Neural Network

We assume dataset X, and its dimension is m imes n. and hypothesis $h_{ heta}(x) = g(heta x), \, g(z) = rac{1}{1 + e^{-z}}.$

Cost Function

$$J(heta) = -rac{1}{m} \sum_{i=1}^m (y^{(i)} \cdot log(h_ heta(x^{(i)})) + (1-y^{(i)}) \cdot log(1-h_ heta(x^{(i)}))$$

SO

$$X = egin{bmatrix} x^{(1)^T} \ x^{(2)^T} \ dots \ x^{(m)^T} \end{bmatrix}, heta = egin{bmatrix} heta^{(0)} \ heta^{(1)} \ dots \ heta^{(n)} \end{bmatrix}$$

here expand X, now $x^{(i)}=\{x_0^{(i)},x_1^{(i)},\dots,x_n^{(i)}\}$, $\theta^{(i)}$ and $x^{(1)^T}$ both are vector, so

$$X heta = egin{bmatrix} heta^T x^{(1)} \ heta^T x^{(2)} \ dots \ heta^T x^{(m)} \end{bmatrix}$$

Gradient

$$rac{\partial J(heta)}{\partial heta_j} = rac{1}{m} \sum_{i=1}^m ((h_ heta(x^{(i)}) - y^{(i)}) \cdot x_j^{(i)})$$

To vectorize this operation over the dataset, we can write

$$\begin{bmatrix} \frac{\partial J(\theta)}{\partial \theta_0} \\ \frac{\partial J(\theta)}{\partial \theta_1} \\ \vdots \\ \frac{\partial J(\theta)}{\partial \theta_n} \end{bmatrix} = \frac{1}{m} \begin{bmatrix} \sum_{i=1}^m ((h_{\theta}(x^{(i)}) - y^{(i)}) \cdot x_0^{(i)}) \\ \sum_{i=1}^m ((h_{\theta}(x^{(i)}) - y^{(i)}) \cdot x_1^{(i)}) \\ \vdots \\ \sum_{i=1}^m ((h_{\theta}(x^{(i)}) - y^{(i)}) \cdot x_i^{(i)}) \end{bmatrix} = \frac{1}{m} \sum_{i=1}^m ((h_{\theta}(x^{(i)}) - y^{(i)}) \cdot x^{(i)}) = \frac{1}{m} X^T \cdot (h_{\theta(x)} - y)$$

where

$$h_{ heta(x)} - y = egin{bmatrix} h_{ heta}(x^{(0)}) - y^{(0)} \ h_{ heta}(x^{(1)}) - y^{(1)} \ dots \ h_{ heta}(x^{(m)}) - y^{(m)} \end{bmatrix}$$

Vectorizing regularized logistic regression

$$J(heta) = -rac{1}{m} \sum_{i}^{m} (y^{(i)} \cdot log(h_{ heta}(x^{(i)})) + (1-y^{(i)}) \cdot log(1-h_{ heta}(x^{(i)}))) + rac{\lambda}{2m} \sum_{j=1}^{m} heta_{j}^{2}$$

SO

$$rac{\partial J(heta)}{\partial heta_j} = egin{cases} rac{1}{m} \sum_{i=1}^m (h_ heta(x^{(i)} - y^{(i)}) \cdot x_j^{(i)}) & j = 0 \ rac{1}{m} \sum_{i=1}^m (h_ heta(x^{(i)} - y^{(i)}) \cdot x_j^{(i)}) + rac{\lambda}{m} heta_j & j > 0 \end{cases}$$