Assignment 1 - Q2

1. The model will not get stuck in local minimum, because the Hessian ~~matr~~ matrix of loss function is positive definit, such that the local ~~min~~ minimum is the same as global minimum.

2. The cross entropy loss is defined as :

$$L = \sum_{i=1}^{I} L_i = \sum_{i=1}^{I} y_i \log[\sigma(a_i)] + \sum_{i=1}^{I} (1-y_i) \log[1-\sigma(a_i)]$$

$$\frac{\partial L}{\partial w} = \sum_{i=1} \frac{\partial L_i}{\partial a_i} \frac{\partial a_i}{\partial w}, \quad \text{where } a_i = -w^T x_i .$$

$$\frac{\partial L_i}{\partial a_i} = y_i \frac{1}{\cancel{\sigma(a_i)}} \left[\cancel{\sigma(a_i)} \overset{(1-\sigma(a_i))}{\cancel{(1-\sigma(a_i))}}\right] + (1-y_i)\left[\frac{1}{\cancel{1-\sigma(a_i)}} \cdot (-1)\left[\sigma(a_i)(\cancel{1-\sigma(a_i)})\right]\right].$$

$$= y_i [1-\sigma(a_i)] - (1-y_i)\sigma(a_i)$$

$$= y_i - \cancel{y_i \sigma(a_i)} - \sigma(a_i) + \cancel{y_i \sigma(a_i)}$$

$$= y_i - \sigma(a_i)$$

$$\frac{\partial a_i}{\partial w} = -x_i$$

Thus,

$$\frac{\partial L}{\partial w} = \sum_{i=1}^{I} \frac{\partial L_i}{\partial a_i} \frac{\partial a_i}{\partial w} = -(y_i - \sigma(a_i))(x_i)$$

$$= (\sigma(a_i) - y_i) x_i, \quad \text{where } \sigma(a_i) = \frac{1}{1+\exp(a_i)}$$