Regularization - Cost Function

⇒ penalize θ_3 and θ_4

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

Intuition: having small values for parameters makes the hypothesis simpler and less prone to overfitting

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

penalizando qualquer parâmetro que queira ser o manda-chuva

REGULARIZATION FOR LINEAR REGRESSION

~~FOR~~ IN GRADIENT DESCENT:

$$\theta_j := \theta_j - \frac{\alpha}{m}\left[\sum_{i=1}^{m}\left(h_\theta(x^{(i)})-y^{(i)}\right)x_j^{(i)} + \lambda\theta_j\right]$$

$$j = 1, 2, 3, \dots, n$$

OBS: $\theta_0 := \theta_0 - \frac{\alpha}{m}\sum_{i=1}^{m}\left(h_\theta(x^{(i)})-y^{(i)}\right)x_0^{(i)}$

Podemos também escrever θ_j assim:

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

~~FOR~~ IN NORMAL EQUATION:

$$\theta_j = \left(X^TX + \lambda\underbrace{\begin{bmatrix} 0 & & & 0 \\ & 1 & & \\ & & 1 & \\ 0 & & & \ddots\ 1 \end{bmatrix}}_{(n+1)\times(n+1)}\right)^{-1}X^T y$$

always invertible!

↳ n = # features

$X_{m\times(n+1)}$

$X^T_{(n+1)\times m}$

$\Theta_{(n+1)\times 1}$

$Y_{m\times 1}$

$$X\Theta = Y$$

$$(X^TX + \lambda L)\,\Theta = X^T y$$

$$X(X^TX + \lambda L)\,\Theta = X.X^T y$$

$$\left[(XX^T)^{-1}.X.(X^T.X) + (XX^T)^{-1}.X.\lambda L\right]\Theta = Y.$$