# Machine Learning : Assignment 1

By: Dorian Buijse (11155647)

Amsterdam University College

## Question 1

### 1.a

Since $h_\theta(x)$ in multivariate linear regression is represented by the function: $h_\theta(x) = \theta_0 * x_0 + \theta_1 * x_1 + \theta_2 * x_2 + ... + \theta_n * x_n$ in vector notation this can be seen as:

$$h_\theta(x) = \begin{bmatrix} \theta_0 \ \theta_1 \ \theta_2 \cdots \theta_{n-1} \ \theta_n \end{bmatrix} \cdot \begin{bmatrix} x_0 \\ x_1 \\ x_2 \\ \vdots \\ x_{n-1} \\ x_n \end{bmatrix} = \theta^T X \qquad (Where\ x_0 = 1) \tag{1}$$

### 1.b

The vectorized $J(\theta)$ is represented by the function:

$$J(\theta) = \frac{(h_\theta(x)^T - y)^T \cdot (h_\theta(x)^T - y)}{2m} \qquad \text{(Where y is a 1 x m vector)} \tag{2}$$

**Note:** x is a rowvector while y is a collumnvector, hence the transpose of x

### 1.c

The vectorized expression for the gradient of the cost function is:

$$\frac{\partial J(\theta)}{\partial \theta} = \frac{(h_\theta(x)^T - y)}{m} \circ x \qquad \text{(Where y is a 1 x m vector)} \tag{3}$$

### 1.d

Since $\frac{\partial J(\theta)}{\partial \theta}$ is a vectorized expression, we can use a single update. Since we use vector computations $\theta$ can be updated in one computation.

$$\theta := \theta - \alpha \frac{\partial J(\theta)}{\partial \theta} \tag{4}$$

## Question 2

Derive an equation that can be used to find the optimal value of the parameter $\theta_1$ for univariate linear regression without doing gradient descent. This can be done by setting the value of the derivative equal to 0. You may assume that the value of $\theta_0$ is fixed.

$$\frac{\partial J(\theta)}{\partial \theta_1} = 0 \tag{5}$$

$$\frac{1}{2m} * \frac{\partial}{\partial \theta_1}(\sum_{i=1}^{m}(h_\theta(x^i) - y^i)^2) = 0 \tag{6}$$

$$\frac{1}{m} * \sum_{i=1}^{m}(h_\theta(x^i) - y^i) * x^i = 0 \tag{7}$$

$$\sum_{i=1}^{m}(h_\theta(x^i) - y^i) * x^i = 0 \tag{8}$$

$$\sum_{i=1}^{m}(\theta_0 * x^i + \theta_1 * (x^i)^2 - y^i * x^i) = 0 \tag{9}$$

$$\sum_{i=1}^{m}\theta_0 * x^i = \sum_{i=1}^{m}\theta_1 * (x^i)^2 - \sum_{i=1}^{m} y^i * x^i \tag{10}$$

$$\sum_{i=1}^{m}\theta_1 * (x^i)^2 = \sum_{i=1}^{m}\theta_0 * x^i - \sum_{i=1}^{m} y^i * x^i \tag{11}$$

$$\sum_{i=1}^{m}\theta_1 * x^i = \sum_{i=1}^{m}\theta_0 - \sum_{i=1}^{m} y^i \tag{12}$$

$$m * \theta_1 = \sum_{i=1}^{m}\frac{\theta_0}{x^i} - \sum_{i=1}^{m}\frac{y^i}{x^i} \tag{13}$$

$$\theta_1 = \frac{1}{m} * \sum_{i=1}^{m}\frac{\theta_0 - y^i}{x^i} \tag{14}$$

14 is the equation which gives the optimal parameter for $\theta_1$