Cajamar, 23 de outubro de 2018

$$y = \theta_0 x^0 + \theta_1 x^1 + \theta_2 x^2 + \theta_3 x^3$$

Modelo: $$\boxed{y = \theta_0 x_0 + \theta_1 x_1 + \theta_2 x_2 + \theta_3 x_3}$$

$$\theta^T X = Y$$

$$J_{\theta j} = \frac{1}{2m} \sum_{i=1}^{m} \left[\sum_{j=0}^{3} \left(\theta_j x_j^{(i)} \right) - \hat{y}_i \right]^2$$

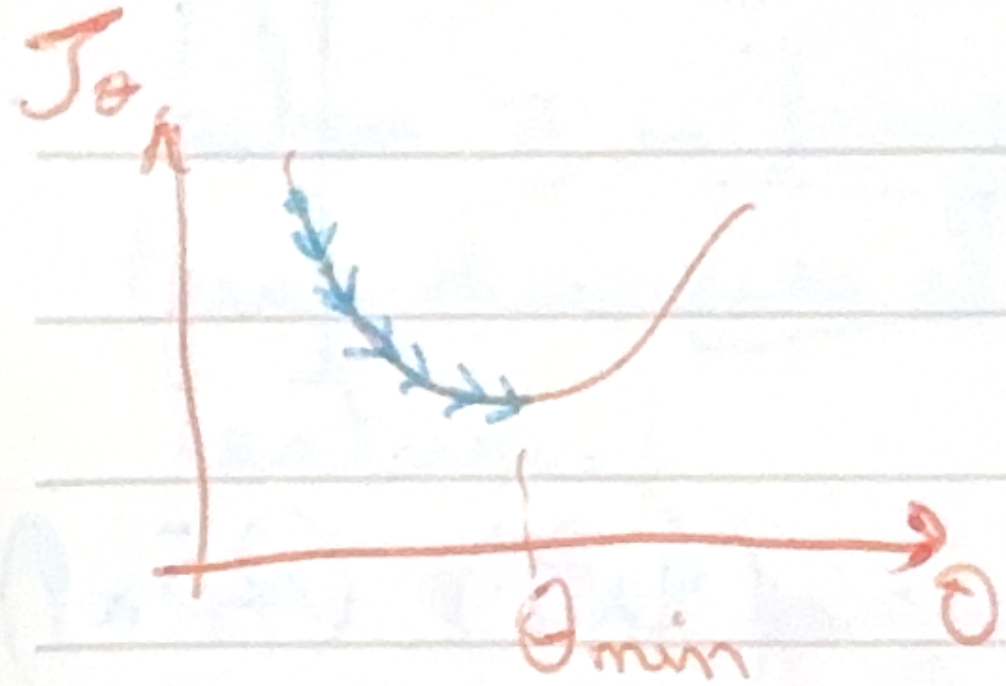
$$\boxed{\frac{\partial J}{\partial \theta_j} = \frac{1}{m} \sum_{i=0}^{m} \left[\sum_{k=0}^{3} \theta_k x_k^{(i)} - \hat{y}_i \right] x_j^{(i)}}$$

$$\theta_j := \theta_j - \alpha \frac{\partial J}{\partial \theta_j}$$

Gradient Descent x Normal Equation

↓ small steps till convergence to θ min

↓ solve analytically the equation for θ



$$\underset{(n+1)\times m}{X^T} \cdot \underset{m\times(n+1)}{X} = \underset{(n+1)\times(n+1)}{A}$$

$(m \times (n+1)) \times ((n+1) \times m)$

$m \times m$

$$X^T \cdot \theta = Y$$

$$(X^T)^{-1} \cdot X^T \theta = (X^T)^{-1} \cdot Y$$

$$\theta = (X^T)^{-1} \cdot Y$$

pera, só vale se X já for quadrada

Então:

$$X\theta = Y$$

$$X^T X \theta = X^T Y$$

$$(X^T X)^{-1} X^T X \theta = (X^T X)^{-1} X^T Y$$

$$\boxed{\theta = (X^T X)^{-1} X^T Y}$$

Nesse caso, não é preciso usar feature scaling $0 < x_i < 1$