Name- Sunayana Gupta
ASU ID- 1222389090

1) $l(\theta) = y \log(h(\theta^T x)) + (1-y)(\log(1-h(\theta^T x)))$

$h(z) = \dfrac{1}{1+e^{-z}}$, $\theta = \begin{bmatrix} \theta_0 \\ \theta_1 \\ \theta_2 \end{bmatrix}$ and $x = \begin{bmatrix} 1 \\ x_1 \\ x_2 \end{bmatrix}$

find $\nabla_\theta l(\theta) = \begin{pmatrix} \frac{\partial l(\theta)}{\partial \theta_0} \\ \frac{\partial l(\theta)}{\partial \theta_1} \\ \frac{\partial l(\theta)}{\partial \theta_2} \end{pmatrix}$

$\dfrac{\partial l(\theta)}{\partial \theta_0} = y \dfrac{\partial \log(h(\theta^T x))}{\partial \theta_0} + (1-y) \dfrac{\partial (\log 1 - h(\theta^T x))}{\partial \theta_0}$

Using the chain rule,

$\dfrac{\partial l(\theta)}{\partial \theta_0} = y \dfrac{\partial \log(z_1)}{\partial z_1} \cdot \dfrac{h(y)}{\partial y} \cdot \dfrac{(\theta^T x)}{\partial \theta_0} + (1-y) \dfrac{\partial \log(z_2)}{\partial z_2} \dfrac{\partial(1-h(y))}{\partial y} \dfrac{(\theta^T x)}{\partial \theta_0}$

where $y = \theta^T x$ and $z_1 = h(\theta^T(x))$ and $z_2 = 1-h(\theta^T(x))$

Now, using $\dfrac{\partial \log(x)}{\partial x} = \dfrac{1}{x}$ and derivation of $h(x) = \dfrac{1}{1+e^{-x}}$

$\dfrac{\partial h(x)}{\partial x} = h(x)(1-h(x))$ and $\theta^T(x) = \theta_0 + \theta_1 x + \theta_2 x$

$\dfrac{\partial l(\theta)}{\partial \theta_0} = \dfrac{y}{h(\theta^T x)} h(\theta^T x)(1-h(\theta^T x))(1) - (1-y)(h(\theta^T x))$

$\dfrac{\partial l(\theta)}{\partial \theta_0} = y - y\, h\theta^T(x) + \cancel{y\, h(\theta^T x)} - h\theta^T(x)$

$= y - h(\theta^T(x))$

Similarly,

$$\frac{\partial l(\theta)}{\partial \theta_1} = x_1 y\,(1 - h(\theta^T x)) - (1-y)x_1\, h\theta^T(x)$$

$$= x_1 y - x_1 y\, \cancel{h(\theta^T x)} - h\theta(x)x_1 + \cancel{y x_1 h\theta^T(x)}$$

$$\frac{\partial l(\theta)}{\partial \theta_1} = x_1 y - h(\theta^T x)x_1$$

$$\frac{\partial l(\theta)}{\partial \theta_2} = x_2(y - h\theta^T x)$$

$$\therefore \quad \nabla_\theta\, l(\theta) = \begin{bmatrix} \frac{\partial l(\theta)}{\partial \theta_0} \\ \frac{\partial l(\theta)}{\partial \theta_1} \\ \frac{\partial l(\theta)}{\partial \theta_2} \end{bmatrix} = \begin{bmatrix} y - h(\theta^T x) \\ x_1(y - h(\theta^T x)) \\ x_2(y - h(\theta^T x)) \end{bmatrix}$$

2.)

$$p(y=1) = \phi_1,$$

$$p(y=2) = \phi_2,$$

$$\vdots$$

$$p(y=k-1) = \phi_{k-1}$$

$$p(y=k) = 1 - \sum_{j=1}^{k-1} \phi_j,$$

which can be written as,

$$p(y) = \phi_1^{1[y=1]}\phi_2^{1[y=2]} \cdots \phi_{k-1}^{1[y=k-1]}\phi_k^{1[y=k]} \quad (x),$$

to be written as,

$$f(y) = b(y)\exp\left(\eta^T T(y) - a(\eta)\right)$$

$T(y)$ assumed as:

$$T(1) = \begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix} \quad T(2) = \begin{bmatrix} 0 \\ 1 \\ \vdots \\ 0 \end{bmatrix} \quad T(3) = \begin{bmatrix} 0 \\ 0 \\ 1 \\ \vdots \\ 0 \end{bmatrix} \quad T(k) = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}$$

$$b(y,\varphi) = \varphi_1^{1\{y=1\}} \cdot \varphi_2^{1\{y=2\}} \cdots \varphi_k^{1\{y=k\}}$$

$$= \varphi_1^{1\{y=1\}} \cdot \varphi_2^{1\{y=2\}} \cdots \varphi_k^{1-\sum_{i=1}^{k-1} 1\{y=i\}}$$

$$= \varphi_1^{(T(y))_1} \varphi_2^{(T(y))_2} \cdots \varphi_k^{1-\sum_{i=1}^{k-1}(T(y))_i}$$

Taking exp and log simultaneously,

$$f(y,\varphi) = \exp\Big((T(y))_1 \log(\varphi_1) + (T(y))_2 \log(\varphi_2) + \cdots + \Big(1 - \sum_{i=1}^{k-1}(T(y))_i\Big)\log(\varphi_k)\Big)$$

$$= \exp\Big((T(y))_1 \log\Big(\frac{\varphi_1}{\varphi_k}\Big) + (T(y))_2 \log\Big(\frac{\varphi_2}{\varphi_k}\Big) + \cdots + (T(y))_{k-1}\log\Big(\frac{\varphi_{k-1}}{\varphi_k}\Big) + \log(\varphi_k)\Big)$$

$$= b(y)\exp\left(\eta^T T(y) - a(\eta)\right)$$

where $\eta = \begin{bmatrix} \log(\varphi_1/\varphi_k) \\ \log(\varphi_2/\varphi_k) \\ \vdots \\ \log(\varphi_{k-1}/\varphi_k) \end{bmatrix}$, $a(\eta) = -\log(\varphi_k)$

$b(y) = 1$

where $\eta_i = \log\left(\frac{\phi_i}{\phi_k}\right)$

$$e^{\eta_i} = \frac{\phi_i}{\phi_k}$$

$$\phi_k e^{\eta_i} = \phi_i \quad \text{--- (1)}$$

$$\phi_k \sum_{i=1}^{k} e^{\eta_i} = \sum_{i=1}^{k} \phi_i = 1 \quad \text{--- (2)}$$

This implies that $\phi_k = \frac{1}{\sum_{i=1}^{k} e^{\eta_i}}$,

substituting this back into (1).

$$\phi_i = \frac{e^{\eta_i}}{\sum_{i=1}^{k} e^{\eta_i}}$$

Given hypothesis, $h_\theta(x) = E[T(y) | x; \theta]$

And it is assumed that η_i's are linearly related to the x's.

$$\therefore \eta_i = \theta_i^T x$$

$$\therefore p(y = i / x; \theta) = \phi_i$$

$$= \frac{e^{\eta_i}}{\sum_{j=1}^{k} e^{\eta_j}} = \frac{e^{\theta_i^T x}}{\sum_{j=1}^{k} e^{\theta_j^T x}}$$

$$h_\theta(x) = E[T(y) \mid x; \theta]$$

$$= E\left[\begin{pmatrix} 1\{y=1\} \\ 1\{y=2\} \\ \vdots \\ 1\{y=k-1\} \end{pmatrix} \middle| x;\theta\right]$$

$$= \begin{bmatrix} \phi_1 \\ \phi_2 \\ \vdots \\ \phi_{k-1} \end{bmatrix}$$

$$h_\theta(x) = \begin{bmatrix} \frac{\exp(\theta_1^T x)}{\sum_{j=1}^{k} \exp(\theta_j^T x)} \\ \frac{\exp(\theta_2^T x)}{\sum_{j=1}^{k} \exp(\theta_j^T x)} \\ \vdots \\ \frac{\exp(\theta_{k-1}^T x)}{\sum_{j=1}^{k} \exp(\theta_j^T x)} \end{bmatrix}$$

3) $P(\text{play} = \text{yes}) = 5/8$

$P(\text{play} = \text{no}) = 3/8$

Outlook	play = yes	play = no
Sunny	2/5	1/3
Overcast	2/5	0/3
Rain	1/5	2/3

Wind	Play = yes	play = no
Weak	4 / 5	0 / 3
Strong	1 / 5	3 / 3

$p(\text{yes} \mid \text{sunny}, \text{strong}) = p(\text{sunny} \mid \text{yes}) * p(\text{strong} \mid \text{yes}) * p(\text{yes})$

$= \frac{2}{5} \times \frac{1}{5} \times \frac{5}{8}$

$= \frac{\cancel{2}}{5} \times \frac{1}{\cancel{5}} \times \frac{\cancel{5}}{\cancel{8}\,4}$

Thus $p(\text{yes} \mid \text{sunny}, \text{strong}) = \frac{1}{20} = 0.05$

$p(\text{no} \mid \text{sunny}, \text{strong}) = \frac{1}{8} = 0.125$

Therefore, the prediction is that they should not play