Q2 - a

To derive the gradiant of cost function. The cost function is the log likelihood

$$l(\theta) = \sum_{i=1}^{n} \{ -\log(1+e^{-\theta x_i}) + (y_i - 1)\theta x_i \}$$

$$\frac{\partial l(\theta)}{\partial \theta} = \sum_{i=1}^{n} \left(y_i x_i - x_i - \left(\frac{e^{-\theta x_i}}{1+e^{\theta x_i}} \right) x_i \right)$$

$$= \sum_{i=1}^{n} \left(y_i - \frac{1}{1+e^{\theta x_i}} \right) x_i = Fdx \quad \text{this is our gradiant decent steps.}$$

for gradiant decent, we need define the steps which is the formula above
And we need the learning rate λ.
the pseudo code will be:

Proceduce GD (Fdx , $\theta(0)$)

$\theta \leftarrow \theta(0)$
while not converged do
$\theta \leftarrow \theta + \lambda \, Fdx(\theta)$
return θ

Fdx is a function that returns a matrix of form $Fdx(\theta) = \begin{bmatrix} \frac{\partial l(\theta)}{\partial \theta_1} \\ \frac{\partial l(\theta)}{\partial \theta_2} \\ \\ \frac{\partial l(\theta)}{\partial \theta_n} \end{bmatrix}$

$\theta \in R$

$\theta(0)$ is a Random initialization

the returned θ is a point in which the log likelihood is the optimun. We run the Algorithm for each and points.