

**Artigo:**

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

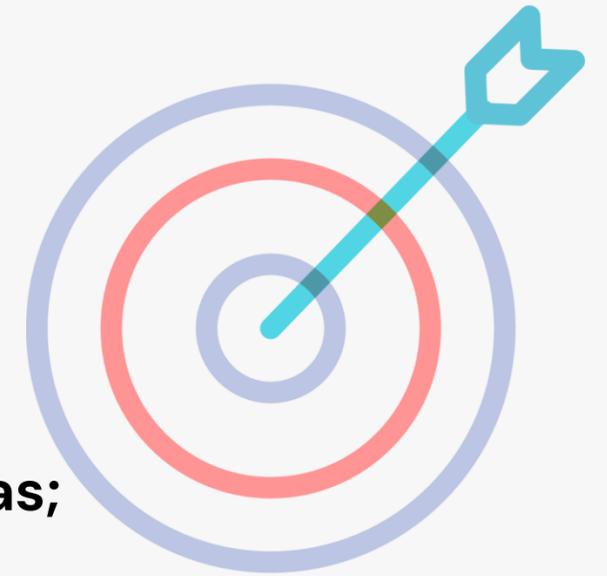
Autor: Sebastian Raschka\*

\*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

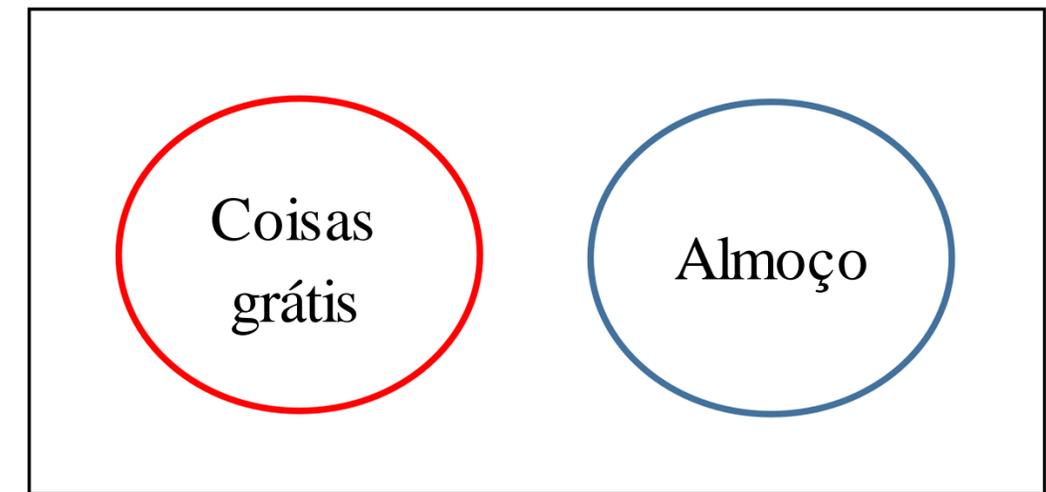
1. Introdução: Termos e técnicas essenciais de avaliação do modelo
2. Bootstrapping e Incerteza
3. Validação cruzada e otimização de hiperparâmetros

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*

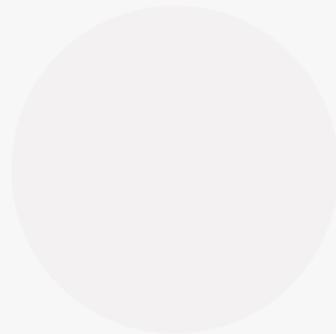
- *com dados reais, alguns costumam ganhar mais vezes (random forests e xgboost).*



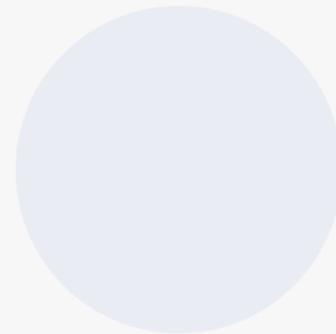
# 1 Introdução: Termos e técnicas essenciais de avaliação do modelo



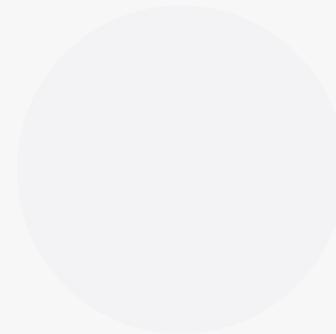
Estimação de performance:  
Desempenho de generalização  
vs  
Seleção de modelo



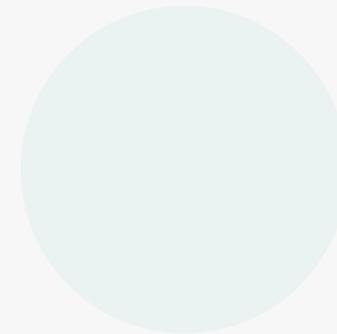
Premissas e Terminologia



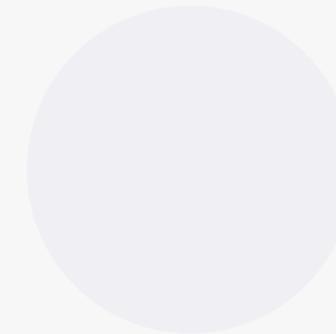
Validação de ressubstituição e Método Holdout



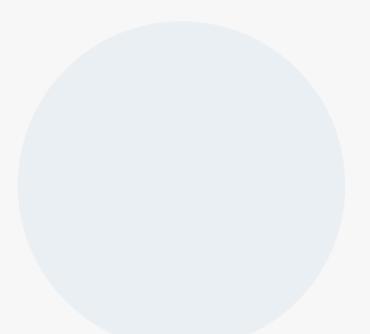
Estratificação



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

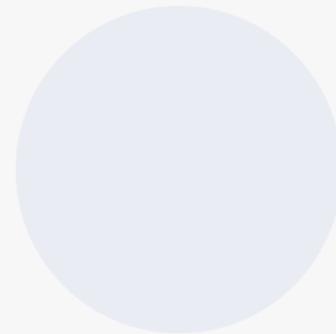
# 1 Introdução: Termos e técnicas essenciais de avaliação do modelo



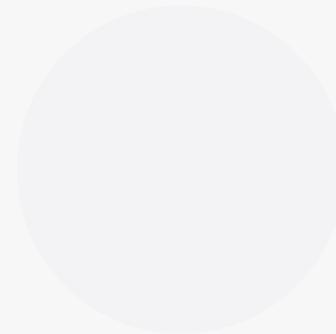
Estimação de performance:  
Desempenho de generalização  
vs  
Seleção de modelo



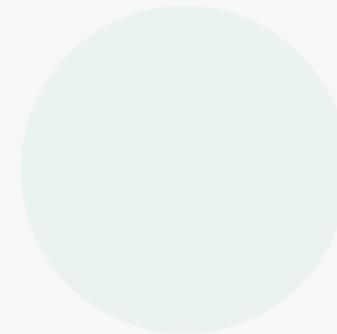
Hipótese e Terminologia



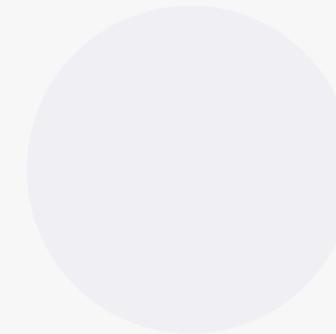
Validação de ressubstituição  
e Método Holdout



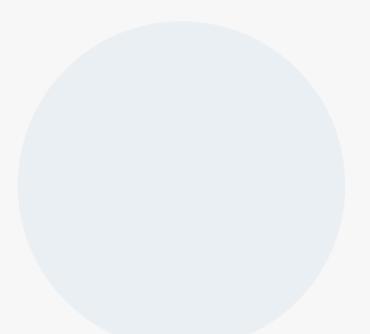
Estratificação



Validação Holdout



Viés pessimista



Intervalos de confiança via  
aproximação normal

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



Hipótese e  
Terminologia

- Assumir a hipótese de que:

Exemplos de treinamento são

{ independentes  
- distribuídos de forma idêntica

\*ex. não independente: séries temporais

- Foco do artigo em aprendizado supervisionado: desfecho conhecido

- Problemas de classificação: desfecho categórico

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



Hipótese e Terminologia

- Acurácia =  $\frac{\text{n}^\circ \text{ predições corretas}}{\text{n}^\circ \text{ de casos}}$

- Acurácia =  $1 - \text{Erro de predição}$        $\text{ACC} = 1 - \text{ERR}$

- Erro de predição =  $\frac{\text{valor esperado da perda}}{\text{n}^\circ \text{ de exemplos}}$

- Perda:  $L(\hat{y}_i, y_i) = \begin{cases} 0 & \text{if } \hat{y}_i = y_i \\ 1 & \text{if } \hat{y}_i \neq y_i \end{cases}$

- Objetivo é minimizar a probabilidade de erro:  $C(h) = \Pr_{(x,y) \sim \mathcal{D}} [h(x) \neq y]$

- Viés (viés de um estimador):  $\text{Bias} = E[\hat{\beta}] - \beta$

Se  $E[\hat{\beta}] - \beta = 0$ : estimador não viesado.

# 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

# PREDIÇÃO COM MACHINE LEARNING

## Dados

Preferencialmente muitos e bons (preenchidos corretamente e preditores fortes). Realizar o pré-processamento das variáveis.

## Algoritmos

Inserir os dados no algoritmo de machine learning para aprender os parâmetros que melhor mapeiam os preditores aos resultados.

## Testar no futuro

Inserir no algoritmo novos dados para testar a qualidade desse algoritmo para prever dados futuros.

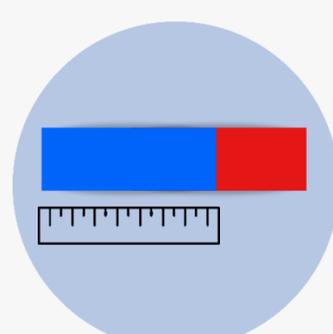
# 1 Introdução: Termos e técnicas essenciais de avaliação do modelo



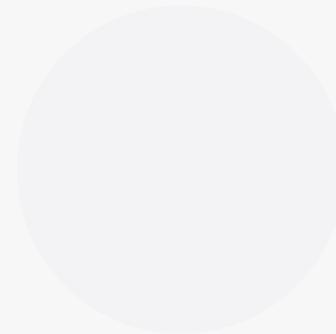
Estimação de performance:  
Desempenho de generalização  
vs  
Seleção de modelo



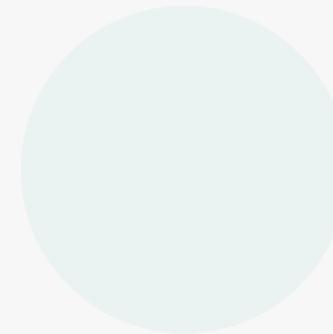
Hipótese e Terminologia



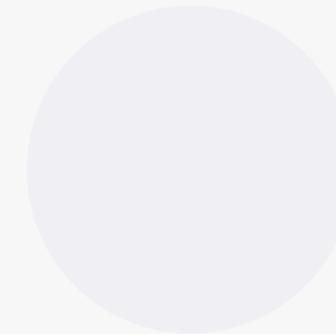
Validação de re-substituição e Método Holdout



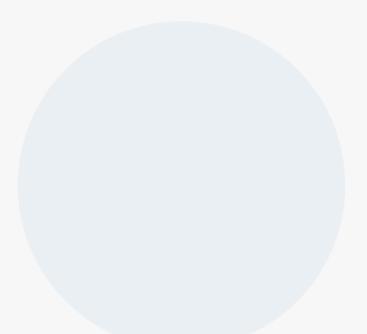
Estratificação



Validação Holdout

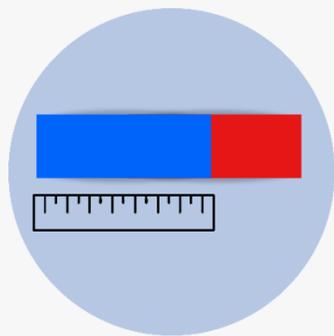


Viés pessimista



Intervalos de confiança via aproximação normal

# 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)

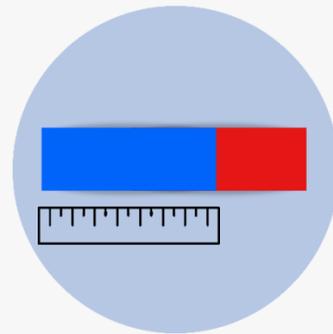
# 1 Introdução: Termos e técnicas essenciais de avaliação do modelo



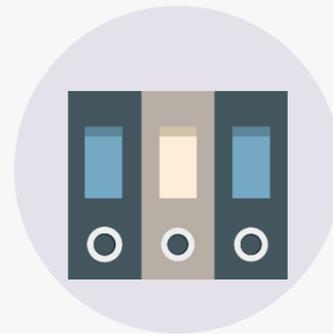
Estimação de performance:  
Desempenho de generalização  
vs  
Seleção de modelo



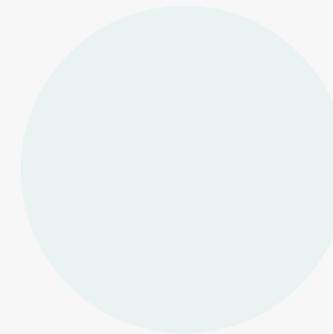
Hipótese e Terminologia



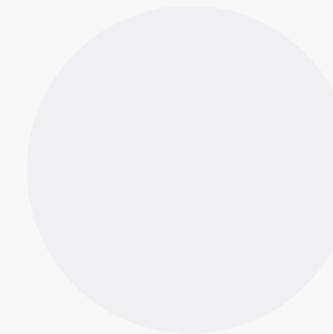
Validação de re-substituição e Método Holdout



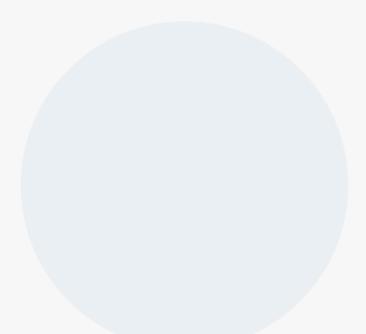
Estratificação



Validação Holdout

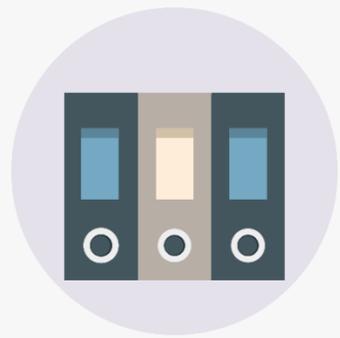


Viés pessimista



Intervalos de confiança via aproximação normal

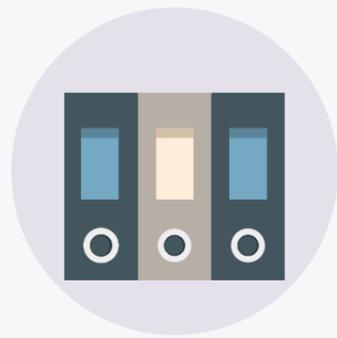
# 1 Introdução: Termos e técnicas essenciais de avaliação do modelo



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



Estratificação

- Nesse caso as classes estarão proporcionalmente representadas nos subconjuntos (treino e teste)
- Serve para manter a proporção nos subconjuntos
- Importante que sempre seja realizada, principalmente quando as classes originais já possuem um alto desequilíbrio

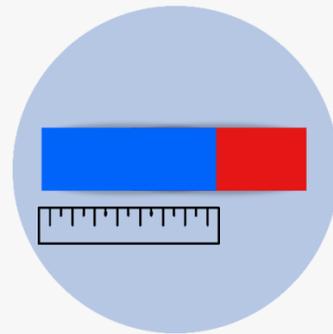
# 1 Introdução: Termos e técnicas essenciais de avaliação do modelo



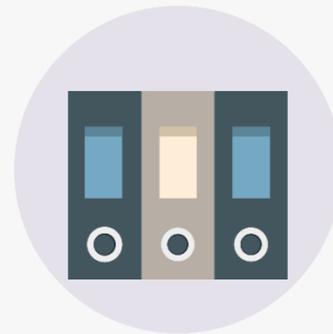
Estimação de performance:  
Desempenho de generalização  
vs  
Seleção de modelo



Hipótese e Terminologia



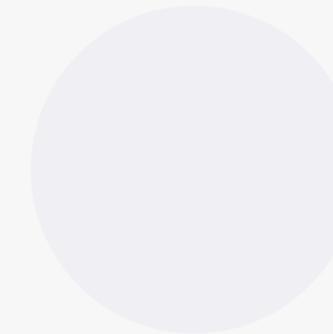
Validação de re-substituição e Método Holdout



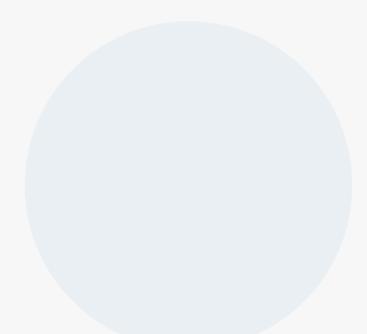
Estratificação



Validação Holdout



Viés pessimista



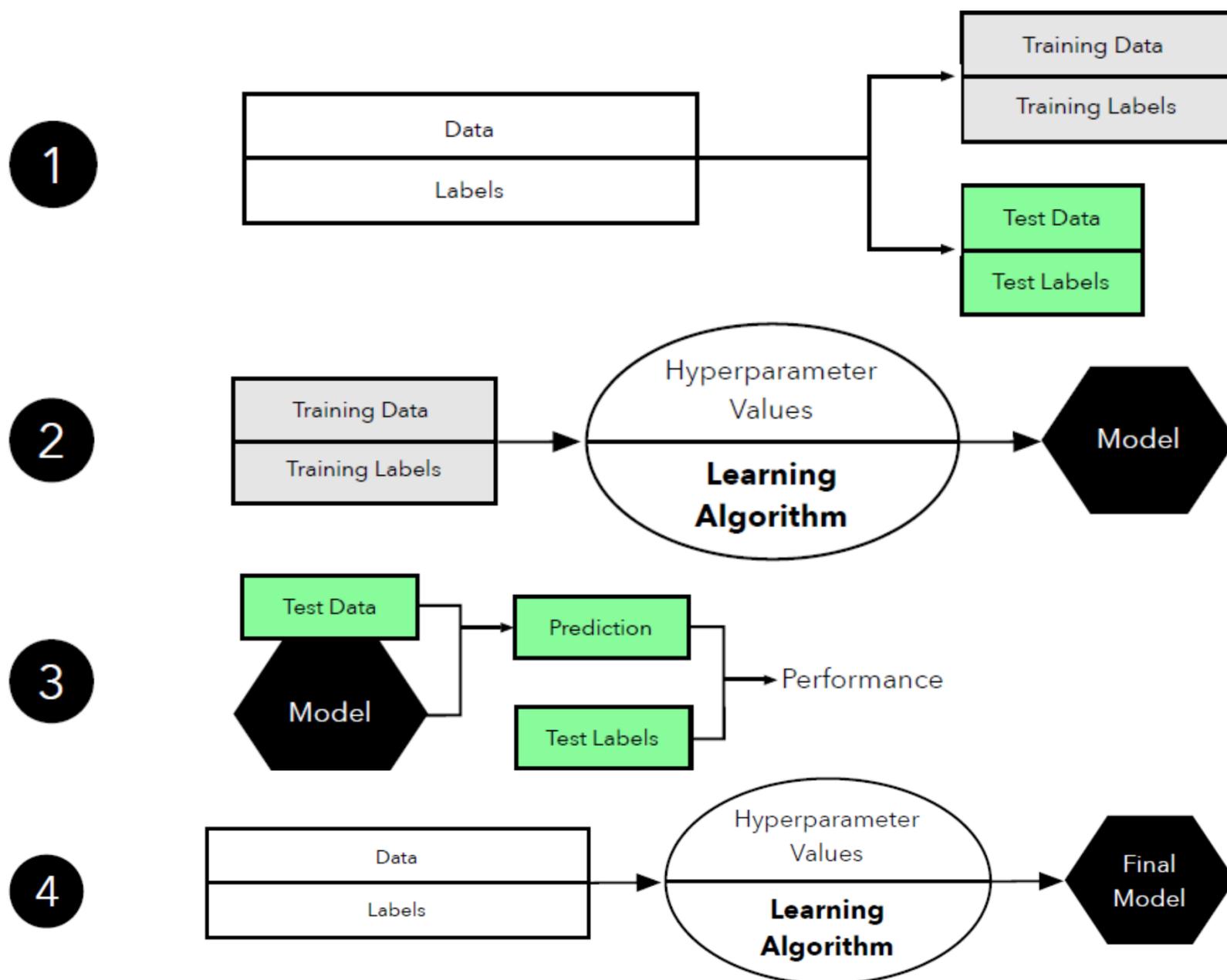
Intervalos de confiança via aproximação normal

# 1

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



Validação Holdout



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

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

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

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

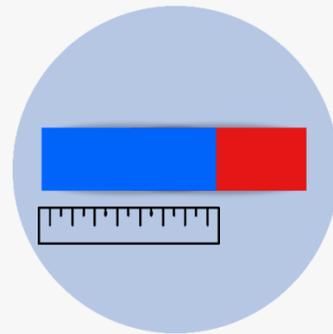
# 1 Introdução: Termos e técnicas essenciais de avaliação do modelo



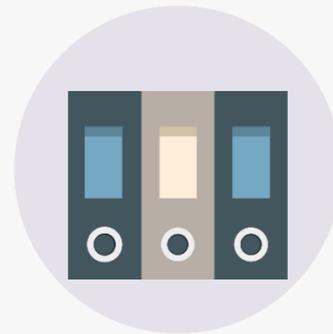
Estimação de performance:  
Desempenho de generalização  
vs  
Seleção de modelo



Hipótese e Terminologia



Validação de re-substituição e Método Holdout



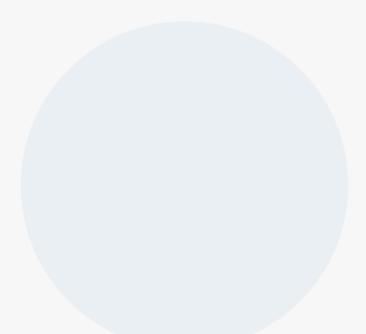
Estratificação



Validação Holdout



Viés pessimista



Intervalos de confiança via aproximação normal

# 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.

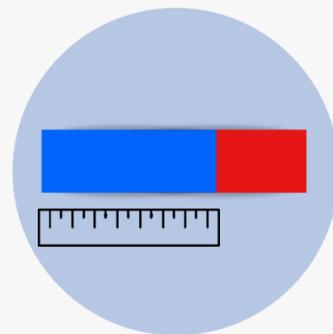
# 1 Introdução: Termos e técnicas essenciais de avaliação do modelo



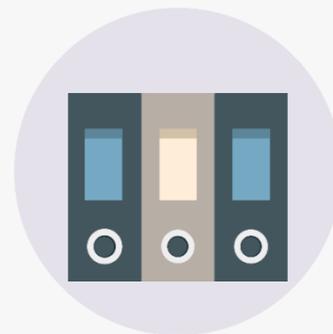
Estimação de performance:  
Desempenho de generalização  
vs  
Seleção de modelo



Hipótese e Terminologia



Validação de re-substituição e Método Holdout



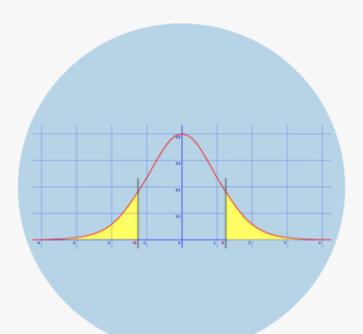
Estratificação



Validação Holdout

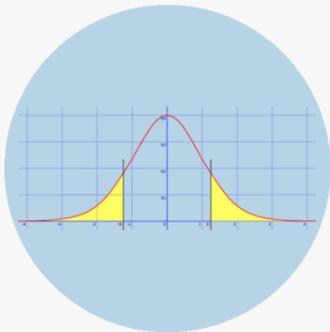


Viés pessimista



Intervalos de confiança via aproximação normal

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



Intervalos de  
confiança via  
aproximação  
normal

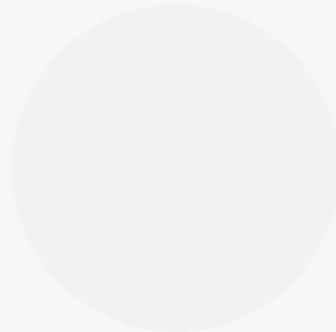
- Com o método holdout calculamos uma estimativa pontual de desempenho de generalização;
- A estimativa pontual pode ser sensível (ex.: alta variância) à divisão dos dados;
- Com a aproximação normal, assume-se normalidade na distribuição e calcula-se o intervalo de confiança na média em uma única divisão de treino-teste sob o teorema do limite central;

# 2

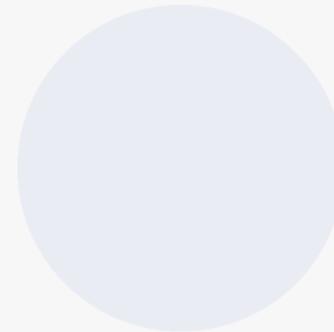
## Bootstrapping e Incerteza



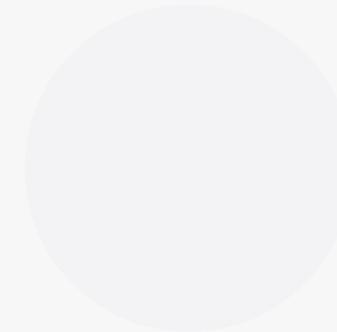
Visão geral



Reamostragem



Validação  
Holdout  
repetida



Método Bootstrap e  
intervalos de  
confiança empíricos

## 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

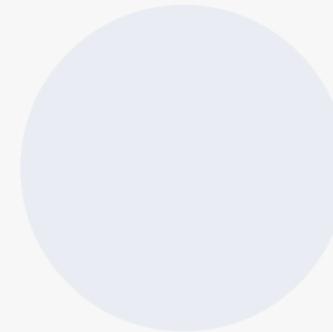
## Bootstrapping e Incerteza



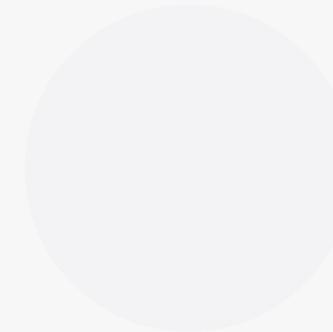
Visão geral



Reamostragem



Validação  
Holdout  
repetida



Método Bootstrap e  
intervalos de  
confiança empíricos

# 2

## Bootstrapping e Incerteza

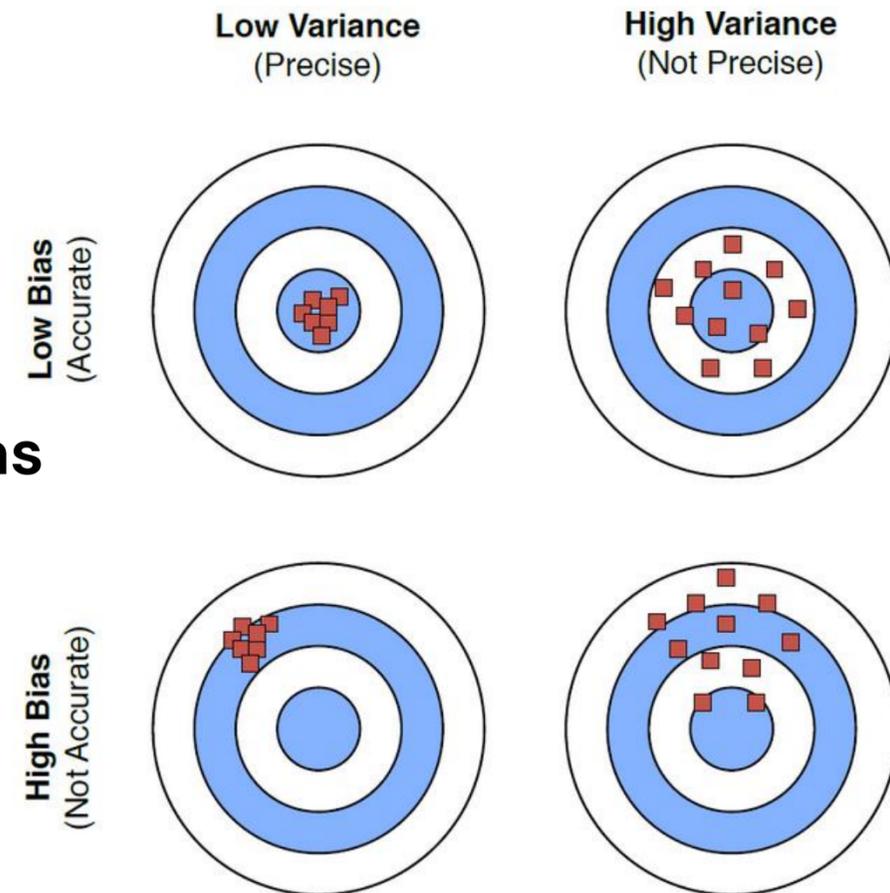


Reamostragem

- Relação viés x variância →

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

- Viés é o erro consistente do modelo.

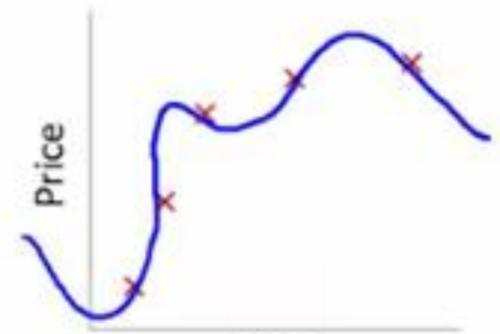


# 2

## Bootstrapping e Incerteza

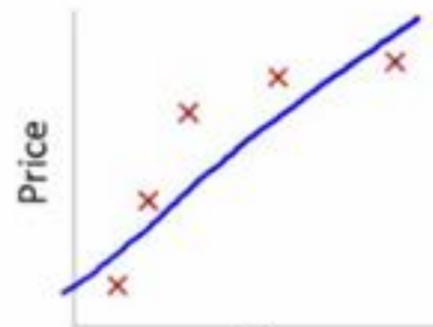


Reamostragem



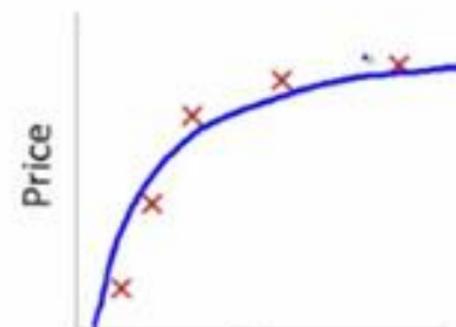
$$\theta_0 + \theta_1x + \theta_2x^2 + \theta_3x^3 + \theta_4x^4$$

High variance  
(overfit)



$$\theta_0 + \theta_1x$$

High bias  
(underfit)



$$\theta_0 + \theta_1x + \theta_2x^2$$

“Just right”

## 2

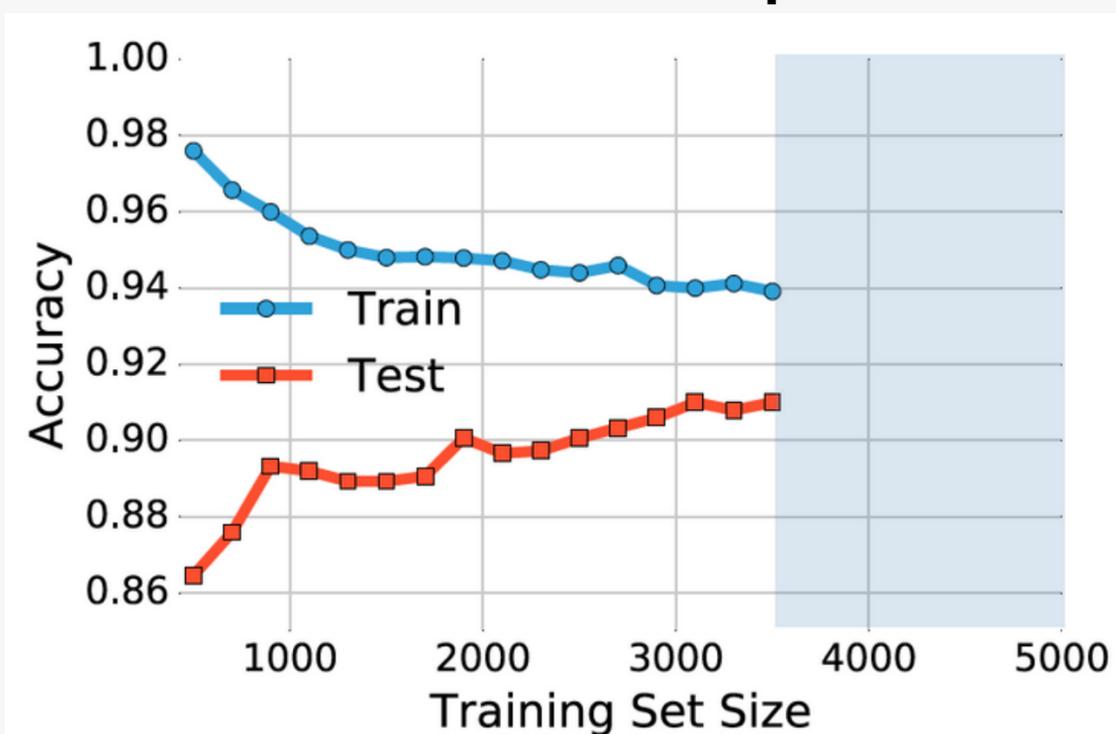
# Bootstrapping e Incerteza



Reamostragem

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

Ex.: MNIST5  
Classificadores  
Softmax



## 2

# Bootstrapping e Incerteza



Reamostragem

- Para grandes conjuntos de dados:
  - A performance de treino diminui à medida que o número de amostras aumenta (menos risco sobreajuste)
  - A performance no teste melhora com o tamanho do conjunto de treino
- Para pequenos conjuntos de dados:
  - Algoritmo capta o ruído no treino (não consegue generalizar)

# 2

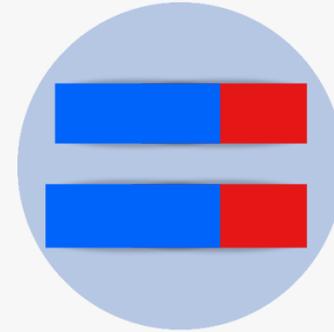
## Bootstrapping e Incerteza



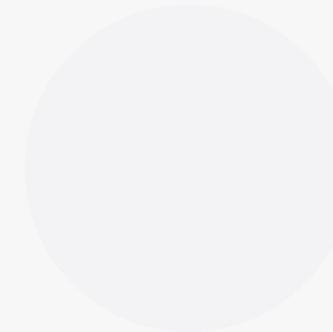
Visão geral



Reamostragem



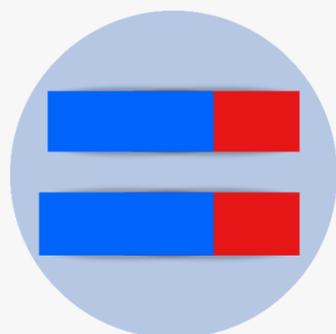
Validação  
Holdout  
repetida



Método Bootstrap e  
intervalos de  
confiança empíricos

## 2

# Bootstrapping e Incerteza



Validação  
Holdout  
repetida

- Estimativa de desempenho mais robusta repetindo o método holdout k vezes com diferentes sementes aleatórias;

- Resultado será o desempenho médio dado por:

$$ACC_{avg} = \frac{1}{k} \sum_{j=1}^k ACC_j$$

- Mostra quão bem o modelo funciona e informa sobre a estabilidade do modelo com diferentes divisões de treino e teste

- Amostragem sem reposição

# 2

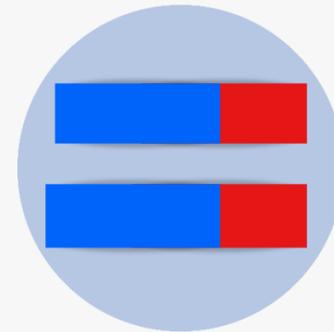
## Bootstrapping e Incerteza



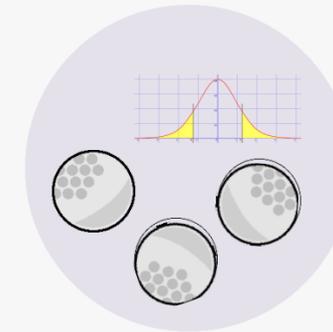
Visão geral



Reamostragem



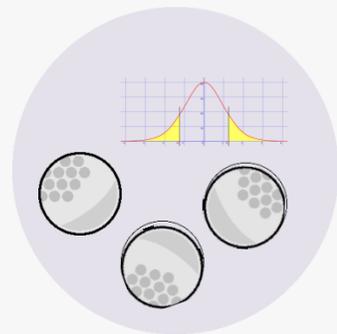
Validação  
Holdout  
repetida



Método Bootstrap e  
intervalos de  
confiança empíricos

# 2

## Bootstrapping e Incerteza

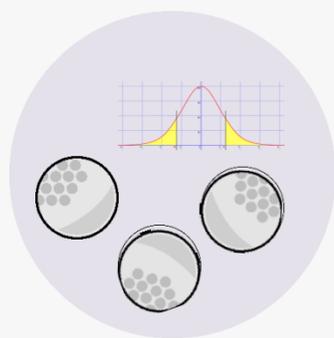


Método Bootstrap e intervalos de confiança empíricos

- **Bootstrapping: técnica para gerar novas amostras por meio de reamostragem com reposição (mesma pessoa pode ser sorteada mais de uma vez).**
- **A ideia é gerar "novas dados" de uma população por amostragem repetida com reposição.**

## 2

# Bootstrapping e Incerteza



Método Bootstrap e intervalos de confiança empíricos

1º Recebe-se o conjunto de dados ( $n$ )

2º Para  $b$  rodadas de bootstrap:

- Sorteia-se uma observação do conjunto de dados e a atribui à  $j$ -ésima amostra de bootstrap;
- Repete-se isso até que a amostra de bootstrap tenha o tamanho  $n$ ;
- Com isso, certos exemplos podem aparecer mais de uma vez em uma amostra de bootstrap;

3º Ajusta-se o modelo a cada amostra de bootstrap e calcula-se a performance estimada de  $b$ .

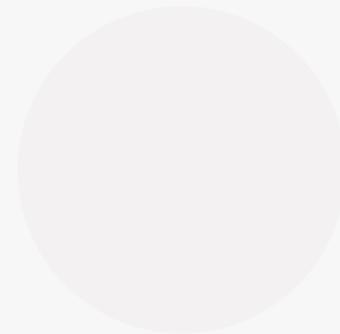
$$ACC_{boot} = \frac{1}{b} \sum_{j=1}^b \frac{1}{n} \sum_{i=1}^n \left( 1 - L(\hat{y}_i, y_i) \right)$$

# 3

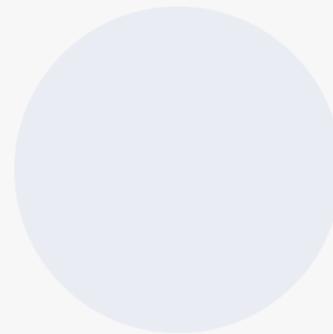
## Validação cruzada e otimização de hiperparâmetros



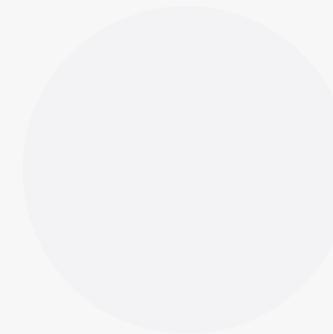
Visão geral



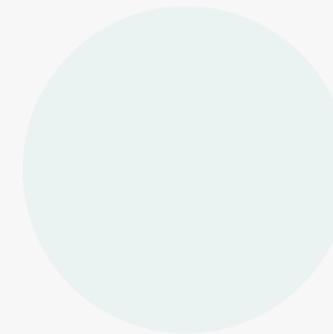
Hiperparâmetros e seleção de modelo



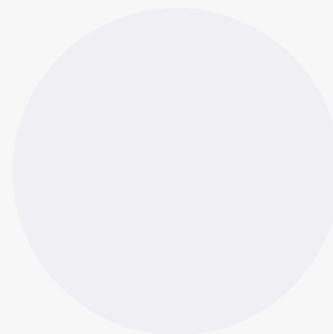
Método Holdout de 3 vias para ajuste de hiperparâmetros



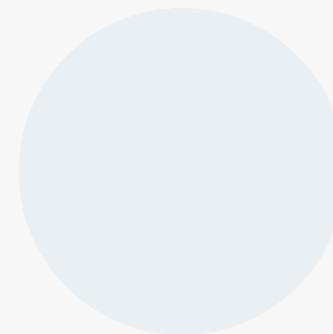
Introdução à validação cruzada k-fold



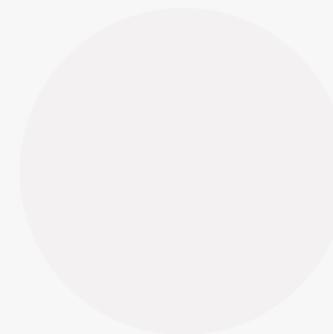
Casos especiais: Validação cruzada 2-folds e leave-one-out



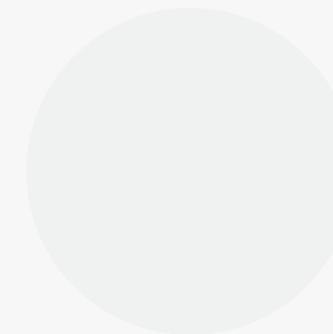
Validação cruzada k-fold e o trade-off viás x variância



Seleção de modelo via validação cruzada k-fold



Uma nota sobre seleção do modelo e grandes conjuntos de dados



Uma nota sobre a seleção de variáveis durante a seleção do modelo

### 3

## Validação cruzada e otimização de hiperparâmetros



Visão geral

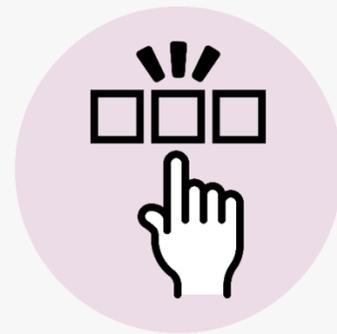
- Quase todos os algoritmos tem configurações a serem especificadas;
- Os "botões de ajuste" são os hiperparâmetros (em geral regularizadores – tentativas de segurar a complexidade);
- Eles ajudam a otimizar o desempenho encontrando o equilíbrio entre o viés e a variância;

# 3

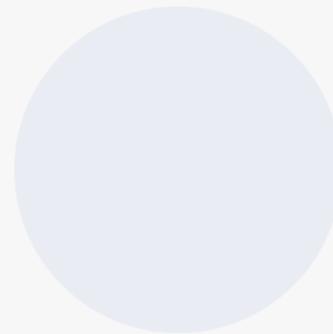
## Validação cruzada e otimização de hiperparâmetros



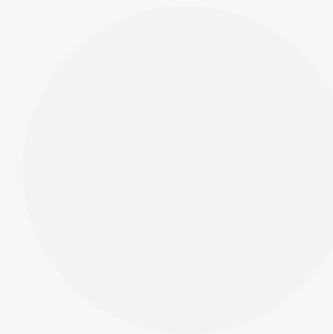
Visão geral



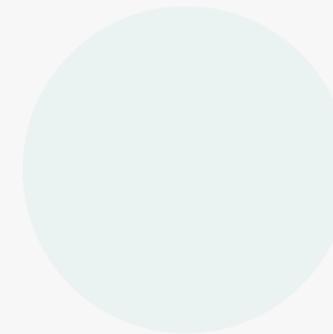
Hiperparâmetros e seleção de modelo



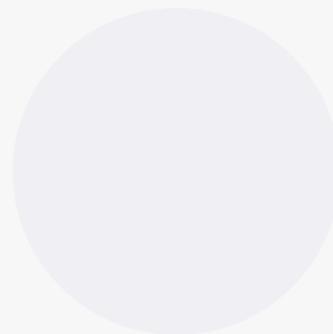
Método Holdout de 3 vias para ajuste de hiperparâmetros



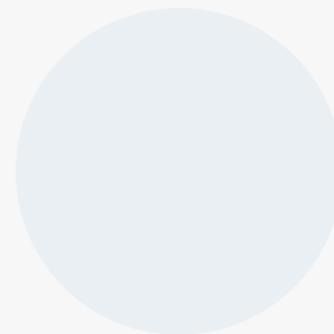
Introdução à validação cruzada k-fold



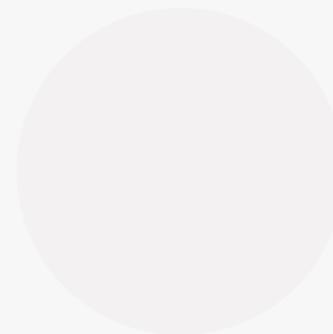
Casos especiais: Validação cruzada 2-folds e leave-one-out



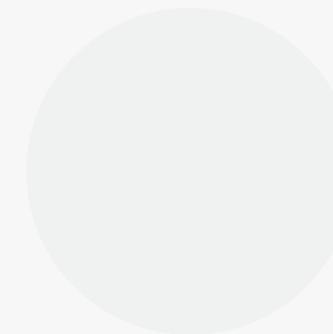
Validação cruzada k-fold e o trade-off viás x variância



Seleção de modelo via validação cruzada k-fold



Uma nota sobre seleção do modelo e grandes conjuntos de dados



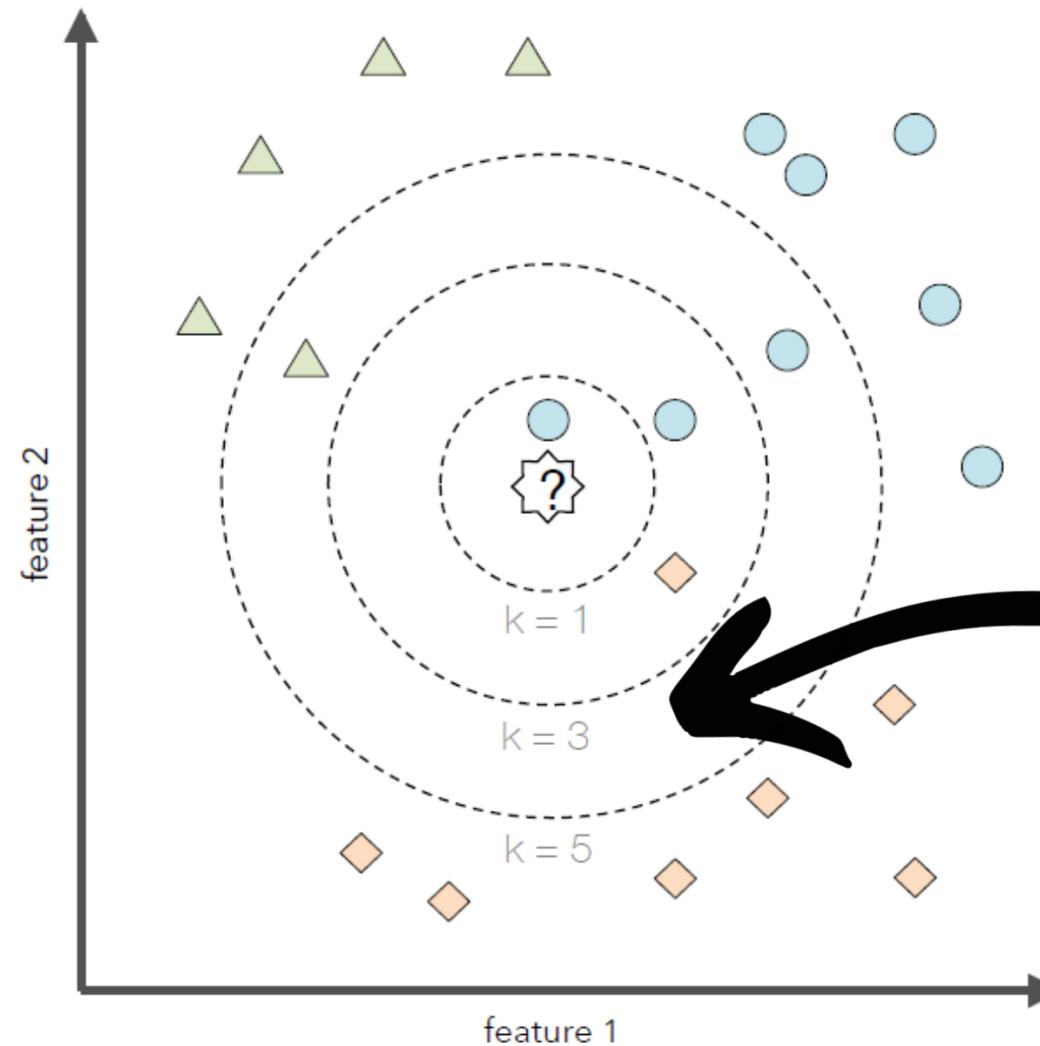
Uma nota sobre a seleção de variáveis durante a seleção do modelo

### 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



Hiperparâmetros  
e seleção de  
modelo

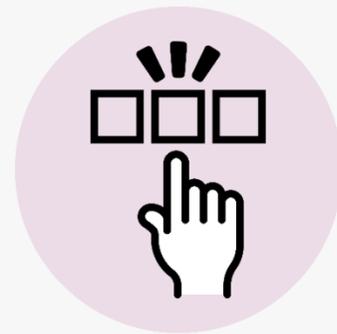
- **Alterar valores de hiperparâmetros durante o treinamento resulta em diferentes modelos;**
- **Seleção do modelo: processo para encontrar o modelo com melhor desempenho, comparando o desempenho de modelos produzidos com diferentes hiperparâmetros;**

# 3

## Validação cruzada e otimização de hiperparâmetros



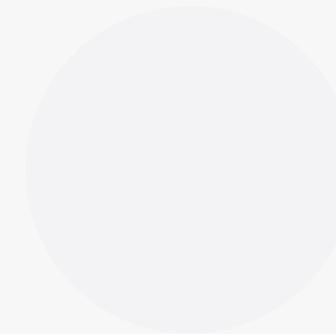
Visão geral



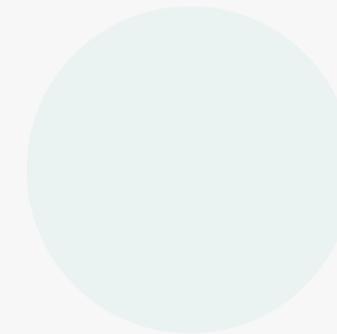
Hiperparâmetros e seleção de modelo



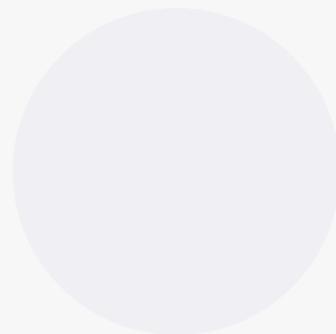
As 3 vias do Método Holdout para ajuste de hiperparâmetros



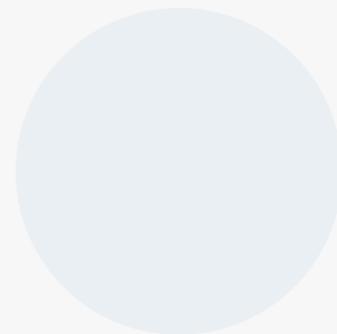
Introdução à validação cruzada k-fold



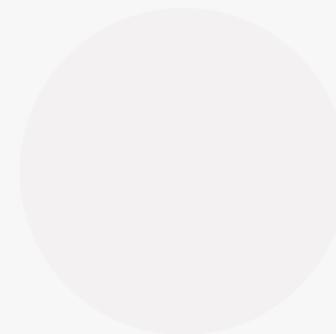
Casos especiais: Validação cruzada 2-folds e leave-one-out



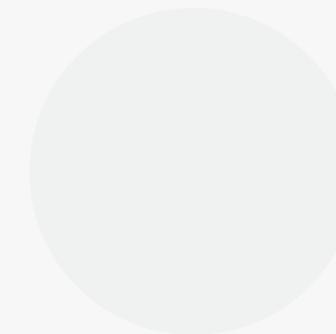
Validação cruzada k-fold e o trade-off viás x variância



Seleção de modelo via validação cruzada k-fold



Uma nota sobre seleção do modelo e grandes conjuntos de dados



Uma nota sobre a seleção de variáveis durante a seleção do modelo

### 3

## Validação cruzada e otimização de hiperparâmetros

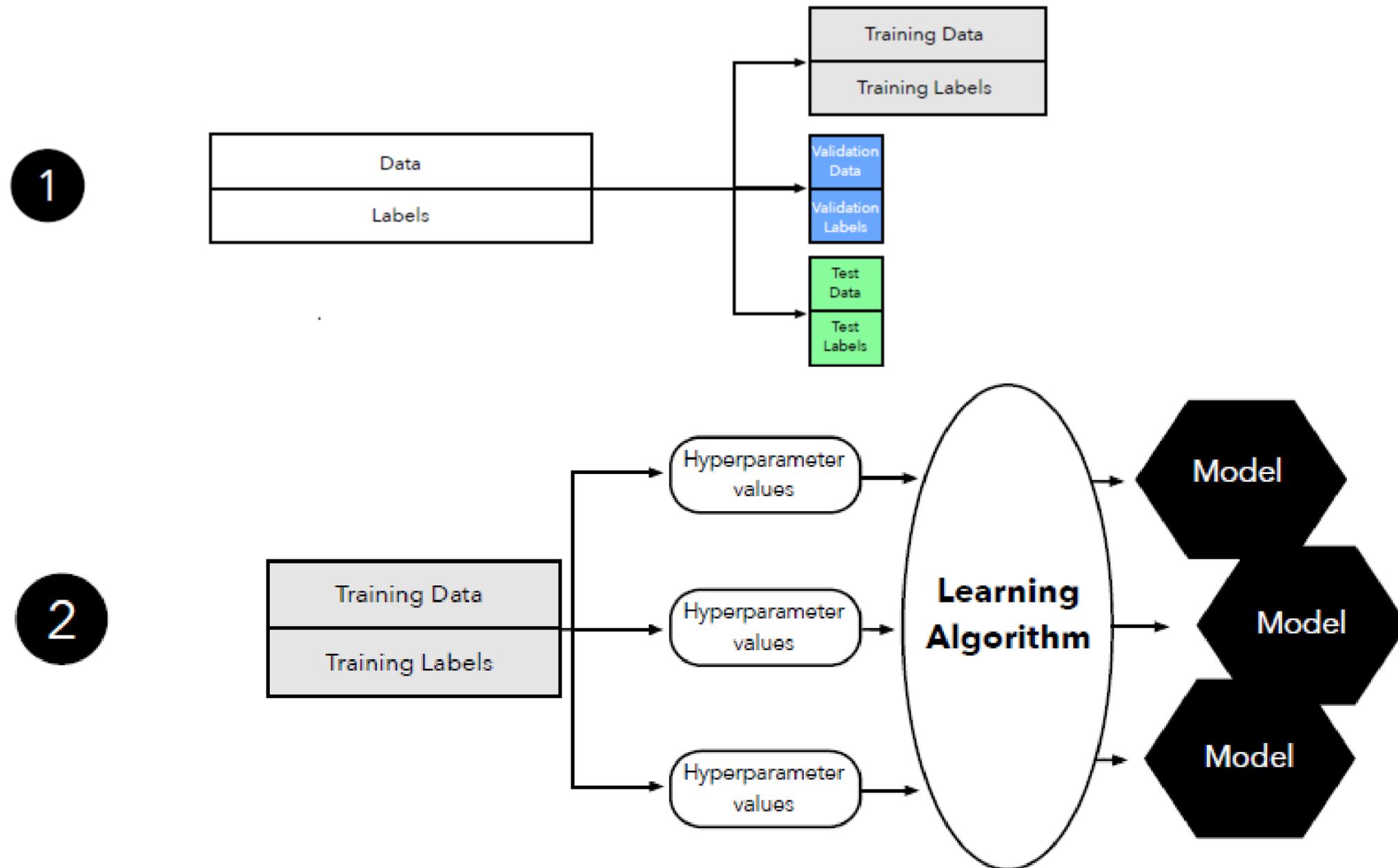


As 3 vias do Método Holdout para ajuste de hiperparâmetros

- 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

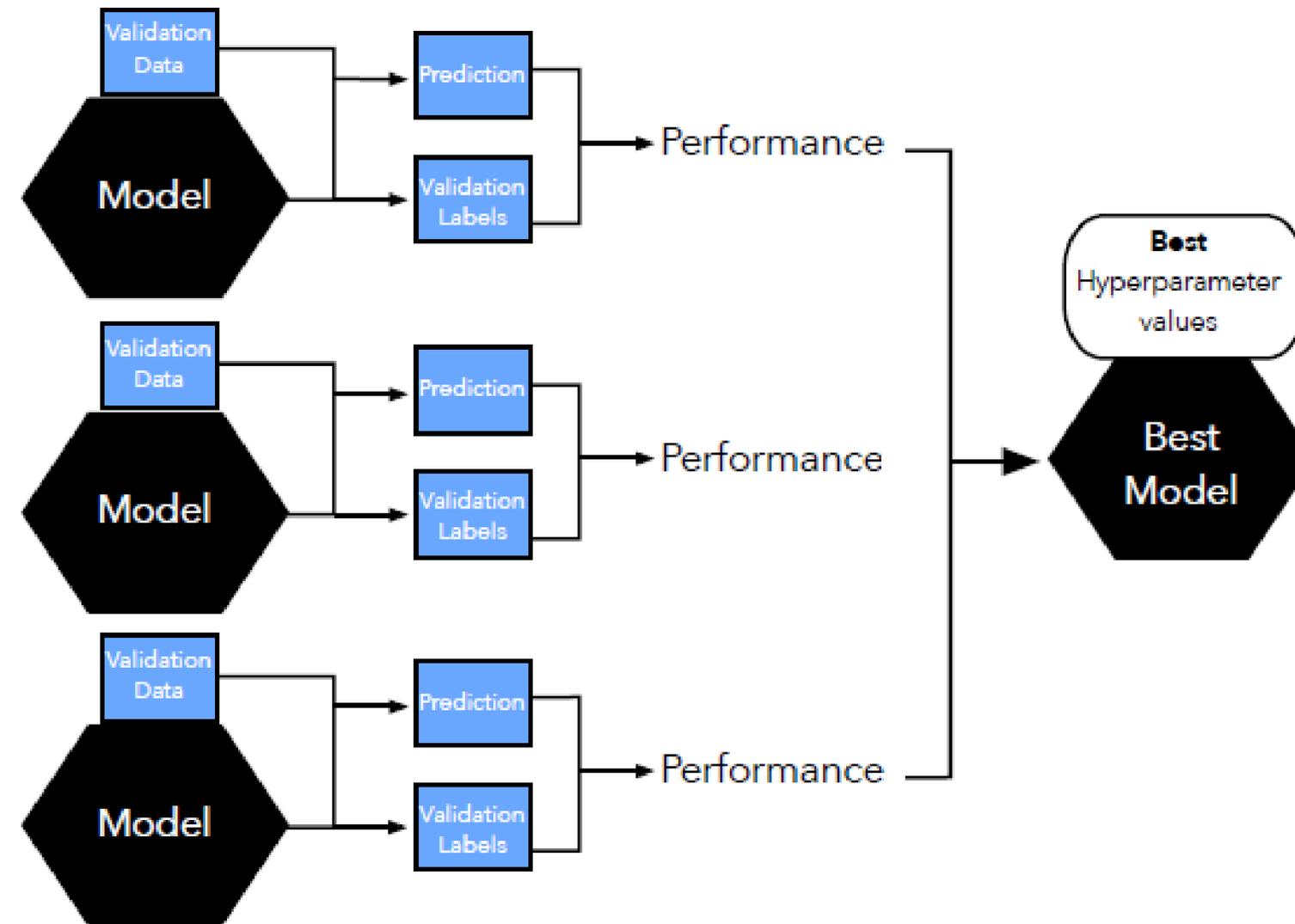
## Validação cruzada e otimização de hiperparâmetros



# 3

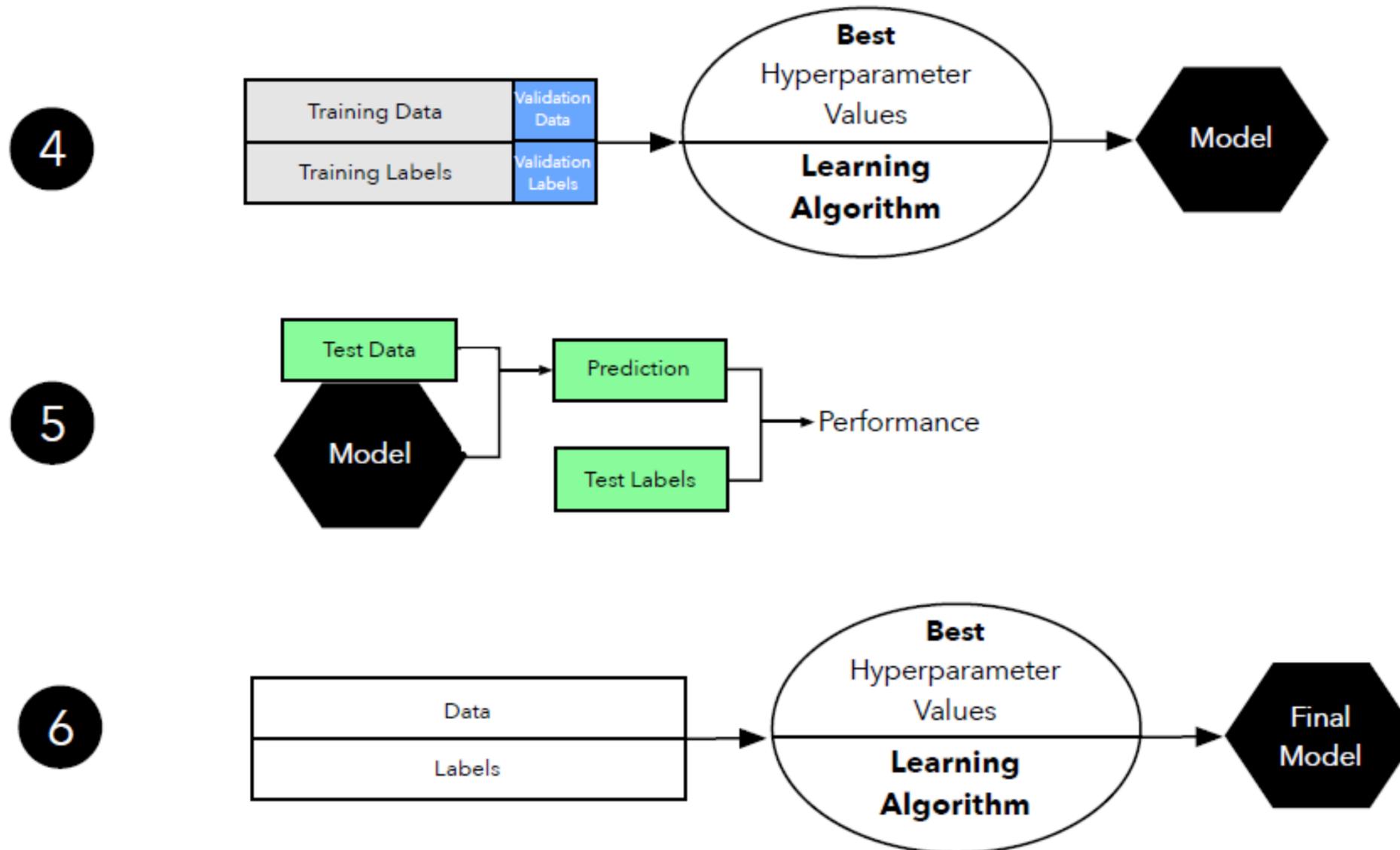
## Validação cruzada e otimização de hiperparâmetros

3



# 3

## Validação cruzada e otimização de hiperparâmetros

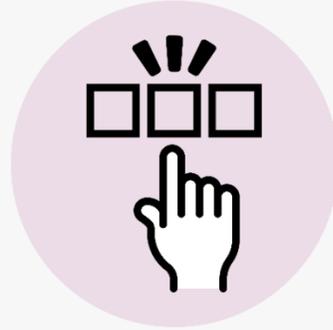


# 3

## Validação cruzada e otimização de hiperparâmetros



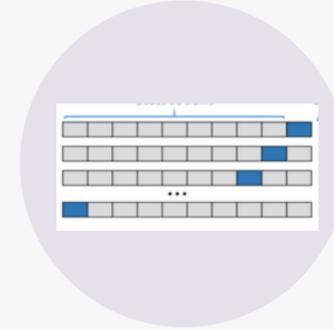
Visão geral



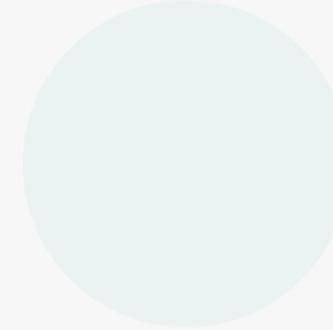
Hiperparâmetros e seleção de modelo



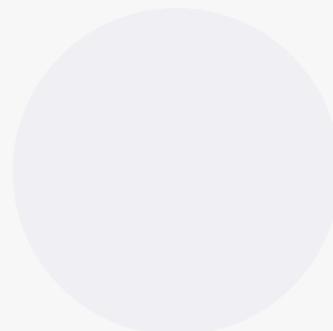
As 3 vias do Método Holdout para ajuste de hiperparâmetros



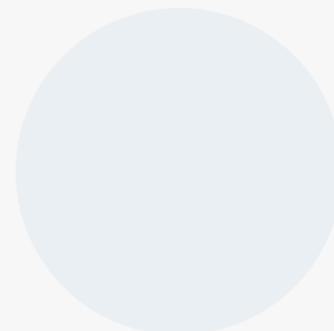
Introdução à validação cruzada k-fold



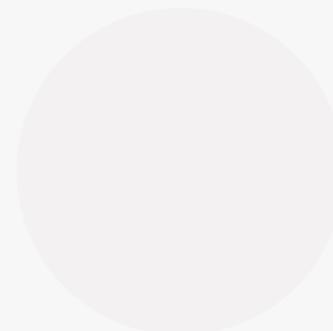
Casos especiais: Validação cruzada 2-folds e leave-one-out



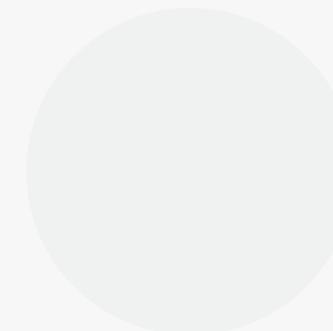
Validação cruzada k-fold e o trade-off viás x variância



Seleção de modelo via validação cruzada k-fold



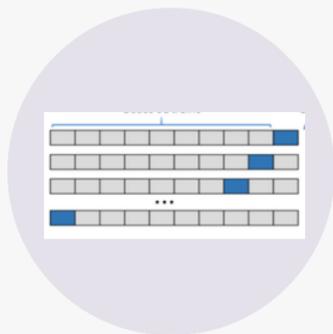
Uma nota sobre seleção do modelo e grandes conjuntos de dados



Uma nota sobre a seleção de variáveis durante a seleção do modelo

### 3

## Validação cruzada e otimização de hiperparâmetros



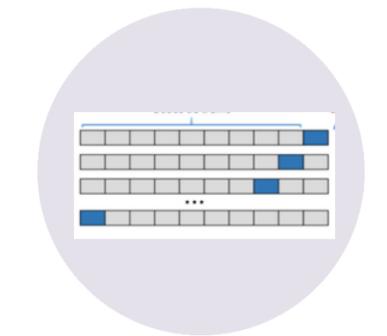
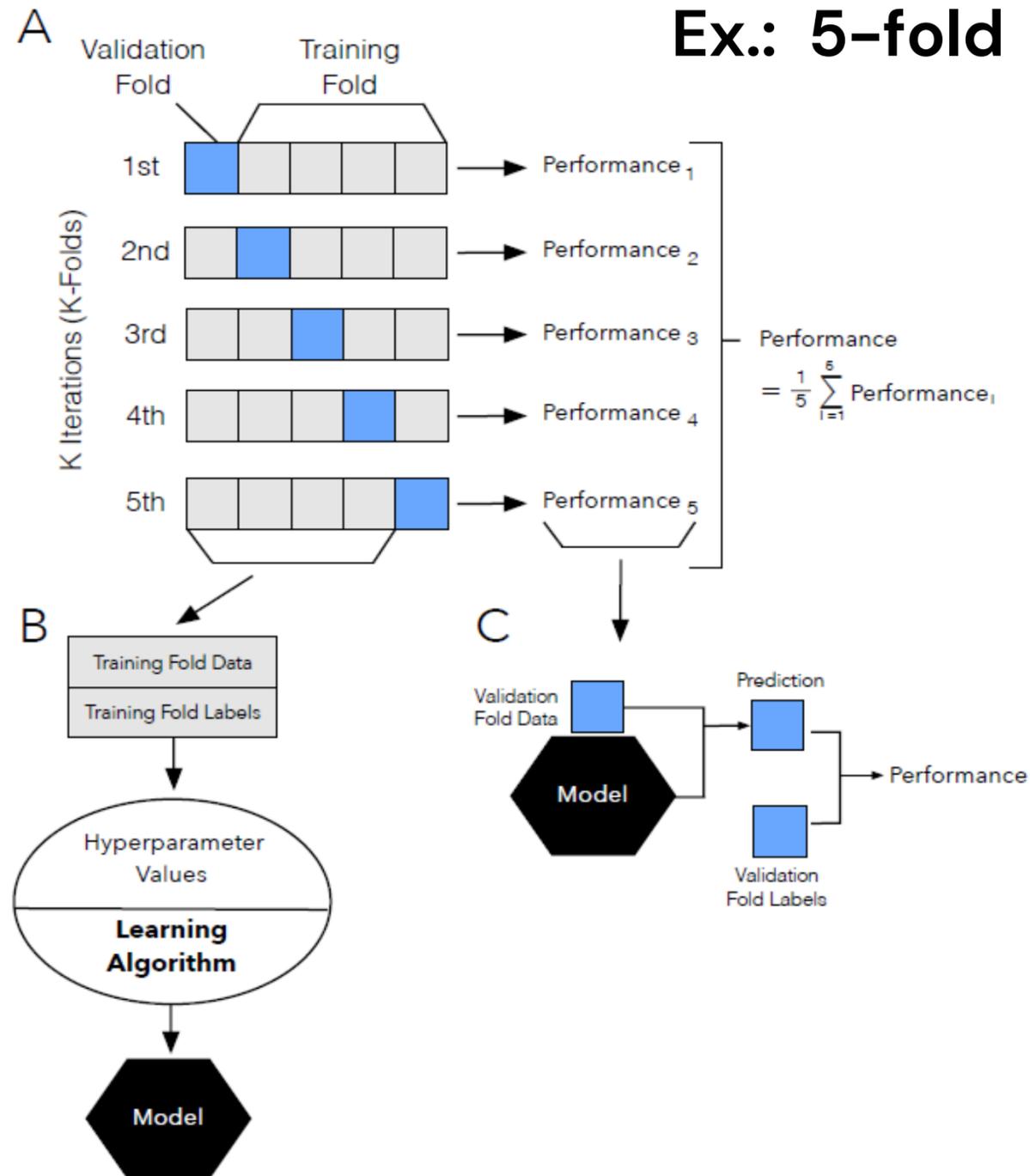
Introdução à  
validação cruzada  
k-fold

- **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



Introdução à validação cruzada k-fold

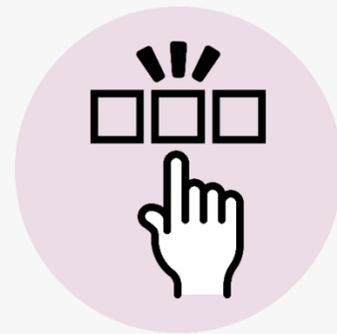
- 5 modelos diferentes;
- Ajustados para conjuntos de treinamento distintos e parcialmente sobrepostos;
- Avaliados em conjuntos de validação não sobrepostos;
- Desempenho: média aritmética dos k desempenhos na validação.

# 3

## Validação cruzada e otimização de hiperparâmetros



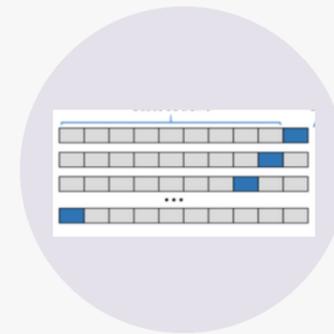
Visão geral



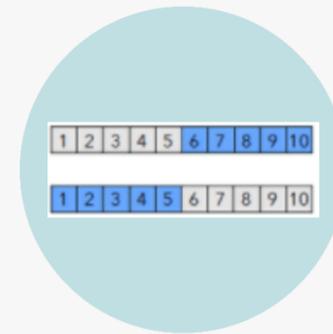
Hiperparâmetros e seleção de modelo



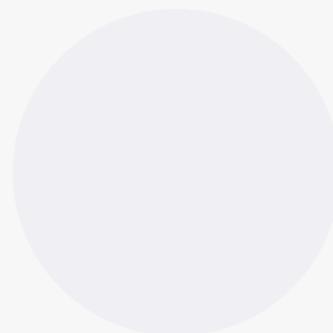
As 3 vias do Método Holdout para ajuste de hiperparâmetros



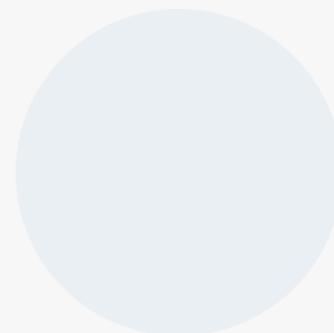
Introdução à validação cruzada k-fold



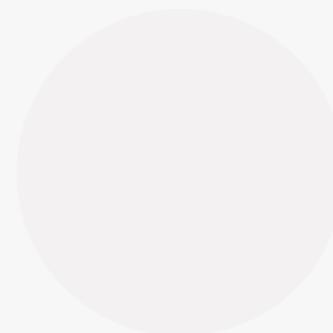
Casos especiais: Validação cruzada 2-folds e leave-one-out



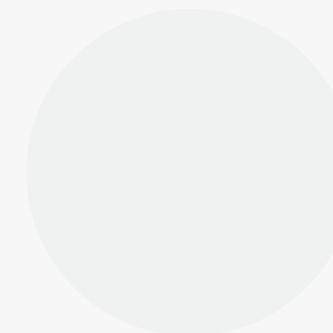
Validação cruzada k-fold e o trade-off viás x variância



Seleção de modelo via validação cruzada k-fold



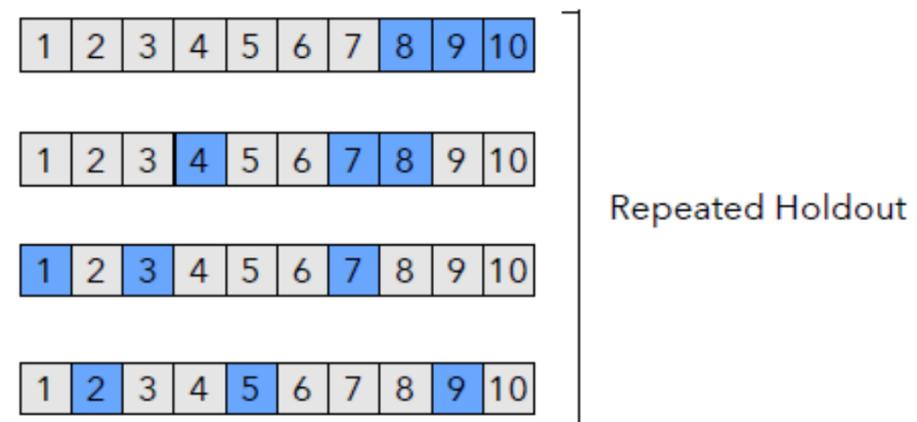
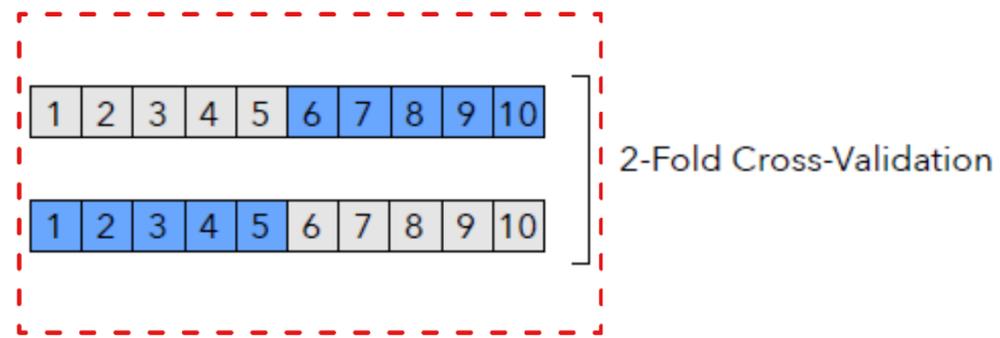
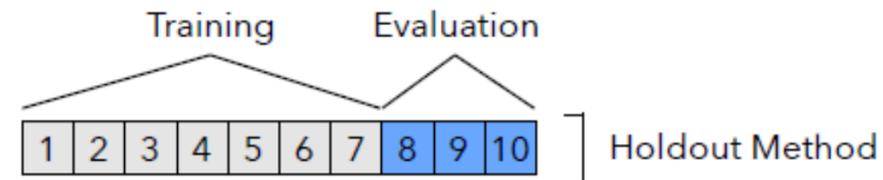
Uma nota sobre seleção do modelo e grandes conjuntos de dados



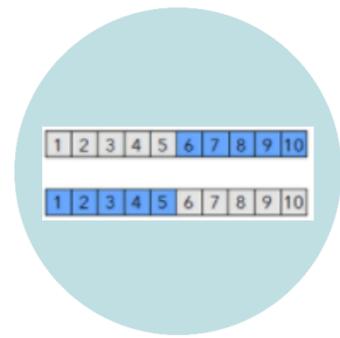
Uma nota sobre a seleção de variáveis durante a seleção do modelo

# 3

## Validação cruzada e otimização de hiperparâmetros



...



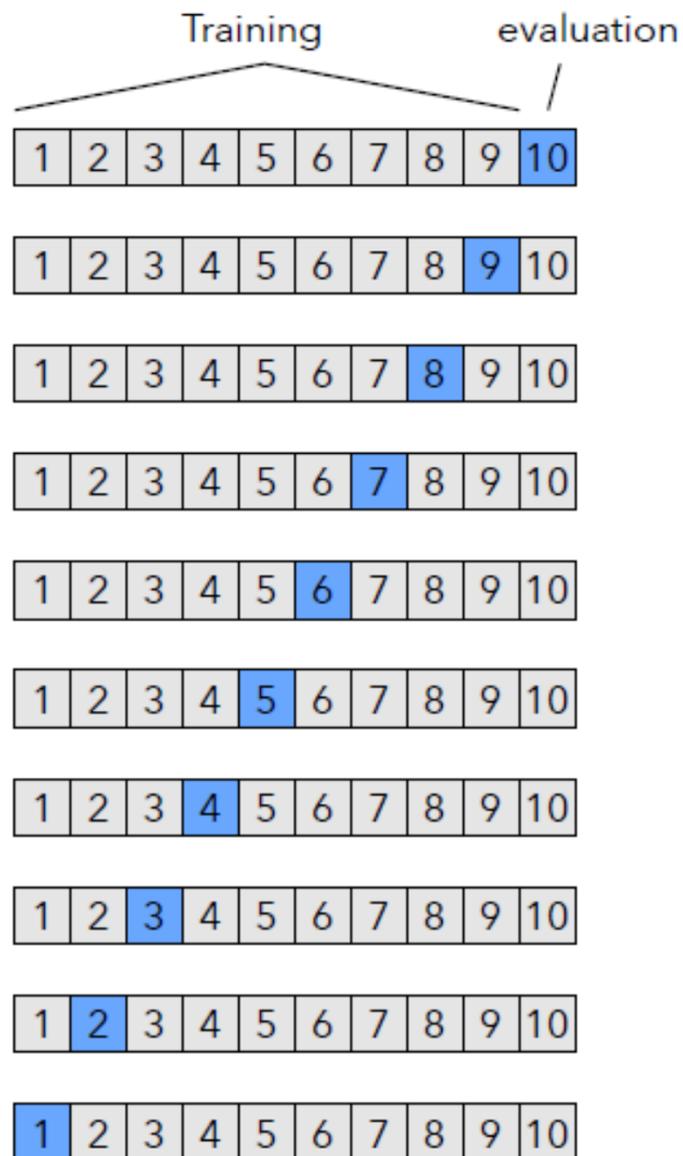
Casos especiais:  
Validação cruzada  
2-folds e leave-one-out

- Literatura descreve como sendo igual ao método holdout;

- Só seria verdade se o método de holdout fosse aplicado com rotação dos dados de treino e validação em 2 rodadas:

### 3

## Validação cruzada e otimização de hiperparâmetros



- Se  $k = n$ , então refere-se a validação cruzada  $k$ -fold como validação cruzada Leave-One-Out (LOOCV);
- Modelo ajustado a  $n-1$  pessoas;
- Avaliado na pessoa remanescente;
- Processo computacionalmente caro;
- Útil para pequenos conjuntos de dados.

1 2 3 4 5 6 7 8 9 10  
1 2 3 4 5 6 7 8 9 10

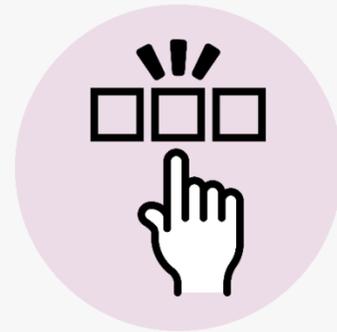
Casos especiais:  
Validação cruzada  
2-folds e leave-one-out

# 3

## Validação cruzada e otimização de hiperparâmetros



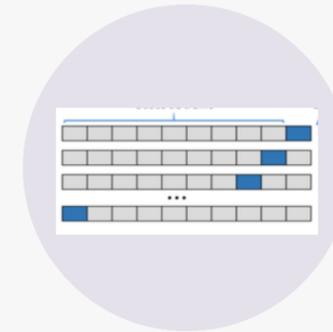
Visão geral



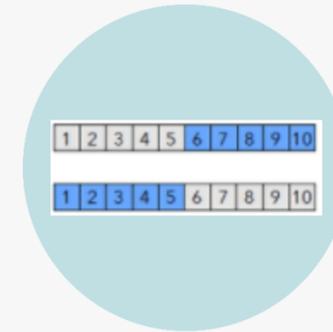
Hiperparâmetros e seleção de modelo



As 3 vias do Método Holdout para ajuste de hiperparâmetros



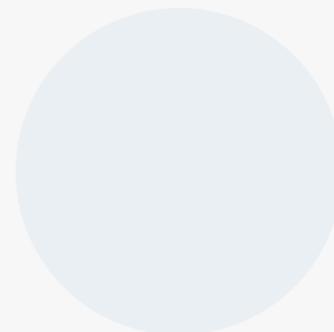
Introdução à validação cruzada k-fold



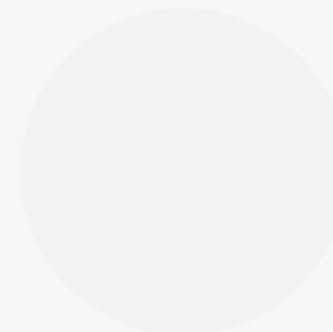
Casos especiais: Validação cruzada 2-folds e leave-one-out



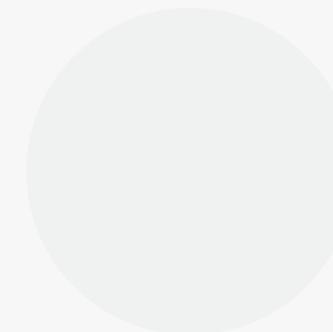
Validação cruzada k-fold e o trade-off viás x variância



Seleção de modelo via validação cruzada k-fold



Uma nota sobre seleção do modelo e grandes conjuntos de dados



Uma nota sobre a seleção de variáveis durante a seleção do modelo

### 3

## Validação cruzada e otimização de hiperparâmetros

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;



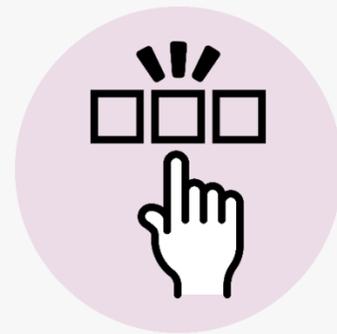
Validação cruzada  
k-fold e o trade-off  
viés x variância

# 3

## Validação cruzada e otimização de hiperparâmetros



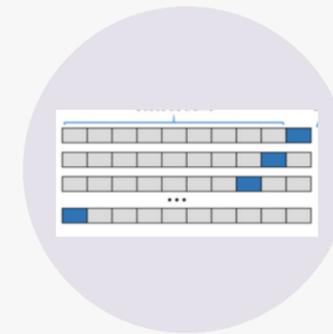
Visão geral



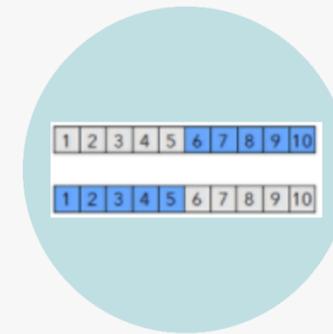
Hiperparâmetros e seleção de modelo



As 3 vias do Método Holdout para ajuste de hiperparâmetros



Introdução à validação cruzada k-fold



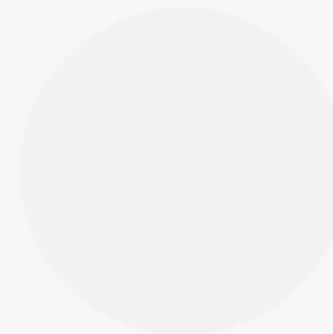
Casos especiais: Validação cruzada 2-folds e leave-one-out



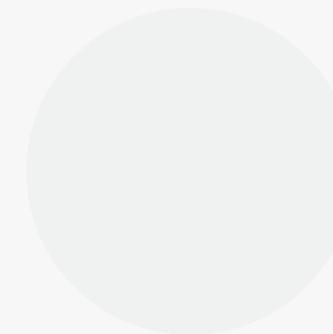
Validação cruzada k-fold e o trade-off viás x variância



Seleção de modelo via validação cruzada k-fold



Uma nota sobre seleção do modelo e grandes conjuntos de dados



Uma nota sobre a seleção de variáveis durante a seleção do modelo

### 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;

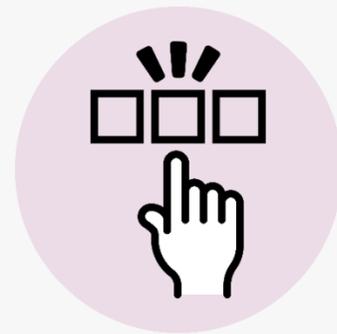
4º Utiliza-se então o conjunto de teste para avaliar o desempenho.

# 3

## Validação cruzada e otimização de hiperparâmetros



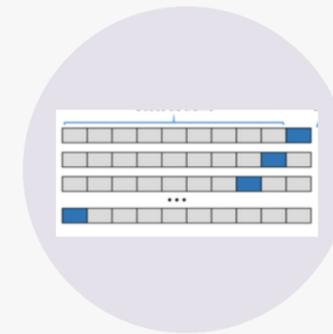
Visão geral



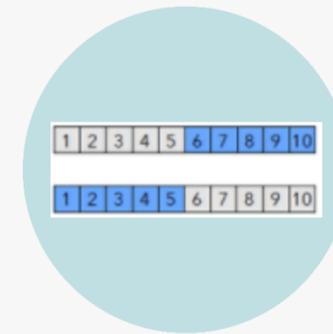
Hiperparâmetros e seleção de modelo



As 3 vias do Método Holdout para ajuste de hiperparâmetros



Introdução à validação cruzada k-fold



Casos especiais: Validação cruzada 2-folds e leave-one-out



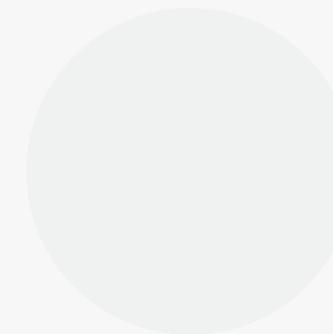
Validação cruzada k-fold e o trade-off viás x variância



Seleção de modelo via validação cruzada k-fold



Uma nota sobre seleção do modelo e grandes conjuntos de dados



Uma nota sobre a seleção de variáveis durante a seleção do modelo

### 3

## Validação cruzada e otimização de hiperparâmetros



Uma nota sobre seleção do modelo e grandes conjuntos de dados

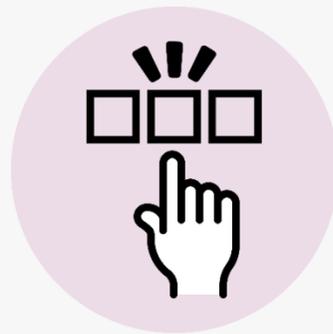
- Algoritmos de deep learning (aprendizado profundo) são utilizados em geral apenas em amostras grandes;
- Cenário onde não há preocupação com alta variância;
- O método com um treinamento, validação e teste é bom comparado a validação cruzada k-fold para a seleção de modelos se o conjunto de dados for grande (computacionalmente mais viável).

# 3

## Validação cruzada e otimização de hiperparâmetros



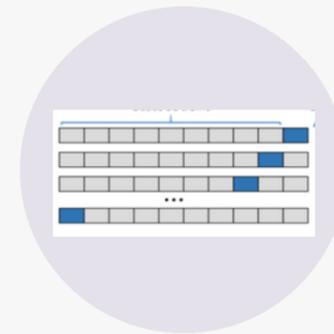
Visão geral



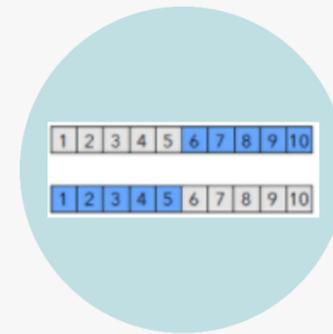
Hiperparâmetros e seleção de modelo



As 3 vias do Método Holdout para ajuste de hiperparâmetros



Introdução à validação cruzada k-fold



Casos especiais: Validação cruzada 2-folds e leave-one-out



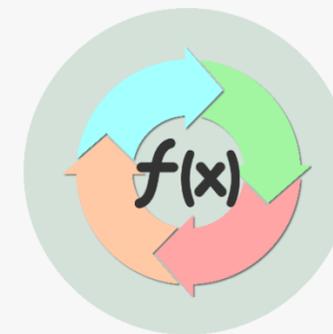
Validação cruzada k-fold e o trade-off viás x variância



Seleção de modelo via validação cruzada k-fold



Uma nota sobre seleção do modelo e grandes conjuntos de dados



Uma nota sobre a seleção de variáveis durante a seleção do modelo

### 3

## Validação cruzada e otimização de hiperparâmetros



Uma nota sobre a seleção de variáveis durante a seleção do modelo

- Normaliza-se os dados e seleciona-se variáveis dentro do loop de validação cruzada k-fold (depois de separados os dados de treinamento e teste);
- Isso reduz o overfitting;
- Pode levar a uma estimativa de desempenho pessimista

## Referência

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

# Data-driven advice for applying machine learning to bioinformatics

Randal S. Olson, William La Cava, Zairah Mustahsan, Akshay Varik, Jason H. Moore. arXiv, 2018.

Análise da performance preditiva (acurácia de VC) de 13 algoritmos em 165 bancos de dados diferentes.

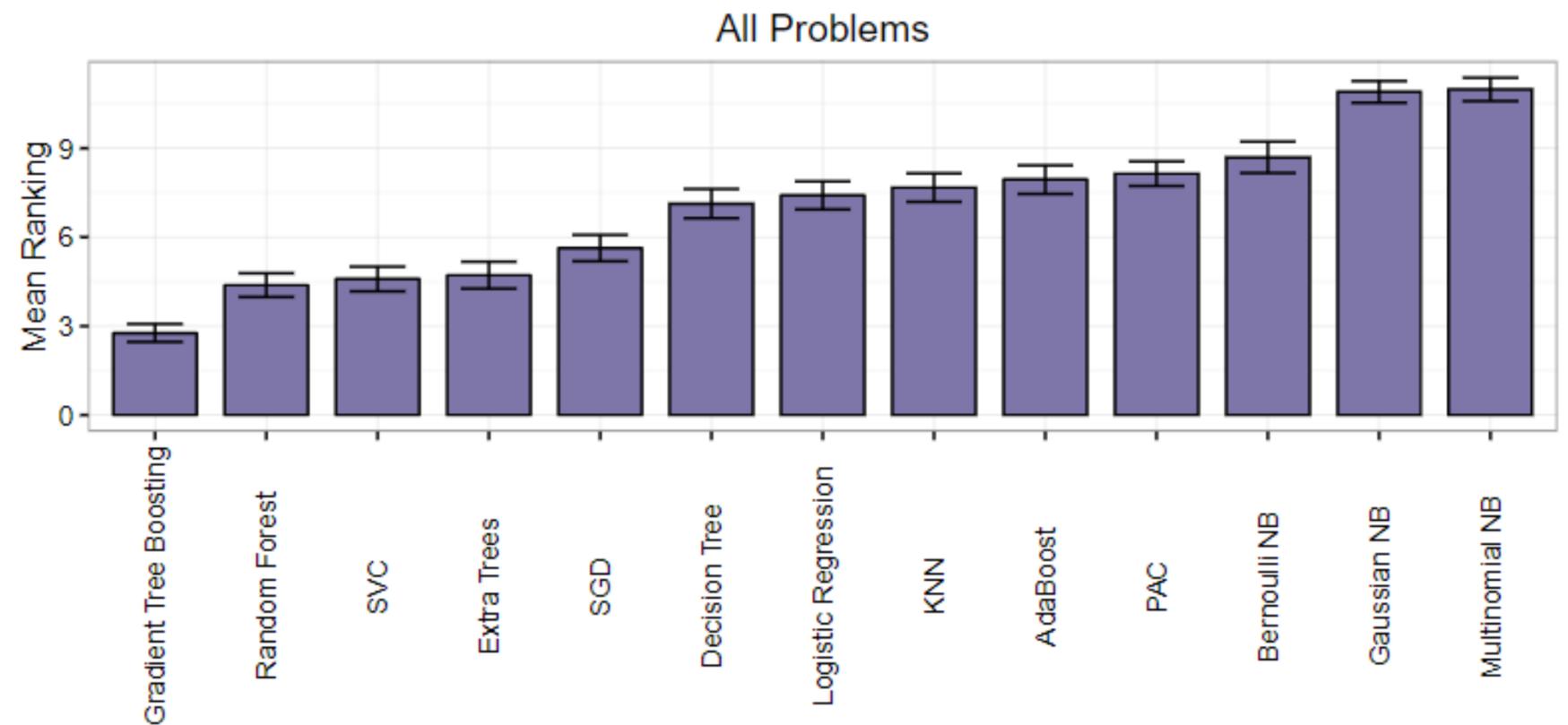


Fig. 1. Average ranking of the ML algorithms over all datasets. Error bars indicate the 95% confidence interval.

# Data-driven advice for applying machine learning to bioinformatics

Randal S. Olson, William La Cava, Zairah Mustahsan, Akshay Varik, Jason H. Moore. arXiv, 2018.

- Ganhos de performance ao selecionar hiperparâmetros via validação cruzada de 10-folds (vs. usar o hiperparâmetro default do scikit-learn): média 3-4%, em alguns casos 50%.

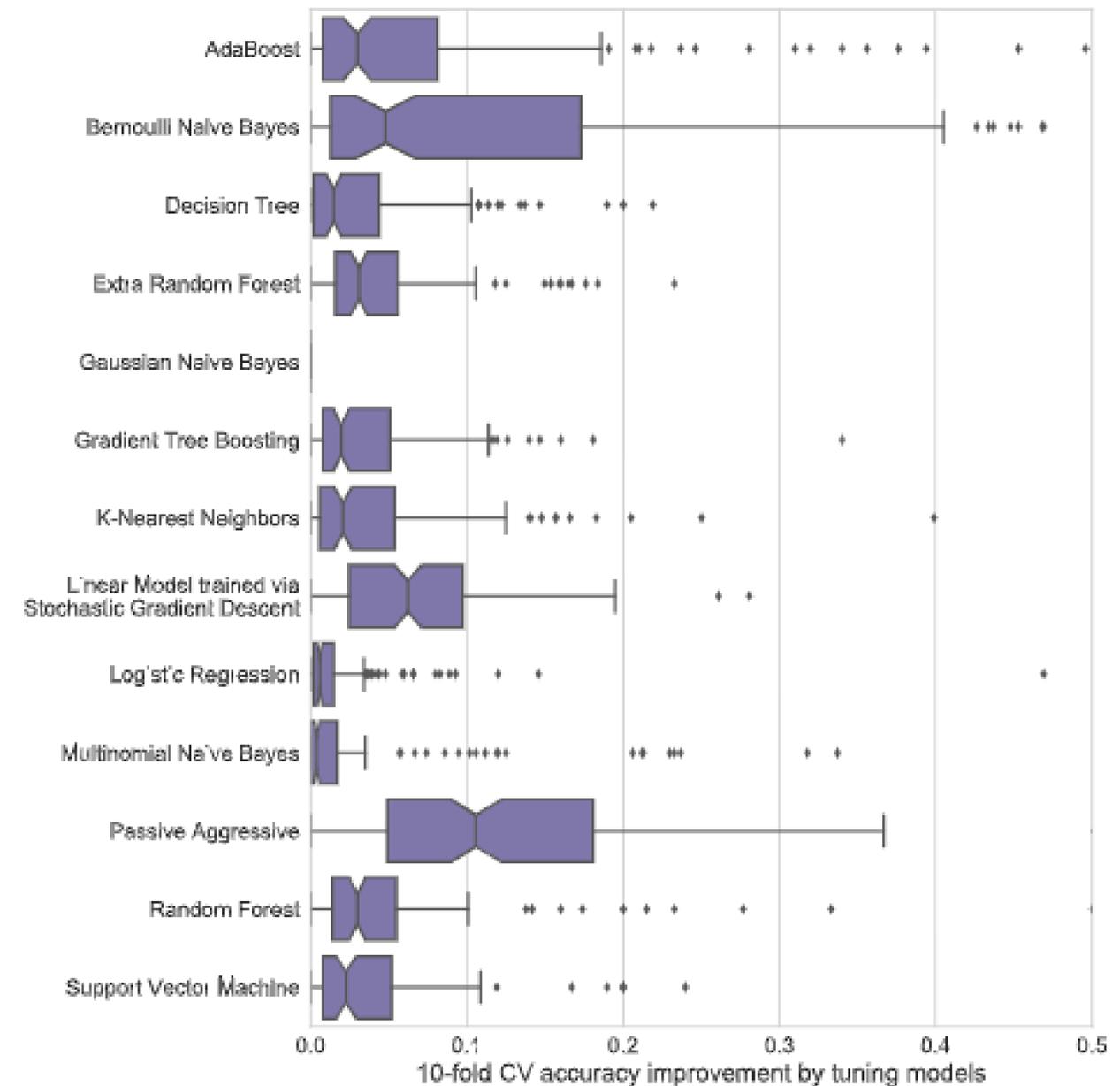


Fig 3. Improvement in 10-fold CV accuracy by tuning each ML algorithm's parameters instead of using the default parameters from scikit-learn

# Data-driven advice for applying machine learning to bioinformatics

Randal S. Olson, William La Cava, Zairah Mustahsan, Akshay Varik, Jason H. Moore.  
arXiv, 2018.

- Ganhos de performance nos 165 bancos de selecionar hiperparâmetros com VC 10-folds + seleção do melhor algoritmo (vs. performance média de todos algoritmos sem tuning): média de 20%, em alguns casos 60%.

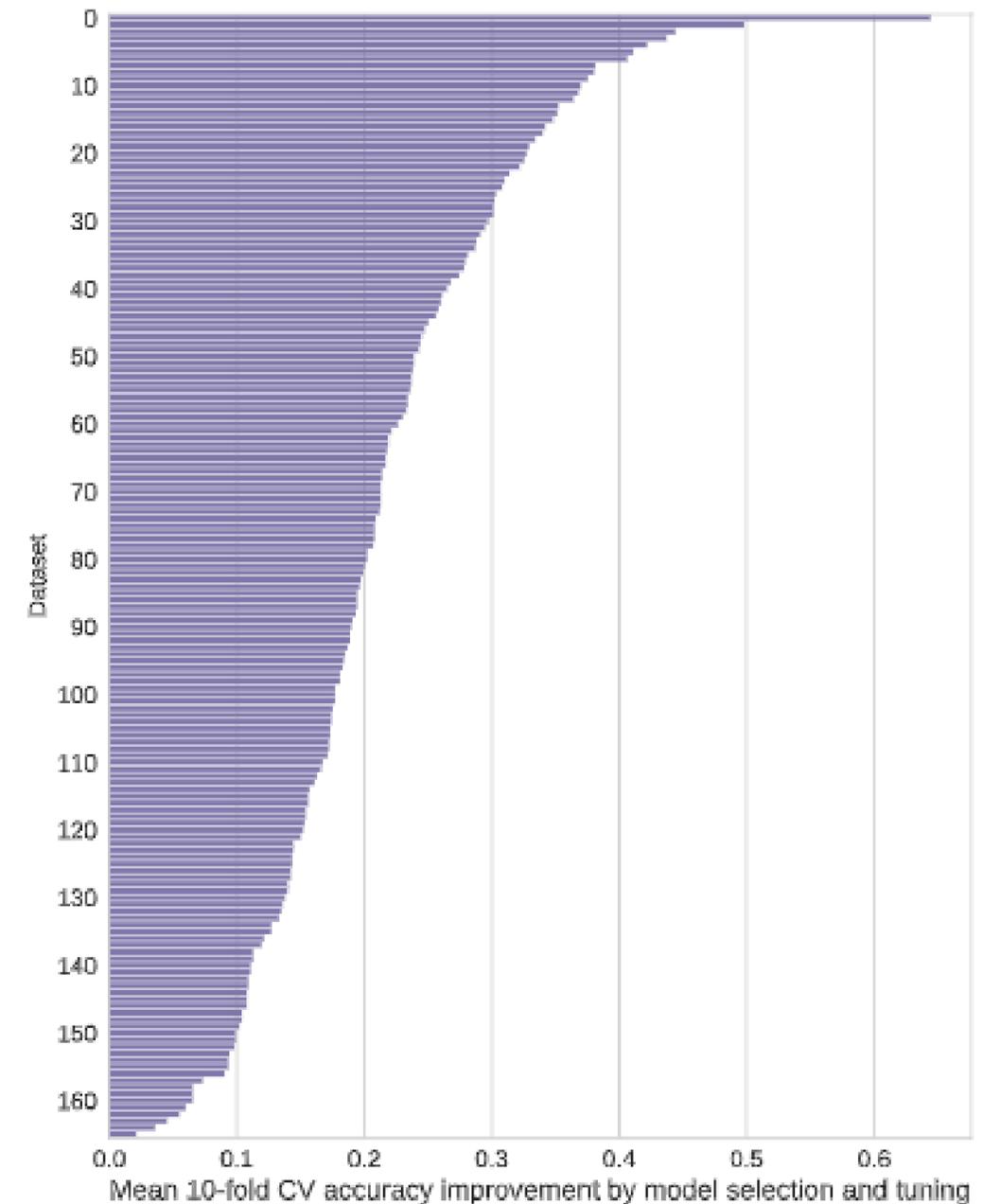


Fig 4 Improvement in 10-fold CV accuracy by model selection and tuning, relative to the average performance on each dataset