

# Introdução a Ciências de Dados

## Aula 2: Selecionando Métodos e Ajustando Modelos

Francisco A. Rodrigues  
ICMC/USP  
francisco@icmc.usp.br



# Aula 2: Selecionando Métodos e Ajustando Modelos

- Modelos preditivos
- Overfitting
- Bias-variance tradeoff
- Escolhendo modelos

# Aprendizado supervisionado

- No aprendizado supervisionado, o objetivo é ajustar um modelo preditivo a partir de um conjunto de exemplos de modo que o modelo seja capaz de prever dados não observados.
- Matematicamente, modelos preditivos são função que, dado um conjunto de exemplos rotulados, constrói um estimador.

$$y = f(X, \theta) + \epsilon$$

Erro ou ruído  
representa a  
informação que não  
está presente no  
modelo.

Estimadores podem ser vistos como funções.

# Classificação

**Definição formal:** Dado um conjunto de observações:

$$D = \{\mathbf{X}, \mathbf{y}, i = 1, \dots, N\}$$

- **f** representa uma função desconhecida (função objetivo).

$$y_i = f(X_i, \theta) + \epsilon_i$$

- Essa função mapeia as entradas nas saídas correspondentes.
- O algoritmo preditivo aprende a aproximação, que permite estimar valores de **f** para novos valores de **X**.

## Classificação

$$y_i \in \{C_1, C_2, \dots, C_n\}$$

# Classificação

## Classificação de documentos:



Sports  
Science  
News



# Regressão

**Definição formal:** Dado um conjunto de observações:

$$D = \{\mathbf{X}, \mathbf{y}, i = 1, \dots, N\}$$

- **f** representa uma função desconhecida (função objetivo).

$$y_i = f(X_i, \theta) + \epsilon_i$$

- Essa função mapeia as entradas nas saídas correspondentes.
- O algoritmo preditivo aprende a aproximação, que permite estimar valores de **f** para novos valores de **X**.

## Regressão

$$y_i \in \mathbb{R}$$

# Regressão

## Exemplo:



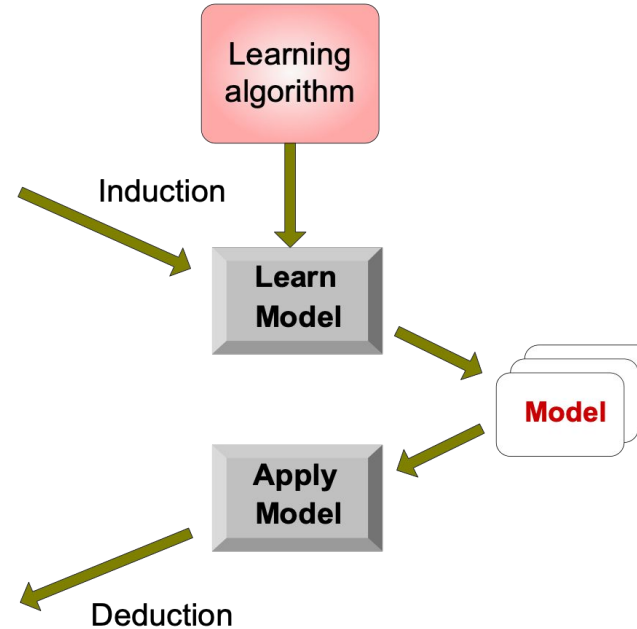
# Etapas

Tid	Attrib1	Attrib2	Attrib3	Class
1	Yes	Large	125K	No
2	No	Medium	100K	No
3	No	Small	70K	No
4	Yes	Medium	120K	No
5	No	Large	95K	Yes
6	No	Medium	60K	No
7	Yes	Large	220K	No
8	No	Small	85K	Yes
9	No	Medium	75K	No
10	No	Small	90K	Yes

Training Set

Tid	Attrib1	Attrib2	Attrib3	Class
11	No	Small	55K	?
12	Yes	Medium	80K	?
13	Yes	Large	110K	?
14	No	Small	95K	?
15	No	Large	67K	?

Test Set

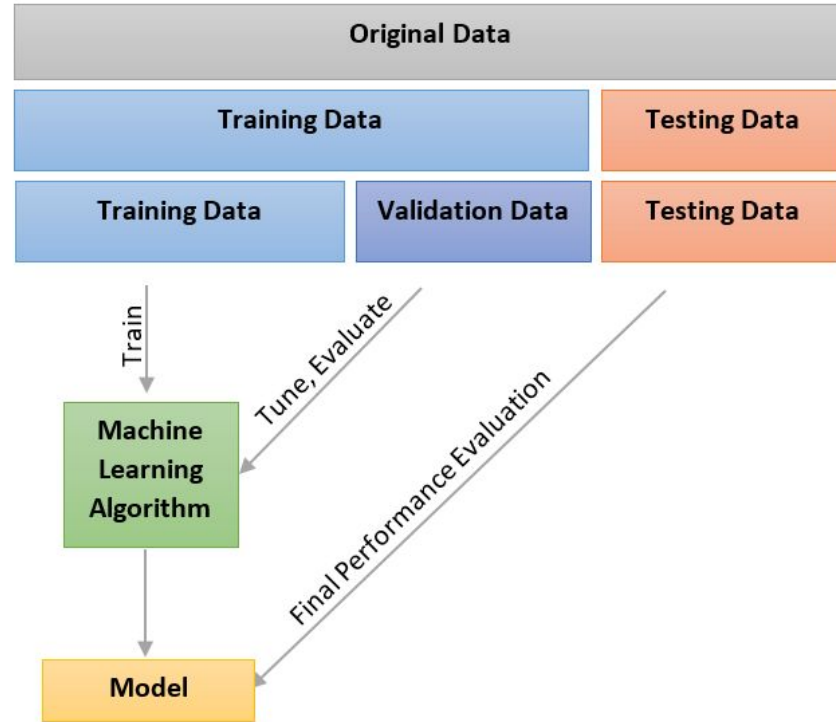




# Método hold-out

No método hold-out, os dados rotulados são divididos em dois grupos:

- $p\%$  são usados no treinamento
- $(1-p)\%$  para teste.



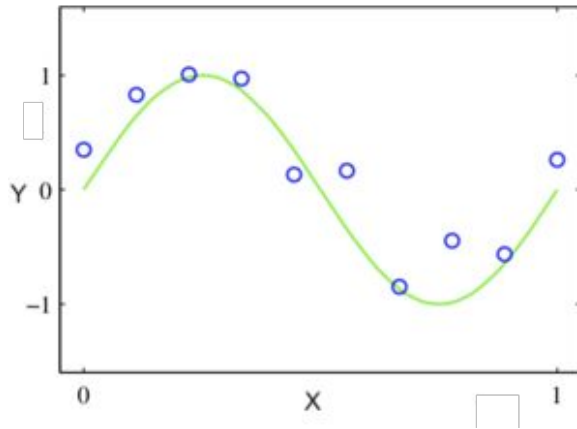
# Aprendizado de máquina

- Muitos problemas em aprendizado de máquina consideram os mesmos ingredientes:
  1. O primeiro ingrediente é um conjunto de dados  $D=(\mathbf{X}, \mathbf{y})$ , onde  $\mathbf{X}$  é uma matriz de variáveis independentes e  $\mathbf{y}$  é o vetor de variáveis dependentes.
  2. O segundo ingrediente é modelo  $f(x, \theta)$ , onde  $f$  é uma função dos parâmetros  $\theta$ .
  3. O terceiro ingrediente é a função custo  $C(\mathbf{y}, f(x, \theta))$  que permite determinar o quanto o modelo  $f$  é adequado para prever  $\mathbf{y}$ .

# Aprendizado de máquina

## Exemplo:

1.  $D(\mathbf{X}, y)$



2. Modelo

$$f(\mathbf{x}, \theta) = \sum_{j=0}^M \theta_j x^j$$

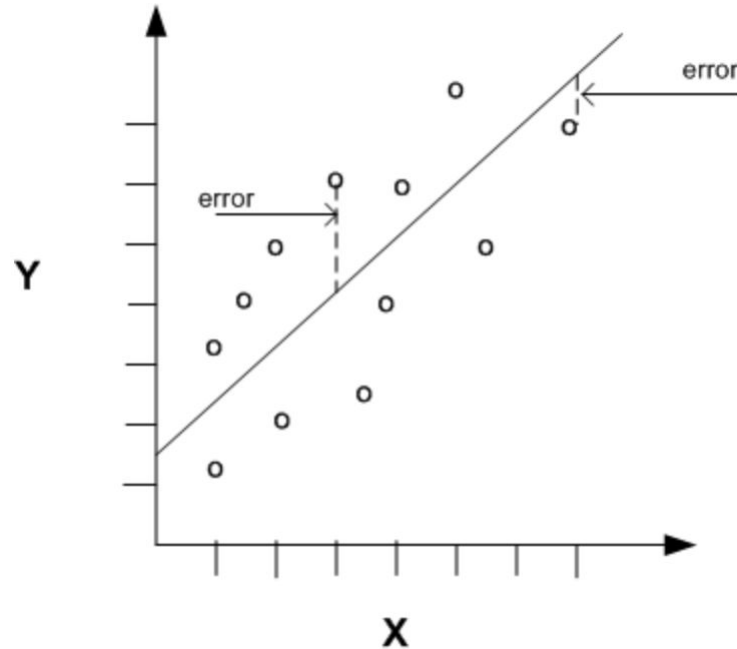
3 Função Custo

$$E(\theta) = \sqrt{\frac{1}{N} \sum_{n=1}^N \{f(x_n, \theta) - y_n\}^2}$$

# Aprendizado de máquina

## Erro quadrático médio:

$$E(\theta) = \sqrt{\frac{1}{N} \sum_{n=1}^N \{f(x_n, \theta) - y_n\}^2}$$



# Aprendizado de máquina

- O modelo é ajustado encontrando-se os valores de  $\theta$  que minimizem a função custo.
- Uma pergunta básica se refere ao ajuste dos parâmetros do modelo:

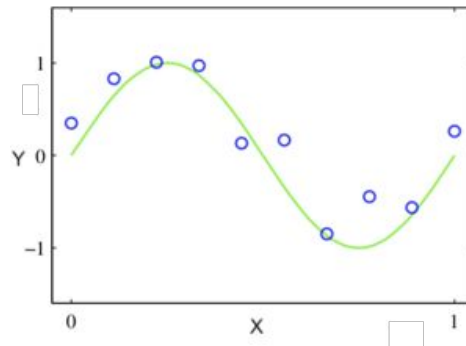
**Quais são os melhores valores dos parâmetros do modelo que permitam generalizar e prever dados desconhecidos com precisão?**



# Regressão Polinomial

- **Exemplo:**
- Os dados foram gerados a partir da função:

$$y = \sin(2\pi x) + \epsilon$$



- Onde  $\epsilon$  tem distribuição uniforme com média zero e desvio padrão  $\sigma$ .
- Vamos supor que temos acesso apenas aos pontos em azul e não conhecemos a curva original

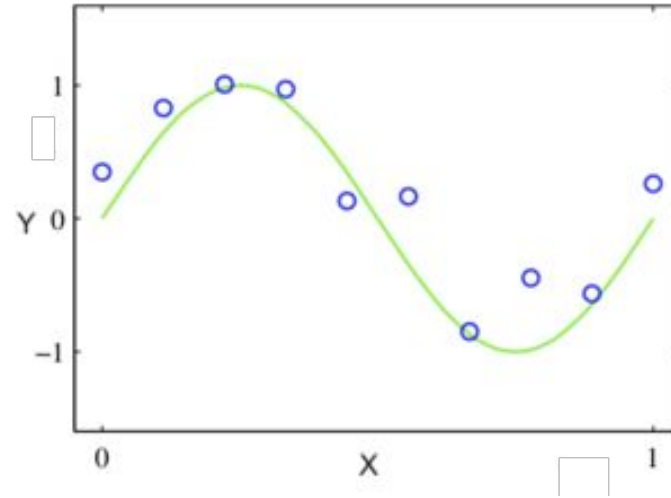
# Regressão Polinomial

- Vamos considerar algumas possíveis curvas para modelar os dados:

Polinômio de grau M:

$$f(\mathbf{x}, \theta) = \theta_0 + \theta_1 x_1 + \dots + \theta_M x_M = \sum_{j=0}^M \theta_j x^j$$

$$E(\theta) = \sqrt{\frac{1}{N} \sum_{n=1}^N \{f(x_n, \theta) - y_n\}^2}$$

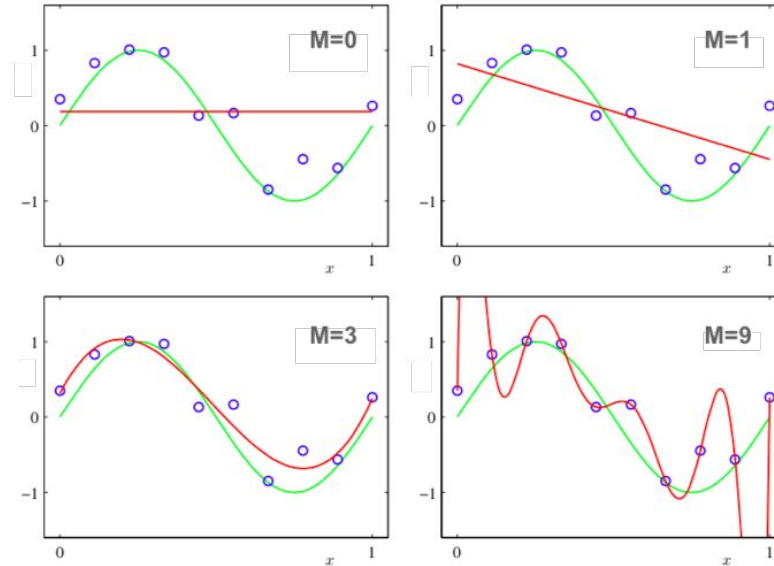


Qual o melhor valor de M?

# Regressão Polinomial

$$f(\mathbf{x}, \theta) = \theta_0 + \theta_1 x_1 + \dots + \theta_M x_M = \sum_{j=0}^M \theta_j x^j$$

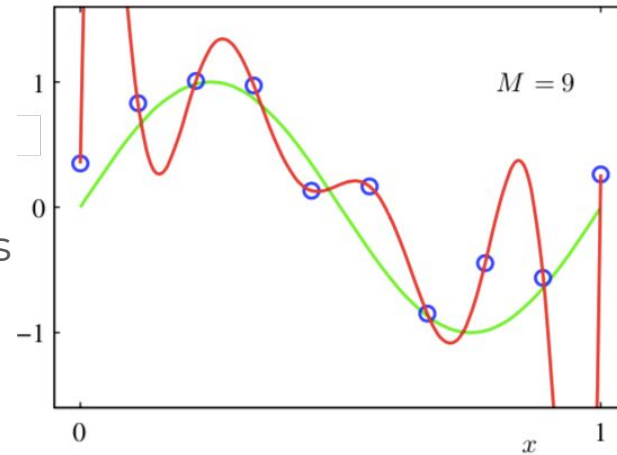
N = 10  
observações



# Overfitting

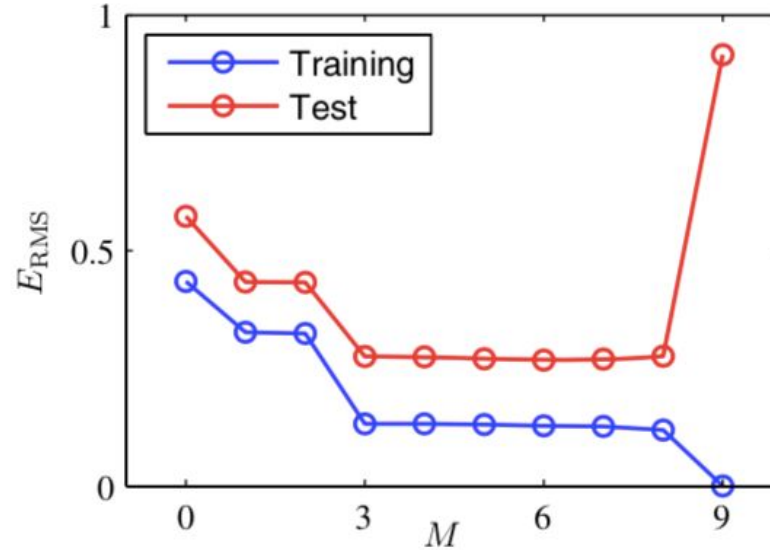
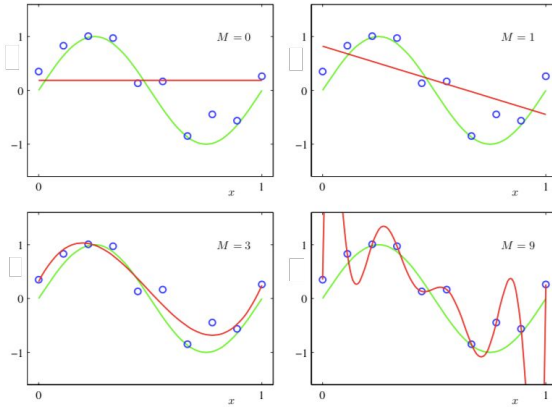
**Overfitting:** Ocorre quando um modelo se ajusta muito bem ao conjunto de dados anteriormente observado, mas se mostra ineficaz para prever novos resultados.

- Nesse exemplo, temos 10 pontos e um polinômio de grau 9.
- O modelo está super adaptado aos dados de treinamento.



# Regressão Polinomial

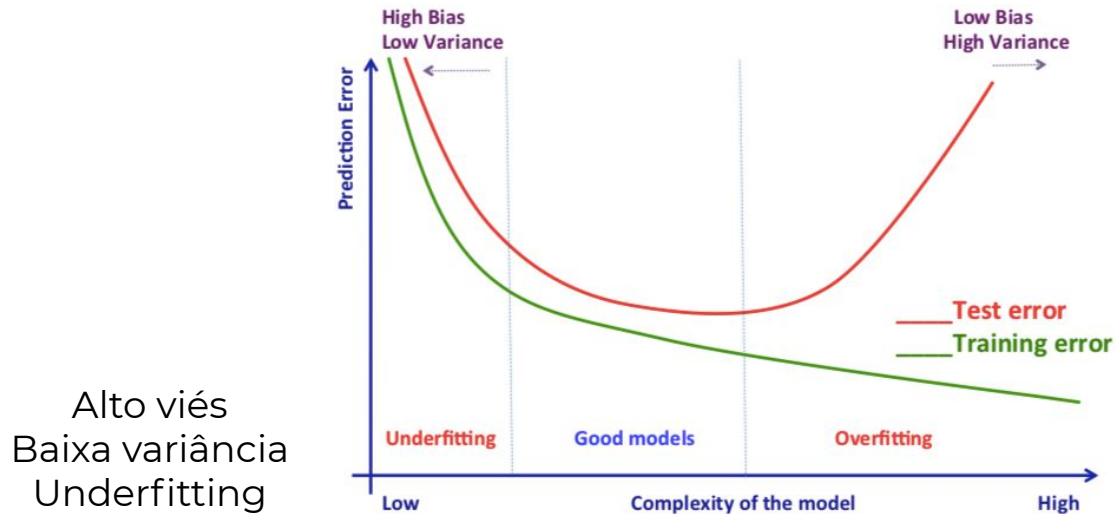
$N = 10$   
observações



- Quando aumentamos a complexidade do modelo, ocorre um super ajuste.

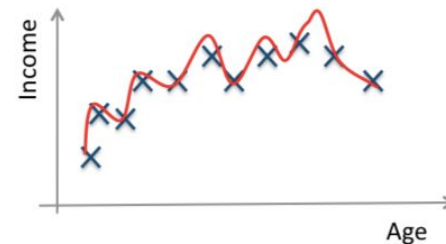
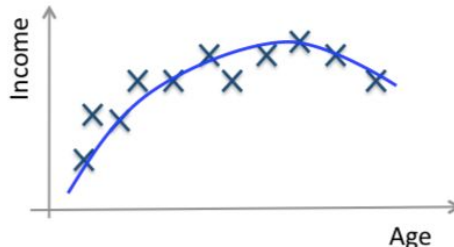
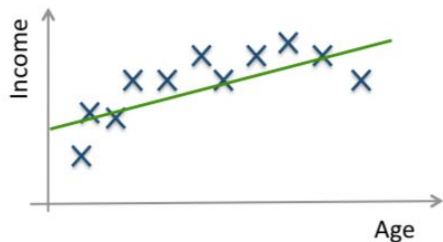


# Viés-variância (bias-variance tradeoff)



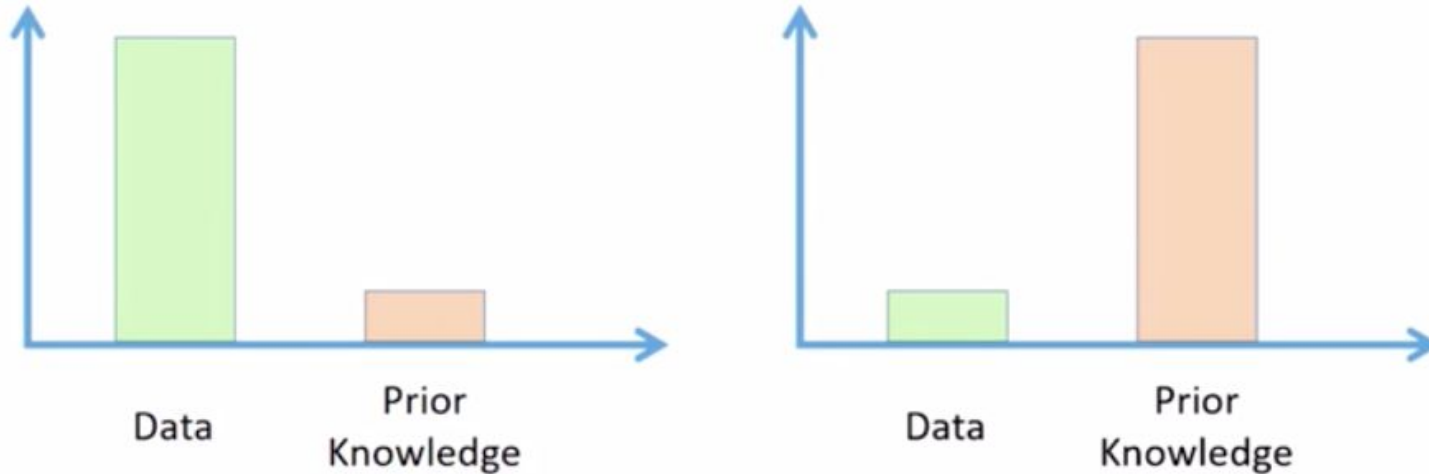
Alto viés  
Baixa variância  
Underfitting

Baixo viés  
Alta variância  
Overfitting

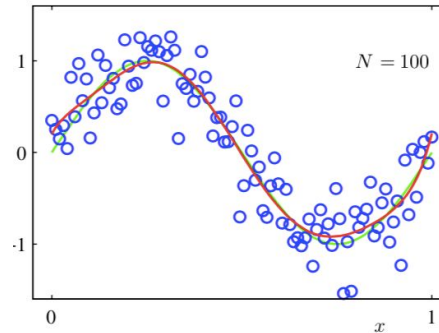
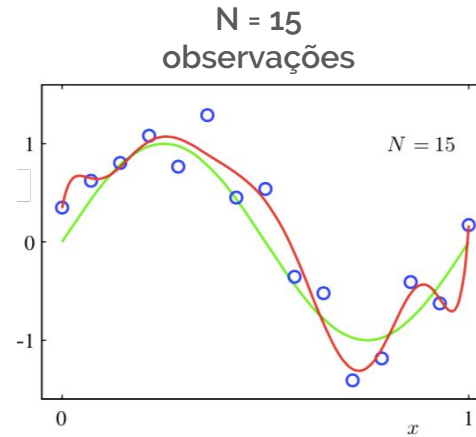
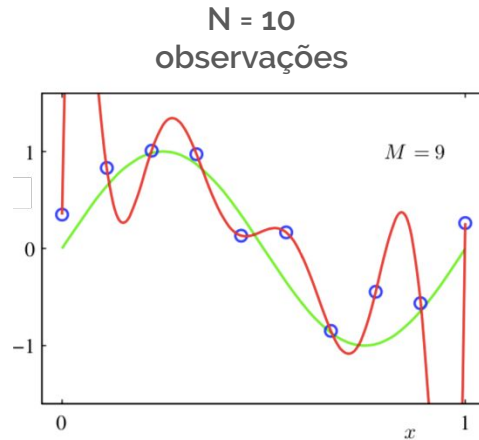


# Escolha dos Modelos

Há uma relação entre a quantidade de dados e a complexidade do modelo.



# Regressão polinomial



N = 100  
observações

# Escolha do Modelo

Há relação entre o viés e variância e isso influencia na escolha do modelo.

Assim, questões fundamentais são:

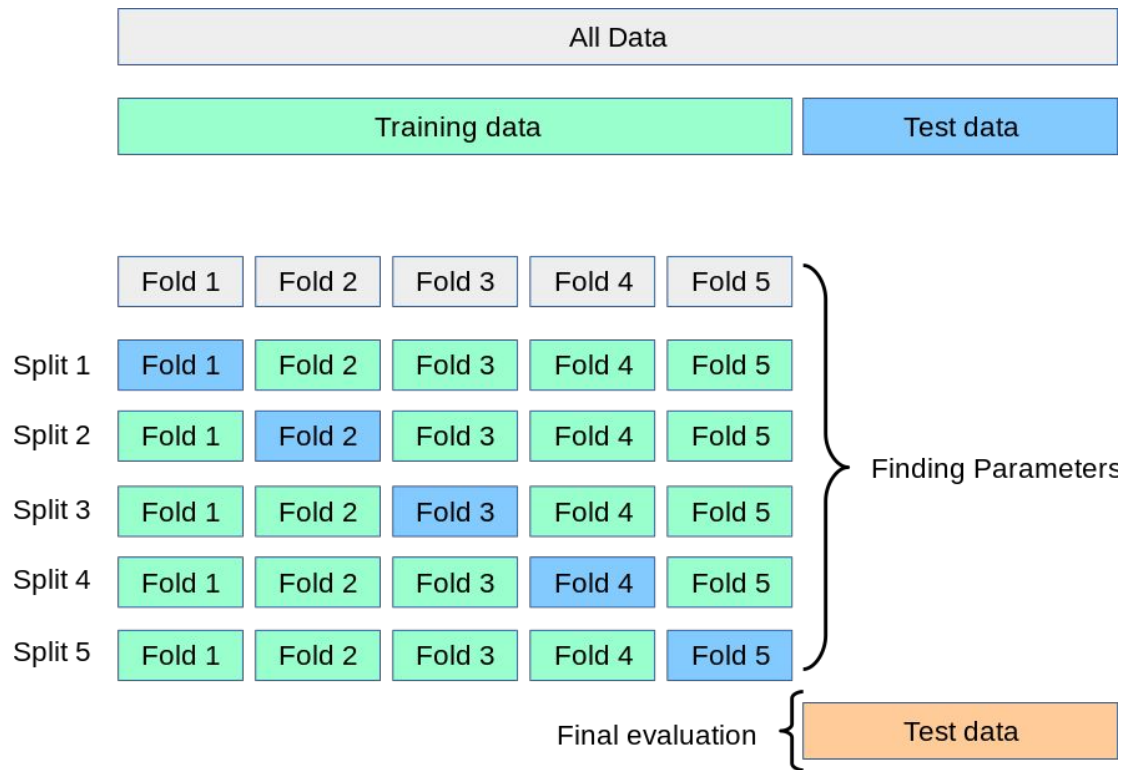
- Como escolher um modelo?
- Quais parâmetros usar?
- Como validar os modelos?

# Escolha do Modelo

- Escolhendo o melhor modelo envolve a redução da variância e viés.
- Infelizmente: não há nenhum método científico padrão para isso.
- Como escolher complexidade ótima e conseguir erro mínimo no conjunto de teste?
- Erro no treinamento não é uma boa estimativa do erro no conjunto de teste.
- Podemos usar validação cruzada.



# Validação cruzada



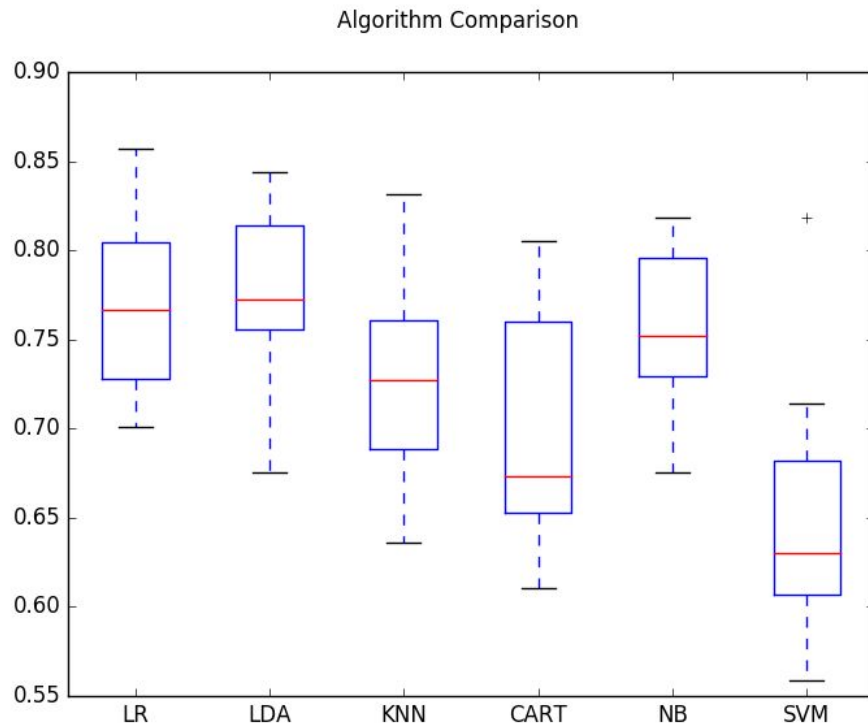
# Validação cruzada

- Na validação cruzada, **todos os dados rotulados são usados**.
- Pode haver **grande variação** nos resultados de cada classificação, pois uma fração  $1/k$  dos dados são colocadas no conjunto de teste.
- A média de todas as classificações **reduz a variância** de todo o processo.
- **Validação não serve para determinar a precisão do modelo**, mas para escolher os atributos e modelos.
- Após a validação, usamos **tudo o conjunto de dados** para ajustar o método de classificação ou regressão, para aplicar no conjunto de teste.

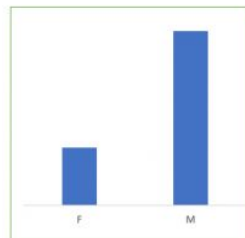
# Validação cruzada

Usamos validação para as seguintes tarefas:

- Comparar modelos.
- Escolha dos parâmetros do modelo (ex. grau do polinômio).



# Casos estratificados



Class Distributions



Manter a distribuição das classes em cada fold.

# Sumário

## Selecionando Métodos e Ajustando Modelos

- Modelos preditivos
- Overfitting
- Bias-variance tradeoff
- Escolhendo modelos



# Leitura Complementar

- Bishop, **Pattern Recognition and Machine Learning**, Springer (capítulo 1).
- Online:  
<https://www.microsoft.com/en-us/research/uploads/prod/2006/01/Bishop-Pattern-Recognition-and-Machine-Learning-2006.pdf>
- Lindholm et al., **Supervised Machine Learning**, 2019.  
[http://www.it.uu.se/edu/course/homepage/sml/literature/lecture\\_not  
es.pdf](http://www.it.uu.se/edu/course/homepage/sml/literature/lecture_notes.pdf)