## Derivation of Logistic Regression

Sahilbharodiya2002@outlook.com

Our hypothesis function can be defined as;

$$h_{\theta}(x) = g(\theta^T x)$$

Where;

$$g(z) = \frac{1}{1+e^{-z}}$$

Cost function will be;

$$J(\theta) = -\frac{1}{m} \, \sum_{i=1}^m \left[ y^{(i)} \log \left( h_{\theta}(x^{(i)}) \right) + \left( 1 - y^{(i)} \right) \log \left( 1 - h_{\theta}(x^{(i)}) \right) \right]$$

Cost function's vectorized implementation is;

$$h = q(X\theta)$$

$$J(\theta) = \frac{1}{m} \ \left( -y^T \log(h) - (1-y)^T \log(1-h) \right)$$

Now, coming to Gradient Descent

Remember the general form of the gradient descent is:

$$\theta_j = \theta_j - \alpha \frac{\partial}{\partial \theta_j} J(\theta)$$
 }

So, gradient descent for the logistic regression is;

$$\theta_j = \theta_j - \frac{\alpha}{m} \; \sum_{i=1}^m \bigl(h_\theta(x^{(i)}) - y^{(i)}\bigr) x_j^{(i)}$$
 }

And its vectorized form is;

$$\theta = \theta - \frac{\alpha}{m} \ X^T(g(X\theta) - y)$$

So, lets come to derivation;

Derivative of sigmoid function;

$$\begin{split} \sigma(x) &= \frac{1}{1 + e^{-x}} \\ \sigma(x)' &= \left(\frac{1}{1 + e^{-x}}\right)' \\ &= -\frac{(1 + e^{-x})'}{(1 + e^{-x})^2} = \frac{e^{-x}}{(1 + e^{-x})^2} = \sigma(x)\sigma(1 - x) \end{split}$$

Now take partial derivative of cost function;

$$\begin{split} \frac{\partial}{\partial \theta_j} J(\theta) &= \frac{\partial}{\partial \theta_j} - \frac{1}{m} \, \sum_{i=1}^m \left[ y^{(i)} \log \left( h_\theta(x^{(i)}) \right) + (1-y^{(i)}) \log \left( 1 - h_\theta(x^{(i)}) \right) \right] \\ &= -\frac{1}{m} \, \sum_{i=1}^m \left[ y^{(i)} \frac{\partial}{\partial \theta_j} \log \left( h_\theta(x^{(i)}) \right) + (1-y^{(i)}) \frac{\partial}{\partial \theta_j} \log \left( 1 - h_\theta(x^{(i)}) \right) \right] \\ &= -\frac{1}{m} \, \sum_{i=1}^m \left[ \frac{y^{(i)} \frac{\partial}{\partial \theta_j} h_\theta(x^{(i)})}{h_\theta(x^{(i)})} + \frac{(1-y^{(i)}) \frac{\partial}{\partial \theta_j} \left( 1 - h_\theta(x^{(i)}) \right)}{1 - h_\theta(x^{(i)})} \right] \\ &= -\frac{1}{m} \, \sum_{i=1}^m \left[ \frac{y^{(i)} \frac{\partial}{\partial \theta_j} \sigma(\theta^T x^{(i)})}{h_\theta(x^{(i)})} + \frac{(1-y^{(i)}) \frac{\partial}{\partial \theta_j} \left( 1 - \sigma(\theta^T x^{(i)}) \right)}{1 - h_\theta(\theta^T x^{(i)})} \right] \\ &= -\frac{1}{m} \, \sum_{i=1}^m \left[ \frac{y^{(i)} \sigma(\theta^T x^{(i)}) \left( 1 - \sigma(\theta^T x^{(i)}) \right) \frac{\partial}{\partial \theta_j} \theta^T x^{(i)}}{h_\theta(x^{(i)})} \right] \\ &= -\frac{1}{m} \, \sum_{i=1}^m \left[ y^{(i)} \left( 1 - h_\theta(x^{(i)}) \right) x_j^{(i)} - (1 - y^{(i)}) h_\theta(x^{(i)}) x_j^{(i)} \right] \\ &= -\frac{1}{m} \, \sum_{i=1}^m \left[ y^{(i)} - y^{(i)} h_\theta(x^{(i)}) - h_\theta(x^{(i)}) + y^{(i)} h_\theta(x^{(i)}) \right] x_j^{(i)} \\ &= -\frac{1}{m} \, \sum_{i=1}^m \left[ h_\theta(x^{(i)}) - y^{(i)} \right] x_j^{(i)} \\ &= \frac{1}{m} \, \sum_{i=1}^m \left[ h_\theta(x^{(i)}) - y^{(i)} \right] x_j^{(i)} \end{split}$$