LEAST SQUARES

$$\hat{\Theta} = \underset{\Theta}{\operatorname{argmin}} \| \underline{X}\Theta - y \|$$

$\underline{X}$: $M \times N$ datos de input $M > N$ *

$\underline{\Theta}$: $N \times 1$ parámetros

y : $M \times 1$ output

$$J(\Theta) = \frac{1}{2} (X\Theta - y)^T (X\Theta - y)$$

$$= \frac{1}{2} (\Theta^T X^T - y^T)(X\Theta - y)$$

$$= \frac{1}{2} (\Theta^T X^T X \Theta - \Theta^T X^T y - y^T X \Theta + y^T y)$$

REPASO

$$\boxed{\frac{\partial J}{\partial \Theta} (\Theta^T S \Theta) = 2S\Theta \qquad \frac{\partial}{\partial \Theta} (\Theta^T a) = a \qquad \frac{\partial}{\partial \Theta} (a^T \Theta) = a^T}$$

$$\frac{\partial J}{\partial \Theta} = \frac{1}{2} (2X^T X \Theta - X^T y - (y^T X)^T) = X^T X \Theta - X^T y \overset{!}{=} 0$$

$$\boxed{\hat{\Theta} = [X^T X]^{-1} X^T y}$$

* si $M = N \longrightarrow \hat{\Theta} = X^{-1} y$ si $M < N$ no hay solución.

ESTIMACIÓN NO-LINEAL

$$\hat{y} = f(\theta)$$

$$J(\theta) = \frac{1}{2}\| f(\theta) - y\|^2 \to \text{Min}$$

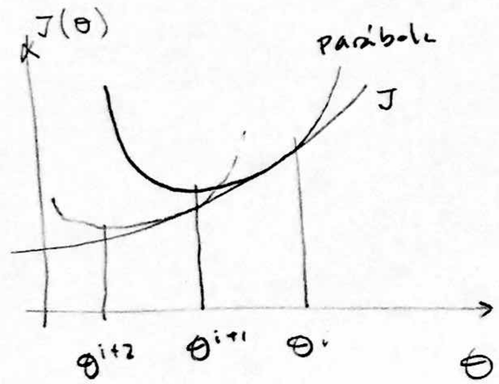
Métodos iterativos: θ_0 : valor inicial, $\theta^{i+1} = \theta^i + \Delta\theta^i$

$$J(\theta^{i+1}) = J(\theta^i + \Delta\theta^i)$$

Series de Taylor:

- si $\theta \in \mathbb{R}$ (una dimensión) $J(\theta^i + \Delta\theta^i) = \sum_{n=0}^{+\infty} \frac{J^n(\theta^i)}{n!}(\Delta\theta^i)^n$

$$J(\theta^i + \Delta\theta^i) \simeq \underbrace{J(\theta^i) + J'(\theta^i)\Delta\theta^i + \frac{J''(\theta^i)[\Delta\theta^i]^2}{2}}_{P}$$



Parábola que es buena aproximación de J en θ^i

Es facil buscar el mínimo de la parábola P

$$\frac{\partial P}{\partial \Delta\theta^i} = 0 \to \Delta\theta^i = -\frac{J'(\theta^i)}{J''(\theta^i)}$$

Método de Newton-Raphson

$$\theta^{i+1} = \theta^i - \frac{J'(\theta^i)}{J''(\theta^i)} \quad \text{para } \theta \in \mathbb{R}$$

- si $\theta \in \mathbb{R}^d$ $d > 1$

$$J(\theta^i + \Delta\theta^i) \simeq J(\theta^i) + [\Delta\theta^i]^T J_\theta(\theta^i) + \frac{1}{2}[\Delta\theta^i]^T J_{\theta\theta}(\theta^i)\Delta\theta^i$$

$$J_\theta = \left[\frac{\partial J}{\partial\theta_1} \cdots \frac{\partial J}{\partial\theta_N}\right] \qquad J_{\theta\theta} = \begin{bmatrix} \frac{\partial^2 J}{\partial\theta_1^2} & \frac{\partial^2 J}{\partial\theta_1\partial\theta_2} & \\ & & \ddots \\ & & \frac{\partial^2 J}{\partial\theta_N^2}\end{bmatrix}$$

$$\frac{\partial J(\theta^{i+1})}{\partial\Delta\theta} = 0 + J_\theta(\theta^i) + J_{\theta\theta}(\theta^i)\Delta\theta^i = 0 \to \theta^{i+1} = \theta^i - [J_{\theta\theta}(\theta^i)]^{-1}J_\theta(\theta^i)$$

para $\theta \in \mathbb{R}^d$