Assignment 2

Wendy Nieuwkamer

1 Question 1

This question is about *vectorization*,i.e. writing expressions in matrix-vector form. The goal is to vectorize the update rule for multivariate linear regression.

Let θ be the parameter vector $\theta = \begin{pmatrix}\theta_0 & \theta_1 & \cdots & \theta_n\end{pmatrix}^T$ and let the i-th data vector be: $x^{(i)} = \begin{pmatrix}x_0 & x_1 & \cdots & x_n\end{pmatrix}^T$ where $x_0 = 1$.

1.1 Write the hypothesis function $h_\theta(x)$ as a vectorial expression.

The summation notation for the hypothesis function is:

$$h_\theta(x^{(i)}) = \theta_0 x_0^{(i)} + \theta_1 x_1^{(i)} + \theta_2 x_2^{(i)} + ... + \theta_n x_n^{(i)}$$

This is the same as the result of the following matrix multiplication:

$$h_\theta(x^{(i)}) = \theta^T x^{(i)}, \tag{1}$$

which is a vectorial expression.

1.2 What is the vectorized expression for the cost function: $J(\theta)$?

The cost function in the notation used up to now:

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

If we simply insert the vectorial notation of the hypothesis function from last question (1) we get:

$$J(\theta) = \frac{1}{2m}\sum_{i=1}^{m}(\theta^T x^{(i)} - y^{(i)})^2.$$

1.3 What is the vectorized expression for the gradient of the cost function?

i.e. what is:

$$\frac{\delta J(\theta)}{\delta \theta} = \begin{pmatrix} \frac{\delta J(\theta)}{\delta \theta_0} \\ \vdots \\ \frac{\delta J(\theta)}{\delta \theta_n} \end{pmatrix}$$

Again the explicit summation over the data vectors from the learning set is allowed here.

The notation we used up until now is:

$$\frac{\delta J(\theta)}{\delta \theta_j} = \frac{1}{m} \sum_{i=1}^{m} (h_\theta(x^{(i)}) - y^{(i)}) x_j^{(i)}.$$

If we integrate the vectorized notation of the hypothesis (1) again we get:

$$\frac{\delta J(\theta)}{\delta \theta_j} = \frac{1}{m} \sum_{i=1}^{m} (\theta^T x^{(i)} - y^{(i)}) x_j^{(i)}.$$

Substituting this summation for the δ notation will give us the following vector:

$$\frac{\delta J(\theta)}{\delta \theta} = \frac{1}{m} \begin{pmatrix} \sum_{i=1}^{m} (\theta^T x^{(i)} - y^{(i)}) \\ \sum_{i=1}^{m} (\theta^T x^{(i)} - y^{(i)}) x_1^{(i)} \\ \vdots \\ \sum_{i=1}^{m} (\theta^T x^{(i)} - y^{(i)}) x_n^{(i)} \end{pmatrix} \tag{2}$$

1.4 What is the vectorized expression for the θ update rule in the gradient descent procedure?

The original notation for the update rule for one theta was:

$$\theta_j = \theta_j - \alpha \frac{1}{m} \sum_{i=1}^{m} (h_\theta(x^{(i)}) - y^{(i)}) x_i^{(i)}$$

By writing it in vector notation we update the entire theta instead of each element separately. Using the formulas from (1) and (2) we get:

$$\theta = \theta - \alpha \frac{1}{m} \begin{pmatrix} \sum_{i=1}^{m} (\theta^T x^{(i)} - y^{(i)}) \\ \sum_{i=1}^{m} (\theta^T x^{(i)} - y^{(i)}) x_1^{(i)} \\ \vdots \\ \sum_{i=1}^{m} (\theta^T x^{(i)} - y^{(i)}) x_n^{(i)} \end{pmatrix}$$

2 Question 2

Consider events with two binary outcomes, X and Y. We encode the two values as 0 and 1. We can represent the outcomes of an experiment in a 2 by 2 frequency table:

	X=0	X=1
Y=0	a	b
Y=1	c	d

Express the following as a function of a/b/c/d:

2.1 $P(X = 0)$

$X = 0$ for both the combinations $X = 0, Y = 0$, and $X = 0, Y = 1$. Thus, we are looking at a and c. Then, according to the standard formula:

$$P(X = 0) = \frac{a + c}{a + b + c + d}$$

2.2 $P(X = 1|Y = 0)$