Page No.
Date

The Derivative of Cost function :-

$$h_\theta(x) = \frac{1}{1+e^{-\theta^T x}}$$

$$\sigma(x) = \frac{1}{1+e^{-x}}$$

$$\frac{d(\sigma(x))}{dx} = \frac{1}{1+e^{-x}} \times \left(1 - \frac{1}{1+e^{-x}}\right) = \sigma(x)\left(1-\sigma(x)\right)$$

Apply chain rule and ~~[illegible]~~ in terms of partial Derivative

$$\frac{\partial(J(\theta))}{\partial(\theta_j)} = -\frac{1}{m} \times \Sigma_{i=1}^{m} \left[y^{(i)} \times \frac{1}{h_\theta(x^{(i)})} \times \frac{\partial(h_\theta(x^{(i)}))}{\partial(\theta_j)} \right]$$

$$+ \sum_{i=1}^{m} \left[(1-y^{(i)}) \times \frac{1}{(1-h_\theta(x^{(i)}))} \times \frac{\partial(1-h_\theta(x^{(i)}))}{\partial(\theta_j)} \right]$$

$$\frac{\partial(J(\theta))}{\partial(\theta_j)} = -\frac{1}{m} \times \left(\Sigma_{i=1}^{m} \left[y^{(i)} \times \frac{1}{h_\theta(x^{(i)})} \times \sigma(z)(1-\sigma(z)) \times \frac{\partial(\theta^T x)}{\partial(\theta_j)} \right] \right.$$

$$\left. + \sum_{i=1}^{m} \left[(1-y^{(i)}) \times \frac{1}{(1-h_\theta(x^{(i)}))} \times \left(-\sigma(z)(1-\sigma(z)\right) \times \frac{\partial(\theta^T x)}{\partial(\theta_j)} \right] \right)$$

Evaluate partial Derivative as a derivative of sigmoid function.

$$\frac{\partial(J(\theta))}{\partial(\theta_j)} = -\frac{1}{m} \times \left(\Sigma_{i=1}^{m} \left[y^{(i)} \times \frac{1}{h_\theta(x^{(i)})} \times \sigma(z)(1-\sigma(z)) \times \frac{\partial(\theta^T x)}{\partial(\theta_j)} \right] \right.$$

$$\left. + \sum_{i=1}^{m} \left[(1-y^{(i)}) \times \frac{1}{(1-h_\theta(x^{(i)}))} \times \left(-\sigma(z)(1-\sigma(z)\right) \times \frac{\partial(\theta^T x)}{\partial(\theta_j)} \right] \right)$$

Page No.
Date

$$\frac{\partial(J(\theta))}{\partial(\theta_j)} = -\frac{1}{m} \times \left(\sum_{i=1}^{m}\left[y^{(i)}\frac{1}{h_\theta(x^{(i)})}\cdot h_\theta(x^{(i)})\left(1-h_\theta(x^{(i)})\right)\times x_j^i\right]\right.$$

$$\left.+\sum_{i=1}^{m}\left[(1-y^{(i)})\times\frac{1}{(1-h_\theta(x^{(i)}))}\times\left(-h_\theta(x^{(i)})(1-h_\theta(x^{(i)})\right)\times x_j^i\right]\right)$$

Simplify :-

$$\frac{\partial(J(\theta))}{\partial(\theta_j)} = -\frac{1}{m}\times\left(\sum_{i=1}^{m}\left[y^{(i)}\times(1-h_\theta(x^{(i)}))\times x_j^i-(1-y^{(i)})\times\right.\right.$$

$$\left.h_\theta(x^{(i)})\times x_j^i\right]$$

$$\frac{\partial(J(\theta))}{\partial(\theta_j)} = \frac{-1}{m}\left(\sum_{i=1}^{m}\left[y^{(i)}-y^{(i)}\times h_\theta(x^{(i)})-h_\theta(x^{(i)})+\right.\right.$$

$$\left.\left.y^{(i)}\times h_\theta(x^{(i)})\right]\times x_j^i\right)$$

$$\frac{\partial(J(\theta))}{\partial(\theta_j)} = -\frac{1}{m}\times\left(\sum_{i=1}^{m}\left[y^{(i)}-h_\theta(x^{(i)})\right]\times x_j^i\right)$$

Convert this into matrix form for the gradient w.r.t. all the weights include the bias term.

$$\frac{\partial(J(\theta))}{\partial(\theta)} = \frac{1}{m}X^T\left[h_\theta(x)-y\right]$$