Name: Amina Tabassum.

HW # 02 ID: 7615

Question No: 01

Since,

$$E(x^q) = \frac{1}{2}\sum_{i=1}^{m}(y^i - h_\theta(x^i))^2 (x^i - x^q)^{-2}$$

Here

$$(w^i)^2 = (x^i - x^q)^{-2}$$

So, cost function depends on x^q.

let w is $m \times m$ feature vector given by:

$$\vec{W} = \begin{bmatrix} w_{11} & 0 & 0 & \cdots & 0 \\ w_{21} & w_{22} & 0 & \cdots & \vdots \\ 0 & 0 & w_{33} & \cdots & 0 \\ \vdots & \vdots & \vdots & & \\ 0 & 0 & 0 & & w_{mm} \end{bmatrix} \text{ and } \vec{Y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_m \end{bmatrix}$$

So,

$$\vec{W}\vec{Y} = \begin{bmatrix} w_1 y_1 \\ w_2 y_2 \\ \vdots \\ w_m y_m \end{bmatrix} \text{ or } \overrightarrow{WY} = [w^1 y^1, w^2 y^2, \ldots, w^m y^m]^T$$

and

$$\vec{X} = \begin{bmatrix} x_{10} & \cdots & x_{1j} & \cdots & x_{1m-1} \\ \vdots & & \vdots & & \vdots \\ x_{k,0} & & x_{kj} & \cdots & x_{k,m-1} \\ \vdots & & \vdots & & \vdots \\ x_{N,0} & & x_{N,j} & \cdots & x_{N,M-1} \end{bmatrix}$$

So,

$$\overline{WX} = \vec{W} \times \vec{X}$$

Putting original $\vec{y}$ and $\vec{x}$ in cost function gives

$$E(x^q) = \frac{1}{2}\sum_{i=1}^{m}\left(w^i y^i - w^i \theta^T x^i\right)^2$$

$$= (WY - WX\theta)^2$$

Now, Normal equation is given by:

$$\theta = (X^T X)^{-1} \cdot (X^T Y)$$

So, replacing $Y = WY$ and $X = WX$ into above equation gives:

$$\boxed{\theta_{ML} = \left((WX)^T WX\right)^{-1}\left((WX)^T (WY)\right)}$$

Question No: 02

The maximum likelihood solution is obtained by computing μ_{ML}: value that maximizes the likelihood function.

$$\frac{d}{d\mu}\left(\sum_{i=1}^{M} x^{(i)} \ln\mu + (1-x^i)\ln(1-\mu)\right) = 0.$$

$$\sum_{i=1}^{m} x^{(i)} \frac{d}{d\mu}\ln(\mu) + (1-x^i)\frac{d}{d\mu}\ln(1-\mu) = 0$$

$$\sum_{i=1}^{m} x^i \times \frac{1}{\mu} + (1-x^i) \times \frac{1}{1-\mu}\frac{d}{d\mu}(1-\mu) = 0$$

$$\sum_{i=1}^{m} x^i \times \frac{1}{\mu} + \sum_{i=1}^{m} (1-x^i) \times \frac{1}{1-\mu} \times (-1) = 0$$

$$\sum_{i=1}^{m} x^i \times \frac{1}{\mu} = \sum_{i=1}^{m} (1-x^i) \frac{1}{1-\mu}$$

Since, $x_i = 1$ (heads), So, using it here gives

$$\frac{h}{\mu} = \frac{m-h}{1-\mu} \qquad \because h = \sum_{i=1}^{m} x^i$$

$$h(1-\mu) = \mu(m-h) \qquad \because \sum_{i=1}^{m} (1-x)^i = m-h$$

$$h - \cancel{\mu h} = \mu m - \cancel{\mu h}$$

$$h = \mu m$$

$$\boxed{\mu = \frac{h}{m}}$$ Hence, proved.

Question NO: 03

Likelihood of data given parameters is given by:

$$L(\theta) = \prod_{i=1}^{m} h(x^i)^{y^i} (1-h(x^i))^{(1-y^i)}$$

Taking log both sides:

$$\ln L(\theta) = \ln\left[\prod_{i=1}^{m} h(x^i)^{y^i} + (1-h(x^i))^{(1-y^i)}\right]$$

$$= \sum_{i=1}^{m} h(x^i) y^i + (1-h(x^i))(1-y^i)$$

which is same as logistic regression cost.