$A = (\sqrt{ra})$

Question 5

The cost function of the logistic regression is updated to penalize high values of the parameters and is given by:

$$L(\theta) = -\frac{1}{m}\left(y^{(i)} \log(h_\theta(x^{(i)})) + (1-y^{(i)})\log_e(1-h_\theta(x^{(i)}))\right) + \frac{\lambda}{2m}\sum_{j=1}^{n}\theta_j^2$$

Here $h_\theta(x) = \sigma(\theta_0 + \theta_1 x_1 + \theta_2 x_2 + \theta_3 x_3 + \dots + \theta_n x_n)$

$$= \sigma(\theta^T x)$$

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

$$\frac{\partial L}{\partial \theta} = \begin{bmatrix} \frac{1}{m}\sum_{i=1}^{m}(h_\theta(x^{(i)}) - y^{(i)})x_0^{(i)} \\ \frac{1}{m}\sum_{i=1}^{m}(h_\theta(x^{(i)}) - y^{(i)})x_1^{(i)} + \frac{\lambda}{m}\theta_1 \\ \frac{1}{m}\sum_{i=1}^{m}(h_\theta(x^{(i)}) - y^{(i)})x_2^{(i)} + \frac{\lambda}{m}\theta_2 \\ \frac{1}{m}\sum_{i=1}(h_\theta(x^{(i)}) - y^{(i)})x_3^{(i)} + \frac{\lambda}{m}\theta_3 \\ \vdots \\ \frac{1}{m}\sum_{i=1}^{m}(h_\theta(x^{(i)}) - y^{(i)})x_n^{(i)} + \frac{\lambda}{m}\theta_n \end{bmatrix}$$

Now

$\frac{\partial}{\partial\theta}\left(\frac{\partial L}{\partial \theta}\right)$ can be calculated using Hessian H, i.e.

$$H(L) = \begin{bmatrix} \frac{\partial^2 L}{\partial\theta_0^2} & \frac{\partial^2 L}{\partial\theta_0\partial\theta_1} & \frac{\partial^2 L}{\partial\theta_0\partial\theta_2} & \cdots & \frac{\partial^2 L}{\partial\theta_0\partial\theta_n} \\ \frac{\partial^2 L}{\partial\theta_1\partial\theta_0} & \frac{\partial^2 L}{\partial\theta_1^2} & \frac{\partial^2 L}{\partial\theta_1\partial\theta_2} & \cdots & \frac{\partial^2 L}{\partial\theta_1\partial\theta_n} \\ \vdots & \vdots & \vdots & & \vdots \\ \frac{\partial^2 L}{\partial\theta_n\partial\theta_0} & \frac{\partial^2 L}{\partial\theta_n\partial\theta_1} & \frac{\partial^2 L}{\partial\theta_n\partial\theta_2} & \cdots & \frac{\partial^2 L}{\partial\theta_n^2} \end{bmatrix}$$

Apart from diagonal entities, the generalized ~~form~~ of every other entity is

$$\frac{\partial^2 L}{\partial \theta_i \partial \theta_j} = \frac{1}{m}\left[\sum_{i=1}^{m} h_\theta(x^{(i)})\left(1 - h_\theta(x^{(i)})\right) x_i\, x_j\right]$$

For diagonal entities we can write

$$\frac{\partial^2 L}{\partial \theta_i^2} = \begin{cases} \frac{1}{m}\left[\sum_{i=1}^{m} h_\theta(x^{(i)})\left(1 - h_\theta(x^{(i)})\right) x_0^2 & \text{for } j=0 \\ \frac{1}{m}\left[\sum_{i=1}^{m} h_\theta(x^{(i)})\left(1 - h_\theta(x^{(i)})\right) x_j^2 + \frac{\lambda}{m} & \text{for } j = 1, \ldots, n \end{cases}$$

Combing above equaties, we can simply write

$$H(L) = \frac{1}{m}\left[\sum_{i=1}^{m} h_\theta(x^{(i)})\left(1 - h_\theta(x^{(i)})\right)\right] x^{(i)} (x^{(i)})^T + \frac{\lambda}{m}\left[\begin{smallmatrix} 0 & \\ & 1 \\ & & \ddots \\ & & & 1 \end{smallmatrix}\right]$$

Now we ~~can~~ have everything to write updated gradient descent and Newton's method in regularized logistic regression.

→ updated gradient descent

$$\theta^{(t+1)} = \theta^{(t)} - \lambda\left(\frac{\partial L}{\partial \theta^{(t)}}\right)$$

Here λ is the learning rate and $\frac{\partial L}{\partial \theta}$ is given in eq (1)

→ updated Newton's method

$$\theta^{(t+1)} = \theta^{(t)} - \left(\frac{\partial L}{\partial \theta}\right) = \theta^{(t)} - \frac{\partial L / \partial \theta}{H(L)} = \theta^{(t)} \, H(L)^{-1} \frac{\partial L}{\partial \theta}$$

Again $\frac{\partial l}{\partial \theta}$ is given is eq (i) and ~~H~~ (L) is given in eq (ii)