

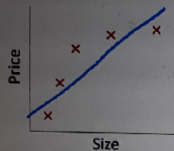
Overfitting

→ Regularization

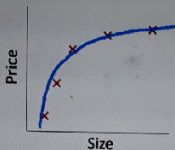
Custo mt baixo - pouca capacidade
de generalização a novos exemplos

→ overfit → high variance,
low bias

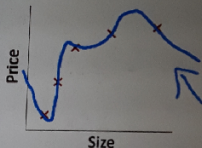
Example: Linear regression (housing prices)



→ $\theta_0 + \theta_1 x$
"Underfit" "High bias"



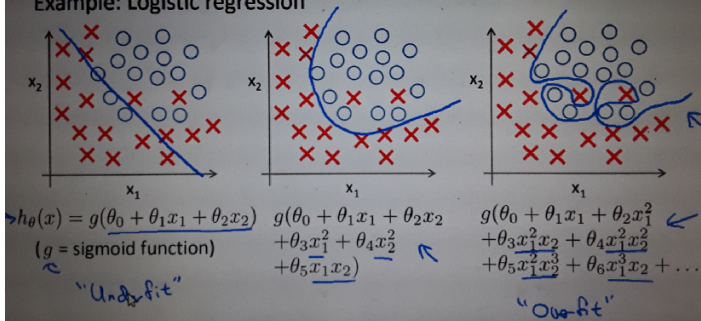
→ $\theta_0 + \theta_1 x + \theta_2 x^2$
"Just right"



→ $\theta_0 + \theta_1 x + \theta_2 x^2 + \theta_3 x^3 + \theta_4 x^4$
"Overfit" "High variance"

Overfitting: If we have too many features, the learned hypothesis may fit the training set very well ($J(\theta) = \frac{1}{2m} \sum_{i=1}^m (h_{\theta}(x^{(i)}) - y^{(i)})^2 \approx 0$), but fail to generalize to new examples (predict prices on new examples).

Example: Logistic regression

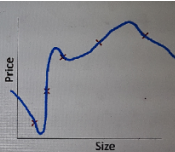


Como resolver overfitting:

→ se tem muitas features

Addressing overfitting:

x_1 = size of house
 x_2 = no. of bedrooms
 x_3 = no. of floors
 x_4 = age of house
 x_5 = average income in neighborhood
 x_6 = kitchen size
 \vdots
 x_{100}



2 opções:

- ① Reduce # of features
 - manually select which features to keep
 - model selection algorithm

② Regularization

- keep all the features but reduce magnitude/values of θ_j
- E' bom qd temos várias features e

Todas contribuem p/predizer y.

Regularização

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

Parâmetro
de regularização

- Se for muito alto,

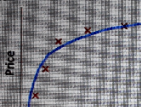
$\theta_1, \theta_2, \dots, \theta_n = 0$ e $h_{\theta}(x) = \theta_0$

Reduzir a influência de uma feature \rightarrow modifico minha cost function.

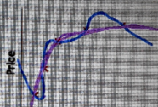
$$\min_{\theta} \frac{1}{2m} \sum_{i=1}^m (h_{\theta}(x^{(i)}) - y^{(i)})^2 + 1.000 \theta_3^2 + 1000 \theta_4^2$$

Adicionei 2 termos extras θ_3 e θ_4 e agora, p/a θ_j se aproximar de zero, preciso reduzir os valores de θ_3 e θ_4 p/ quase zero.

Intuition



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



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

Suppose we penalize and make θ_3, θ_4 really small.

$$\rightarrow \min_{\theta} \frac{1}{2m} \sum_{i=1}^m (h_{\theta}(x^{(i)}) - y^{(i)})^2 + \underbrace{1000 \theta_3^2}_{\theta_3 \rightarrow 0} + \underbrace{1000 \theta_4^2}_{\theta_4 \rightarrow 0}$$

Posso regularizar todos os parâmetros,

Como:

$$\min_{\Theta} \frac{1}{2m} \sum_{i=1}^m (h_{\Theta}(x^{(i)}) - y^{(i)})^2 + \lambda \sum_{j=1}^n \Theta_j^2$$

se λ mt grande, esmago mt
o peso de cada feature e pode
causar underfitting.

E se $\lambda=0$? Ou mt pequeno?
overfitting!

