death and destruction” that malicious AI use could cause. How can the medical community navigate AI’s substantial challenges to realise its health potential?

AI in medicine is nothing new. Non-generative machine learning can already perform impressively at discrete tasks, such as interpreting medical images. *The Lancet Oncology* recently published one of the first randomised controlled trials of AI-supported mammography, demonstrating a similar cancer detection rate and nearly halved screen-reading workload compared with unassisted reading. AI has driven progress in infectious diseases and molecular medicine and has enhanced field-deployable diagnostic tools. But the medical applications of generative AI remain largely speculative. Automation of evidence synthesis and identification of de novo drug candidates could expedite clinical research. AI-enabled generation of medical notes could ease the administrative burden for health-care workers, freeing up time to see patients. Initiatives such as the Bill & Melinda Gates Foundation’s Global Grand Challenges seek innovative uses of large language models in low-income and middle-income countries (LMICs).

These advances come with serious risks. AI performs best at well defined tasks and when models can easily augment rather than replace human judgement. Applying generative AI to heterogeneous data is complicated. The black box nature of many models makes it challenging to appraise their suitability and generalisability. Large language models can make mistakes easily missed by humans or hallucinate non-existent sources. Transfer of personal data to technology firms without adequate regulation could compromise patient privacy. Health equity is a particularly serious concern. Algorithms trained on health-care datasets that reflect bias in health-care spending, for example, worsened racial disparities in access to care in the USA. Most health data come from high-income countries, which could bias models

exacerbating historical injustice and discrimination when used elsewhere. These issues all risk eroding patient trust.

How then to ensure that AI is a force for good in medicine? The scientific community has a key role in rigorous testing, validation, and monitoring of AI. The UN is assembling a high-level advisory body to build global capacity for trustworthy, safe, and sustainable AI; it is crucial that health and medicine are well represented. An equitable approach will require a diversity of local knowledge. WHO has partnered with the International Digital Health and AI Research Collaborative to boost participation from LMICs in the governance of safe and ethical AI in health through cross-border collaboration and common guidance. But without investment in local infrastructure and research, LMICs will remain reliant on AI developed in the USA and Europe, and costs could be prohibitive without open access alternatives. At present, the pace of technological progress far outstrips the guidance, and the power imbalance between the medical community and technology firms is growing.

Allowing private entities undue influence is dangerous. The UN Secretary General has urged the Security Council to help ensure transparency, accountability, and oversight on AI. Regulators must act to ensure safety, privacy, and ethical practice. The EU’s AI Act, for example, will require high risk AI systems to be assessed before approval and subjected to monitoring. Regulation should be a key concern of the first major global summit on AI safety, being held in the UK later this year. Although technology companies should be part of the regulatory conversation, there are already signs of resistance. Amazon, Google, and Epic have objected to proposed US rules to regulate AI in health technologies. The tension between commercial interests and transparency risks compromising patient wellbeing, and marginalised groups will suffer first.

There is still time for us to create the future we want. AI could continue to bring benefits if integrated cautiously. It could change practice for the better as an aid—not a replacement—for doctors. But doctors cannot ignore AI. Medical educators must prepare health-care workers for a digitally augmented future. Policy makers must work with technology firms, health experts, and governments to ensure that equity remains a priority. Above all, the medical community must amplify the urgent call for stringent regulation. ■ *The Lancet*



See [World Report](#) page 517

For the AI-assisted mammography trial see [Articles Lancet Oncology 2023; 24: 936–44](#)

For more on AI in infectious diseases see [Science 2023; 381: 164–70](#)

For more on AI in molecular medicine see [N Engl J Med 2023; 388: 2456–65](#)

For more on generative AI in medicine see [Comment Lancet Digit Health 2023; 5: e107–8](#)

For more on the Global Grand Challenges see <https://gch.grandchallenges.org/challenge/catalyzing-equitable-artificial-intelligence-ai-use>

For more on the dangers of biased health data see [Science 2019; 366: 447–53](#)

For more on WHO’s efforts to improve access to AI see <https://www.who.int/news/item/06-07-2022-who-and-i-dair-to-partner-for-inclusive-impactful-and-responsible-international-research-in-artificial-intelligence-and-digital-health>

For more on the UN Secretary General’s remarks see <https://press.un.org/en/2023/sgsm21880.doc.htm>

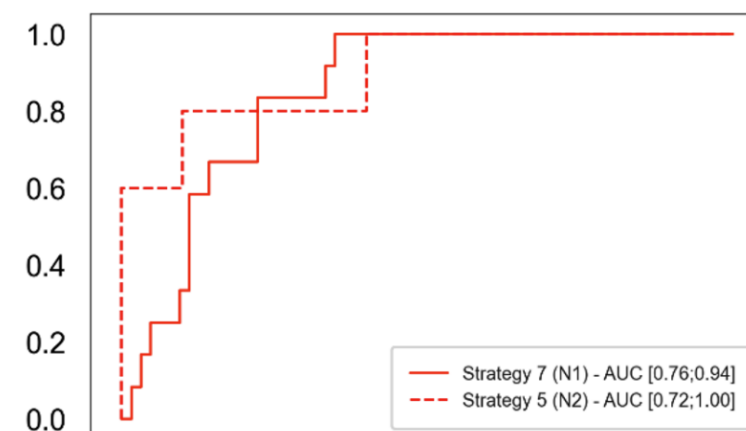
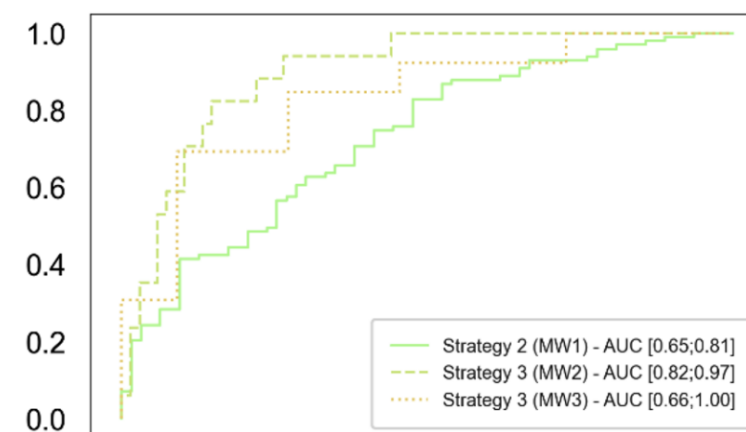
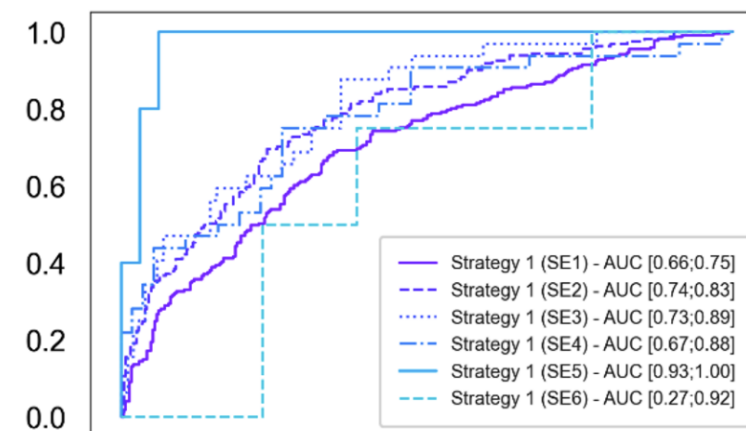
For more on the AI Act see <https://artificialintelligenceact.eu>

For more on the global summit on AI see <https://www.gov.uk/government/news/uk-to-host-first-global-summit-on-artificial-intelligence>

# Improving the performance of machine learning algorithms for health outcomes predictions in multicentric cohorts

Roberta Moreira Wichmann, Fernando Timoteo Fernandes, Alexandre Dias Porto Chiavegatto Filho

Scientific Reports 2023; 13, 1022.



- Coorte multicêntrica de pacientes com RT-PCR positivo para covid-19 (n = 8.477) em 18 hospitais das cinco regiões brasileiras.
- Oito estratégias diferentes foram usadas para treinar e avaliar o desempenho de algoritmos de machine learning para prever óbito.
- Os melhores desempenhos preditivos foram obtidos ao usar dados de treinamento do mesmo hospital, que foi a estratégia vencedora para 11 (61%) dos 18 hospitais participantes.
- O uso de mais dados de pacientes de outras regiões diminuiu o desempenho preditivo.



# PRÉ-PROCESSAMENTO DOS DADOS

- Técnicas de pré-processamento de dados
  - Seleção das variáveis.
  - Vazamento de dados.
  - Padronização.
  - Redução de dimensão.
  - Colinearidade.
  - Valores missing.
  - One-hot encoding.

PRÉ-PROCESSAMENTO DOS  
DADOS

PADRONIZAÇÃO

REDUÇÃO DE DIMENSÃO

VARIÁVEIS COLINEARES

VARIÁVEIS MISSING

ONE-HOT ENCODING



# PRÉ-PROCESSAMENTO DOS DADOS

PRÉ-PROCESSAMENTO DOS  
DADOS

PADRONIZAÇÃO

REDUÇÃO DE DIMENSÃO

VARIÁVEIS COLINEARES

VARIÁVEIS MISSING

ONE-HOT ENCODING

Preditores plausíveis:

- Pré-selecionar variáveis que sejam preditoras plausíveis (bom senso do pesquisador).
- Coincidências acontecem em análises de big data e pode ser que o algoritmo dê muita importância para associações espúrias.



# PRÉ-PROCESSAMENTO DOS DADOS

PRÉ-PROCESSAMENTO DOS  
DADOS

PADRONIZAÇÃO

REDUÇÃO DE DIMENSÃO

VARIÁVEIS COLINEARES

VARIÁVEIS MISSING

ONE-HOT ENCODING

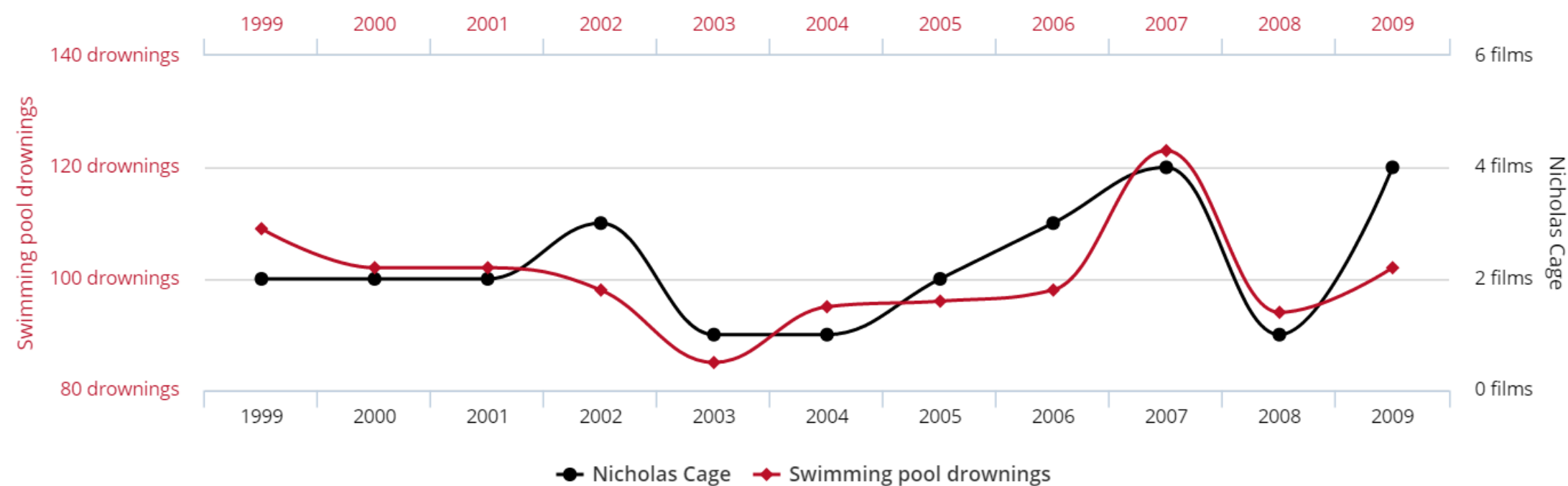
tylervigen.com

Number of people who drowned by falling into a pool

correlates with

Filmas Nicolas Cage appeared in

Correlation: 66,6% ( $r=0,666004$ )



# PRÉ-PROCESSAMENTO DOS DADOS

Cuidado com vazamento de informação (“data leakage”).

- Acontece quando os dados de treino apresentam informação escondida que faz com que o modelo aprenda padrões que não são do seu interesse.
- Uma variável preditora tem escondida o resultado certo:
  - Não é a variável que está predizendo o desfecho, mas o desfecho que está predizendo ela.

PRÉ-PROCESSAMENTO DOS  
DADOS

PADRONIZAÇÃO

REDUÇÃO DE DIMENSÃO

VARIÁVEIS COLINEARES

VARIÁVEIS MISSING

ONE-HOT ENCODING

JOURNAL OF MEDICAL INTERNET RESEARCH

Chiavegatto Filho et al

Letter to the Editor

Data Leakage in Health Outcomes Prediction With Machine Learning. Comment on “Prediction of Incident Hypertension Within the Next Year: Prospective Study Using Statewide Electronic Health Records and Machine Learning”

Alexandre Chiavegatto Filho, PhD; André Filipe De Moraes Batista, MSc, PhD; Hellen Geremias dos Santos, MPH, PhD

Department of Epidemiology, School of Public Health, University of São Paulo, São Paulo, Brazil



# PRÉ-PROCESSAMENTO DOS DADOS

## Exemplo

Incluir o número  
identificador do  
paciente como variável  
preditora

PRÉ-PROCESSAMENTO DOS  
DADOS

PADRONIZAÇÃO

REDUÇÃO DE DIMENSÃO

VARIÁVEIS COLINEARES

VARIÁVEIS MISSING

ONE-HOT ENCODING



# PRÉ-PROCESSAMENTO DOS DADOS

PRÉ-PROCESSAMENTO DOS  
DADOS

PADRONIZAÇÃO

REDUÇÃO DE DIMENSÃO

VARIÁVEIS COLINEARES

VARIÁVEIS MISSING

ONE-HOT ENCODING

## Exemplo

Incluir o número  
identificador do  
paciente como variável  
preditora

## Problema

Se pacientes de hospital  
especializado em câncer  
tiverem números  
semelhantes.  
Se o objetivo for prever  
câncer, algoritmo irá dar  
maior probabilidade a esses  
pacientes.  
Esse algoritmo aprendeu algo  
interessante para o sistema  
de saúde?



# PRÉ-PROCESSAMENTO DOS DADOS

PRÉ-PROCESSAMENTO DOS DADOS

PADRONIZAÇÃO

REDUÇÃO DE DIMENSÃO

VARIÁVEIS COLINEARES

VARIÁVEIS MISSING

ONE-HOT ENCODING

Exemplo

Incluir o número  
identificador do  
paciente como variável  
preditora

Problema

Se pacientes de hospital  
especializado em câncer  
tiverem números  
semelhantes.  
  
Se o objetivo for prever  
câncer, algoritmo irá dar  
maior probabilidade a esses  
pacientes.  
  
Esse algoritmo aprendeu algo  
interessante para o sistema  
de saúde?

Motivo

Motivo pelo qual os  
dados e os algoritmos de  
machine learning  
precisam ser abertos.  
  
Sempre analisar  
importância preditora  
das variáveis (Shapley).





# PADRONIZAÇÃO

- A escala das variáveis pode afetar muito a qualidade das predições.
- Alguns algoritmos dão preferência para utilizar variáveis com valores muito alto.

PRÉ-PROCESSAMENTO DOS  
DADOS

► PADRONIZAÇÃO

REDUÇÃO DE DIMENSÃO

VARIÁVEIS COLINEARES

VARIÁVEIS MISSING

ONE-HOT ENCODING

# PADRONIZAÇÃO

PRÉ-PROCESSAMENTO DOS  
DADOS

► PADRONIZAÇÃO

REDUÇÃO DE DIMENSÃO

VARIÁVEIS COLINEARES

VARIÁVEIS MISSING

ONE-HOT ENCODING

- Padronizar as variáveis contínuas para todas terem média de 0 e desvio-padrão de 1.

$$z_i = \frac{x_i - \mu}{\sigma}$$

- Ou seja, é feita a subtração da média e a divisão pelo desvio padrão dos valores da variável.

# REDUÇÃO DE DIMENSÃO

- Quanto maior a dimensão dos dados (número de variáveis) maior o risco de o algoritmo encontrar e utilizar associações espúrias.

PRÉ-PROCESSAMENTO DOS  
DADOS

PADRONIZAÇÃO

► REDUÇÃO DE DIMENSÃO

VARIÁVEIS COLINEARES

VARIÁVEIS MISSING

ONE-HOT ENCODING

# REDUÇÃO DE DIMENSÃO

PRÉ-PROCESSAMENTO DOS  
DADOS

PADRONIZAÇÃO

► REDUÇÃO DE DIMENSÃO

VARIÁVEIS COLINEARES

VARIÁVEIS MISSING

ONE-HOT ENCODING

## - Análise de Componentes Principais

Técnica de aprendizado  
não supervisionado.

O objetivo é encontrar  
combinações lineares das  
variáveis preditoras que  
incluam a maior quantidade  
possível da variância original.

O primeiro componente  
principal irá preservar a maior  
combinação linear possível dos  
dados, o segundo a maior  
combinação linear possível não  
correlacionada com o primeiro  
componente, etc.

# VARIÁVEIS COLINEARES

Uma das razões pela qual a ACP é tão utilizada, é o fato de que cria componentes principais não correlacionados.

- Na prática, alguns algoritmos conseguem melhor performance preditiva com variáveis com baixa correlação.

PRÉ-PROCESSAMENTO DOS  
DADOS

PADRONIZAÇÃO

REDUÇÃO DE DIMENSÃO

► VARIÁVEIS COLINEARES

VARIÁVEIS MISSING

ONE-HOT ENCODING



# VARIÁVEIS COLINEARES

Uma outra forma de diminuir a presença de variáveis com alta correlação é excluí-las.

- Variáveis colineares trazem informação redundante (tempo perdido).
- Além disso, aumentam a instabilidade dos modelos.
- Estabelecer um limite de correlação com alguma outra variável (0,75 a 0,90).

PRÉ-PROCESSAMENTO DOS  
DADOS

PADRONIZAÇÃO

REDUÇÃO DE DIMENSÃO

► VARIÁVEIS COLINEARES

VARIÁVEIS MISSING

ONE-HOT ENCODING

# VARIÁVEIS MISSING

*É importante entender por que valores de uma variável estão faltantes.*

PRÉ-PROCESSAMENTO DOS  
DADOS

PADRONIZAÇÃO

REDUÇÃO DE DIMENSÃO

VARIÁVEIS COLINEARES

► VARIÁVEIS MISSING

ONE-HOT ENCODING

Motivo sistemático → INFORMAÇÃO PREDITIVA.

Grande diferença em relação a estudos de inferência, em que valores missing devem ser evitados.

Não conseguiu responder a uma pergunta sobre o seu passado → INFORMAÇÃO PREDITIVA.

Pode ajudar na predição de problemas cognitivos graves no futuro

Em variáveis categóricas adicionar uma categoria para missing.

Imputação com machine learning para valores contínuos (adicionar nova variável indicativa de missing).

# ONE-HOT ENCODING

Alguns algoritmos têm dificuldade em entender variáveis que têm mais do que uma categoria.

Acham que é uma variável contínua (0, 1, 2, 3...) → porém não têm significado contínuo.

A solução é transformar todas as categorias em uma variável diferente de valores 0 e 1 (one-hot encoding).

Variável com n categorias → criadas n variáveis.

PRÉ-PROCESSAMENTO DOS DADOS

PADRONIZAÇÃO

REDUÇÃO DE DIMENSÃO

VARIÁVEIS COLINEARES

VARIÁVEIS MISSING

► ONE-HOT ENCODING

# ONE-HOT ENCODING

Pode trazer problemas em alguns modelos, como na regressão linear.

Solução: criar dummies.  
n-1 variáveis (deixar a mais frequente como categoria de referência).

PRÉ-PROCESSAMENTO DOS  
DADOS

PADRONIZAÇÃO

REDUÇÃO DE DIMENSÃO

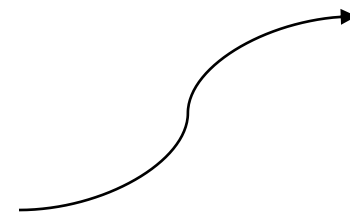
VARIÁVEIS COLINEARES

VARIÁVEIS MISSING

► ONE-HOT ENCODING

# ONE-HOT ENCODING

Cada vez mais popular: **target encoding**



Transformar categóricas em contínuas

O valor da categoria vira o seu valor médio do desfecho entre os indivíduos dessa categoria

PRÉ-PROCESSAMENTO DOS DADOS

PADRONIZAÇÃO

REDUÇÃO DE DIMENSÃO

VARIÁVEIS COLINEARES

VARIÁVEIS MISSING

► ONE-HOT ENCODING



**Artigo:**

# **Model Evaluation, Model Selection, and Algorithm Selection in Machine Learning**

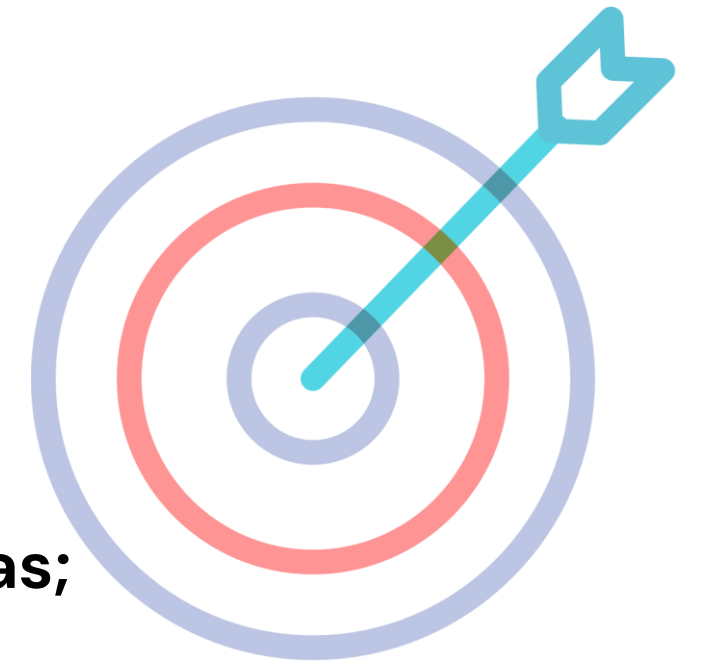
Autor: Sebastian Raschka\*

\*Department of Statistics, University of Wisconsin-Madison

# Objetivo



- Revisar diferentes técnicas para cada uma dessas sub tarefas;
- Vantagens e desvantagens de cada técnica;
- Recomendações para melhores práticas em ML;



---

Discussão



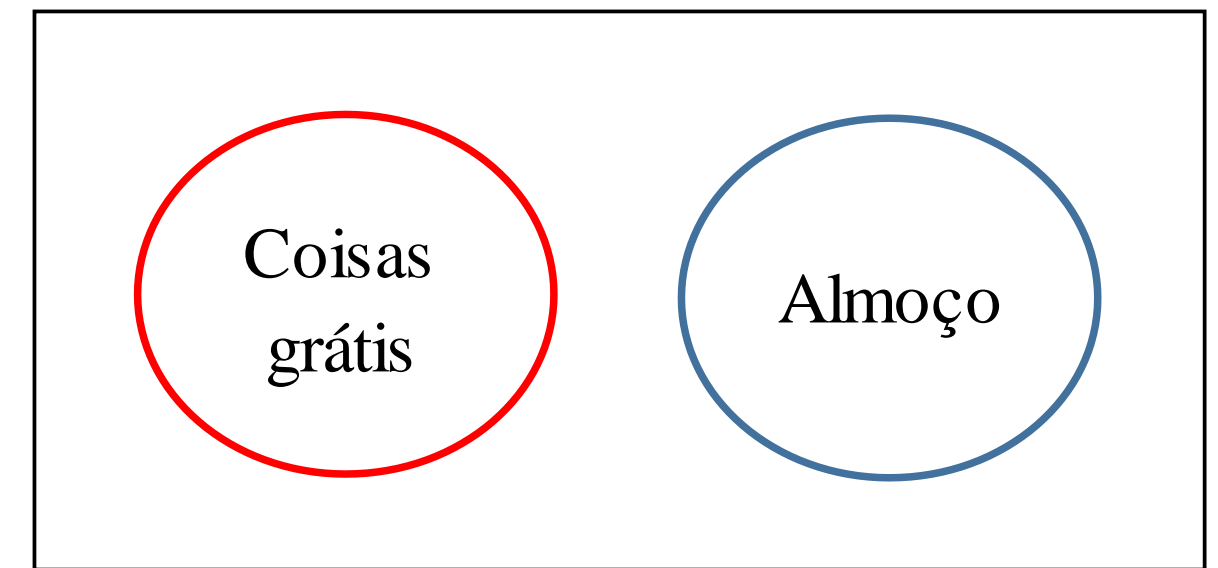
1. Introdução: Termos e técnicas essenciais de avaliação do modelo
2. Bootstrapping e Incerteza
3. Validação cruzada e otimização de hiperparâmetros

Teorema do “não há almoço grátis”:

- Em infinitos conjuntos diferentes de dados, nenhum algoritmo é garantido a priori de ter melhor performance.
- A única forma de saber qual vai ter melhor performance é testar todos.

*Segredo*

- *com dados reais, alguns costumam ganhar mais vezes (gradiente boosting e random forests).*



# 1

## Introdução: Termos e técnicas essenciais de avaliação do modelo



Estimação de  
performance:  
Desempenho de  
generalização  
vs  
Seleção de  
modelo



Premissas e  
Terminologia



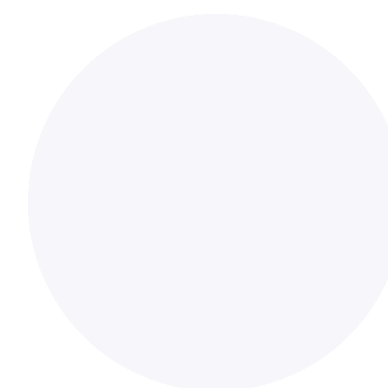
Validação de  
ressubstituição  
e Método  
Holdout



Estratificação



Validação  
Holdout



Viés  
pessimista



Intervalos de  
confiança via  
aproximação  
normal

# 1 Introdução: Termos e técnicas essenciais de avaliação do modelo



Estimação de  
performance:  
Desempenho de  
generalização  
vs  
Seleção de  
modelo

- Estimar o desempenho não é trivial
- Configurações diferentes do mesmo algoritmo resultam em performances diferentes
- Necessidade de comparação do desempenho
- **Motivos para avaliar o desempenho de algoritmos:**
  - Estimar a capacidade de generalização (futuro)
  - Selecionar o modelo com melhor desempenho
  - Comparar e selecionar o melhor algoritmo de ML



# How Machine Learning Algorithms Work

Supervised learning is the computerized task of inferring a function from labeled training data.

$$Y = f(x),$$

# 1 Introdução: Termos e técnicas essenciais de avaliação do modelo



Hipótese e  
Terminologia

- **Função alvo (target):** verdadeira função  $f(x)=y$ , a qual queremos aprender ou aproximar;
- **Modelo:** função que acreditamos (esperamos) que seja semelhante a verdadeira;
- **Algoritmo de aprendizado:** conjunto de instruções para modelar uma função alvo usando dados de treinamento.

# 1 Introdução: Termos e técnicas essenciais de avaliação do modelo



Hipótese e  
Terminologia

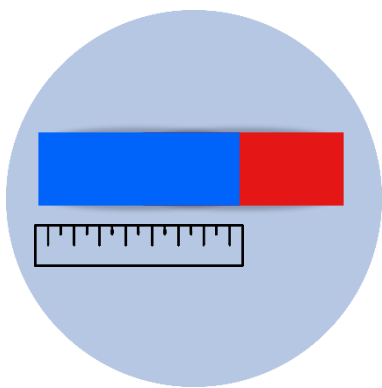
- **Hiperparâmetros:** são os parâmetros (meta-parâmetros) de *ajuste* de um algoritmo de ML.

Ex.: o valor que define a profundidade máxima em uma árvore de decisão.

- **Parâmetros do modelo:** parâmetros que o algoritmo ajusta aos dados de treinamento.

Ex.: inclinação em uma regressão linear

# 1 Introdução: Termos e técnicas essenciais de avaliação do modelo



Validação de  
re-substituição e  
Método Holdout

– Holdout: técnica de avaliação mais simples

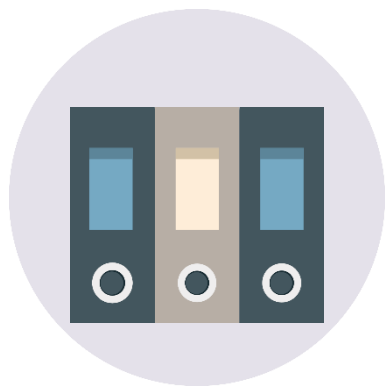
1º divide-se o conjunto de dados rotulados em duas partes: treino e teste (subamostragem aleatória) para evitar sobreajuste

2º ajusta-se o modelo aos dados de treinamento para efetuar a predição no teste

3º estimativa da performance é dada pelos acertos

O teste deve ocorrer em um conjunto de novos dados  
(não "vistos" pelo modelo)

# 1 Introdução: Termos e técnicas essenciais de avaliação do modelo



Estratificação

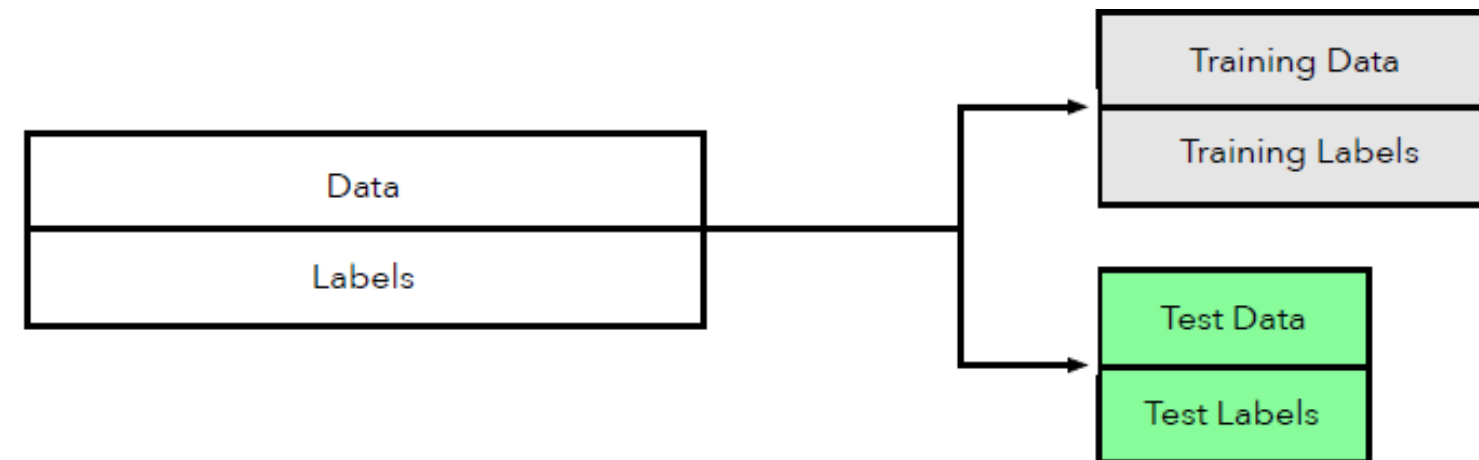
- Banco de dados → amostra aleatória (distr. probabilidade)
- Assume-se que é representativa da população real a ser predita
- Subamostragem sem reposição modifica estatísticas da amostra
- Corrigido com o uso da divisão aleatória estratificada

# 1 Introdução: Termos e técnicas essenciais de avaliação do modelo



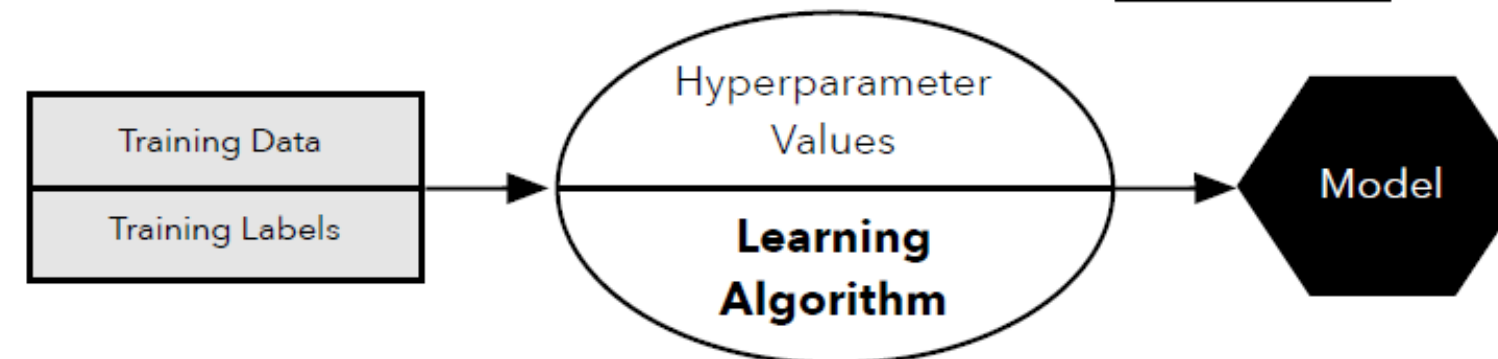
Validação  
Holdout

1



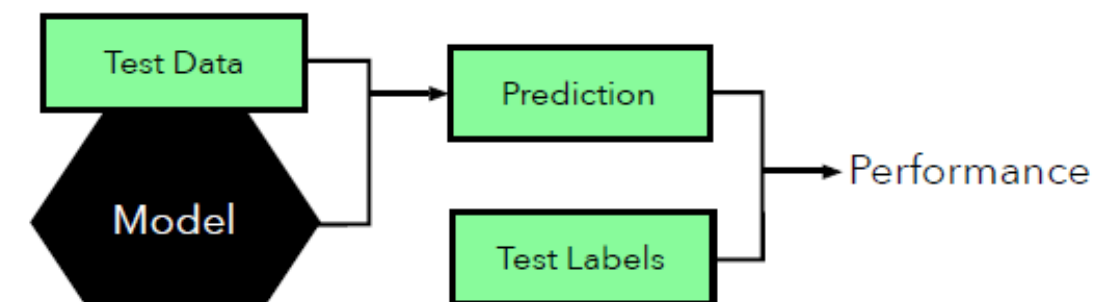
1º Divide-se os dados  
Teste = novos dados (mundo real)  
O teste deve ser usado **apenas uma vez!**

2



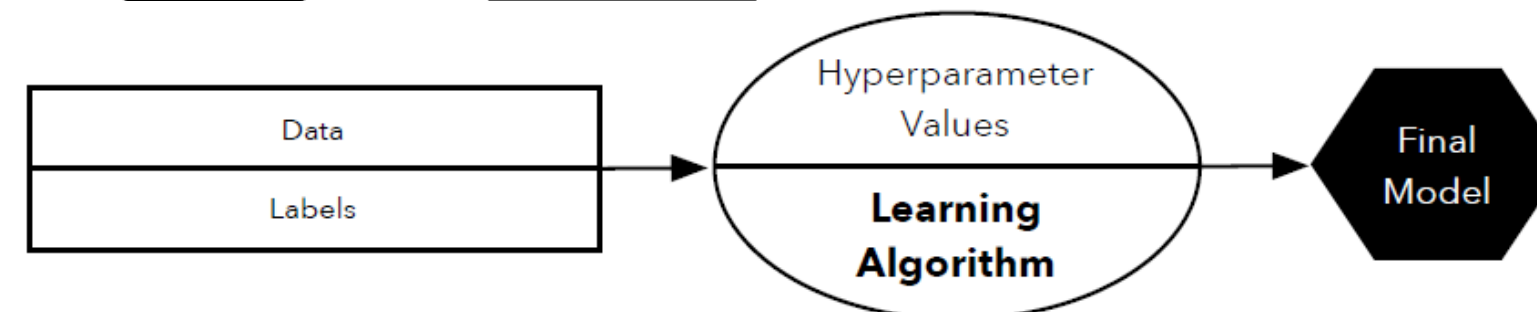
2º Escolhemos os algoritmos  
Procedimento para otimizar os  
hiperparâmetros ou utilizar os  
padrões.

3



3º Quão bom é o modelo?  
Desempenho é medido pela  
predição nos dados de teste e  
comparação com realidade.

4



4º Resultado da performance no  
dados não vistos anteriormente.  
Assume-se mesma performance  
do que dados futuros.



# 1 Introdução: Termos e técnicas essenciais de avaliação do modelo



Viés  
pessimista

- Se o modelo não atingiu a sua plena capacidade então a estimativa do desempenho obtém um viés pessimista;
- O algoritmo poderia aprender um modelo melhor se tivesse mais dados;
- Solução: ajustar o modelo a todo o banco de dados após estimar o desempenho de generalização;
- Problema: não se pode estimar o desempenho de generalização do modelo reajustado.

## 2

# Bootstrapping e Incerteza



Visão geral

**– Por que queremos avaliar o modelo?**

**1º Queremos estimar o desempenho preditivo em dados futuros**

**2º Queremos aumentar o desempenho ajustando o algoritmo e selecionando o melhor modelo**

**3º Queremos identificar o algoritmo mais adequado para o problema em questão, ou seja, queremos comparar diferentes modelos e algoritmos**



## 2

# Bootstrapping e Incerteza



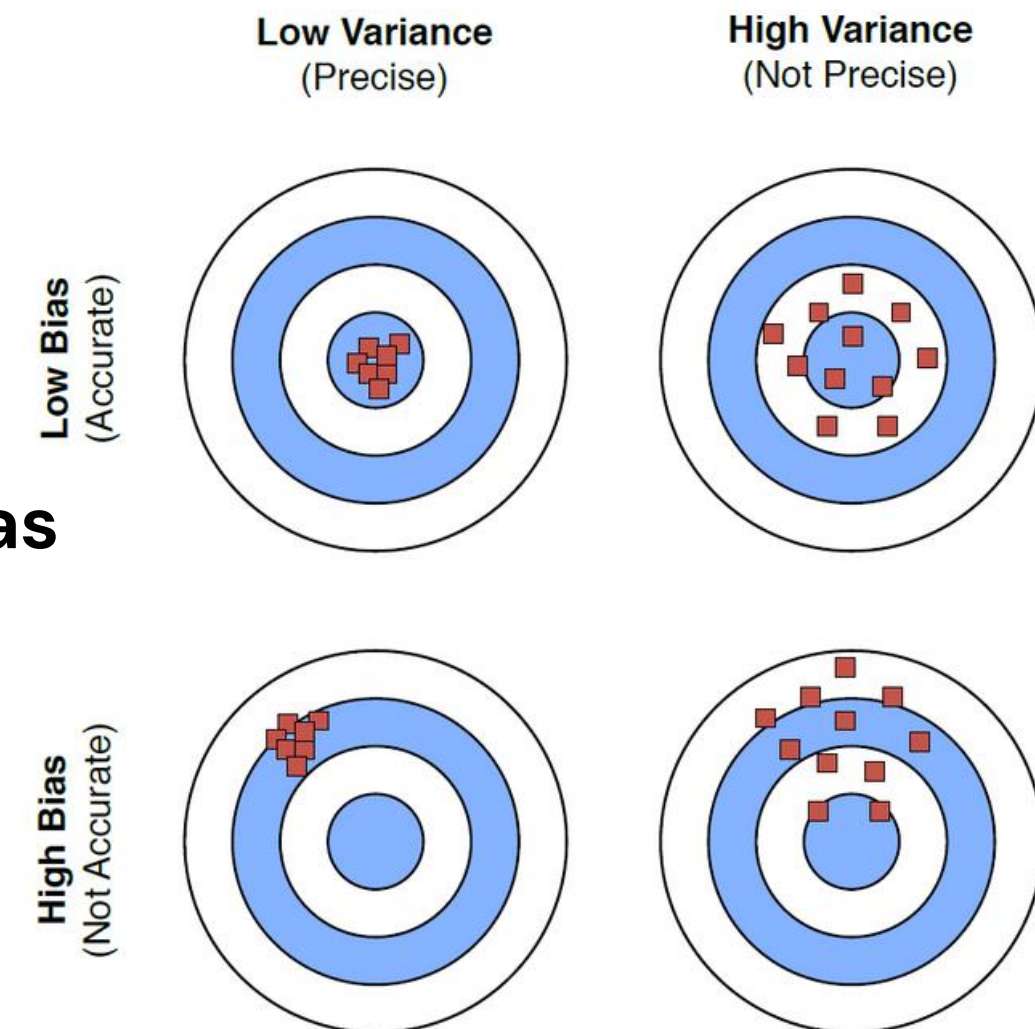
Reamostragem

– Relação viés x variância



– Variância é a mudança do modelo devido a pequenas mudanças nos dados.

– Viés é o erro consistente do modelo.

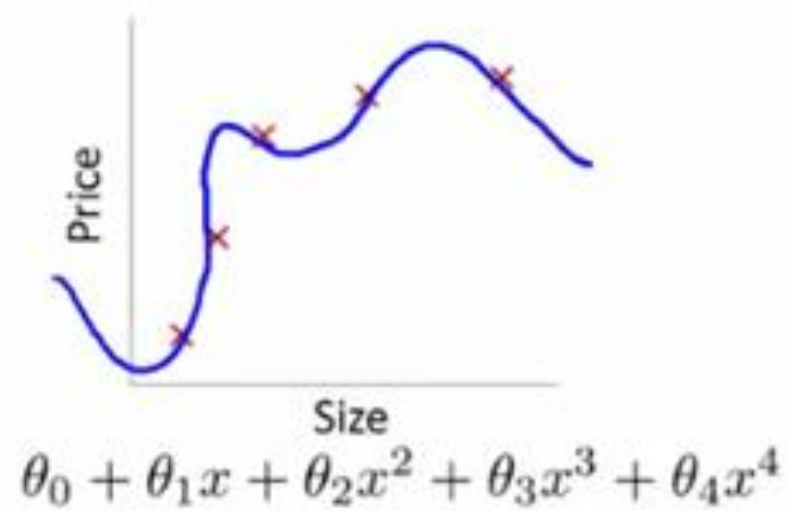


## 2

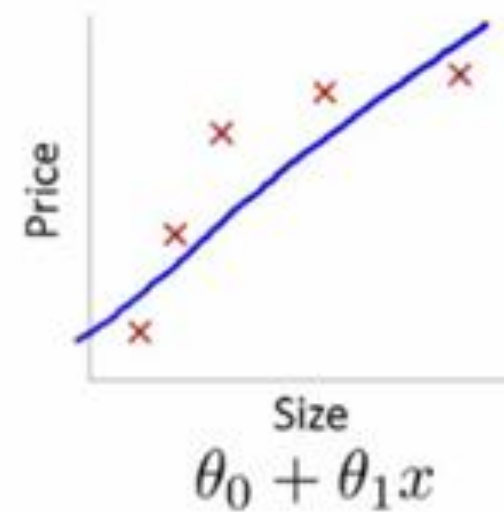
## Bootstrapping e Incerteza



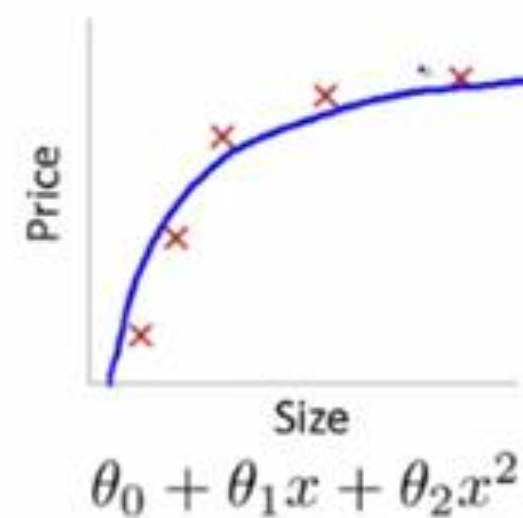
Reamostragem



High variance  
(overfit)



High bias  
(underfit)



"Just right"

## 2

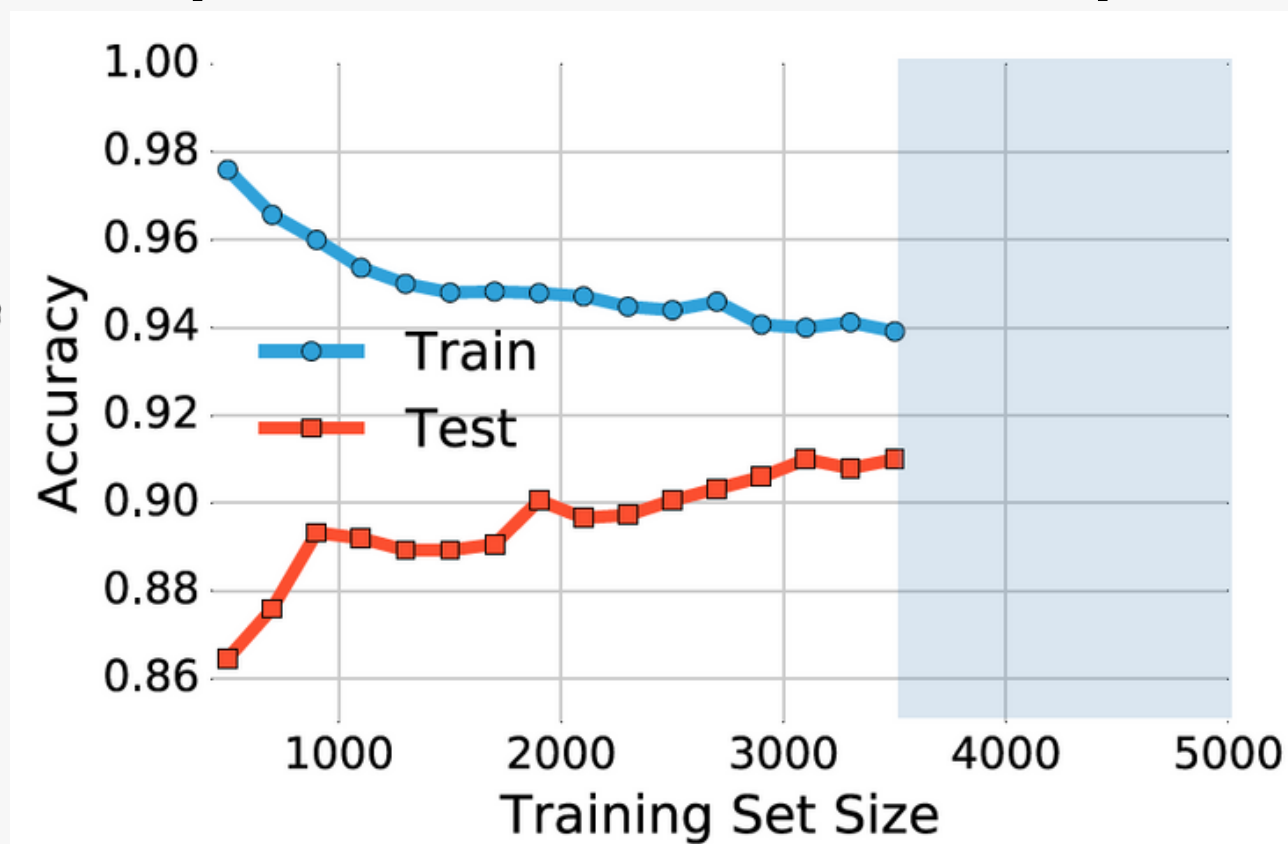
## Bootstrapping e Incerteza



Reamostragem

- Pequenos conjuntos de dados de teste aumentam o viés pessimista, pois seria possível formular uma hipótese mais poderosa e generalizável.

Ex.: MNIST5  
Classificadores  
Softmax



### 3

## Validação cruzada e otimização de hiperparâmetros

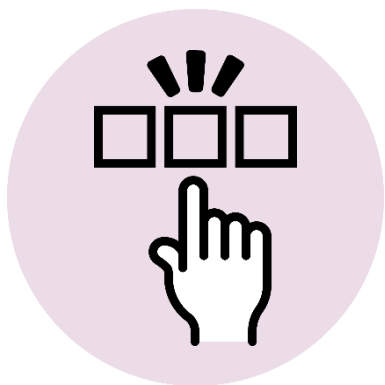


Visão geral

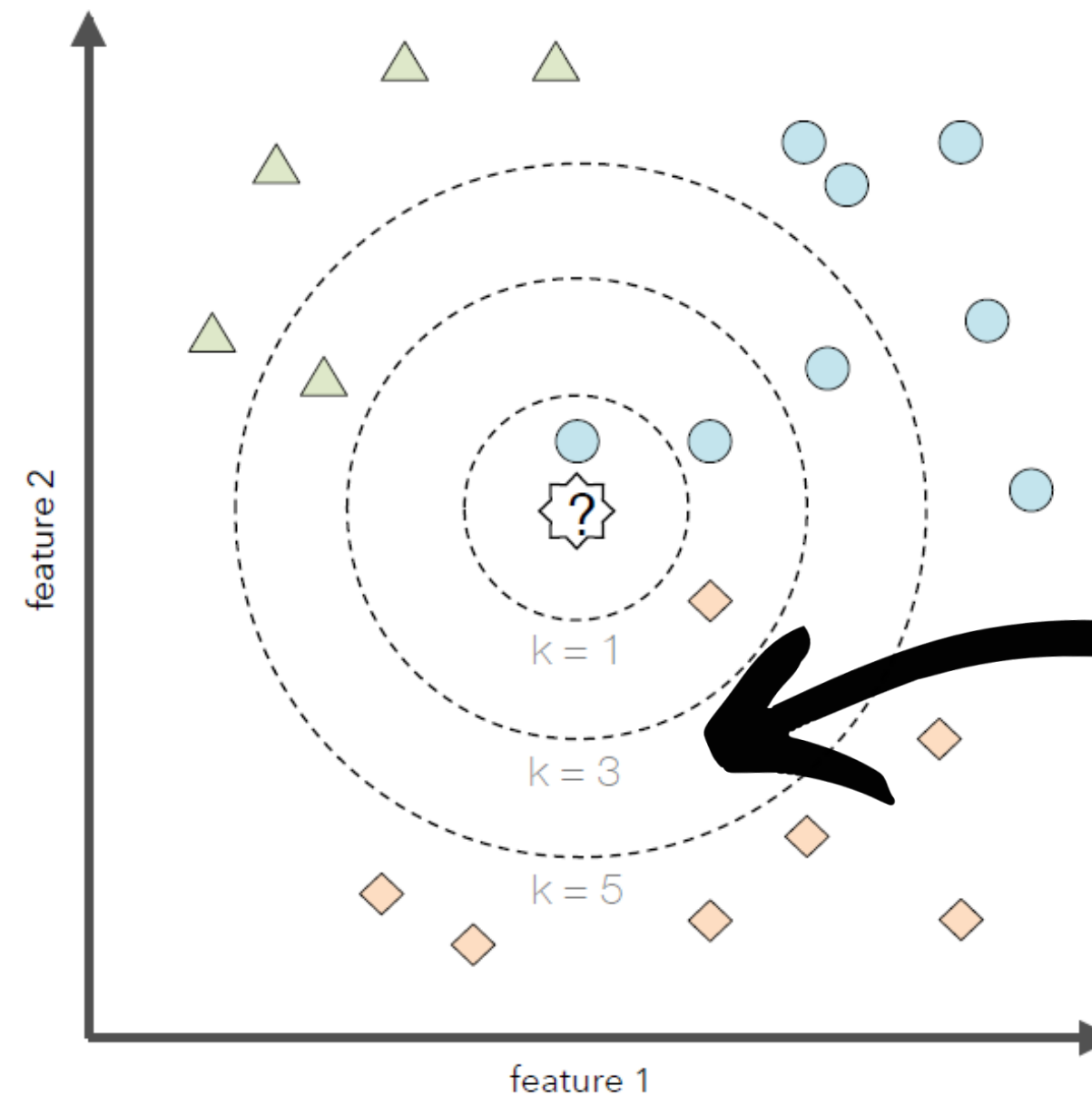
- Quase todos os algoritmos tem configurações a serem especificadas;
- Os "botões de ajuste" são os hiperparâmetros (em geral regularizadores – tentativas de segurar a complexidade);
- Eles ajudam a otimizar o desempenho encontrando o equilíbrio entre o viés e a variância;

### 3

## Validação cruzada e otimização de hiperparâmetros



Hiperparâmetros  
e seleção de  
modelo



Hiperparâmetro  
do algoritmo  
K-nearest neighbors

### 3

## Validação cruzada e otimização de hiperparâmetros

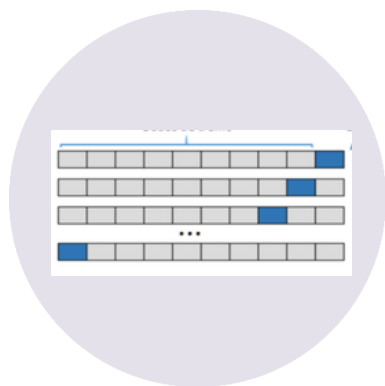


As 3 vias do Método Holdout para ajuste de hiperparâmetros

- Como selecionar hiperparâmetros se o teste só pode ser olhado uma única vez?
- Três vias: Treinamento/ Validação / Teste;
- Treinamento e validação para ajuste dos hiperparâmetros e seleção dos modelos;
- Entretanto, quanto menor o conjunto de dados, maior será o viés pessimista do desempenho e a variância.

### 3

## Validação cruzada e otimização de hiperparâmetros



Introdução à  
validação cruzada  
k-fold

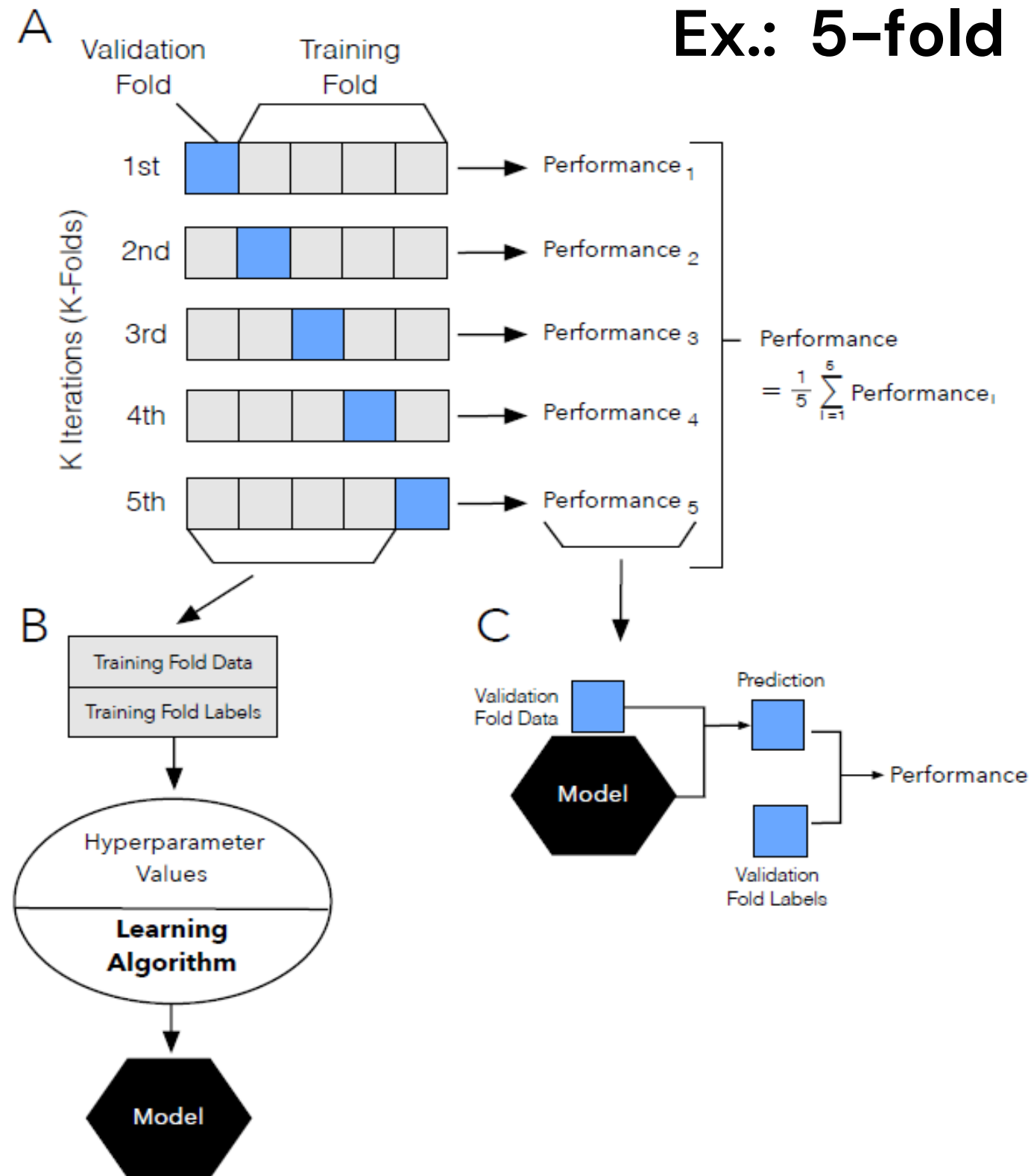
- **Validação cruzada:** cruzamento de estágios de treinamento e validação em rodadas sucessivas;
- **Cada amostra do conjunto de dados tem a oportunidade de ser testada;**
- **K-fold:** itera-se sobre o conjunto de dados k vezes e a cada rodada uma parte é usada na validação e o restante é mesclado para treinamento;
- **Usa todos os dados para treinamento e uma vez para teste.**



# 3

## Validação cruzada e otimização de hiperparâmetros

Ex.: 5-fold

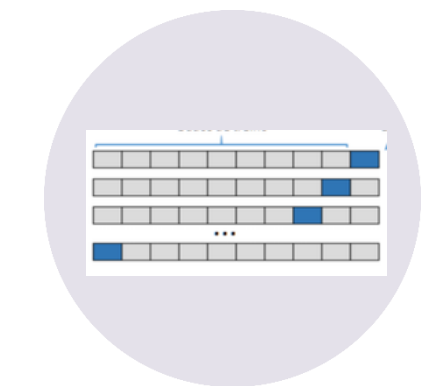


- 5 modelos diferentes;

- Ajustados para conjuntos de treinamento distintos e parcialmente sobrepostos;

- Avaliados em conjuntos de validação não sobrepostos;

- Desempenho: média aritmética dos k desempenhos na validação.



Introdução à  
validação cruzada  
k-fold



### 3

## Validação cruzada e otimização de hiperparâmetros



Validação cruzada  
k-fold e o trade-off  
viés x variância

Outros autores demonstram que 10-folds otimiza o tradeoff entre viés e variância e repetir a validação cruzada melhora performance.

Quando aumentamos o número de folds em geral:

- O viés de desempenho diminui;
- A variância aumenta;
- O custo computacional aumenta;

### 3

## Validação cruzada e otimização de hiperparâmetros



Seleção de modelo  
via validação  
cruzada k-fold

**1º Divide-se o conjunto de dados em duas partes (treino e teste);**

**2º Experimenta-se várias configurações de hiperparâmetros (Bayesian optimization, randomized search, ou grid search). Para cada configuração de hiperparâmetro aplica-se o método de validação cruzada k-fold no treino;**

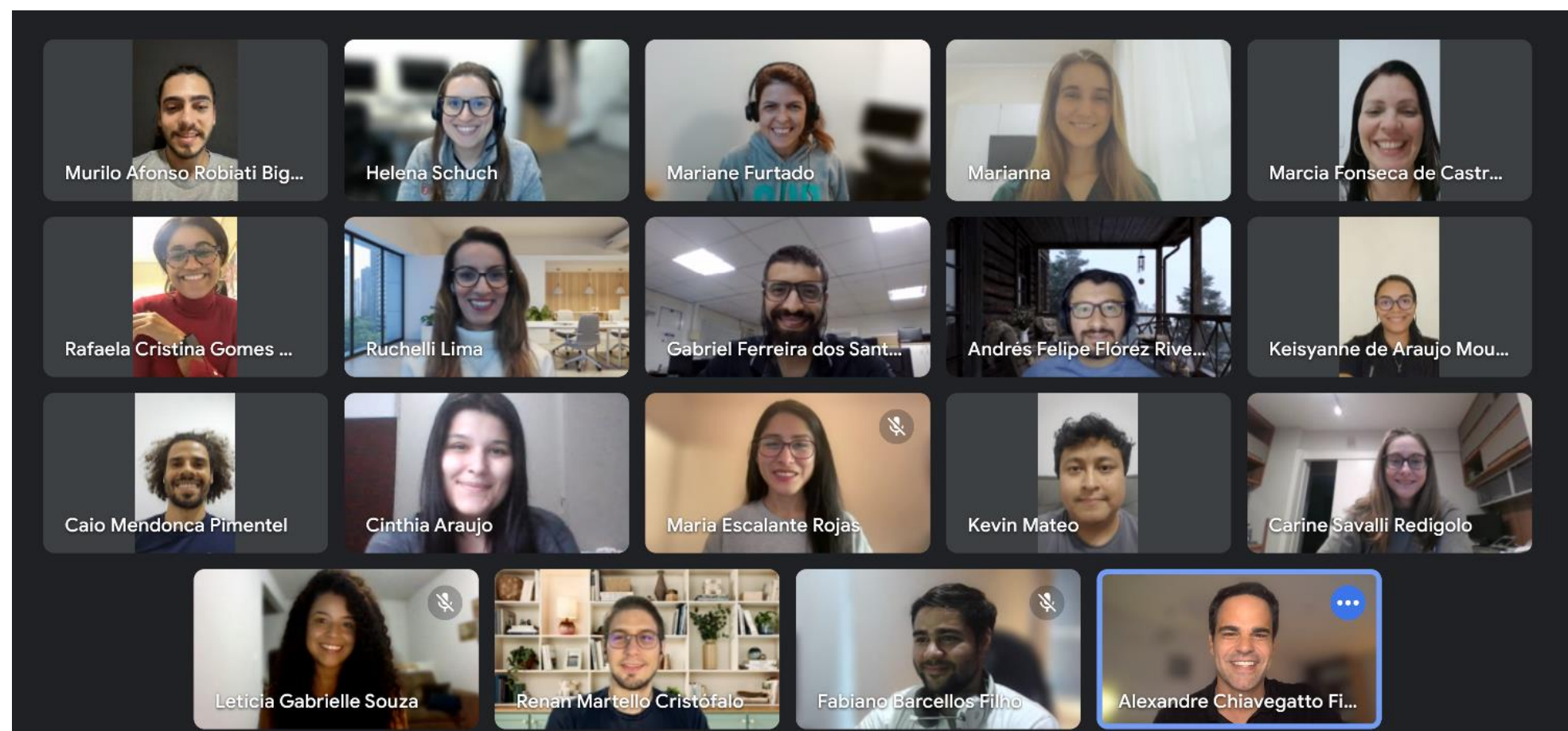
**3º Com as melhores configurações de hiperparâmetros resultantes da validação cruzada k-fold, utiliza-se então o conjunto de treinamento completo para o ajuste do modelo;**

**4º Utiliza-se então o conjunto de teste para avaliar o desempenho.**

## Referência



**RASCHKA, Sebastian. Model Evaluation, Model Selection, and Algorithm Selection in Machine Learning. arXiv:1811.12808v3 [cs.LG] 11 Nov 2020.**



*Obrigado!*

Alexandre Chiavegatto Filho



Alexandre Chiavegatto Filho



@SaudenoBR



@labdaps



alexdiasporto@usp.br

