Machine Learning – Written Assignment 1

Philip Hartout

September 15, 2016

2. (a) Given the following data, manually (using only a calculator) calculate two iterations of the gradient descent algorithm for univariate linear regression function. Initialize the parameters such that the regression function passes through the origin (0, 0) and has an angle of 45 degrees. Use a learning rate of 0.1. Give the intermediate results of your calculations and also compute the mean-squared error of the function after 2 iterations.

x	**y**
3	6
5	7
6	10

Table 1: Training sample

Gradient descent algorithm application

We set our learning rate $\alpha = 0.1$. Recall that $J(\theta_0, \theta_1) = \frac{1}{2m}\sum_{i=1}^{m}(h_\theta(x^{(i)}) - y^{(i)})^2$
Initialisation
$\theta_0 = 0$, $\theta_1 = 1$. So, $h_\theta(x) = 0 + 1 \cdot x$
Variable update #1
Now we update θ_0 and θ_1

$$\begin{aligned}
&\theta_0 := \theta_0 - \alpha \cdot \frac{\partial}{\partial \theta_0} J(\theta_0) \\
\Leftrightarrow &\theta_0 := \theta_0 - \alpha \frac{1}{m} \sum_{i=1}^{m} h_\theta(x^{(i)}) - y^{(i)} \\
\Leftrightarrow &\theta_0 := 0 - 0.1 \frac{1}{3}\left[(3-6) + (5-7) + (6-10)\right] \\
\Leftrightarrow &\theta_0 := 0.3
\end{aligned}$$

Now, we update θ_1 simultaneously, i.e. we still use our hypothesis function we had in the beginning: $h(x) = 0 + 1 \cdot x$. This yields:

$$\begin{aligned}
&\theta_1 := \theta_1 - \alpha \cdot \frac{\partial}{\partial \theta_1} J(\theta_1) \cdot x^{(i)} \\
\Leftrightarrow &\theta_1 := \theta_1 - \alpha \frac{1}{m} \sum_{i=1}^{m} (h_\theta(x^{(i)}) - y^{(i)}) \cdot x^{(i)} \\
\Leftrightarrow &\theta_1 := 1 - 0.1 \frac{1}{3}\left[(3-6) \cdot 3 + (5-7) \cdot 5 + (6-10) \cdot 6 +\right] \\
\Leftrightarrow &\theta_1 := \frac{73}{30}
\end{aligned}$$

Variable update #2

Now we update θ_0 and θ_1 with $h_\theta(x) = 0.3 + \frac{73}{30} \cdot x$

$$\begin{aligned}
&\theta_0 := \theta_0 - \alpha \cdot \frac{\partial}{\partial \theta_0} J(\theta_0) \\
\Leftrightarrow &\theta_0 := \theta_0 - \alpha \frac{1}{m} \sum_{i=1}^{m} h_\theta(x^{(i)}) - y^{(i)} \\
\Leftrightarrow &\theta_0 := 0.3 - 0.1 \frac{1}{3} \left[(7.6 - 6) + (\frac{187}{15} - 7) + (\frac{149}{10} - 10) \right] \\
\Leftrightarrow &\theta_0 := -\frac{89}{900}
\end{aligned}$$

Now, we update θ_1 simultaneously, i.e. we still use our hypothesis function we had in the beginning: $h(x) = 0 + 1 \cdot x$. This yields:

$$\begin{aligned}
&\theta_1 := \theta_1 - \alpha \cdot \frac{\partial}{\partial \theta_1} J(\theta_1) \cdot x^{(i)} \\
\Leftrightarrow &\theta_1 := \theta_1 - \alpha \frac{1}{m} \sum_{i=1}^{m} (h_\theta(x^{(i)}) - y^{(i)}) \cdot x^{(i)} \\
\Leftrightarrow &\theta_1 := \frac{73}{30} - 0.1 \frac{1}{3} \left[(7.6 - 6) \cdot 3 + (\frac{187}{15} - 7) \cdot 5 + (\frac{149}{10} - 10) \cdot 6 \right] \\
\Leftrightarrow &\theta_1 := \frac{86}{225}
\end{aligned}$$

Conclusion

We end up with the following hypothesis fitting our data: $h_\theta(x) = -\frac{89}{900} + \frac{86}{225} \cdot x$.
The mean square error (MSE) for this hypothesis is calculated as follows:

$$\begin{aligned}
&\text{MSE} = \frac{1}{m} \sum_{i=1}^{m} \left(h_\theta(x) - y^{(i)} \right)^2 \\
\Leftrightarrow &\text{MSE} = \frac{1}{3} \left[\left(\frac{943}{900} - 6 \right)^2 + \left(\frac{1631}{900} - 7 \right)^2 + \left(\frac{79}{36} - 10 \right)^2 \right] \\
\Leftrightarrow &\text{MSE} = \frac{6067669}{162000} \approx 37.45
\end{aligned}$$

(b) Convert the data to z-scores (with mean = 0, sd = 1) and repeat the calculations above. Compare the results with those for the original data.
Given the data is exactly the same[1], we will not observe any difference with the results of the previous iterations of the gradient descent algorithm.

x	**y**
3	6
5	7
6	10

Table 2: *z*-scores of the training sample

4. Derive an equation that can be used to find the optimal value of the parameter θ_1 for univariate linear regression without doing gradient descent. This can be done by setting the value of the derivative equal to 0. You may assume that the value of θ_0 is fixed.
We can seek to minimise the cost function $J(\theta_0, \theta_1) = \frac{1}{2m} \sum_{i=1}^{m} (h_\theta(x^i) - y^{(i)})^2$ by setting its partial derivative with

[1]Indeed, $\frac{x-0}{1} = x$

respect to θ_1 to 0. As instructed, we consider θ_0 to be a fixed constant, so that we consider $J(\theta_1)$. This yields:

$$\begin{aligned} 0 &= \frac{\partial}{\partial\theta_1} J(\theta_1) \\ \Leftrightarrow 0 &= \frac{\partial}{\partial\theta_1} \frac{1}{2m} \sum_{i=1}^{m} (h_\theta(x^i) - y^{(i)})^2 \cdot x^{(i)} \\ \Leftrightarrow 0 &= \frac{1}{m} \sum_{i=1}^{m} (h_\theta(x^i) - y^{(i)}) \cdot x^{(i)}. \end{aligned}$$