Assignment 2

Wendy Nieuwkamer

1 Question 1

This question is about *vectorization*, writing in vector form, of the update rule for multivariate linear regression.

1.1 Write out the update rule

Give the update rule for a a single θ and data $x_1, x_2, ..., x_n$ using scalars and the dots notation that is used in this sentence for the x variables.

Gradient Descent Algorithm for Multivariate Linear Regression

Repeat until convergence {

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

}

The hypothesis h_θ is:

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

So, the desired notation is:

$$\theta_j := \theta_j - \alpha \frac{1}{m} \sum_{i=1}^{m} (x_j^{(i)} \theta_0 x_0^{(i)} + x_j^{(i)} \theta_1 x_1^{(i)} + x_j^{(i)} \theta_2 x_2^{(i)} + ... + x_j^{(i)} \theta_n x_n^{(i)} - x_j^{(i)} y^{(i)})$$

1.2 Write the hypothesis in vector notation

The hypothesis h_θ is:

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

The hypothesis written in vector notation is:

$$h_\theta(x^{(i)}) = \theta^T x^{(i)} = \begin{bmatrix} \theta_0 & \theta_1 & \theta_2 & \cdots & \theta_n \end{bmatrix} \begin{bmatrix} x_0^{(i)} \\ x_1^{(i)} \\ x_2^{(i)} \\ \vdots \\ x_n^{(i)} \end{bmatrix}$$

1.3 Write the gradient for θ_0 in vector notation

The gradient in standard notation is:

$$\frac{\delta}{\delta\theta_0} J(\theta) = \frac{1}{m} \sum_{i=1}^{m} (h_\theta(x^{(i)}) - y^{(i)})$$

The same in vector notation is:

$$\begin{aligned} \frac{\delta}{\delta\theta_0} J(\theta) &= \frac{1}{m} \begin{bmatrix} 1 & 1 & 1 & \cdots & 1 \end{bmatrix} \cdot (\theta^T x^{(i)} - y^{(i)}) \\ &= \frac{1}{m} x_0^T \cdot (\theta^T x^{(i)} - y^{(i)}) \end{aligned}$$

1.4 Write the gradient for θ_j in vector notation for $i > 0$

The gradient in standard notation is:

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

The same in vector notation is:

$$\begin{aligned} \frac{\delta}{\delta\theta_i} J(\theta) &= \frac{1}{m} \begin{bmatrix} x_i^{(1)} & x_i^{(2)} & x_i^{(3)} & \cdots & x_i^{(n)} \end{bmatrix} \cdot (\theta^T x^{(i)} - y^{(i)}) \\ &= \frac{1}{m} x_i^T \cdot (\theta^T x^{(i)} - y^{(i)}) \end{aligned}$$

1.5 Write the entire update rule in vector notation

Combining the previous answers gives:

Gradient Descent Algorithm for Multivariate Linear Regression

Repeat until convergence {

$$\theta := \theta - \alpha \frac{1}{m} x^T \cdot (\theta \cdot x - y)$$

}