Linear Regression with One Variable

Or Yuet Tung Esther

1 Squared Error Cost Function

The error when parameters w and b are set to a particular value respectively is calculated by the squared error cost function as follows:

$$J(w,b) = \frac{1}{2m}\sum_{i=0}^{m-1}(\hat{y}^{(i)} - y^{(i)})^2$$

Intuitively, the larger the value of $J(w,b)$, the larger the error. Thus, our goal is to minimize $J(w,b)$ by choosing the most suitable values of w and b.

2 Gradient Descent

To minimize $J(w,b)$, we may use an optimization algorithm called gradient descent.

2.1 Analogy

Imagine if you are standing on a mountain and your goal is to get to the lowest point. You look around and see which paths are downward sloping. After that, you take a step towards the downward-sloping path. Then, you repeat the process above until you cannot go down further anymore. The ultimate location is the point where you will settle.

2.2 Implementation

Firstly, you start with some w and b. Then, you repeat the following until convergence:

$$tmp_w = w - \alpha\frac{\partial}{\partial w}J(w,b) \tag{1}$$

$$tmp_b = b - \alpha\frac{\partial}{\partial b}J(w,b) \tag{2}$$

$$w = tmp_w \tag{3}$$

$$b = tmp_b \tag{4}$$

where α is the learning rate and $\alpha > 0$

2.3 Intuition

$\frac{\partial}{\partial b}J(w,b)$ and $\frac{\partial}{\partial w}J(w,b)$ are slopes. For easier understanding, let us assume that $b=0$. If the "path" is upward sloping, the slope would be positive. Thus, $tmp_w = w - \alpha \times (positive\,number)$, and w would decrease so that $J(w,b)$ is smaller. (See Figure 1)

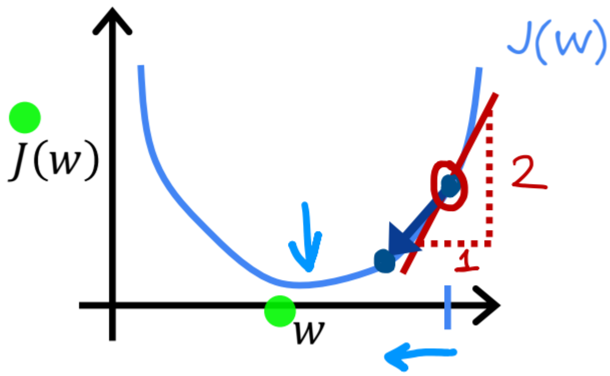

Figure 1: Upward-sloping path

In contrast, the slope would be negative if the "path" is downward sloping. Therefore, $tmp_w = w - \alpha \times (negative\,number)$, and w would increase.

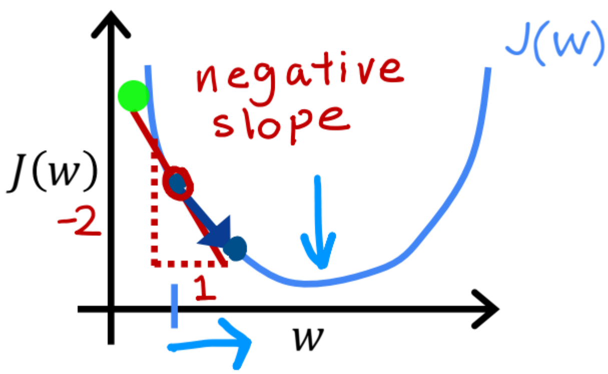

Figure 2: Downward-sloping path

3 Remarks

- Figures are taken from Professor Andrew Ng's course
- My assignment on Linear Regression: Github Link