2.a) We assume that the labels are either 1 or -1 dependending on the result of the classifier.

$$\hat{y_i} = 1 \quad \text{if} \quad x_i^T w > 0$$

$$\hat{y_i} = -1 \quad \text{if} \quad x_i^T w < 0$$

Thus, $\hat{y_i} = \text{sign}(x_i^T w)$

When a point is easy to classify, if it is far from the decision boundary, then it will be correctly classify.

$$\hat{y_i} = \text{sign}(x_i^T w) = y_i$$

Given,

$$l_i(w) = \log(1 + e^{-y_i x_i^T w})$$

as $y_i = \text{sign}(x_i^T w)$

Thus, $l_i(w) = \log\left(1 + \frac{1}{e^{|x_i^T w|}}\right)$

As the point is easy to classify. $|x_i^T w|$ is very large and thus

$$l_i(w) = \log\Big(1 + \underbrace{\frac{1}{e^{|x_i^T w|}}}_{\to 1}\Big)$$ becomes small.

thus log tends to 0.

b) Given,

$$f(w) = \underbrace{\sum_{i=1}^{n} \log(1+e^{-y_i x_i^T w})}_{l_1(w)} + \underbrace{\lambda \|w\|_2^2}_{r(w)}$$

If we take the derivative,

$$\frac{\partial f(w)}{\partial w_j} \left(\sum_{i=1}^{n} \log(1+e^{-y_i x_i^T w}) + \lambda \|w\|_2^2 \right)$$

where w_j is an element from vector w

Thus we have,

$$\sum_{i=1}^{n} \frac{d(l_1(w))}{\partial w_j} + \lambda \frac{\partial(\|w\|_2^2)}{\partial w_j}$$

as the elements are seperable.

Now, $\|w\|_2^2 = w_1^2 + w_2^2 + w_3^2 + \dots + w_m^2$

$$\frac{\partial \|w\|_2^2}{\partial w_j} = 2w_j$$

Thus, $\nabla_w \|w\|_2^2 = \begin{bmatrix} 2w_1 \\ 2w_2 \\ 2w_3 \\ \vdots \\ 2w_n \end{bmatrix} = 2w$

Let $x_i^T w = x_{i1} w_1 + x_{i2} w_2 + \dots + x_{in} w_n$

$$\frac{\partial l_i(w)}{\partial w_j} = \frac{1}{1+e^{-y_i x_i^T w}} \left(e^{-y_i x_i^T w}\right)\left(-y_i x_{ij}\right)$$

$$l_i(w) = \log\left(1+e^{-y_i x_i^T w}\right)$$

$$e^{l_i(w)} = 1 + e^{-y_i x_i^T w}$$

$$\frac{\partial l_i(w)}{\partial w_j} = \frac{1}{e^{l_i(w)}}\left(e^{l_i(w)} - 1\right)\left(-y_i x_{ij}\right)$$

$$\nabla_w l_i(w) = \left(1 - \frac{1}{e^{l_i(w)}}\right)\left(-y_i x_{ij}\right)$$

$$= -y_i \left(1 - \frac{1}{e^{l_i(w)}}\right) x_i$$

$$f(w) = \sum_{i=1}^{N} l_i(w) + \lambda \|w\|_2^2 \ 10$$

$$\nabla_w f(w) = \sum_{i=1}^{N} -y_i\left(1 - \frac{1}{e^{l_i(w)}}\right) x_i + \lambda(2w)$$

$$= \sum_{i=1}^{N} -y_i\left(1 - \frac{1}{e^{\log(1+e^{-y_i x_i^T w})}}\right) x_i + \lambda(2w)$$

P.T.O

Scanned with CamScanner

$$= \sum_{i=1}^{n} (-y_i)\left(1 - \frac{1}{1+e^{-y_i x_i^T w}}\right) x_i + \lambda(2w)$$