Machine Learning - Assignment 1

Vincent Roest 10904816

September 14, 2016

Question 2

2a) We can use the following formulas, the hypothesis function, cost function and theta-update function respectively

$$\begin{aligned}
h_\theta(x^{(i)}) &= \theta_0 + theta_1 x^{(i)} \\
J(\theta_0, \theta_1) &= \frac{1}{2m}\sum_{i=1}^{m}(h_\theta(x^{(i)}) - y^{(i)})^2 \\
\theta_{(j)} &= \theta_{(j)} - \alpha\frac{\partial}{\partial\theta_{(j)}}J(\theta_j), \text{with j} = \{0, 1\}
\end{aligned} \tag{1}$$

From the question we can initialize θ_0 such that the regression function passes the origin, namely by choosing it to be 0: $\theta_0 = 0$. Moreover, if we choose θ_1 to be 1, we have an angle of 45°, because the function h_θ now resembles the function $y = x$, which passes through the origin with a 45 degree angle. Also, we know that $\alpha = 0.1$ and $m = 3$.

We can now also calculate the values for the regression function using the given data set and the initial values described above:

$$\begin{aligned}
&h_\theta(x^{(i)}) &&= \theta_0 + \theta_1 x^{(i)} \\
&\text{Which we fill in for every value of x in the data set:} && \\
&h_\theta(3) &&= 0 + 1 * 3 = 3 \\
&h_\theta(5) &&= 0 + 1 * 5 = 5 \\
&h_\theta(6) &&= 0 + 1 * 6 = 6
\end{aligned} \tag{2}$$

Then, we can do the first iteration:

$$\begin{aligned}
\theta_0 &:= \theta_0 - \alpha\frac{1}{m}\sum_{i=1}^{3}(h_\theta(x^{(i)}) - y^{(i)}) \\
&:= 0 - 0.1 * \tfrac{1}{3}((3-6) + (5-7) + (6-10)) \\
&:= 0 - 0.1 * \tfrac{1}{3}(-9) \\
&:= 0 - 0.1 * -3 \\
&:= 0.3 \\
\theta_1 &:= \theta_1 - \alpha\frac{1}{m}\sum_{i=1}^{3}(h_\theta(x^{(i)}) - y^{(i)}) * x^{(i)} \\
&:= 1 - 0.1 * \tfrac{1}{3}((3-6)*3 + (5-7)*5 + (6-10)*6) \\
&:= 1 - 0.1 * \tfrac{1}{3}(-43) \\
&:= 1 - \tfrac{43}{30} = \tfrac{73}{30}
\end{aligned} \tag{3}$$

Now we update:

$$\begin{aligned}
&h_\theta(x^{(i)}) &&= \theta_0 + \theta_1 x^{(i)} \\
&\text{Which we fill in for every value of x in the data set:} && \\
&h_\theta(3) &&= 0.3 + \tfrac{73}{30} * 3 = 7.59 \\
&h_\theta(5) &&= 0.3 + \tfrac{73}{30} * 5 = 12.45 \\
&h_\theta(6) &&= 0.3 + \tfrac{73}{30} * 6 = 14.88
\end{aligned} \tag{4}$$

Repeat for second iteration:

$$\begin{aligned}
\theta_0 &:= \theta_0 - \alpha\tfrac{1}{m}\textstyle\sum_{i=1}^{3}(h_\theta(x^{(i)}) - y^{(i)}) \\
&:= 0.3 - 0.1 * \tfrac{1}{3}((7.59-6) + (12.45-7) + (14.88-10)) \\
&:= 0.3 - 0.1 * \tfrac{1}{3}(11.92) \\
&:= 0.3 - 0.1 * 3.97 \\
&:= -0.097 \\
\theta_1 &:= \theta_1 - \alpha\tfrac{1}{m}\textstyle\sum_{i=1}^{3}(h_\theta(x^{(i)}) - y^{(i)}) * x^{(i)} \\
&:= \tfrac{73}{30} - 0.1 * \tfrac{1}{3}((7.59-6)*3 + (12.45-7)*5 + (14.88-10)*6) \\
&:= \tfrac{73}{30} - 0.1 * \tfrac{1}{3}(61.3) \\
&:= \tfrac{73}{30} - 0.1 * 20.43 \\
&:= 0.39
\end{aligned} \tag{5}$$

Now we can finalize the values for h_θ to compute the MSE:

$$\begin{aligned}
h_\theta(3) &= -0.097 + 0.39 * 3 = 1.07 \\
h_\theta(5) &= -0.097 + 0.39 * 5 = 1.85 \\
h_\theta(6) &= -0.097 + 0.39 * 6 = 2.24
\end{aligned} \tag{6}$$

Then, we calculate the MSE from the formula we know:

$$\begin{aligned}
MSE &= \tfrac{1}{2m}\textstyle\sum_{i=1}^{3}(h_\theta(x^{(i)}) - y^{(i)})^2 \\
&= \tfrac{1}{6}((1.07-6)^2 + (1.85-7)^2 + (2.24-10)^2) \\
&= \tfrac{1}{6} * 111.05 \\
&= 18.51
\end{aligned} \tag{7}$$

2b) By definition, Z-scores are given by: $Z = \frac{X-\mu}{\sigma}$. Since the question specifies $\mu = 0$ and $\sigma = 1$, we basically do not change Z every time we calculate the Z-value, they are simply the old values of for X! Therefore, any calculations won't be necessary.

Question 4

We know that the cost function is given by:

$$J(\theta_0, \theta_1) = \frac{1}{2m}\sum_{i=1}^{m}(h_\theta(x^{(i)}) - y^{(i)})^2 \tag{8}$$

For the hypothesis function, we use the usual definition: $h_\theta(x^{(i)}) = \theta_0 + \theta_1 x^{(i)}$. Taking the derivative with respect to θ_0 of the cost function yields:

$$\frac{\partial}{\partial\theta_0}J(\theta_0, \theta_1) = \frac{1}{m}\sum_{i=1}^{m}(h_\theta(x^{(i)}) - y^{(i)}) \tag{9}$$

We set this to zero:

$$\begin{aligned}
\tfrac{\partial}{\partial\theta_0}J(\theta_0, \theta_1) &= \tfrac{1}{m}\textstyle\sum_{i=1}^{m}(h_\theta(x^{(i)}) - y^{(i)}) = 0 \\
\tfrac{1}{m}\textstyle\sum_{i=1}^{m}((\theta_0 + \theta_1 * x^{(i)}) - y^{(i)}) &= 0 \\
\theta_0 + \theta_1 * \textstyle\sum_{i=1}^{m} x^{(i)}) - \sum_{i=1}^{m} y^{(i)} &= 0 \\
\textstyle\sum_{i=1}^{m} y^{(i)} - \theta_1 * \sum_{i=1}^{m} x^{(i)}) &= \theta_0 \\
\theta_0 &= \bar{y} - \theta_1 * \bar{x}
\end{aligned} \tag{10}$$

Where the bars represent the mean values of the variables.Then we can do the same for the θ_1:

$$\begin{aligned}
\frac{\partial}{\partial\theta_1}J(\theta_0,\theta_1) &= \frac{1}{m}\textstyle\sum_{i=1}^{m}((\theta_0+\theta_1 * x^{(i)} - y^{(i)}) * x^{(i)}) = 0 \\
\textstyle\sum_{i=1}^{m}(\theta_0 x^{(i)} + \theta_1 x^{(i)} x^{(i)} - y^{(i)} x^{(i)}) &= 0 \\
\text{Substitution for } \theta_0 \text{ since this is fixed:} & \\
\textstyle\sum_{i=1}^{m}((\bar{y} - \theta_1 * \bar{x}) * x^{(i)} + \theta_1 x^{(i)} x^{(i)} - y^{(i)} x^{(i)}) &= 0 \\
\textstyle\sum_{i=1}^{m}(\bar{y} * x^{(i)} - \theta_1 * \bar{x} * x^{(i)} + \theta_1 x^{(i)} x^{(i)} - y^{(i)} x^{(i)}) &= 0 \\
\textstyle\sum_{i=1}^{m}(\bar{y} - \theta_1 * \bar{x} + \theta_1 x^{(i)} - y^{(i)}) * x^{(i)} &= 0 \\
\textstyle\sum_{i=1}^{m}(\bar{y} + (x^{(i)} - \bar{x}) * \theta_1 - y^{(i)}) &= 0 \\
\textstyle\sum_{i=1}^{m}(\bar{y} - y^{(i)}) &= -\theta_1 \textstyle\sum_{i=1}^{m}(x^{(i)} - \bar{x}) \\
\theta_1 &= -\frac{\sum_{i=1}^{m}(\bar{y} - y^{(i)})}{\sum_{i=1}^{m}(x^{(i)} - \bar{x})} \\
\theta_1 &= \frac{\sum_{i=1}^{m}(y^{(i)} - \bar{y})}{\sum_{i=1}^{m}(x^{(i)} - \bar{x})}
\end{aligned} \tag{11}$$

So, now we have an expression for the optimal value of θ_1.