DL-ENG-015-JUN25

Module 1 Assignment: Regresion Model Optimization

Student: Jia Pei, jiapeish@gmail.com

First, let us do some manually calculations:

(a) let us assume that $Y = X + 2$, then we can compute the
$Y_{estimated} = [3, 0, 5, 6.5, 2, -2, 1, 6, 1]^T$,
according to the loss function $J(\theta) = \dfrac{1}{9} \|Y - Y_{estimated}\|^2$, we can calculate that $J(\theta) = 2.944$

(b) let us assume that $Y = X + 3$, then we can compute the
$Y_{estimated} = [4, 1, 6, 7.5, 3, -1, 2, 7, 2]^T$,
we can calculate that $J(\theta) = 2.833$

(c) let us assume that $Y = 0.5X + 3$, then we can compute the
$Y_{estimated} = [3.5, 2, 4.5, 5.25, 3, 1, 2.5, 5, 2.5]^T$,
we can calculate that $J(\theta) = 5.813$

(d) let us assume that $Y = 1.5X + 2$, then we can compute the
$Y_{estimated} = [3.5, -1, 6.5, 8.75, 2, -4, 0.5, 8, 0.5]^T$,
we can calculate that $J(\theta) = 3.267$

Since 2.833 is the smallest, we can resonable assume that $Y = X + 3$ is near the ideal value of θ.

Now let's compute the gradient, since $J_{\mathrm{OLS}}(\theta, \theta_0) = \frac{1}{n}\sum_{i=1}^{n}\left(\mathbf{x}^{(i)}\theta^\top + \theta_0 - y^{(i)}\right)^2$, we can note $\hat{y}^{(i)} = x^{(i)}\theta^\top + \theta_0$, then we can get the following partial value

$$\frac{\partial J}{\partial \theta_0} = \frac{1}{n}\sum_{i=1}^{n} 2(\hat{y}^{(i)} - y^{(i)}) \cdot \frac{\partial \hat{y}^{(i)}}{\partial \theta_0} = \frac{2}{n}\sum_{i=1}^{n}(\hat{y}^{(i)} - y^i)$$

$$\frac{\partial J}{\partial \theta} = \frac{1}{n}\sum_{i=1}^{n} 2(\hat{y}^{(i)} - y^{(i)}) \cdot \frac{\partial \hat{y}^{(i)}}{\partial \theta} = \frac{2}{n}\sum_{i=1}^{n}(\hat{y}^{(i)} - y^i) \cdot x^{(i)}$$

With Excel, let we set $\theta_0 = 3$ and $\theta = [\theta_1] = 1$, then we can get the gradient is $\frac{\partial J}{\partial \theta_0} = 0.098765432$ and $\frac{\partial J}{\partial \theta_1} = -0.240740741$, so it's time to update the θ_0 and θ_1, $\theta_0^{new} = \theta_0^{old} - \eta\frac{\partial J}{\partial \theta_0}$ and $\theta_1^{new} = \theta_1^{old} - \eta\frac{\partial J}{\partial \theta_1}$.

1st iteration: $\theta_0^{new} = 2.999012346, \theta_1^{new} = 1.002407407, J_\theta = 2.833333333$

2nd iteration: $\theta_0^{new} = 2.99802421, \theta_1^{new} = 1.004776422, J_\theta = 2.827279578$

let's change learning rate from 0.01 to 0.1, then

3rd iteration: $\theta_0^{new} = 2.988138502, \theta_1^{new} = 1.028067534, J_\theta = 2.821395635$

4nd iteration: $\theta_0^{new} = 2.978212466, \theta_1^{new} = 1.047634834, J_\theta = 2.767566167$

5nd iteration: $\theta_0^{new} = 2.968289588, \theta_1^{new} = 1.064015344, J_\theta = 2.72724026$

6nd iteration: $\theta_0^{new} = 2.958405214, \theta_1^{new} = 1.077745659, J_\theta = 2.696362973$

...

We draw the estimated value of $\hat{y}$ in the Excel, they are the red points and red line. We can see that there are some differences:

1. The slope of the lines are a bit different

2. The main difference is due to the left down points, they scatter far from the lines.

3. Google's line seems more balanced than my estimated line, since 2 points are above my line, whereas there are 3 points below my line in the left down region.