First subtask:

Name: Amit Kumar
Roll: 20CS30003.

Average empirical loss function for the logistic regression:-

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

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

& θ is an dimensional parameter vector

find Hessian matrix H of the empirical loss function with respect to θ, and show that the hessian H is positive-semi-definite in nature.

Ans:- $H_{kl} = \frac{\partial^2}{\partial\theta_k\partial\theta_l}(J(\theta))$

$h_\theta(x) = \frac{1}{1+e^{-(\theta^T x)}}$ | $^*\frac{\partial h_\theta(x)}{\partial\theta} = \frac{e^{-(\theta^T x)}\cdot x}{(1+e^{-(\theta^T x)})^2} = h_\theta(x)(1-h_\theta(x))\cdot x$ — (1)

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

$$= \frac{-1}{m}\sum_{i=1}^{m} y^{(i)}\left(\frac{\cancel{h_\theta(x^{(i)})}(1-h_\theta(x^{(i)}))x_k^{(i)}}{\cancel{h_\theta(x^{(i)})}}\right) - (1-y^{(i)})\frac{\cancel{(1-h_\theta(x^{(i)}))}h_\theta(x^{(i)})x_k^{(i)}}{\cancel{(1-h_\theta(x^{(i)}))}}$$

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

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

So, $$\frac{\partial^2 J(\theta)}{\partial\theta_k\partial\theta_l} = \frac{-1}{m}\sum_{i=1}^{m}\frac{\partial}{\partial\theta_l}(-h_\theta(x^{(i)}))x_k^{(i)}$$

$$= \frac{-1}{m}\sum_{i=1}^{m} -h_\theta(x^{(i)})(1-h_\theta(x^{(i)}))x_k^{(i)}\cdot x_l^{(i)}$$

$$\therefore H_{kl} = \frac{\partial^2 J(\theta)}{\partial \theta_k \partial \theta_l} = \frac{1}{m}\sum_{i=1}^{m} h_\theta(x^{(i)})(1-h_\theta(x^{(i)}))\cdot x_l^{(i)}\cdot x_k^{(i)}$$

(Hessian matrix ~~term~~) general term

since $H_{kl} = H_{lk}$ (from the eqn)

$\therefore$ H is a symmetric matrix.

H is a $m\times m$ matrix.

For H to be a positive semidefinite matrix,

$$z^T H z \geq 0$$

where z^T is a $m\times 1$ matrix or $z \in R^m$

$$\text{so, } z^T H z = \sum_l \sum_k \frac{1}{m}\cdot\sum_{i=1}^{m} h_\theta(x^{(i)})(1-h_\theta(x^{(i)}))\cdot x_l^{(i)}\cdot x_k^{(i)}\cdot z_j\cdot z_k$$

$$= \frac{1}{(m)}\sum_{i=1}^{m} h_\theta(x^{(i)})(1-h_\theta(x^{(i)}))\sum_l (x_l^{(i)}\cdot z_l)\sum_k x_k^{(i)} z_k$$

$$= \frac{1}{m}\sum_{i=1}^{m} h_\theta(x^{(i)})(1-h_\theta(x^{(i)}))\left(\sum_l (x_l^{(i)} z_l)\right)^2$$

But $\sum_l x_l^{(i)} z_l = x^{(i)T} z$

$$\therefore z^T H.z = \frac{1}{m}\sum_{i=1}^{m} h_\theta(x^{(i)})(1-h_\theta(x^{(i)}))(x^{(i)T} z)^2 \geq 0$$

since $1 > h_\theta(x^{(i)}) > 0$

$\therefore (1-h_\theta(x^{(i)})) > 0$ & $(x^{(i)T} z)^2 \geq 0, \forall i$

Hence, H is positive semidefinite in nature.