# 1 Analytical solution with Linear Algebra

## 1.1 Algebraic manipulation

let's denote $x_i = \begin{bmatrix} 1 \\ x_{i1} \\ .. \\ .. \\ x_{in} \end{bmatrix}$ and $\theta = \begin{bmatrix} \theta_0 \\ \theta_1 \\ .. \\ .. \\ \theta_n \end{bmatrix}$ the expression: $h(\theta) = \theta_0 + \theta_1 \cdot x_1 + ..\theta_n \cdot x_n$ can be expressed as $\theta \cdot x^T$.
the cost function can be expressed as follows:

$$J = \frac{1}{2m} \sum_{i=1}^{m} (y_i - \dot{y}_i)^2 = \frac{1}{2m} \sum_{i=1}^{m} (y_i - x_i^T \cdot \theta)^2.$$

let's consider $Y = \begin{bmatrix} y_1 \\ y_2 \\ .. \\ .. \\ y_m \end{bmatrix}$ and $X = \begin{bmatrix} 1 & x_{11} & x_{12} & .. & x_{1n} \\ 1 & x_{21} & x_{22} & .. & x_{2n} \\ .. & .. & .. & .. & .. \\ 1 & x_{m1} & x_{m2} & .. & x_{mn} \end{bmatrix}$ then

the following equality holds $\dot{Y} = \begin{bmatrix} \dot{y}_1 \\ \dot{y}_2 \\ .. \\ .. \\ \dot{y}_n \end{bmatrix} = X \cdot \theta$

Therefore our the cost function

$J = \frac{1}{2m} \cdot |(Y - \dot{Y})| = \frac{1}{2m} \cdot (Y - \dot{Y})^T \cdot (Y - \dot{Y}) = \frac{1}{2m} \cdot (Y - X \cdot \theta)^T \cdot (Y - X \cdot \theta)$

## 1.2 solution

Using Linear Algebra techniques, the solution for the problem $J \rightarrow min$ is the vector

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

## 1.3 Note

This solution is valid only when the matrix $X$ have linearly independent columns. Otherwise the matrix $(X^T \cdot X)^{-1}$ might not exist. In such cases, using numerical optimization methods such as Gradient Descent might be the best approach.

## 2 Gradient Descent

The algorithm can be expressed as follows
repeat until convergence {
    $\theta_i := \theta_i - \alpha \cdot \frac{dJ}{d\theta_i}$
    for $i = 0, 1, ... \ n$
}

Having

$$J = \frac{1}{2m} \cdot \sum_{i=1}^{m} (y_i - \sum_{j=0}^{n} (x_{ij} \cdot \theta_j))^2$$

then

$$\begin{aligned} \frac{dJ}{d\theta_k} &= \frac{1}{m} \cdot \sum_{i=1}^{m} (y_i - \sum_{j=0}^{n} (x_{ij} \cdot \theta_j)) \cdot x_{ik} \\ &= \frac{1}{m} \cdot \sum_{i=1}^{m} (y_i - \dot{y}_i) \cdot x_{ik} \end{aligned}$$

The expression above can be vectorized as follows:

$$\frac{\delta J}{\delta \theta_k} = X^T \cdot (X \cdot \theta - y)$$