Hessian Matrix Nitin
Rohit

I have mentioned about the cost/loss function in the sub-task-2 pdf elaborately and I'll be using the same formula in here to explain about Hessian matrix.
The cost function is given as:

$$cost(y_p) = -\frac{1}{m}\sum_{i=1}^{m}[(y)log(y_p) + (1-y)log(1-y_p)]$$

here, $$y_p = h(x) = \sigma(z(x)) = \frac{1}{1+e^{-\left(W^TX+W_0\right)}} = \frac{1}{1+e^{-(W_0+Wx_1+...+W_nx_n)}}$$

We will be using vector calculas to make the calculation look more tidy,

$$\frac{\partial f}{\partial W} = \left[\frac{\partial f}{\partial W_1}\ \frac{\partial f}{\partial W_2}\ ....\right] where\ W_1, W_2\ are\ individual\ components$$

Same way vector differentiation by vetcor is defined similarly. So in vector form Hessian matrix is defined as the second order derivative of a matrix:

$$H = \frac{\partial^2 f}{\partial W^2}$$

Then using above rule we get the following row matrix. The values of derivatives is calculated in sub-task-2 pdf:

$$\frac{\partial f}{\partial W} = [\ x_1(\sigma(z)-y)\ \ x_2(\sigma(z)-y)\ ....\ ]\quad for\ one\ sample$$

Then differentiating again with W, we get the Hessian matrix:

$$H = \frac{\partial^2 f}{\partial W^2} = \begin{bmatrix} x_1x_1 & x_1x_2 & . & . \\ x_2x_1 & . & . & . \\ x_3x_1 & . & . & . \\ . & . & . & x_nx_n \end{bmatrix}\sigma(z)(1-\sigma(z))$$

Then extending it to all the samples we get the following equation, where $z_i(or\ X_i)$ denotes the different sample vectors and x_1, x_2,.. represents the components of that sample. In x_i^j i and j represents feature and sample respectively.

$$H = \frac{1}{m}\sum_{i=1}^{m}\begin{bmatrix} x_1x_1 & x_1x_2 & . & . \\ x_2x_1 & . & . & . \\ x_3x_1 & . & . & . \\ . & . & . & x_nx_n \end{bmatrix}\sigma(z_i)(1-\sigma(z_i))$$

$$\Rightarrow H = \frac{1}{m}\begin{bmatrix} \sum_{i=1}^{m} x_1x_1\sigma(z_i)(1-\sigma(z_i)) & \sum_{i=1}^{m} x_1x_2\sigma(z_i)(1-\sigma(z_i)) & . & . \\ \sum_{i=1}^{m} x_2x_1\sigma(z_i)(1-\sigma(z_i)) & . & . & . \\ \sum_{i=1}^{m} x_3x_1\sigma(z_i)(1-\sigma(z_i)) & . & . & . \\ . & . & . & \sum_{i=1}^{m} x_nx_n\sigma(z_i)(1-\sigma(z_i)) \end{bmatrix}$$

The above matrix looks like product of two matrixs, then lets write it in product form,

$$\Rightarrow \begin{bmatrix} x_1^1\sigma(z_1)(1-\sigma(z_1)) & x_1^2\sigma(z_2)(1-\sigma(z_2)) & . & . \\ x_2^1\sigma(z_1)(1-\sigma(z_1)) & . & . & . \\ x_3^1\sigma(z_1)(1-\sigma(z_1)) & . & . & . \\ . & . & . & x_n^m(\sigma(z_m)(1-\sigma(z_m))) \end{bmatrix}\begin{bmatrix} x_1^1 & x_2^1 & . & . \\ x_1^2 & . & . & . \\ x_1^3 & . & . & . \\ . & . & . & x_n^m \end{bmatrix}$$

The first matrix can be spilt in two parts as following,

$$\Rightarrow \begin{bmatrix} x_1^1 & x_1^2 & . & . \\ x_2^1 & . & . & . \\ x_3^1 & . & . & . \\ . & . & . & x_n^m \end{bmatrix}\begin{bmatrix} \sigma(z_1)(1-\sigma(z_1)) & 0 & 0 & . \\ 0 & \sigma(z_2)(1-\sigma(z_2)) & . & . \\ 0 & . & . & . \\ . & . & . & \sigma(z_m)(1-\sigma(z_m)) \end{bmatrix}\begin{bmatrix} x_1^1 & x_2^1 & . & . \\ x_1^2 & . & . & . \\ x_1^3 & . & . & . \\ . & . & . & x_n^m \end{bmatrix}$$

From above it can seen that the following can be written as $H = X^TDX$ and the elements of the diagonal matrix are non-negative, as sigmoid funtion lies between (0,1). As H is congruent to D hence all the eigen values are non-negative. If bias term is also considered then x_1 for all sample will be 1 and we will get the same result.

Hence it is concluded that the corresponding Hessian matrix is positive semi-definite.