Machine Learning
Written Assignment 1

Simon Fang
10898492

September 16, 2016

Exercise 2

(a) In order to calculate two iterations of the gradient descent algorithm, we are given the following algorithm:

Repeat until convergence{

$$\theta_0 := \theta_0 - \alpha \frac{1}{m} \sum_{i=1}^{m} (H_\theta(x_i) - y_i)$$

$$\theta_1 := \theta_1 - \alpha \frac{1}{m} \sum_{i=1}^{m} (H_\theta(x_i) - y_i) x_i$$

}

where $h_\theta(x) = \theta_0 + \theta_1 x$, α is the learning rate and for each iteration, we update the values of θ simultaneously

In the problem, we are given that $\theta_0 = 0$, $\theta_1 = 1$ and $\alpha = 0.1$ as the initial values. Using these values, we compute the first iteration:

$$\begin{aligned}
\theta_0 &:= \theta_0 - \alpha \frac{1}{m} \sum_{i=1}^{m} (h_\theta(x_i) - y_i) \\
&= 0 - 0.1 \ \frac{1}{3} \ (3 - 6 + 5 - 7 + 6 - 10) \\
&= 0.3
\end{aligned}$$

$$
\begin{aligned}
\theta_1 &:= \theta_1 - \alpha \frac{1}{m} \sum_{i=1}^{m} (h_\theta(x_i) - y_i)x_i \\
&= 1 - 0.1 \ \frac{1}{3} \ [(3-6)3 + (5-7)5 + (6-10)6] \\
&= 2.43
\end{aligned}
$$

We use the results from iteration 1 and we update the values of θ_0 and θ_1. Then we proceed with the second iteration:

$$
\begin{aligned}
\theta_0 &:= \theta_0 - \alpha \frac{1}{m} \sum_{i=1}^{m} (h_\theta(x_i) - y_i) \\
&= 0.3 - 0.1 \ \frac{1}{3}(0.3 + 2.43 \cdot 3 - 6 + 0.3 + 2.43 \cdot 5 - 7 + 0.3 + 2.43 \cdot 6 - 10) \\
&= -0.097
\end{aligned}
$$

$$
\begin{aligned}
\theta_1 &:= \theta_1 - \alpha \frac{1}{m} \sum_{i=1}^{m} (h_\theta(x_i) - y_i)x_i \\
&= 2.43 - 0.1 \ \frac{1}{3}[(0.3 + 2.43 \cdot 3 - 6)3 + (0.3 + 2.43 \cdot 5 - 7)5 + (0.3 + 2.43 \cdot 6 - 10)6] \\
&= 0.39
\end{aligned}
$$

After having computed two iterations of the gradient descent algorithm, we proceed with calculating the Mean Squared Error. We can calculate this using the following formula:

$$
\text{MSE} = \frac{1}{m} \sum_{i=1}^{m} \left(\hat{Y}_i - Y_i\right)^2
$$

In our case, we will have to use the hypothesis function as our function, which is given as: $h_\theta(x) = \theta_0 + \theta_1 x$ and the MSE will be:

$$
\begin{aligned}
\text{MSE} &= \frac{1}{m} \sum_{i=1}^{m} \left(h_\theta(x_i) - y_i\right)^2 \\
&= \frac{1}{3}[(-0.098 + 0.39 \cdot 3 - 6)^2 + (0.098 + 0.39 \cdot 6 - 7)^2 + (-0.097 + 0.39 \cdot 6 - 10)^2] \\
&= 37.2
\end{aligned}
$$

(b) In order to calculate the z score, we will use the following formula:

$$z = \frac{X - \mu}{\sigma}$$

where μ is the mean, σ is the standard deviation and X is the raw score. We are given that $\mu = 0$ and $\sigma = 1$. If we plug in these values into z, then the z score becomes equal to the raw score. Therefore, all the values will remain the same and the calculations will be identical. So see part (a) for the calculations and values.

Exercise 4

Finding the optimal value of the parameter θ_1 for univariate linear regression without doing gradient descent and assuming that θ_0 is fixed is only possible if the following holds:

$$\theta_1 := \theta_1 - \alpha \frac{\partial J}{\partial \theta_1} = \theta_1$$

This implies that the partial derivative of the cost function with respect to θ_1 must be equal to 0, which gives us the following equation:

$$\begin{aligned}\frac{\partial J}{\partial \theta_1} &= \frac{\partial}{\partial \theta_1} \frac{1}{2m} \sum_{i=1}^{m} (h_\theta(x_i) - y_i)^2 \\ &= \frac{1}{m} \sum_{i=1}^{m} (\theta_0 + \theta_1 x_i - y_1) x_i = 0\end{aligned}$$

We may assume that m is nonzero and that θ_0 is fixed. We continue by multiplying both sides with m and we re-arrange the equation, which gives us the following:

$$\theta_0 x_1 + \theta_1 x_1 x_1 - y_1 x_1 + \theta_0 x_2 + \theta_1 x_2 - y_2 x_2 + \ldots + \theta_0 x_m + \theta_1 x_m x_m - y_m x_m = 0$$

Then we factorize the θ's and we get the following:

$$\theta_0 (x_1 + \ldots + x_m) + \theta_1 (x_1 x_1 + \ldots + x_m x_m) - (y_1 x_1 + \cdots + y_m x_m) = 0$$

We keep θ_1 on the left and we move everything else on the right. Then we divide by the factor of θ_1, which gives us the optimal value of θ_1:

$$\begin{aligned}\theta_1^* &= \frac{(y_1 x_1 + \ldots + y_m x_m) - \theta_0 (x_1 + \ldots + x_m)}{x_1 x_1 + \ldots + x_m x_m} \\ &= \frac{\sum_{i=1}^{m} y_i x_i - \theta_0 \sum_{i=1}^{m} x_i}{\sum_{i=1}^{m} x_i x_i}\end{aligned}$$