

Aprendizagem de Máquina

O que é Aprendizagem de Máquina?

"The field of study that gives computers the ability to learn without being explicitly programmed." (Arthur Samuel)

Algumas Aplicações

Reconhecimento de faces

Reconhecimento de voz

Reconhecimento de spam

Natural Language Processing

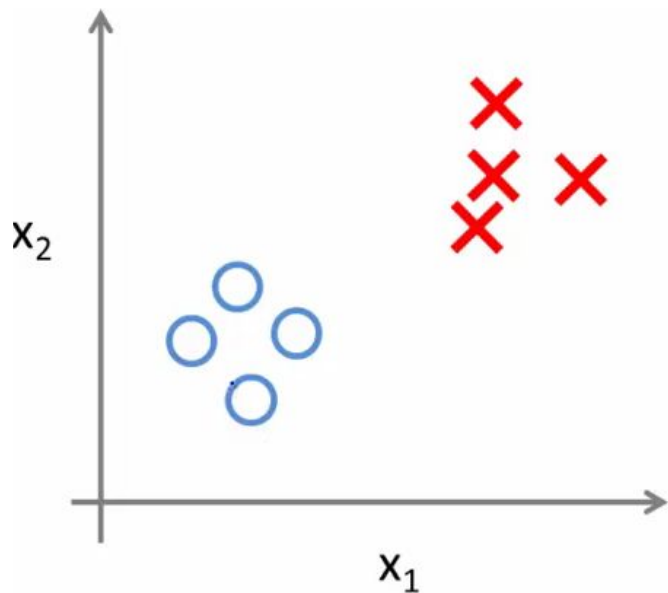
Computer Vision

Sistemas de recomendação

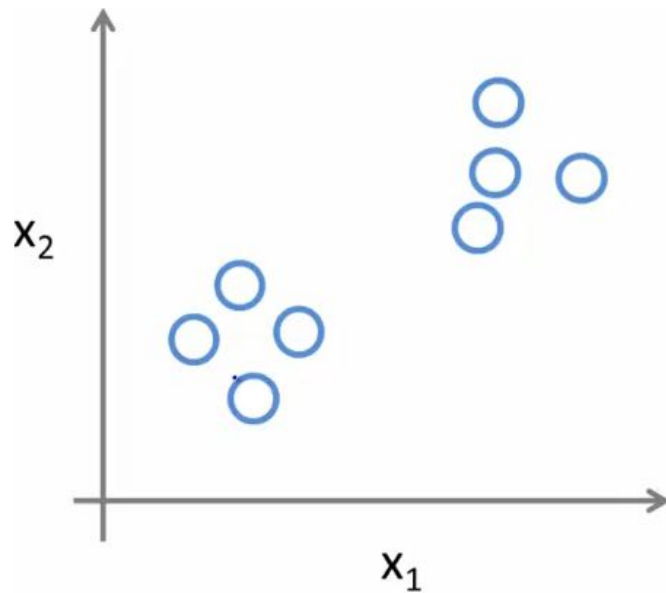
Mineração de dados

etc.

Aprendizagem Supervisionada vs não Supervisionada



Supervisionado



não Supervisionado

Aprendizagem Supervisionada

Em aprendizagem supervisionada, recebemos um conjunto de dados e sabemos **output** (valor ou classe) esperado para cada exemplo do conjunto de dados.

Problemas de aprendizagem supervisionada podem ser categorizados como **regressão** ou **classificação**.

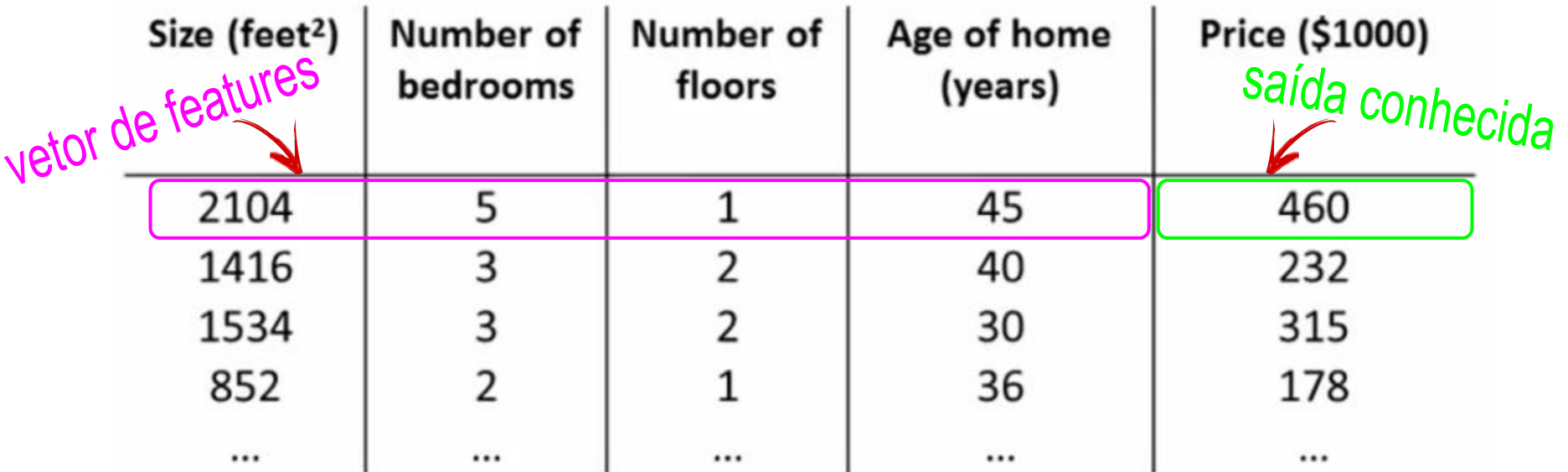
Aprendizagem Supervisionada

Size (feet ²)	Number of bedrooms	Number of floors	Age of home (years)	Price (\$1000)
2104	5	1	45	460
1416	3	2	40	232
1534	3	2	30	315
852	2	1	36	178
...

Aprendizagem Supervisionada

Size (feet ²)	Number of bedrooms	Number of floors	Age of home (years)	Price (\$1000)
2104	5	1	45	460
1416	3	2	40	232
1534	3	2	30	315
852	2	1	36	178
...

Aprendizagem Supervisionada

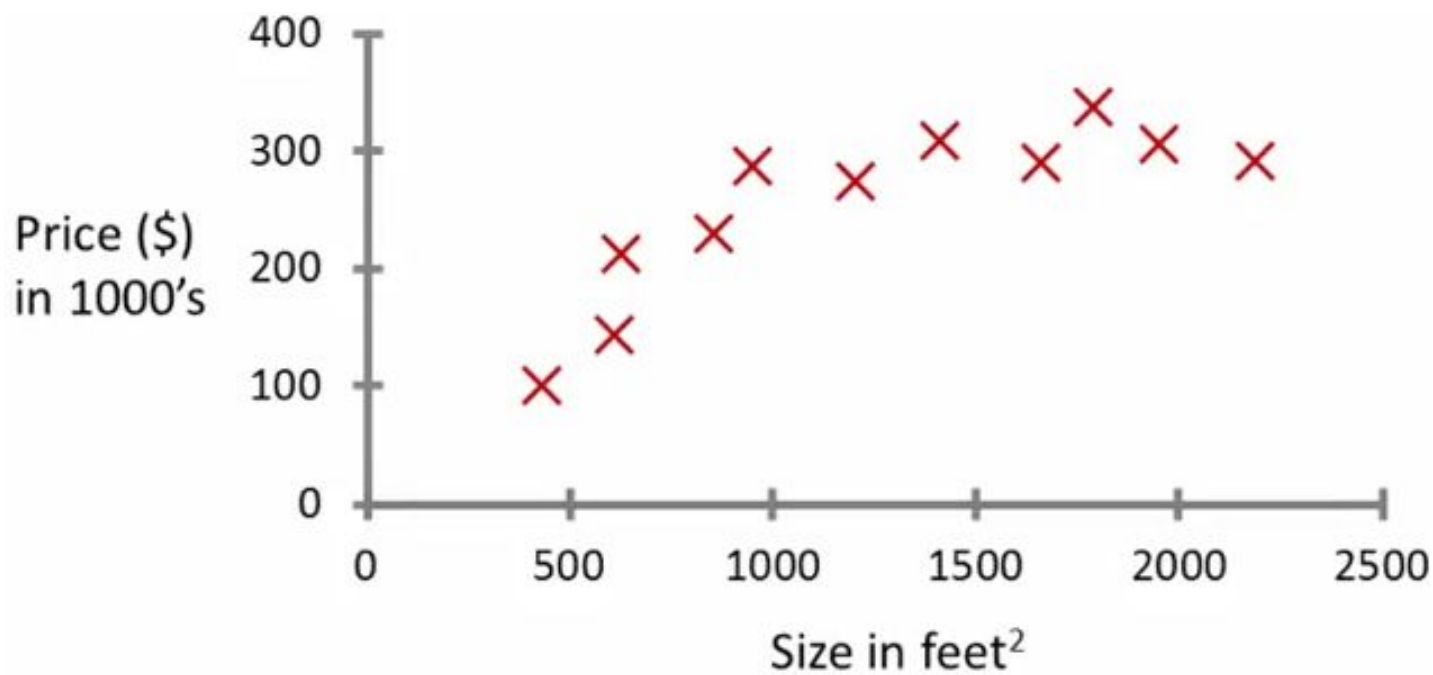


The diagram illustrates a supervised learning dataset. It consists of a table with five columns: 'Size (feet²)', 'Number of bedrooms', 'Number of floors', 'Age of home (years)', and 'Price (\$1000)'. The first four columns are grouped by a magenta box and labeled 'vetor de features' (feature vector) in magenta text. The fifth column is labeled 'saída conhecida' (known output) in green text. The first row of data is highlighted with a magenta box around the features and a green box around the price. The subsequent rows show more data points, and the table ends with an ellipsis in each column.

Size (feet²)	Number of bedrooms	Number of floors	Age of home (years)	Price (\$1000)
2104	5	1	45	460
1416	3	2	40	232
1534	3	2	30	315
852	2	1	36	178
...

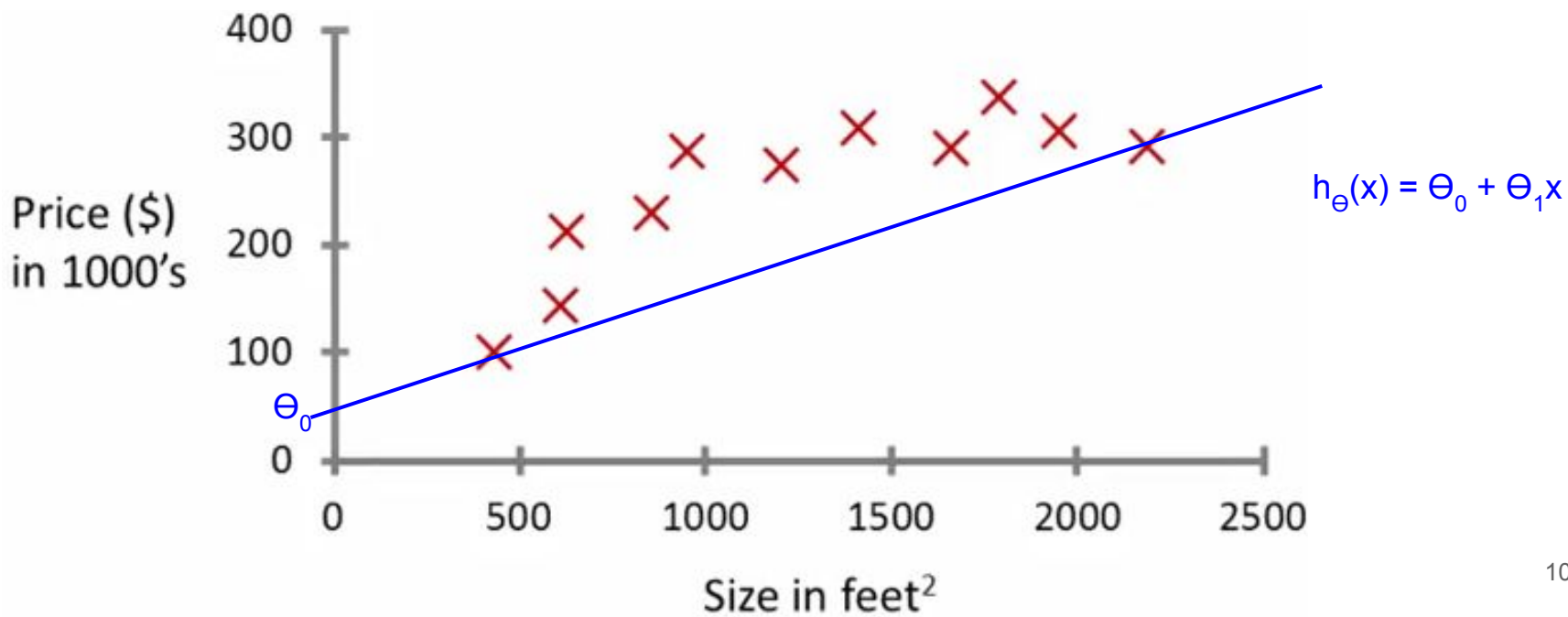
Regressão

(exemplo: predição de valor de casa)



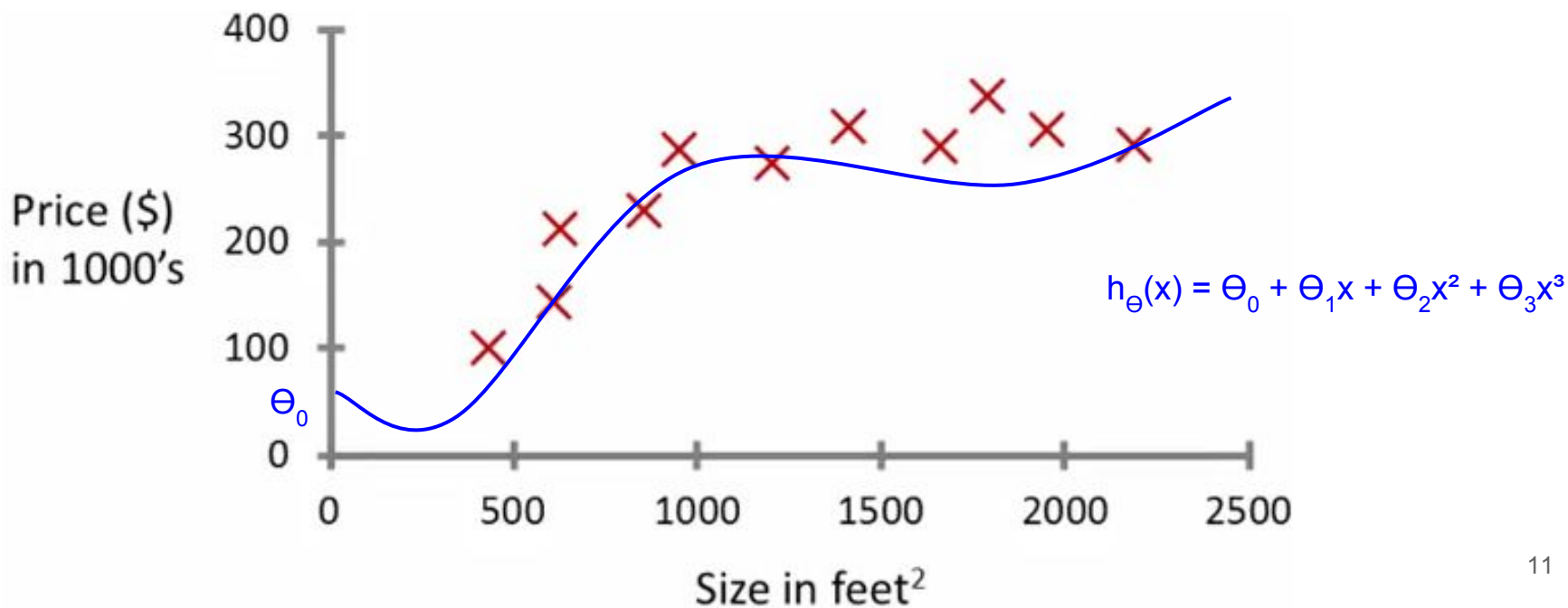
Regressão

(exemplo: predição de valor de casa)



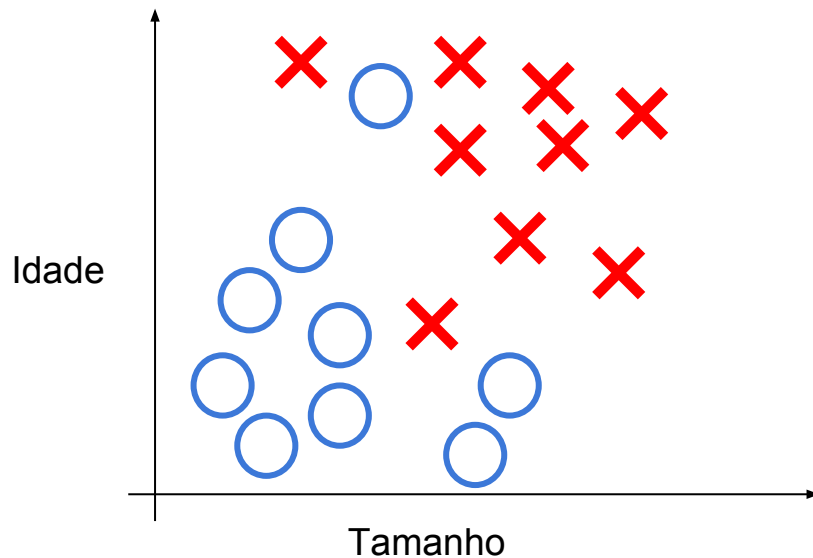
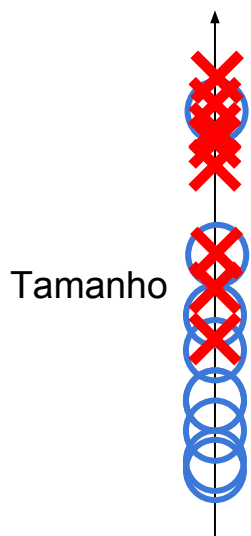
Regressão

(exemplo: predição de valor de casa)



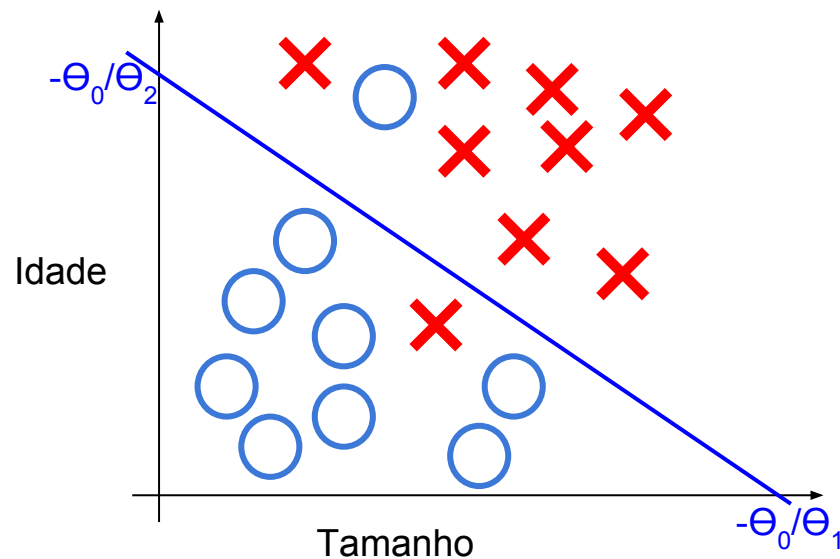
Classificação

(exemplo: classificação de tumor)



Classificação

(exemplo: classificação de tumor)

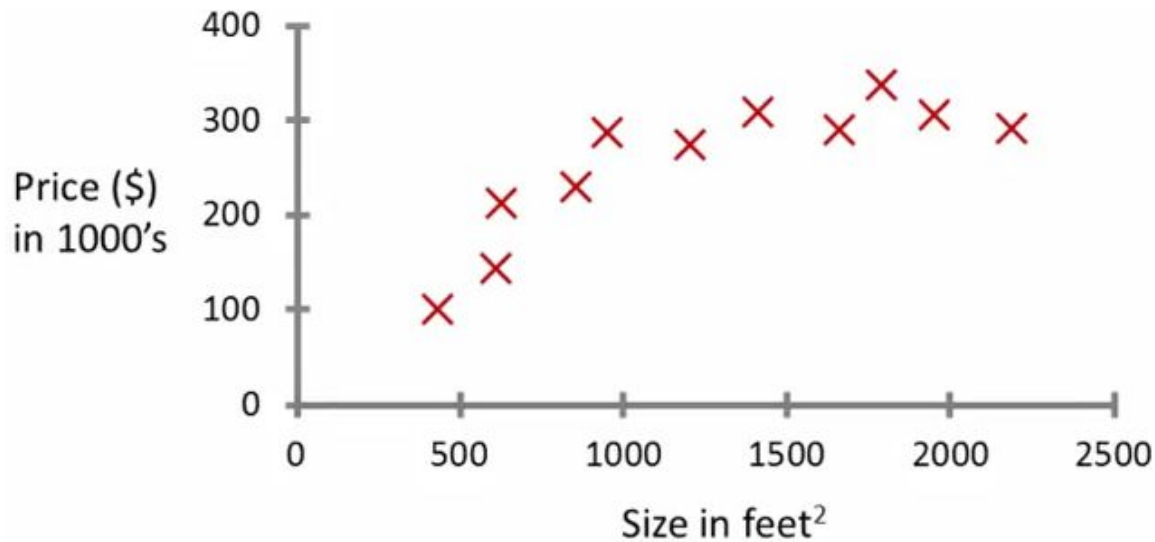


$$h_{\Theta}(\mathbf{x}) = \Theta_0 + \Theta_1 x_1 + \Theta_2 x_2$$

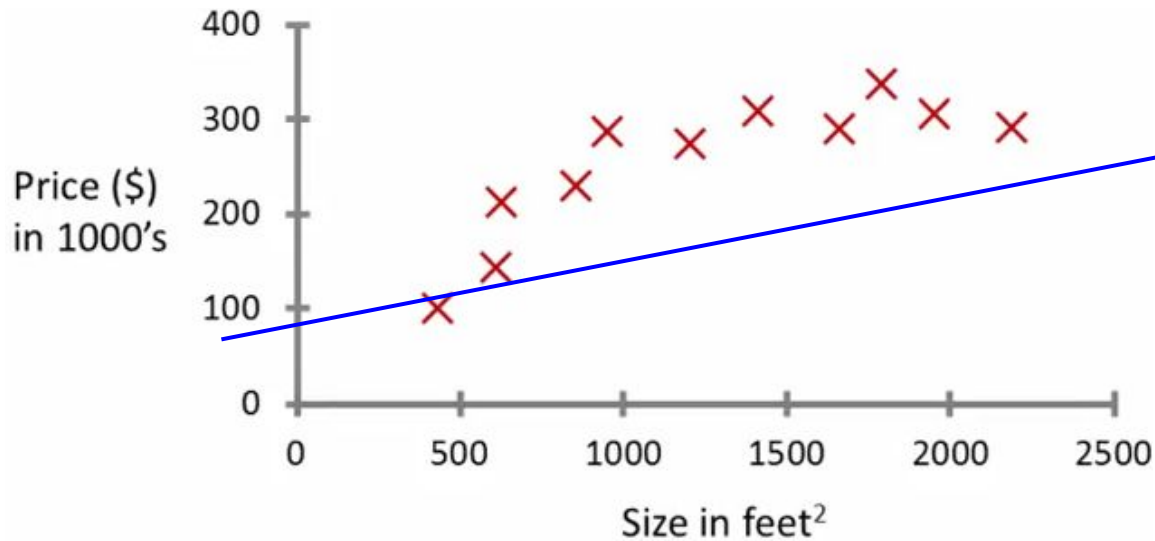
Como encontrar uma hipótese eficiente?

(to fit the model)

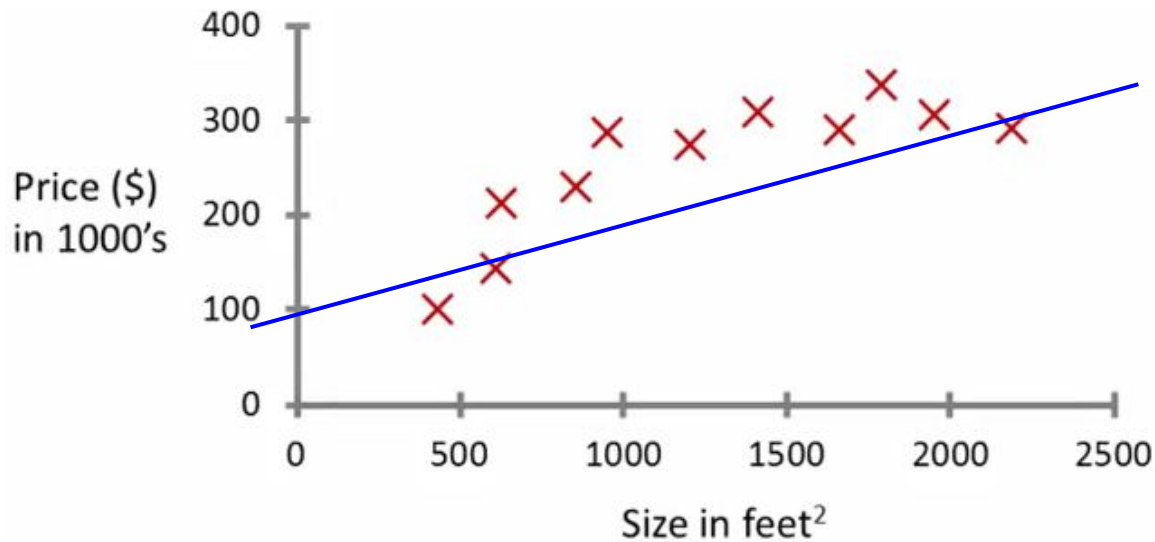
Como encontrar uma hipótese eficiente?



Como encontrar uma hipótese eficiente?



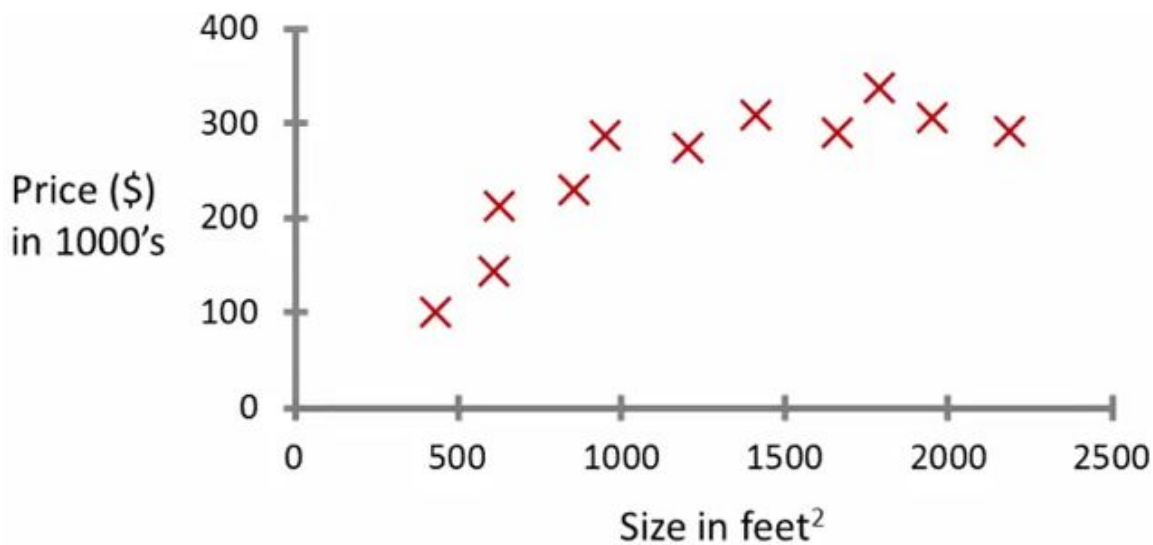
Como encontrar uma hipótese eficiente?



Hipótese: $h_{\theta}(\mathbf{x}) = \Theta_0 + \Theta_1 x_1 + \Theta_2 x_2 + \dots + \Theta_n x_n$

Cost function:

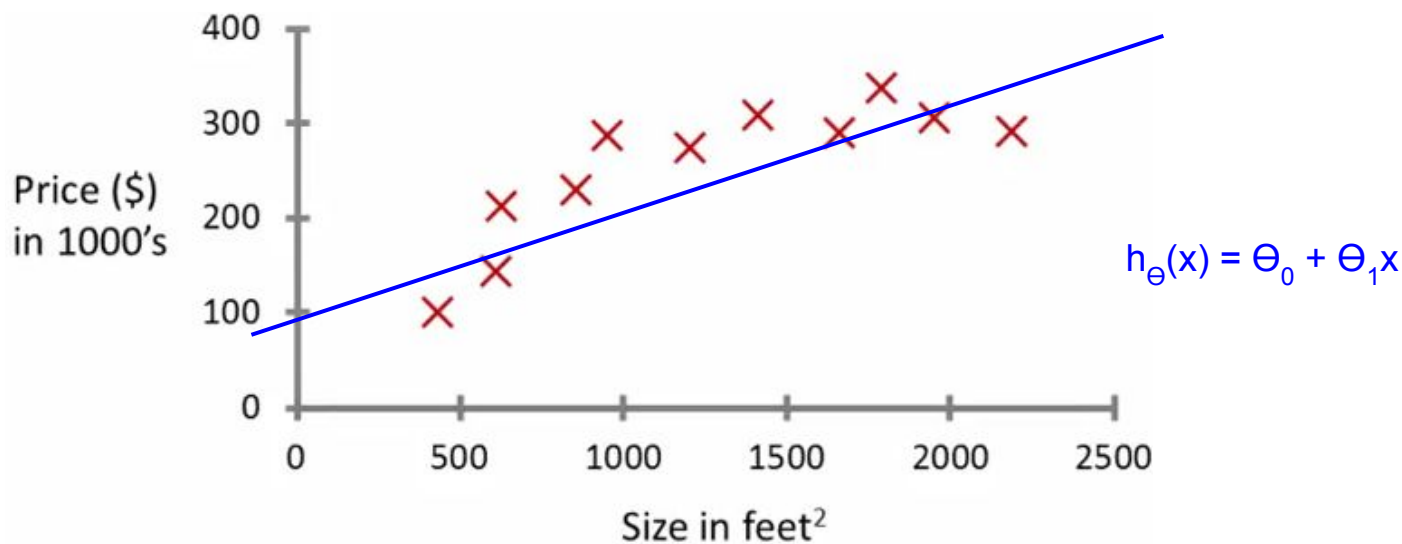
$$J(\theta_0, \theta_1, \dots, \theta_n) = \frac{1}{2m} \sum_{i=1}^m (h_{\theta}(x^{(i)}) - y^{(i)})^2$$



Hipótese: $h_{\Theta}(\mathbf{x}) = \Theta_0 + \Theta_1 x_1 + \Theta_2 x_2 + \dots + \Theta_n x_n$

Cost function:

$$J(\theta_0, \theta_1, \dots, \theta_n) = \frac{1}{2m} \sum_{i=1}^m (h_{\theta}(x^{(i)}) - y^{(i)})^2$$

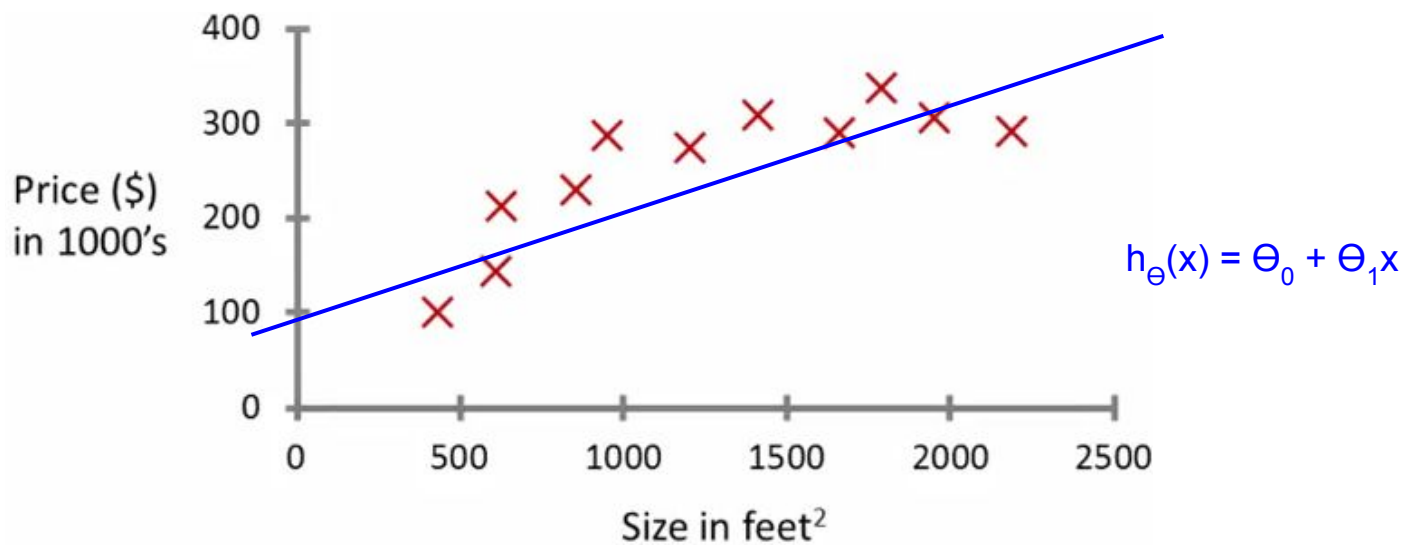


Hipótese: $h_{\theta}(\mathbf{x}) = \Theta_0 + \Theta_1 x_1 + \Theta_2 x_2 + \dots + \Theta_n x_n$

Cost function:

$$J(\theta_0, \theta_1, \dots, \theta_n) = \frac{1}{2m} \sum_{i=1}^m (h_{\theta}(x^{(i)}) - y^{(i)})^2$$

saída esperada para um exemplo (i)

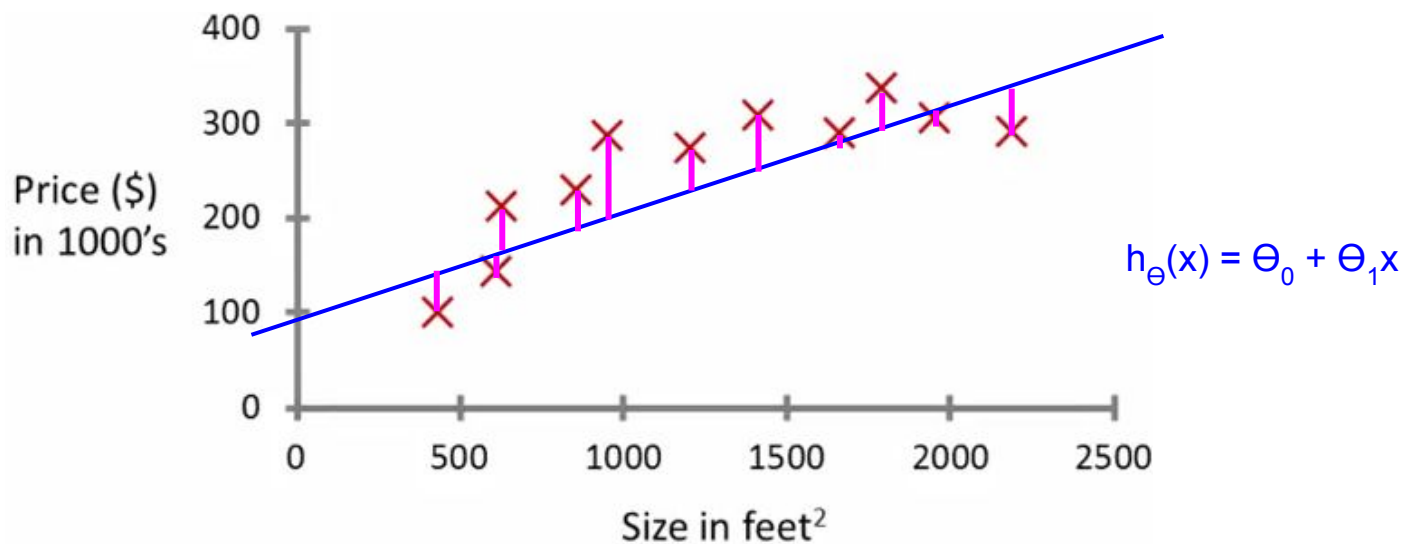


Hipótese: $h_{\theta}(\mathbf{x}) = \Theta_0 + \Theta_1 x_1 + \Theta_2 x_2 + \dots + \Theta_n x_n$

Cost function:

$$J(\theta_0, \theta_1, \dots, \theta_n) = \frac{1}{2m} \sum_{i=1}^m (h_{\theta}(x^{(i)}) - y^{(i)})^2$$

saída esperada para um exemplo (i)



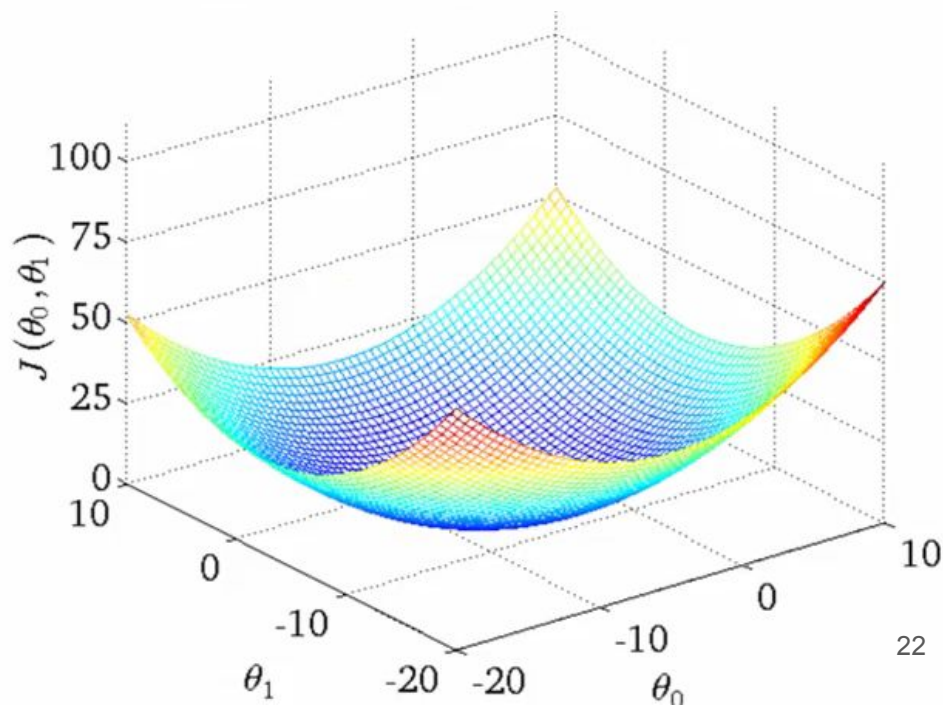
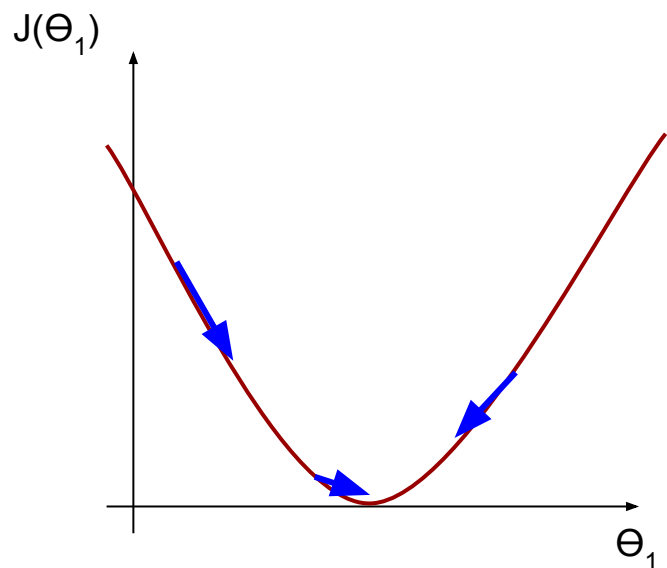
Gradient descent:

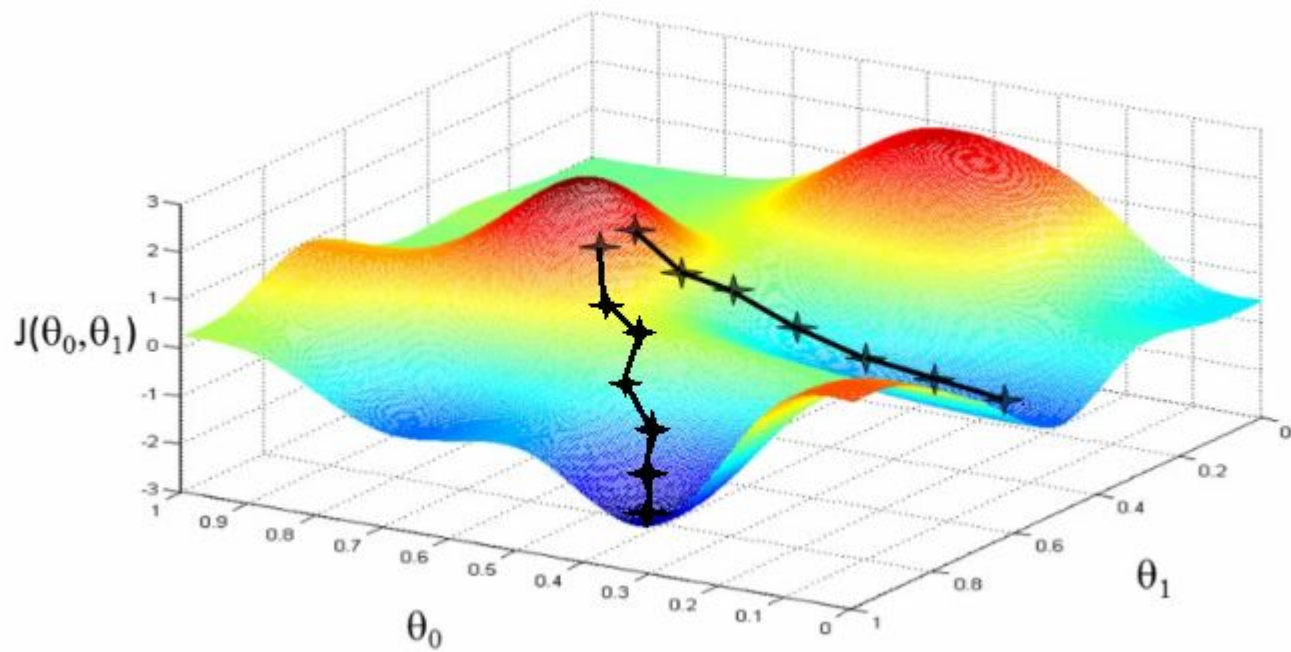
Repeat {

$$\theta_j := \theta_j - \alpha \frac{\partial}{\partial \theta_j} J(\theta_0, \dots, \theta_n)$$

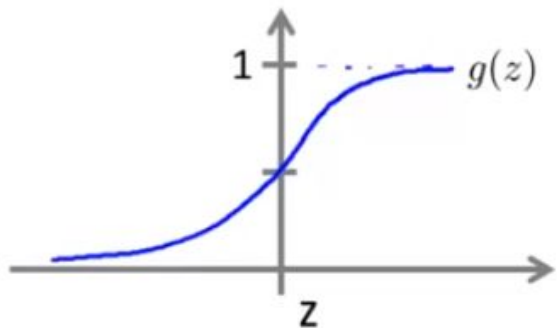
}

(simultaneously update for every $j = 0, \dots, n$)





Custo em classificadores



$$h_{\theta}(x) = g(\theta^T x)$$

$$g(z) = \frac{1}{1+e^{-z}}$$

$$h_{\theta}(x) = \frac{1}{1 + e^{-\theta^T x}}$$

$$\text{Cost}(h_{\theta}(x), y) = \begin{cases} -\log(h_{\theta}(x)) & \text{if } y = 1 \\ -\log(1 - h_{\theta}(x)) & \text{if } y = 0 \end{cases}$$

$$\begin{aligned}
 J(\theta) &= \frac{1}{m} \sum_{i=1}^m \text{Cost}(h_{\theta}(x^{(i)}), y^{(i)}) \\
 &= -\frac{1}{m} \left[\sum_{i=1}^m y^{(i)} \log h_{\theta}(x^{(i)}) + (1 - y^{(i)}) \log (1 - h_{\theta}(x^{(i)})) \right]
 \end{aligned}$$

Want $\min_{\theta} J(\theta)$:

Repeat {

$$\theta_j := \theta_j - \alpha \sum_{i=1}^m (h_{\theta}(x^{(i)}) - y^{(i)}) x_j^{(i)}$$

(simultaneously update all θ_j)
}

Recapitulando...

Hypothesis: $h_{\theta}(x) = \theta^T x = \theta_0 x_0 + \theta_1 x_1 + \theta_2 x_2 + \dots + \theta_n x_n$

Cost function:

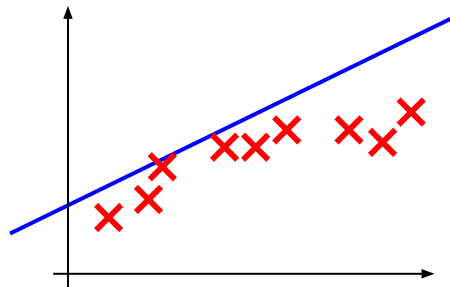
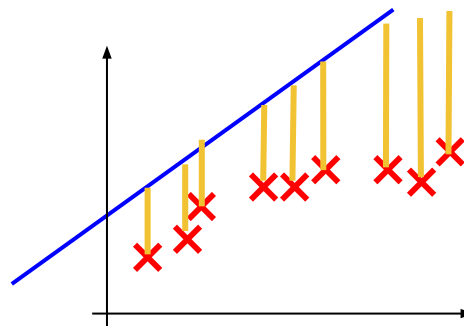
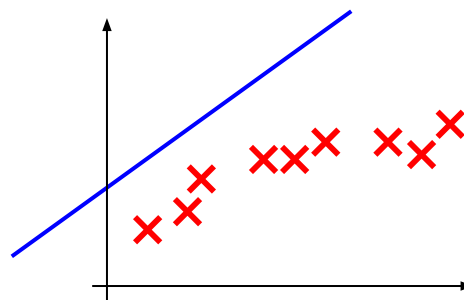
$$J(\theta_0, \theta_1, \dots, \theta_n) = \frac{1}{2m} \sum_{i=1}^m (h_{\theta}(x^{(i)}) - y^{(i)})^2$$

Gradient descent:

Repeat {

$$\theta_j := \theta_j - \alpha \frac{\partial}{\partial \theta_j} J(\theta_0, \dots, \theta_n)$$

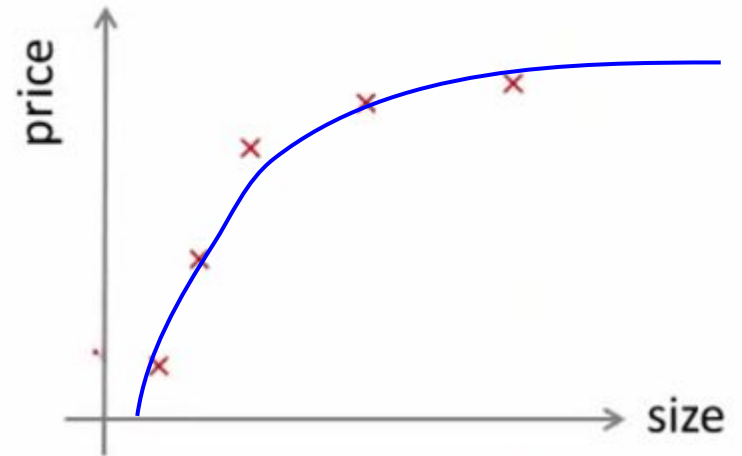
}



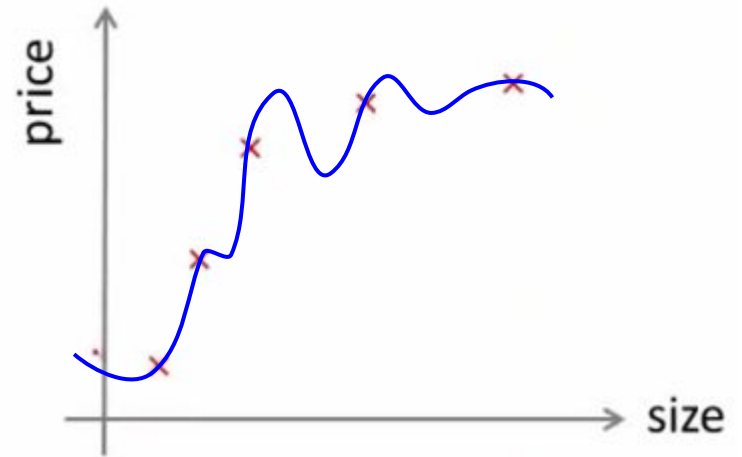
Teste da Hipótese



Teste da Hipótese



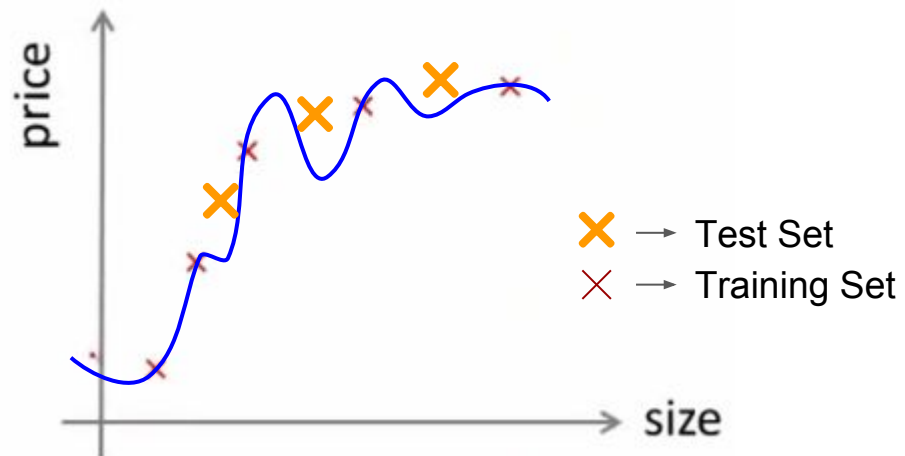
Teste da Hipótese



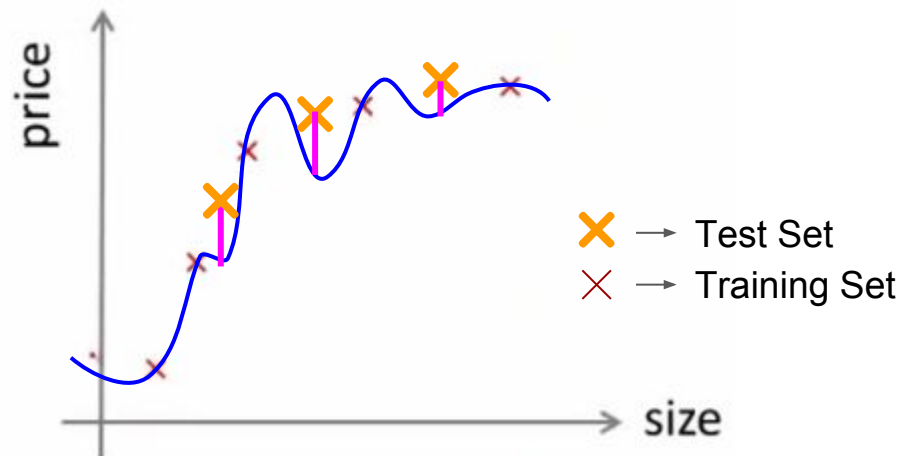
Teste da Hipótese

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

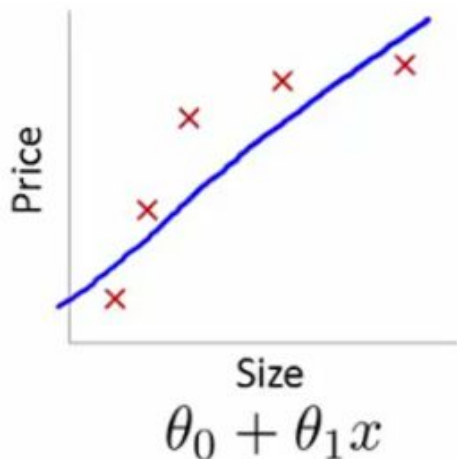
Size	Price
2104	400
1600	330
2400	369
1416	232
3000	540
1985	300
1534	315
1427	199
1380	212
1494	243



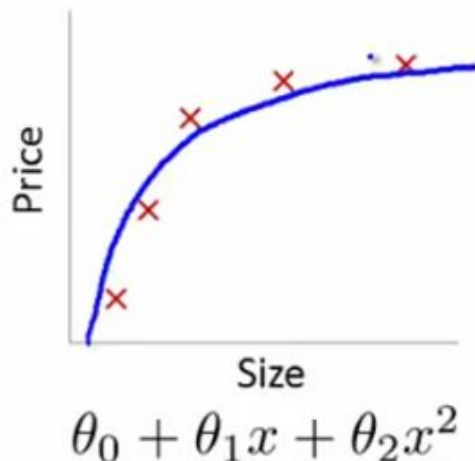
Size	Price
2104	400
1600	330
2400	369
1416	232
3000	540
1985	300
1534	315
1427	199
1380	212
1494	243



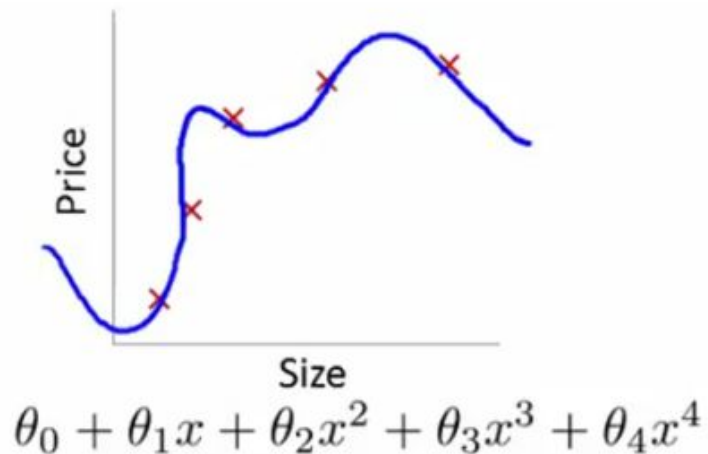
Overfitting e Underfitting (variance e bias)



High bias
(underfit)



“Just right”



High variance
(overfit)

