2. Gradient Descent for Multiple Variables

August 6, 2021

1 Cost function and Gradient Descent with Multple Features

From the previous section we derived a matrix expression of the hypothesis function for multiple features. Given $h_\theta(x) = \Theta^T \times X$ where Θ is given as $\begin{pmatrix} \theta_0 \\ \theta_1 \\ \vdots \\ \theta_n \end{pmatrix}$ (the parameters of the features) and X is given as $\begin{pmatrix} x_0 \\ x_1 \\ \vdots \\ x_n \end{pmatrix}$

Now our cost function can be given in terms of the feature vector Θ as below

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

Substituting this into the gradient descent algorithm, we get

$$\begin{aligned}
\{\theta_j &= \theta_j - \alpha \times \frac{\partial}{\partial\theta_j}J(\Theta) \\
&= \theta_j - \frac{\alpha}{2m}\sum_{i=1}^{m}\frac{\partial}{\partial\theta_j}(h_\theta(x^i) - y^i)^2 \\
&= \theta_j - \frac{\alpha}{2m}\sum_{i=1}^{m}\frac{\partial}{\partial\theta_j}(h_\theta(x^i) - y^i)^2 \\
&= \theta_j - \alpha\frac{1}{m}\sum_{i=1}^{m}(h_\theta(x^i) - y^i)x_j^i; forj = 1, ..., n\}
\end{aligned} \tag{2}$$

because we defined the feature x_0 in the previous section and the vector X contains it, we can arrive at the general definition of the gradient descent function defined in equation 2