Sheet 2

Omar Atef
7619

Q. 1

1) $z = Vec \cdot W = \begin{bmatrix} -1 \\ 1 \\ -1 \end{bmatrix} [-\ln 4, \ln 2, -\ln(3)]$

$z = \ln 4 + \ln 2 + \ln 3$, $y = \frac{e^z}{e^z + 1} = \frac{1}{1 + e^{-z}}$

$y = \frac{1}{1 + e^{-\ln 4} e^{-\ln 2} e^{-\ln 3}} = \frac{1}{1 + \frac{1}{4} \times \frac{1}{2} \times \frac{1}{3}} = \frac{1}{\frac{25}{24}} = \boxed{\frac{24}{25}}$

2) For prob(vec) $\rightarrow 1$

$e^{-z} \rightarrow 0$

$z \rightarrow \infty$

$V \cdot w \rightarrow \infty$

$V^1 \rightarrow -\infty$

$V^2 \rightarrow \infty$

$V^3 \rightarrow -\infty$

$\boxed{vec = [-9999, 9999, -9999]}$

$prob = \frac{1}{1 + e^{-(9999 \times 3)}} \simeq 1$

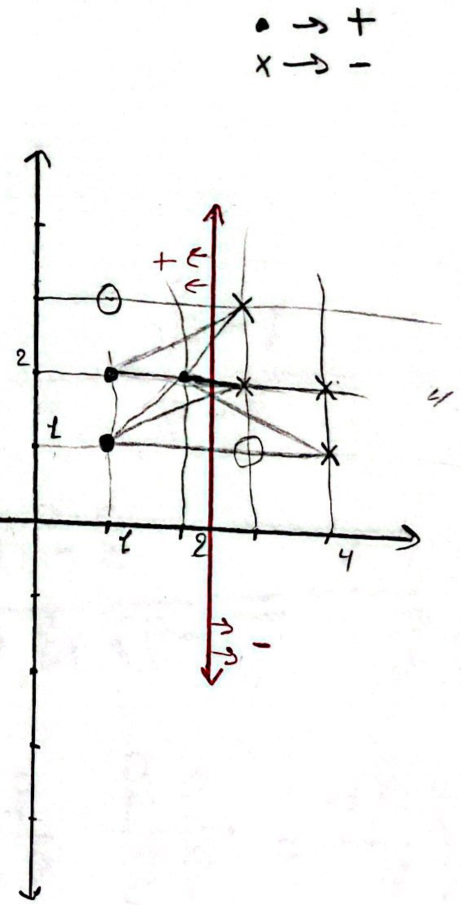
Scanned with CamScanner

3)- a would change dramatically with any change in the data due to very large weights.

This can be fixed by penalizing large weights in the loss function during training

- b would generalize better because the large weights in a probably indicate overfitting

Q.2



a) ~~4+2+1+1~~

$3+3+3+3=12$

infinite number of lines but 12 pairs of opposite samples that have a valid seprator perpendicular to the line between them

b) all pairs can make a perfect seprator

best fit is the red drawn line

c) $(1,3) \rightarrow +ve$

$(3,1) \rightarrow -ve$

Scanned with CamScanner

Q.3

1) $\mathcal{L} = -d \log(y) - (1-d) \log(1-y)$

$$y = \frac{1}{1+e^{-z}} \quad , \quad z = wx$$

$$\frac{d\mathcal{L}}{dw} = \frac{d\mathcal{L}}{dy} \frac{dy}{dz} \frac{dz}{dw} = \left(\frac{-d}{y} + \frac{1-d}{1-y}\right)\left(\frac{e^{-z}}{(1+e^{-z})^2}\right) x$$

$$\frac{d\mathcal{L}}{dw} = \frac{-d + \cancel{dy} + y - \cancel{dy}}{y - y^2} \times \frac{y}{z} \frac{e^{-z}}{1+e^{-z}} \quad x$$

$$= \frac{y-d}{\cancel{1-y}} \frac{e^{-z}}{1+e^{-z}} \quad x$$

$$\frac{1+e^{-z}}{1+e^{-z}} - \frac{1}{1+e^{-z}}$$

$$\cancel{(1-y)}$$

$$\frac{d\mathcal{L}}{dw} = (y-d)x$$

Q. 4

original Entropy $= -\frac{4}{9}\log_2\left(\frac{4}{9}\right) - \frac{5}{9}\log_2\left(\frac{5}{9}\right) = 0.991$

Entropy given a1 is T $= -\frac{3}{4}\log_2\left(\frac{3}{4}\right) - \frac{1}{4}\log_2\left(\frac{1}{4}\right) = 0.811$

" " " " F $= -\frac{1}{5}\log_2\left(\frac{1}{5}\right) - \frac{4}{5}\log_2\left(\frac{4}{5}\right) = 0.722$

Information gain:

$$Ig(y|a1) = 0.991 - \left(\frac{4}{9}0.811 + \frac{5}{9}0.722\right) = 0.23$$

$E(y|a2=T) = H\left(\frac{2}{5}, \frac{3}{5}\right) = 0.971$

$E(y|a2=F) = H\left(\frac{1}{2}, \frac{1}{2}\right) = 1$

$$Ig(y|a2) = 0.007$$

splitting for a1 is better