Graded assignment 1 – Bobbie van Gorp

1. (a) The vectorized expression for the hypothesis function is the result of the matrix-matrix multiplication of the transpose of the theta vector and the feature vector x :

$$h_\vartheta(x) = \vartheta^T x$$

(b) The vectorized expression for the cost function J is as follows:

$$J(\vartheta) = \frac{1}{2m}\sum_{i=1}^{m}(h_\vartheta(x^{(i)}) - y^{(i)})^2$$

(c) The vectorized expression for the gradient of the cost function is:

$$\frac{\partial J(\vartheta)}{\partial \vartheta} = \frac{1}{m}\sum_{i=1}^{m}\left(h_\vartheta(x^{(i)}) - y^{(i)}\right) x^{(i)}$$

(d)

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

(e) If we take the vectorization a step further, so that there is no more summation over training examples necessary, we end up with the following formulas:

Let X be the matrix

$$\begin{matrix} x_0^{(0)} & x_1^{(0)} & \dots & x_n^{(0)} \\ x_0^{(1)} & x_1^{(1)} & \dots & x_n^{(1)} \\ \vdots & \vdots & \dots & \vdots \\ x_0^{(m)} & x_1^{(m)} & \dots & x_n^{(m)} \end{matrix}$$

Then

$$h_\vartheta(x) = \vartheta^T X^T$$

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

$$\frac{\partial J(\vartheta)}{\partial \vartheta} = \frac{1}{m}X^T(h_\vartheta(X)^T - y)$$

$$\vartheta := \vartheta - \alpha\frac{1}{m}X^T(h_\vartheta(X)^T - y)$$

2.

$$\frac{\partial J(\theta)}{\partial \theta_1} = \frac{1}{m}\sum_{i=1}^{m}\left(\theta_0 + \theta_1 x_1^{(i)} - y^{(i)}\right) x_1^{(i)}$$

$$0 = \sum_{i=1}^{m}\left(\theta_0 + \theta_1 x_1^{(i)} - y^{(i)}\right) x_1^{(i)}$$

$$0 = \sum_{i=1}^{m}\left(\theta_0 + \theta_1 x_1^{(i)} - y^{(i)}\right) x_1^{(i)}$$

$$0 = \sum_{i=1}^{m}\left(\theta_0 + \theta_1 x_1^{(i)} - y^{(i)}\right)$$

$$\sum_{i=1}^{m}\left(\theta_1 x_1^{(i)}\right) = \sum_{i=1}^{m}\left(y^{(i)} - \theta_0\right)$$

$$\sum_{i=1}^{m}(\theta_1) = \sum_{i=1}^{m}\left(\frac{y^{(i)} - \theta_0}{x_1^{(i)}}\right)$$