

# Aprendizagem de Máquina: Questões práticas

Prof.: Eric A. Antonelo

Slides baseados no curso de *Machine Learning*  
de Andrew Ng

DAS-UFSC

# Validação Cruzada

Conjunto de dados:

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

# Validação Cruzada

Conjunto de dados:

Size	Price	
2104	400	Conjunto Treinamento
1600	330	
2400	369	
1416	232	
3000	540	
1985	300	
1534	315	Conjunto de validação
1427	199	
1380	212	Teste
1494	243	

# Seleção de Modelos com validação cruzada

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.  $h_{\theta}(x) = \theta_0 + \theta_1 x + \dots + \theta_{10} x^{10}$

Treinar com o **conjunto de treinamento**

e avaliar no **conjunto de validação**.

Escolher o modelo com menor erro de validação.

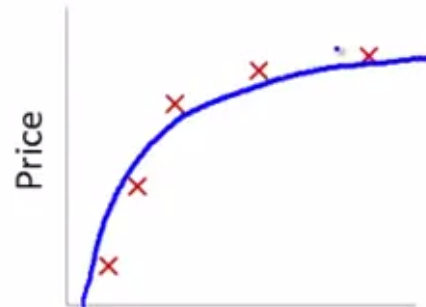
E estimar o erro de generalização/teste deste modelo, usando o **conjunto de teste**.

# Bias/Variância



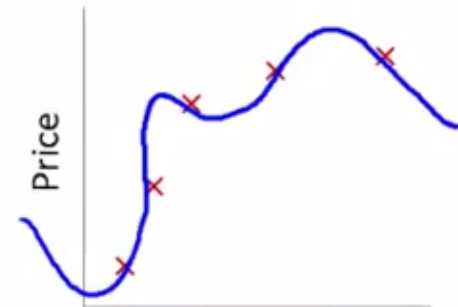
Size  
 $\theta_0 + \theta_1 x$

High bias  
(underfit)



Size  
 $\theta_0 + \theta_1 x + \theta_2 x^2$

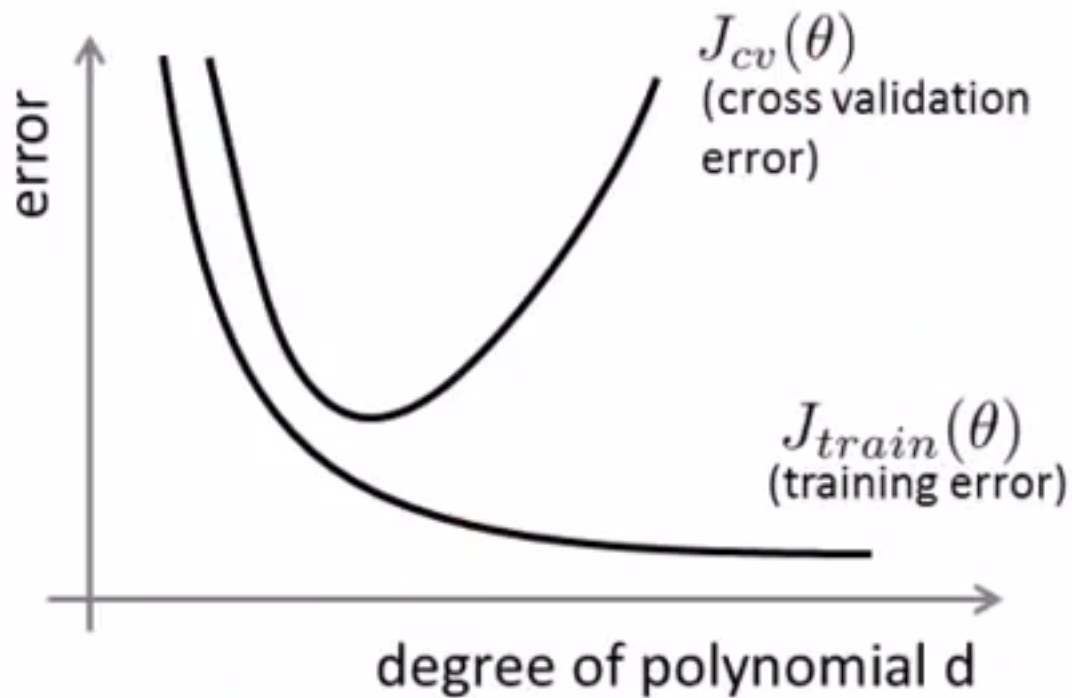
"Just right"



Size  
 $\theta_0 + \theta_1 x + \theta_2 x^2 + \theta_3 x^3 + \theta_4 x^4$

High variance  
(overfit)

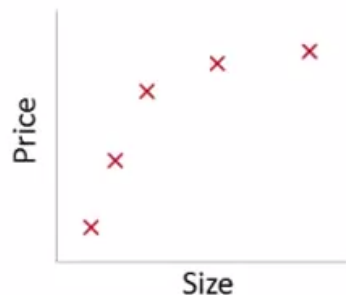
# Bias/Variância



# Parâmetro de regularização

$$h_{\theta}(x) = \theta_0 + \theta_1 x + \theta_2 x^2 + \theta_3 x^3 + \theta_4 x^4$$

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

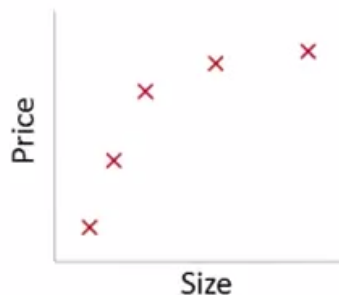


Large  $\lambda$

High bias (underfit)

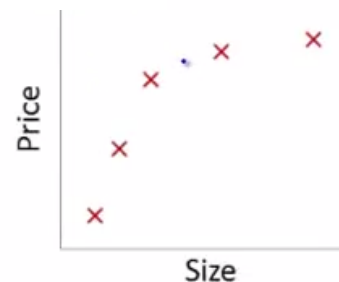
$\lambda = 10000$ .  $\theta_1 \approx 0, \theta_2 \approx 0, \dots$

$h_{\theta}(x) \approx \theta_0$



Intermediate  $\lambda$

"Just right"



Small  $\lambda$

High variance (overfit)

# Parâmetro de regularização

$$h_{\theta}(x) = \theta_0 + \theta_1 x + \theta_2 x^2 + \theta_3 x^3 + \theta_4 x^4$$

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

$$\lambda = 0$$

$$\lambda = 0.01$$

$$\lambda = 0.02$$

$$\lambda = 0.04$$

$$\lambda = 0.08$$

$$\lambda = 10$$

Calcular erro de  
validação:

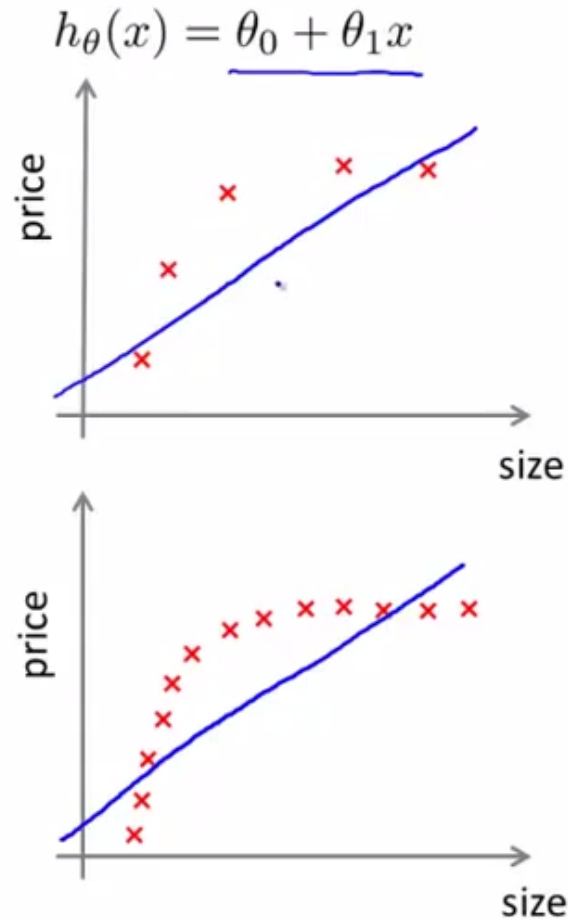
$$J_{cv}(\theta) = \frac{1}{2m_{cv}} \sum_{i=1}^{m_{cv}} (h_{\theta}(x_{cv}^{(i)}) - y_{cv}^{(i)})^2$$

Escolher o valor que minimiza o  
erro de validação.



# Curvas de Aprendizagem

Bias Alto

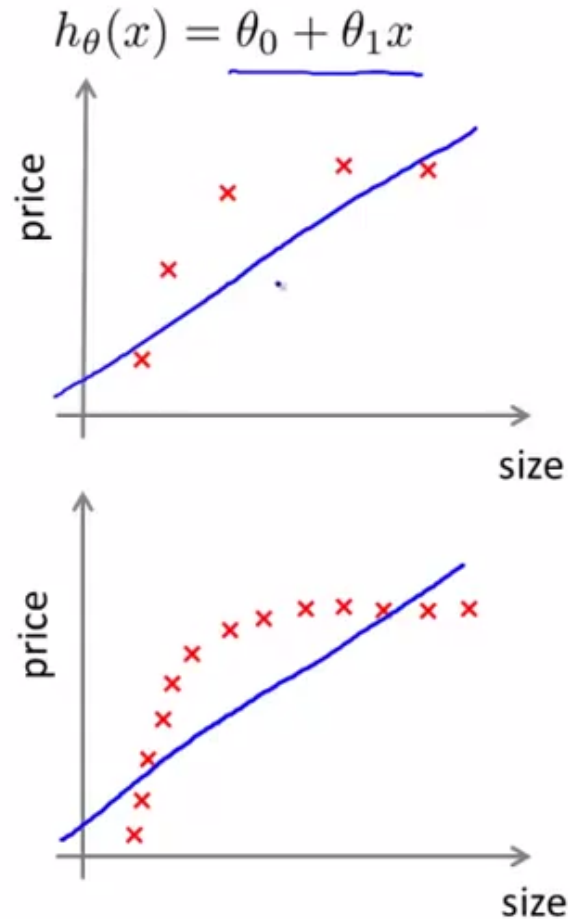


*Gráfico no quadro*

# Curvas de Aprendizagem

Bias Alto

Não adianta  
adicionar  
mais dados



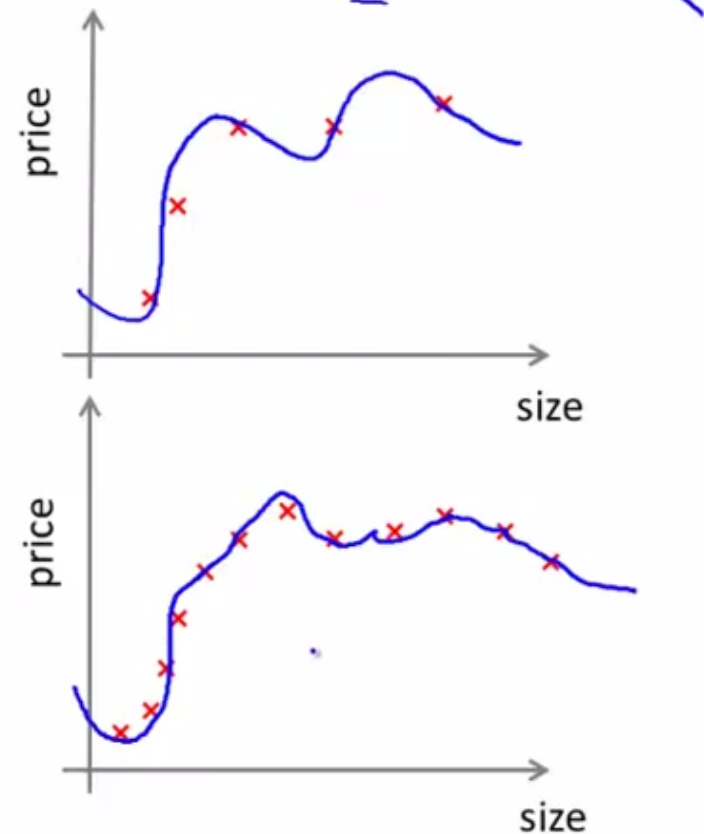
# Curvas de Aprendizagem

Variância  
Alta

*Gráfico no quadro*

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

(and small  $\lambda$ )



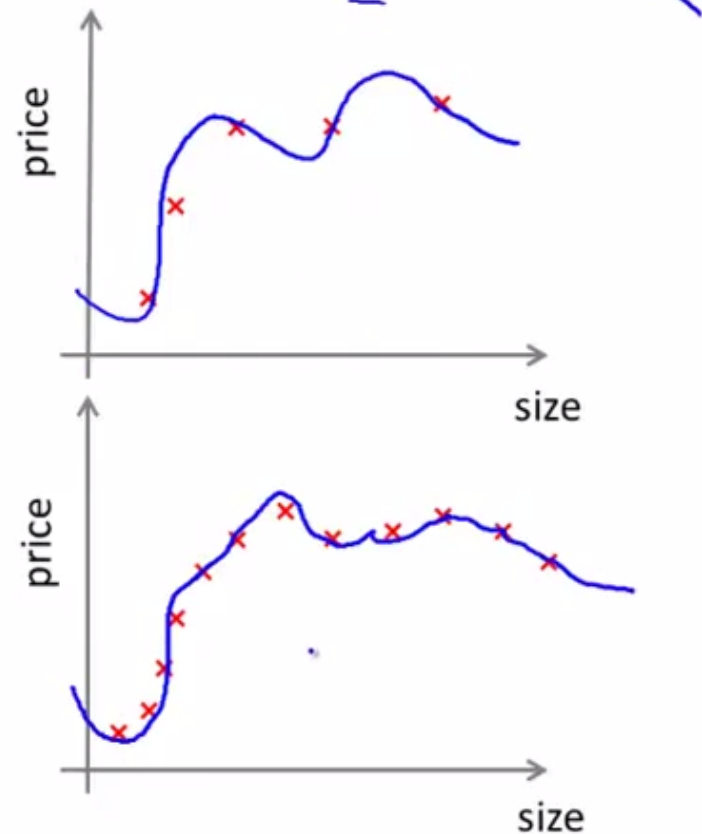
# Curvas de Aprendizagem

Variância  
Alta

Vale a pena  
adicionar  
mais dados

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

(and small  $\lambda$ )



# O que fazer quando o erro de teste é grande?

- Coletar mais dados
- Tentar conjunto menor de variáveis/termos
- Tentar obter termos/variáveis adicionais
- Tentar adicionar termos polinomiais
- Tentar diminuir  $\lambda$
- Tentar aumentar  $\lambda$

# O que fazer quando o erro de teste é grande?

## RESOLVE

- Coletar mais dados
- Tentar conjunto menor de variáveis/termos
- Tentar obter termos/variáveis adicionais
- Tentar adicionar termos polinomiais
- Tentar diminuir  $\lambda$
- Tentar aumentar  $\lambda$

Variância alta

Variância alta

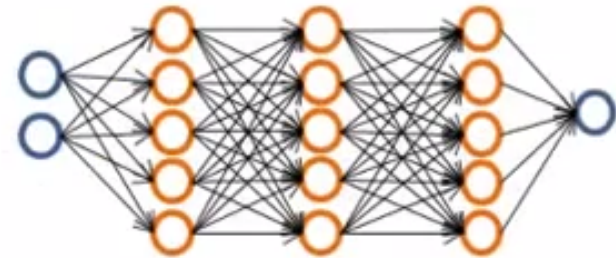
Bias Alto

Bias Alto

Bias Alto

Variância alta

# Redes neurais e “overfitting”



# Redes neurais e “overfitting”

Redes “pequenas”;  
menos parâmetros; mais  
fácil ocorrer “**underfitting**”



Computacionalmente  
mais barato

Redes “grandes”;  
mais parâmetros; mais  
fácil ocorrer “**overfitting**”



Computacionalmente  
mais caro

