Page ____

CSCI 5521
Introduction to Machine Learning
Homework 5

Name : Saksham Goel
ID : goelx029 | 5138568

Q1 Using the backprop handout, I will define the following

let M = # of samples, also will use superscript $^{(m)}$ to denote m^{th} sample.
& instead of using E, I will use $E^{(m)}$ to denote the error at m^{th} sample.
then

$$E^{(m)} = \frac{1}{2}\sum_{i=1}^{h}\left(z_i^{(m)} - t_i^{(m)}\right)^2$$

So we can define total error as E as follows:

$$E = \sum_{m=1}^{M} E^{(m)}$$

Then we can calculate δ_i as follows

$$\delta_i = \frac{\partial E}{\partial \hat{z}_j} = \frac{\partial E}{\partial z_i}\frac{\partial z_i}{\partial \hat{z}_i}$$

$$\text{for } \frac{\partial E}{\partial z_i} = \frac{\partial}{\partial z_i}\sum_{m=1}^{M} E^{(m)} = \sum_{m=1}^{M}\frac{\partial}{\partial z_i}\frac{1}{2}\sum_{i=1}^{h}\left(z_i^{(m)} - t_i^{(m)}\right)^2$$

$$\Rightarrow \frac{\partial E}{\partial z_i} = \sum_{m=1}^{M}\frac{1}{2}\sum_{i=1}^{h}\frac{\partial}{\partial z_i}\left(z_i^{(m)} - t_i^{(m)}\right)^2$$

$$\frac{\partial E}{\partial z_i} = \sum_{m=1}^{M} \cancel{\frac{1}{2}} \times \cancel{2}\left(z_i^{(m)} - t_i^{(m)}\right)$$

$$\boxed{\frac{\partial E}{\partial z_i} = \sum_{m=1}^{M} \left(z_i^{(m)} - t_i^{(m)}\right)}$$

so $$\delta_i = \frac{\partial E}{\partial z_i}\frac{\partial z_i}{\partial \hat{z}_i} = \boxed{\left\{\sum_{m=1}^{M} \left(z_i^{(m)} - t_i^{(m)}\right)\right\} g'\left(\hat{z}_i^{(m)}\right)}$$

Now we can find γ_j

$$\gamma_j = \frac{\partial E}{\partial \hat{y}_j} = \frac{\partial E}{\partial y_j}\frac{\partial y_j}{\partial \hat{y}_j}$$

$$= \left(\sum_{i=1}^{k} \delta_i \, \omega_{ij}\right) g'(\hat{y}_j)$$

$$\boxed{\gamma_j = \left\{\sum_{m=1}^{M} \left[\left\{\sum_{i=1}^{k}\left[z_i^{(m)} - t_i^{(m)}\right]\right\} g'\left(\hat{z}_i^{(m)}\right)\right\} \omega_{ij}^{(m)}\right\} g'\left(\hat{y}_j^{(m)}\right)}$$

Now finding $\frac{\partial E}{\partial w_{ij}}$

$$\boxed{\frac{\partial E}{\partial w_{ij}} = \delta_i y_j = \left\{\left\{\sum_{m=1}^{M}\left(\left(z_i^{(m)} - t_i^{(m)}\right)\right\} g'\left(\hat{z}_i^{(m)}\right)\right\} y_j^{(m)}\right)}$$

for $i \in \{1, k\}$
$j \in \{0, n\}$

Also finding $\frac{\partial E}{\partial v_{jk}}$

$$\frac{\partial E}{\partial v_{jk}} = \delta_j \, x_k$$

$$\boxed{\frac{\partial E}{\partial v_{jk}} = \left\{ \sum_{n=1}^{M} \left[\sum_{i=1}^{h} \left(\left(z_i^{(n)} - t_i^{(n)}\right) g'\left(z_i^{(n)}\right) w_{ij}^{(n)} \right) \right] g'\left(y_j^{(n)}\right) x_k^{(n)} \right\}}$$

b) Log Likelihood $= \log\left(p^t \, (1-p)^{1-t} \right)$

$$Lh = \log(p^t) + \log\left((1-p)^{1-t}\right)$$

$$= t \log(p) + (1-t) \log(1-p)$$

$$\frac{\partial Lh}{\partial w_j} = \frac{\partial}{\partial w_j} \left(t \log(p) + (1-t) \log(1-p) \right)$$

$$= \frac{\partial}{\partial w_j} t \log(p) + \frac{\partial}{\partial w_j} (1-t) \log(1-p)$$

$$= \frac{t}{p} \frac{\partial p}{\partial w_j} + (1-t) \frac{-1}{1-p} \frac{\partial p}{\partial w_j}$$

$$= \left\{ \left(\frac{t}{p}\right) - \left(\frac{1-t}{1-p}\right) \right\} \frac{\partial p}{\partial w_j} = \left(\frac{t - \cancel{tp} - p + \cancel{tp}}{p(1-p)} \right) \frac{\partial p}{\partial w_j}$$

Scanned by CamScanner

finding $\frac{\partial f}{\partial w_j} \rightarrow \frac{\partial}{\partial w_j}\left(\frac{1}{1+e^{-w^Tx}}\right)$

$$= \frac{+1}{(1+e^{-w^Tx})^2} \times e^{-w^Tx} \times +1 \times x_j$$

$$= \frac{e^{-w^Tx}\, x_j}{(1+e^{-w^Tx})^2}$$

$$= \frac{1}{1+e^{-w^Tx}}\left(\frac{e^{-w^Tx}}{1+e^{-w^Tx}}\right) x_j$$

$$= f\,(1-f)\, x_j$$

So $\frac{\partial LL}{\partial w_j} = \frac{t-f}{\cancel{f(1-f)}}\, \cancel{f(1-f)}\, x_j$

$$\boxed{\frac{\partial LL}{\partial w_j} = (t-f)\,x_j}$$

Now because we are trying to maximize the log likelihood, we will use gradient ascent.

Gradient ascent

while not converges
    for each sample in
        $w_j = w_j + (\underbrace{t}_{\text{true label}} - \underbrace{f}_{\text{probability given by sigmoid function}})\,x_j$

Date
Page

Q2 a) For figure 3, the shaded region is marked by the line:

$$x_2 > -1.5 \Rightarrow \boxed{x_2 + 1.5 > 0}$$

we can convert this to the perceptron as follow

$$0x_1 + 1x_2 + 1.5 \cdot 1 > 0$$

$$= \begin{bmatrix} 1.5 \\ 0 \\ 1 \end{bmatrix}^T \begin{bmatrix} 1 \\ x_1 \\ x_2 \end{bmatrix} > 0$$

Hence $\boxed{w = \begin{bmatrix} 1.5 \\ 0 \\ 1 \end{bmatrix}}$

Similarly for figure 4: shaded region is given by

$$x_1 - 2x_2 > 2 \Rightarrow x_1 - 2x_2 - 2 > 0$$

$$= -2 \times 1 + 1 \times x_1 - 2 \times x_2 > 0$$

$$= \begin{bmatrix} -2 \\ 1 \\ -2 \end{bmatrix}^T \begin{bmatrix} 1 \\ x_1 \\ x_2 \end{bmatrix} > 0$$

So $\boxed{w = \begin{bmatrix} -2 \\ 1 \\ -2 \end{bmatrix}}$

Date
Page

b) For this part we can use the weights from previous part to get the result in the hidden layer.

So $V = \begin{bmatrix} 1.5 & 0 & 1 \\ -2 & 1 & -2 \end{bmatrix}$

with the given non linearity this will ensure that the hidden layer will have the four cases

Y_1	Y_2
1	1
1	0
0	1
0	0

Now we know our output should be 1 only if Y_1 & Y_2 both are one. So all that is left is to implement a AND function using W.

to do that if we use weights as $W = \begin{bmatrix} -1.5 & 1 & 1 \end{bmatrix}$

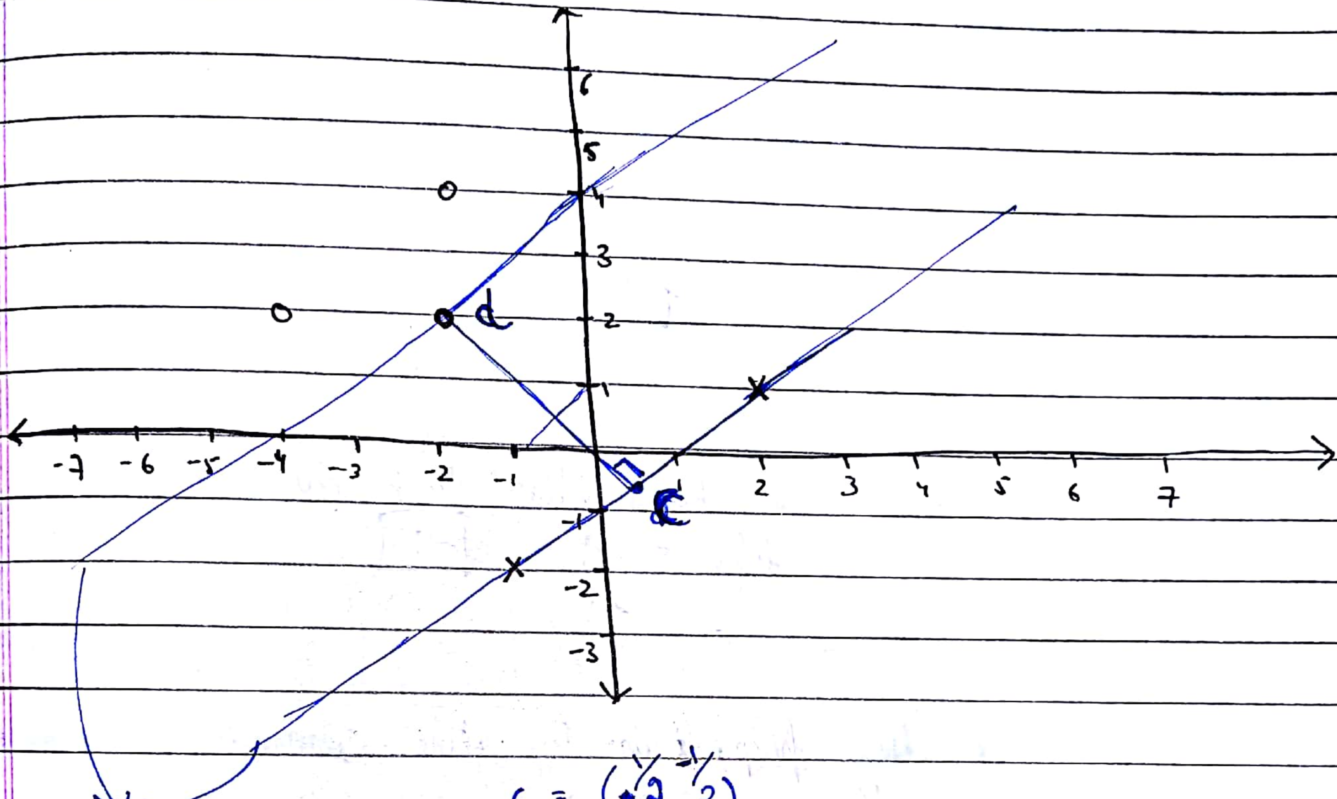
we would get

Y_1	Y_2	$W \begin{bmatrix} 1 \\ Y_1 \\ Y_2 \end{bmatrix}$	Activation
1	1	0.5	1
1	0	-0.5	0
0	1	-0.5	0
0	0	-1.5	0

& this is what we want. hence

$$\boxed{V = \begin{bmatrix} 1.5 & 0 & 1 \\ -2 & 1 & -2 \end{bmatrix} \quad W = \begin{bmatrix} -1.5 & 1 & 1 \end{bmatrix}}$$

Q3 SVM



Supportors

$c = \left(\tfrac{1}{2}, -\tfrac{1}{2}\right)$

$d = \left(-2, 2\right)$

So we can find $w = -1(-2, 2) + \frac{1}{2}(-1, -2) + \frac{1}{2}(2, 1)$

$= (2, -2) + \left(\frac{1}{2}, -\frac{1}{2}\right)$

$$\boxed{w = \left(\frac{5}{2}, -\frac{5}{2}\right)}$$

We found c because it was the closest "-1" point to the "+1" convex hull. to find d we had to do the following calculations:

line connecting two +1 points ⇒

$$y+2 = 1(x+1)$$

$$\boxed{y = x - 1} \quad (i)$$

So $\boxed{\text{Slope} = 1}$

Now the line perpendicular to it will have

$$\text{Slope} = \frac{-1}{1} = \boxed{-1}$$

Now the line perpendicular to line connecting +1 point & passing through -2, 2 is:-

$$(y-2) = -1(x+2)$$

$$\boxed{y = -x} \quad (ii)$$

Now finding the intersection of (i) & (ii) we get d

$$-x = x - 1$$

$$1 = 2x \Rightarrow \boxed{x = 0.5} \quad \text{So } \boxed{y = -0.5}$$

So $\boxed{d = \begin{bmatrix} 0.5 \\ -0.5 \end{bmatrix}}$

so $w = \begin{bmatrix} 0.5 \\ -0.5 \end{bmatrix} - \begin{bmatrix} -2 \\ 2 \end{bmatrix}$

$$\boxed{w = \begin{bmatrix} 2.5 \\ -2.5 \end{bmatrix}}$$

Now to find $b \rightarrow$ such that classifier is given by $w^T x + b > 0$, we ~~can~~ have a formula as follows

$$b = -\frac{\max\limits_{y^{(i)}=-1} w^T x^{(i)} + \min\limits_{y^{(i)}=1} \{ w^T x^{(i)}}{2}$$

$$= -\frac{\begin{bmatrix} 2.5 \\ -2.5 \end{bmatrix}^T \begin{bmatrix} -2 \\ 2 \end{bmatrix} + \begin{bmatrix} 2.5 \\ -2.5 \end{bmatrix}^T \begin{bmatrix} 2 \\ 1 \end{bmatrix}}{2}$$

$$= -\left(\frac{-10 + 2.5}{2}\right)$$

$$= -\left(\frac{-7.5}{2}\right) = \boxed{+3.75}$$

& seperator $\rightarrow$ $\boxed{2.5x_1 - 2.5x_2 + 3.75 > 0} \rightarrow$ if it is greater than 0; predict class label =1 else -1

$$\boxed{\text{Seperator function}(x_1, x_2) = \begin{cases} 1 & 2.5x_1 - 2.5x_2 + 3.75 > 0 \\ -1 & \text{otherwise} \end{cases}}$$

Now we need to scale the values of w so that the values of $w^T x + b = \pm 1$ for support vectors.

we know w is of form $(a, -a)$ then

for support vector $\rightarrow$ $(2, 1)$

$$2a - a + b = 1 = \boxed{a + b = 1} \quad \text{— (iii)}$$

for support vector $(-2, 2)$

$$-2a - 2a + b = -1 \quad \boxed{-4a + b = -1} \quad \text{— (iv)}$$

Solving (iii) & (iv) we get

$$5a = 2, \quad \boxed{a = 2/5}$$

$$\text{so } \boxed{b = 3/5}$$

So now $\boxed{w = \left(\frac{2}{5}, \frac{-2}{5}\right)}$ & $\boxed{b = 3/5}$

So the final separator is as follows:-

$$\text{Separator-function}(x_1, x_2) \begin{cases} 1 & \frac{2}{5}x_1 - \frac{2}{5}x_2 + \frac{3}{5} > 0 \\ -1 & \text{otherwise} \end{cases}$$

& the margin is as follows →

$$\rho = \frac{2}{\|w\|} = \frac{2}{\sqrt{\left(\frac{2}{5}\right)^2 + \left(\frac{-2}{5}\right)^2}}$$

$$= \frac{2}{\sqrt{\frac{4+4}{25}}} = \frac{\cancel{2} \times 5}{\cancel{2}\sqrt{2}}$$

$$= \boxed{\frac{5}{\sqrt{2}}}$$