

# Inteligência Artificial: Uma Abordagem de Aprendizado de Máquina

## Metodologia experimental e Análise

Prof. Tiago A. Almeida

## Motivação

- Suponha que você tenha aplicado o método de regressão linear com regularização para fazer previsões

$$J(\theta) = \frac{1}{2m} \left[ \sum_{i=1}^m (h_{\theta}(x^{(i)}) - y^{(i)})^2 + \lambda \sum_{j=1}^n \theta_j^2 \right]$$

Tiago A. Almeida

## Motivação

- Contudo, ao testar a hipótese em um novo conjunto de amostras você descobriu que o método produz **erros inaceitáveis** nas suas previsões. E agora?
  - Inserir mais amostras de treinamento ?
  - Reduzir a quantidade de atributos ?
  - Aumentar a quantidade de atributos ?
  - Adicionar atributos polinomiais ( $x_1^2, x_2^2, x_1x_2, \dots$ )?
  - Reduzir o valor de  $\lambda$ ?
  - Aumentar o valor de  $\lambda$ ?

Muito tempo pode ser perdido tentando aleatoriamente uma ou mais opções

Tiago A. Almeida

## Análise ou Diagnóstico

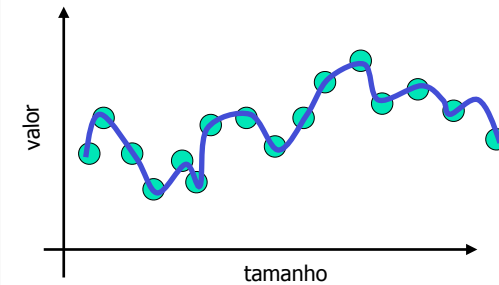
- Diagnóstico:** teste para estimar o que está correto/incorreto com o método de aprendizado de máquina e descobrir qual é a melhor estratégia para aumentar o desempenho.
- A análise pode ser custosa (em termos de implementação), porém pode poupar muito tempo de tentativas erradas.

Tiago A. Almeida

# Inteligência Artificial: Uma Abordagem de Aprendizado de Máquina

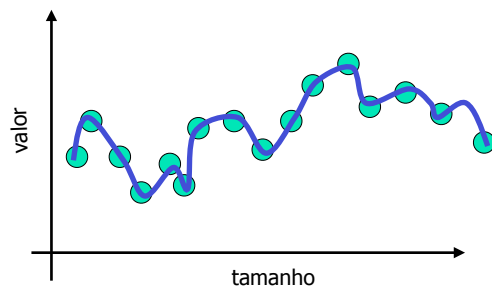
## Avaliando uma hipótese

## Avaliando uma hipótese



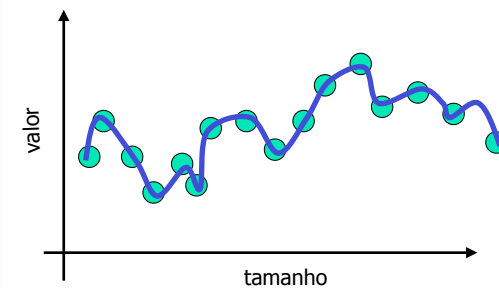
Hipótese resultante produz baixo erro de treinamento:  $J$

## Avaliando uma hipótese



**Overfitting:**  
não consegue generalizar exemplos não treinados

## Avaliando uma hipótese



Como identificar *overfitting* e *underfitting* quando não é possível plotar?

# Avaliando uma hipótese

## Base de dados:

Tamanho	Valor
2104	400
1600	330
2400	369
1416	232
3000	540
1985	300
1534	315
1427	199
1380	212
1494	243

# Avaliando uma hipótese

## Base de dados:

Tamanho	Valor
2104	400
1600	330
2400	369
1416	232
3000	540
1985	300
1534	315
1427	199
1380	212
1494	243

70% - Treinamento

30% - Teste

Divisão da base em duas partes: treinamento e teste

# Avaliando uma hipótese

## Base de dados:

Tamanho	Valor
2104	400
1600	330
2400	369
1416	232
3000	540
1985	300
1534	315
1427	199
1380	212
1494	243

70% - Treinamento

30% - Teste

$(x^{(1)}, y^{(1)})$   
 $(x^{(2)}, y^{(2)})$   
 $\vdots$   
 $(x^{(m)}, y^{(m)})$

$(x_{test}^{(1)}, y_{test}^{(1)})$   
 $(x_{test}^{(2)}, y_{test}^{(2)})$   
 $\vdots$   
 $(x_{test}^{(m_{test})}, y_{test}^{(m_{test})})$

# Avaliando uma hipótese

## Procedimento de treinamento e teste

- Calcula o parâmetro  $\theta$  a partir da base de treinamento (minimizando o erro de treinamento  $J(\theta)$ )
- Calcula a taxa de erro/acerto no conjunto de teste  $J(\theta)$

Observe que a função custo varia de método para método

## Avaliando uma hipótese

- Validação cruzada com  $k$ -partições
  - Escolhe  $k$  e divide a base de dados em  $k$  partes
  - Treina com  $k-1$  partes e testa com 1 parte:
    - Calcula o parâmetro  $\theta$  a partir da base de treinamento (minimizando o erro de treinamento  $J(\theta)$ )
    - Calcula a tx de erro/acerto no conjunto de teste  $J(\theta)$
  - Repete até ter testado com todas as  $k$  partes
  - Calcula média e desvio padrão das taxas de erros/acertos

## Validação cruzada

- $k$ -fold cross-validation

Tamanho	Valor
2104	400
1600	330
2400	369
1416	232
3000	540
1985	300
1534	315
1427	199
1380	212
1494	243

## Inteligência Artificial: Uma Abordagem de Aprendizado de Máquina

### Seleção de modelo

Prof. Tiago A. Almeida

## Seleção de modelo

### Modelos

- 1.  $h_{\theta}(x) = \theta_0 + \theta_1 x$
- 2.  $h_{\theta}(x) = \theta_0 + \theta_1 x + \theta_2 x^2$
- 3.  $h_{\theta}(x) = \theta_0 + \theta_1 x + \dots + \theta_3 x^3$
- $\vdots$
- 10.  $h_{\theta}(x) = \theta_0 + \theta_1 x + \dots + \theta_{10} x^{10}$

**Modelos:** possíveis graus do polinômio que forma a hipótese e/ou diferentes valores do parâmetro de regularização

## Seleção de modelo

### Modelos

- 1.  $h_{\theta}(x) = \theta_0 + \theta_1 x$
- 2.  $h_{\theta}(x) = \theta_0 + \theta_1 x + \theta_2 x^2$
- 3.  $h_{\theta}(x) = \theta_0 + \theta_1 x + \dots + \theta_3 x^3$
- $\vdots$
- 10.  $h_{\theta}(x) = \theta_0 + \theta_1 x + \dots + \theta_{10} x^{10}$

Treina com cada modelo e avalia no conjunto de teste. Escolhe modelo com menor taxa de erro.

## Seleção de modelo

### Modelos

- 1.  $h_{\theta}(x) = \theta_0 + \theta_1 x$
- 2.  $h_{\theta}(x) = \theta_0 + \theta_1 x + \theta_2 x^2$
- 3.  $h_{\theta}(x) = \theta_0 + \theta_1 x + \dots + \theta_3 x^3$
- $\vdots$
- 10.  $h_{\theta}(x) = \theta_0 + \theta_1 x + \dots + \theta_{10} x^{10}$

Modelo escolhido. Como avaliar se o conjunto de teste foi usado para escolher o modelo?

## Seleção de modelo

### Base de dados:

Tamanho	Valor
2104	400
1600	330
2400	369
1416	232
3000	540
1985	300
1534	315
1427	199
1380	212
1494	243

Divisão da base em três partes: treinamento, validação e teste

60% - Treinamento

20% - Validação

20% - Teste

$(x^{(1)}, y^{(1)})$   
 $(x^{(2)}, y^{(2)})$   
 $\vdots$   
 $(x^{(m)}, y^{(m)})$   


---

 $(x_{cv}^{(1)}, y_{cv}^{(1)})$   
 $(x_{cv}^{(2)}, y_{cv}^{(2)})$   
 $\vdots$   
 $(x_{cv}^{(m_{cv})}, y_{cv}^{(m_{cv})})$   


---

 $(x_{test}^{(1)}, y_{test}^{(1)})$   
 $(x_{test}^{(2)}, y_{test}^{(2)})$   
 $\vdots$   
 $(x_{test}^{(m_{test})}, y_{test}^{(m_{test})})$

## Seleção de modelo

### Procedimento de seleção e teste

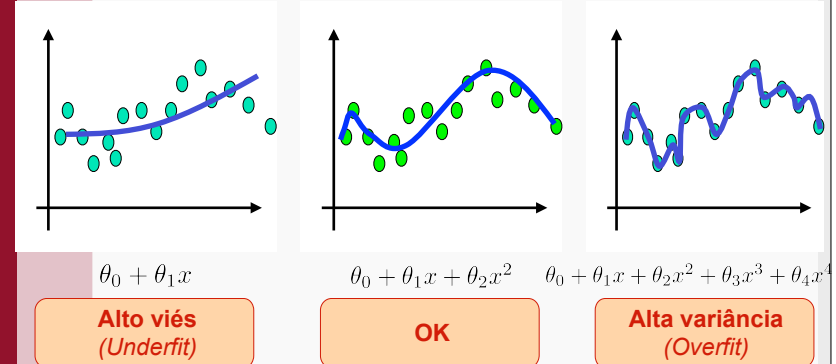
- Calcular o parâmetro  $\theta$  a partir da base de treinamento (minimizando o erro de treinamento  $J_{train}(\theta)$ )
- Calcular o erro/acerto no conjunto de validação cruzada  $J_{cv}(\theta)$  e selecionar o modelo
- Calcular o erro/acerto do modelo no conjunto de teste  $J_{test}(\theta)$

# Inteligência Artificial: Uma Abordagem de Aprendizado de Máquina

## Viés x Variância

Prof. Tiago A. Almeida

## Viés x Variância

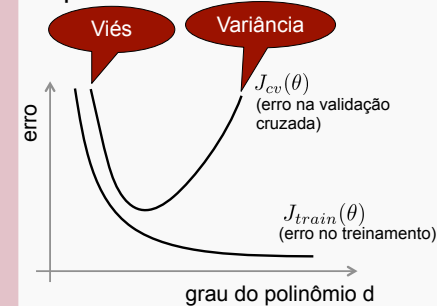


## Viés x Variância

- Suponha que o seu algoritmo apresenta desempenho insatisfatório ( $J_{cv}(\theta)$  ou  $J_{test}(\theta)$  é alto). Será que é por causa de um problema de *viés* ou de *variância*?

## Viés x Variância

- Suponha que o seu algoritmo apresenta desempenho insatisfatório ( $J_{cv}(\theta)$  ou  $J_{test}(\theta)$  é alto). Será que é por causa de um problema de *viés* ou de *variância*?



### Viés (underfit):

- $J_{train}(\theta)$  é alto
- $J_{cv}(\theta) \approx J_{train}(\theta)$

### Variância (overfit):

- $J_{train}(\theta)$  é baixo
- $J_{cv}(\theta) \gg J_{train}(\theta)$

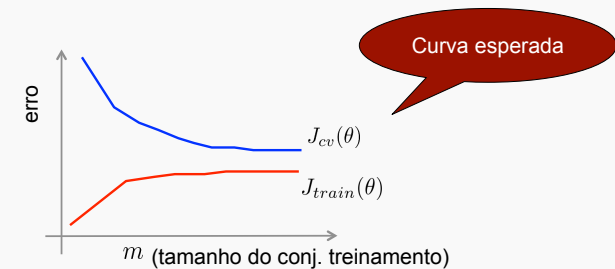
# Inteligência Artificial: Uma Abordagem de Aprendizado de Máquina

## Curvas de aprendizado

Prof. Tiago A. Almeida

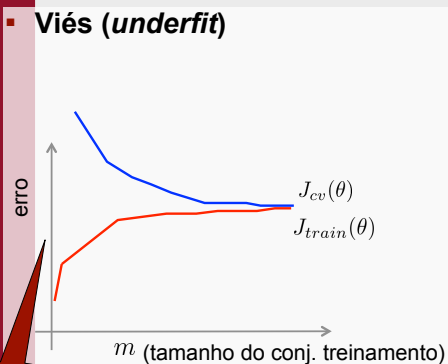
## Curvas de aprendizado

- Gráfico de custo (erro) nas partições de treino e validação (ou teste) por número de amostras de treinamento

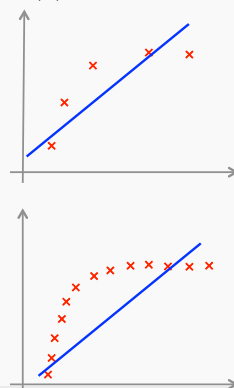


## Curvas de aprendizado

### Viés (*underfit*)

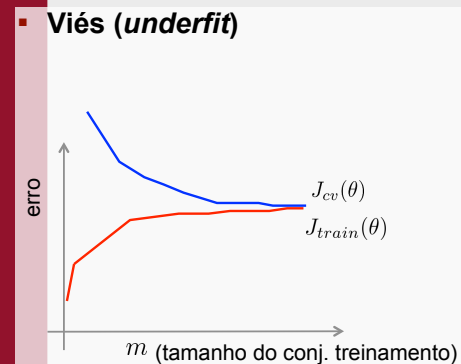


$$h_{\theta}(x) = \theta_0 + \theta_1 x$$

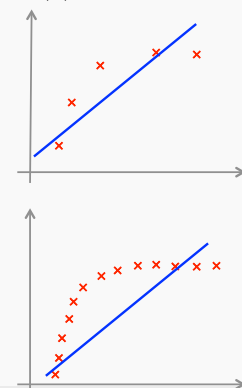


## Curvas de aprendizado

### Viés (*underfit*)



$$h_{\theta}(x) = \theta_0 + \theta_1 x$$



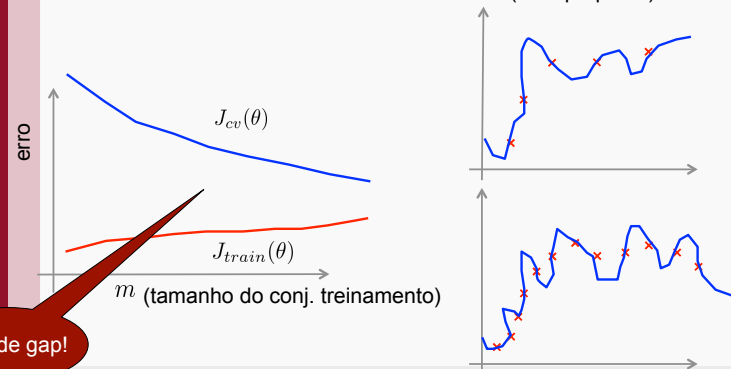
Aumentar o tamanho da base de treinamento não ajudará muito.

## Curvas de aprendizado

### Variância (overfit)

$$h_{\theta}(x) = \theta_0 + \theta_1 x + \dots + \theta_{100} x^{100}$$

(e  $\lambda$  pequeno)

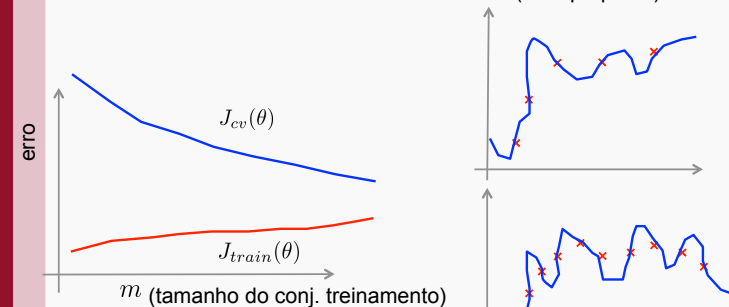


## Curvas de aprendizado

### Variância (overfit)

$$h_{\theta}(x) = \theta_0 + \theta_1 x + \dots + \theta_{100} x^{100}$$

(e  $\lambda$  pequeno)



Aumentar o tamanho da base de treinamento poderá ajudar.

## Inteligência Artificial: Uma Abordagem de Aprendizado de Máquina

### Decisões

Prof. Tiago A. Almeida

## Decisões

- Suponha que você tenha aplicado o método de regressão linear com regularização para fazer previsões
- Contudo, ao testar a hipótese em um novo conjunto de amostras você descobriu que o método produz **erros inaceitáveis** nas suas previsões. E agora?
  - Inserir mais amostras de treinamento (**pode resolver alta variância**)
  - Reduzir a quantidade de atributos (**pode resolver alta variância**)
  - Aumentar a quantidade de atributos (**pode resolver alto viés**)
  - Adicionar atributos polinomiais ( $x_1^2, x_2^2, x_1 x_2, \dots$ ) (**pode resolver alto viés**)
  - Reduzir o valor de  $\lambda$  (**pode resolver alto viés**)
  - Aumentar o valor de  $\lambda$  (**pode resolver alta variância**)