Problem set 1

#1. (a)

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

$y^{(i)} \in \{0,1\}$ $h_\theta(x) = g(\theta^T x)$ $g(z) = \frac{1}{1+e^{-z}}$

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

Find Hessian:

$$H_{ij} = \frac{\partial^2 l(\theta)}{\partial\theta_i \partial\theta_j}$$

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

$$= -\frac{1}{m}\sum_{i=1}^{m}\frac{\partial}{\partial\theta_i}\left\{y^{(i)}\log(h_\theta(x^{(i)})) + (1-y^{(i)})\log(1-h_\theta(x^{(i)}))\right\}$$

$$= -\frac{1}{m}\sum_{i=1}^{m}\left\{y^{(i)} \times \frac{1}{h_\theta(x^{(i)})} \times g(\theta^T x^{(i)})\{1-g(\theta^T x^{(i)})\}\cdot x_i^{(i)}\right.$$

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

$$\left. + (1-y^{(i)}) \times \frac{1}{1-h_\theta(x^{(i)})} \times (-1)g(\theta^T x^{(i)})(1-g(\theta^T x^{(i)}))\cdot x_i^{(i)}\right\}$$

$$= -\frac{1}{m}\sum_{i=1}^{m}\left\{y^{(i)} \times (1-g(\theta^T x^{(i)}))\cdot x_i^{(i)} + (1-y^{(i)})\cdot(-1)g(\theta^T x^{(i)})\cdot x_i^{(i)}\right\}$$

$$= -\frac{1}{m}\sum_{i=1}^{m}\left\{x_i^{(i)} \times \left(y^{(i)} - y^{(i)}g(\theta^T x^{(i)}) + y^{(i)}g(\theta^T x^{(i)}) - g(\theta^T x^{(i)})\right)\right\}$$

$$\nabla_{\theta_i} J(\theta) = -\frac{1}{m}\sum_{i=1}^{m}\left\{x_i^{(i)}\left(y^{(i)} - g(\theta^T x^{(i)})\right)\right\}$$

$$\frac{\partial}{\partial\theta_j}\nabla_{\theta_i} J(\theta) = \frac{\partial}{\partial\theta_j} -\frac{1}{m}\sum_{i=1}^{m}\left\{x_i^{(i)}(y^{(i)} - g(\theta^T x^{(i)}))\right\}$$

$$= -\frac{1}{m}\sum_{i=1}^{m}\frac{\partial}{\partial\theta_j}\left\{x_i^{(i)}(y^{(i)} - g(\theta^T x^{(i)}))\right\}$$

$$= -\frac{1}{m}\sum_{i=1}^{m} x_i^{(i)} \times (-1) \times \frac{\partial}{\partial\theta_j} g(\theta^T x^{(i)})$$

$$= -\frac{1}{m}\sum_{i=1}^{m} x_i^{(i)} \times (-1) \times g(\theta^T x^{(i)})(1-g(\theta^T x^{(i)}))x_j^{(i)}$$

$$\therefore H_{ij} = \frac{\partial^2 J(\theta)}{\partial\theta_i \partial\theta_j} = \frac{1}{m}\sum_{i=1}^{m} g(\theta^T x^{(i)})(1-g(\theta^T x^{(i)}))x_i^{(i)}x_j^{(i)}$$

$$H_{ij} = \frac{\partial^2 J(\theta)}{\partial\theta_i \partial\theta_j} = \frac{1}{m}\sum_{i=1}^{m} g(\theta^T x^{(i)})(1-g(\theta^T x^{(i)}))\, x_i^{(i)} x_j^{(i)}$$

$(i) \neq i$

$$Z^T H Z \geq 0 \rightarrow \text{“}H \geq 0\text{”} \Leftrightarrow \text{positive semi-definite}$$

= J is convex, and has no local minima other than global one.

$$\sum_i \sum_j z_i x_i x_j z_j = (x^T z)^2 \geq 0$$

$$H_{jk} = \frac{1}{m}\sum_{i=1}^{m} g(\theta^T x^{(i)})(1-g(\theta^T x^{(i)}))\, x_j^{(i)} x_k^{(i)}$$

$$Z^T H Z = \frac{1}{m}\sum_{i=1}^{m}\sum_{j=1}^{n}\sum_{k=1}^{n} g(\theta^T x^{(i)})(1-g(\theta^T x^{(i)}))\, x_j^{(i)} x_k^{(i)} z_j z_k$$

$$\downarrow$$

$$\frac{1}{m}\sum_{i=1}^{m}\sum_{j=1}^{n}\sum_{k=1}^{n} g(\theta^T x^{(i)})(1-g(\theta^T x^{(i)}))\{(x^{(i)})^T z\}^2 \geq 0$$

$$\therefore Z^T H Z \geq 0$$