Graded Assignment 1

Exercise 1

(a) Vectorized hypothesis:

$$h_\theta\left(x^{(i)}\right) = \frac{1}{1 + e^{-\theta^{(i)\,T} \cdot x^{(i)}}}$$

As logistic regression is used, the hypothesis function uses this standard fraction model instead of just a dot multiplication. For this to work with vectors, the θ *and* x *need to be dot multiplied.*

(b) Vectorized cost function:

$$J(\theta) = \frac{1}{m}\sum_{i=1}^{m} y^{(i)} \log h_\theta\left(x^{(i)}\right) + \left(1 - y^{(i)}\right) * \log\left(1 - h_\theta\left(x^{(i)}\right)\right)$$

This logistic function is taking into account when y = 1 *or* y = 0*, to prevent going to infinity and causing an overflow. Dividing by matrix* x' *width* m *should provide the cost.*

(c) Vectorized gradient function:

$$\theta \coloneqq \frac{1}{m}\sum_{i=1}^{m}\left[x^{(i)} \cdot (h_\theta\left(x^{(i)}\right) - y^{(i)})\right]$$

This function is similar to the vectorized cost function, where it is simply a dot multiplication of the data matrix and the hypothesis minus y. Similar to (b), the width of matrix x *is used to divide the sum.*

(d) Vectorized update expression:

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

The update expression is the current cost updated: minus the gradient function times the alpha. This results in the new cost.

(e) Removing the explicit summation by putting all data in a matrix:

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

Because matrix dot multiplication already includes summation, it is no longer necessary to include an explicit summation in the cost function formula. This is what is used in the code of the Notebook. Using `sum()` *functions is not necessary because of this.*

Exercise 2

To obtain the optimal values of the hypothesis of a univariate linear regression ($h_\theta = \theta_0 + \theta_1 x$) without gradient descent, the derivative of the cost function must be set to zero. As θ_0 is fixed in this example, we only need to look at θ_1.

By taking the cost function and singling out the summation, we get the following:

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

$$2m * J(\theta) = \sum_{i=1}^{m}\left(h_\theta\left(x^{(i)}\right) - y^{(i)}\right)^2$$

Deriving and taking the numbers out of the sum gives us this formula. This can be done with the chain rule.

$$\frac{\partial S}{\partial C} = -2\sum_{i=1}^{m}\left(h_\theta\left(x^{(i)}\right) - y^{(i)}\right)$$

Now, equal it to zero. This means that the -2 *part is no longer relevant.*

$$\sum_{i=1}^{m} h_\theta\left(x^{(i)}\right) - \sum_{i=1}^{m} y^{(i)} = 0$$

We can describe the left element this as $m * h_\theta\left(x^{(i)}\right)$*:*

$$\sum_{i=1}^{m} h_\theta\left(x^{(i)}\right) = \sum_{i=1}^{m} y^{(i)} = m * h_\theta\left(x^{(i)}\right) = m * y^{(i)}$$

Then, rewriting the function gives us the following:

$$h_\theta\left(x^{(i)}\right) = \frac{\sum_{i=1}^{m} y^{(i)}}{m} = y^{(i)}$$

As h_θ is only dependent on θ_1 and not θ_0 in this case, this formula should provide the minimalized θ_1 without using gradient descent.