Neural Network

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

Cost Function

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

so

$$X = \begin{bmatrix} {x^{(1)}}^T \\ {x^{(2)}}^T \\ \vdots \\ {x^{(m)}}^T \end{bmatrix}, \theta = \begin{bmatrix} \theta^{(0)} \\ \theta^{(1)} \\ \vdots \\ \theta^{(n)} \end{bmatrix}$$

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

$$X\theta = \begin{bmatrix} \theta^T x^{(1)} \\ \theta^T x^{(2)} \\ \vdots \\ \theta^T x^{(m)} \end{bmatrix}$$

Gradient

$$\frac{\partial J(\theta)}{\partial \theta_j} = \frac{1}{m}\sum_{i=1}^{m}((h_\theta(x^{(i)}) - y^{(i)}) \cdot x_j^{(i)})$$

To vectorize this operation over the dataset, we can write

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

where

$$h_{\theta(x)} - y = \begin{bmatrix} h_\theta(x^{(0)}) - y^{(0)} \\ h_\theta(x^{(1)}) - y^{(1)} \\ \vdots \\ h_\theta(x^{(m)}) - y^{(m)} \end{bmatrix}$$

Vectorizing regularized logistic regression

$$J(\theta) = -\frac{1}{m}\sum_{i}^{m}(y^{(i)} \cdot log(h_\theta(x^{(i)})) + (1 - y^{(i)}) \cdot log(1 - h_\theta(x^{(i)}))) + \frac{\lambda}{2m}\sum_{j=1}^{m}\theta_j^2$$

so

$$\frac{\partial J(\theta)}{\partial \theta_j} = \begin{cases} \frac{1}{m}\sum_{i=1}^{m}(h_\theta(x^{(i)} - y^{(i)}) \cdot x_j^{(i)}) & j = 0 \\ \frac{1}{m}\sum_{i=1}^{m}(h_\theta(x^{(i)} - y^{(i)}) \cdot x_j^{(i)}) + \frac{\lambda}{m}\theta_j & j > 0 \end{cases}$$