Gradient Descent Linear Regression

Sahil Bharodiya

Gradient descent is optimization tool in machine learning.

If hypothesis function is $h_\theta(x) = \theta_0 + \theta_1 x$,

Then cost function will be,

$$J(\theta_0, \theta_1) = \frac{1}{2m} \sum_{i=1}^{m} (h_\theta(x_i) - y_i)^2$$

Here notice that I have taken $\frac{1}{2m}$ for sake of convenience in gradient descent.

So gradient descent is,

$$\begin{aligned}
\nabla J(\theta_0, \theta_1) &= \nabla \left(\frac{1}{2m} \sum_{i=1}^{m} (h_\theta(x_i) - y_i)^2 \right) \\
&= \nabla \left(\frac{1}{2m} \sum_{i=1}^{m} (\theta_0 + \theta_1 x_i - y_i)^2 \right) \\
&= \frac{1}{2m} \times 2 \sum_{i=1}^{m} (\theta_0 + \theta_1 x_i - y_i) \nabla (\theta_0 + \theta_1 x_i - y_i) \\
&= \frac{1}{m} \sum_{i=1}^{m} (\theta_0 + \theta_1 x_i - y_i) \nabla (\theta_0 + \theta_1 x_i - y_i)
\end{aligned}$$

Now, take $\nabla J(\theta_0, \theta_1) = \frac{\partial}{\partial \theta_0} J(\theta_0, \theta_1)$ for finding optimal value of θ_0.

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

take $\nabla J(\theta_0, \theta_1) = \frac{\partial}{\partial \theta_1} J(\theta_0, \theta_1)$ for finding optimal value of θ_1.

$$\frac{\partial}{\partial \theta_1} J(\theta_0, \theta_1) = \frac{1}{m} \sum_{i=1}^{m} (\theta_0 + \theta_1 x_i - y_i) \frac{\partial}{\partial \theta_1} (\theta_0 + \theta_1 x_i - y_i)$$

$$= \frac{1}{m}\sum_{i=1}^{m}(\theta_0 + \theta_1 x_i - y_i) \times (0 + x_i - 0)$$

$$= \frac{1}{m}\sum_{i=1}^{m}(\theta_0 + \theta_1 x_i - y_i) x_i$$

$$= \frac{1}{m}\sum_{i=1}^{m}(h_\theta(x_i) - y_i) x_i$$

Example: $f(\theta_0, \theta_1) = \theta_0^2 + \theta_1^3$ then find its gradient.

$\Rightarrow \nabla f = \nabla(\theta_0^2 + \theta_1^3)$

$$\frac{\partial}{\partial \theta_0} f = \frac{\partial}{\partial \theta_0}(\theta_0^2 + \theta_1^3)$$

$$= 2\theta_0 \frac{\partial}{\partial \theta_0}\theta_0 + 0$$

$$= 2\theta_0$$

$$\frac{\partial}{\partial \theta_1} f = \frac{\partial}{\partial \theta_1}(\theta_0^2 + \theta_1^3)$$

$$= 0 + 3\theta_1^2 \frac{\partial}{\partial \theta_1}\theta_1$$

$$= 3\theta_1$$

Similarly, you can apply it in also multi variable linear regression. In multi variable linear regression you have 'n' variables like $\theta_0, \theta_1, \theta_2, \ldots, \theta_n$.
