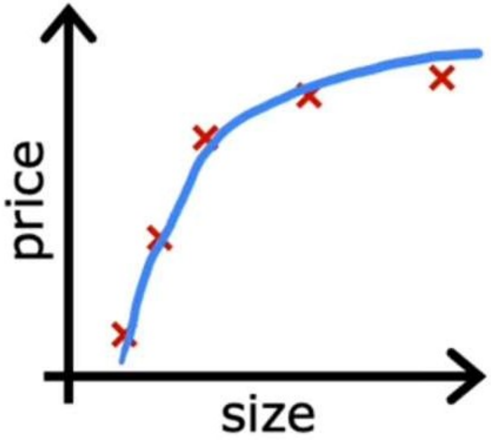
Stanford
ONLINE

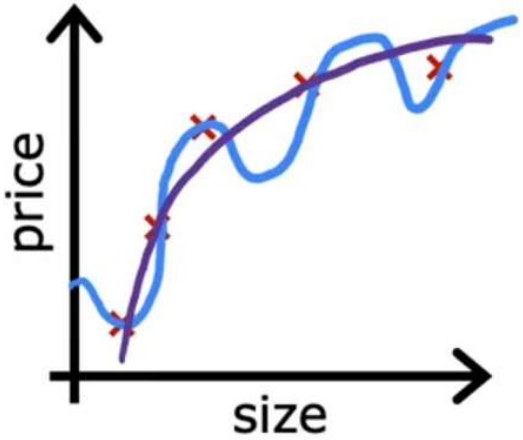
DeepLearning.AI

Regularization to Reduce Overfitting

Cost Function with Regularization

Intuition





$$w_1 x + w_2 x^2 + b$$

$$w_1 x + w_2 x^2 + \cancel{\underbrace{w_3}_{\approx 0} x^3} + \cancel{\underbrace{w_4}_{\approx 0} x^4} + b$$

make w_3, w_4 really small (≈ 0)

$$\min_{\vec{w},b} \frac{1}{2m} \sum_{i=1}^{m} \left(f_{\vec{w},b}\left(\vec{x}^{(i)}\right) - y^{(i)}\right)^2 + \cancel{1000 \underbrace{w_3^2}_{0.001}} + \cancel{1000 \underbrace{w_4^2}_{0.002}}$$

Regularization

small values $w_1, w_2, \cdots, w_n, b$

simpler model $w_3 \approx 0$

less likely to overfit $w_4 \approx 0$

size x_1	bedrooms x_2	floors x_3	age x_4	avg income x_5	...	distance to coffee shop x_{100}	price y

$w_1, w_1, w_2, \cdots, w_{100}, b$

n features

$n = 100$

$$J(\vec{\mathrm{w}}, b) = \frac{1}{2m}\left[\sum_{i=1}^{m}\left(f_{\vec{\mathrm{w}},b}\left(\vec{\mathrm{x}}^{(i)}\right) - y^{(i)}\right)^2 + \overbrace{\underbrace{\frac{\lambda}{2m}}_{\text{"lambda"}}\sum_{j=1}^{n} w_j^2}^{\text{regularization term}} + \frac{\lambda}{2m}b^2\right]$$

can include or exclude b

"lambda"
regularization parameter

$\lambda > 0$

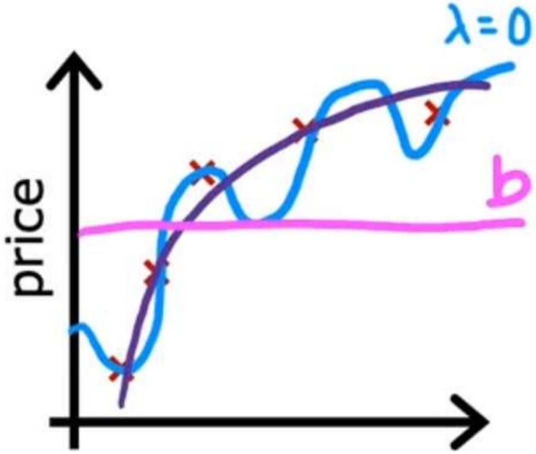
Regularization

$$\min_{\vec{w},b} J(\vec{w},b) = \min_{\vec{w},b}\left[\overbrace{\frac{1}{2m}\sum_{i=1}^{m}\left(f_{\vec{w},b}\left(\vec{x}^{(i)}\right)-y^{(i)}\right)^2}^{\text{mean squared error}} + \overbrace{\frac{\lambda}{2m}\sum_{j=1}^{n} w_j^2}^{\text{regularization term}}\right]$$

fit data

Keep w_j small

λ balances both goals



choose $\lambda = 10^{10}$

$$f_{\vec{w},b}(\vec{x}) = \underset{\approx 0}{\cancel{w_1 x}} + \underset{\approx 0}{\cancel{w_2 x^2}} + \underset{\approx 0}{\cancel{w_3 x^3}} + \underset{\approx 0}{\cancel{w_4 x^4}} + b$$

$f(x) = b$

Choose λ