CIS 472/572 Homework #5 (Written)

Mamtaj Akter

TOTAL POINTS

18 / 21

QUESTION 1

Q1 7 pts

1.1 Part a 2 / 2

✓ - 0 pts Correct

1.2 Part b 2 / 2

✓ - 0 pts Correct

1.3 Part c 3 / 3

✓ - 0 pts Correct

QUESTION 2

Q2 9 pts

2.1 Part a 3 / 3

✓ - 0 pts Correct

2.2 Part b 2 / 3

✓ - 1 pts Not a clear construction method.

2.3 Part c 3 / 3

✓ - 0 pts Correct

QUESTION 3

3 Q3 (your answer should say 472 if you are enrolling in 472) 3 / 5

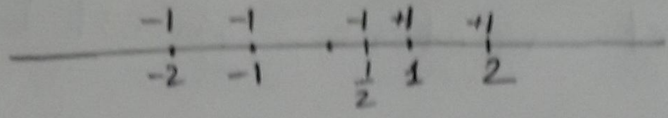
✓ - 1 pts Gradient for bias not calculated.

✓ - 1 pts Partially correct.

You have forgotten that the The log-likelihood function is the sum of what you have here. (Before applying the partial derivation)

Written Homework - 3

1. (a)



The separator line can be any line in between $\frac{1}{2}$ and 1.

So, lets suppose the line goes on $(\frac{1}{2}, -1)$ point

$wx + b = 0$

$\Rightarrow w \cdot \frac{1}{2} + b = 0$ $\quad\Rightarrow \frac{w}{2} = -b$

$\Rightarrow b = -\frac{1}{2}w$ $\quad\Rightarrow w = -2b$

Now, suppose the separator line goes through $(1, +1)$ point.

$wx + b = 0$

$\Rightarrow w \cdot 1 + b = 0$ $\quad\Rightarrow w = -b$

$\Rightarrow b = -w$

So, the parameter space of b can be any in between $-w$ and $-\frac{1}{2}w$, that is

$-w < b < \frac{-1}{2}w$

And the parameter space of w can be any in between $-b$ and $-2b$, that is

$-b < w < -2b$

1.1 Part a 2 / 2

✓ - 0 pts Correct

1 (b) for the maximum - margin separator, the support vectors must be $\frac{1}{2}$ and 1.

so, in point $\frac{1}{2}$,

$wx + b = -1$

$\Rightarrow w \cdot \frac{1}{2} + b = -1$ - - - - - (1)

And in point 1,

$wx + b = 1$

$\Rightarrow w \cdot 1 + b = 1$ - - - - - (2)

After solving equation (1) and (2), we get

$w = 4$ and $b = -3$.

1.2 Part b 2 / 2

✓ **- 0 pts** Correct

<u>case 2:</u> B and D are support vectors
C is violating the margin

here we get $\omega = 1$ and $b = 0$.

After calculating $\sum \max(0, 1 - y(\omega x + b))$
we get 1.5

For $C = 5$, the hinge loss is 8
$C = 0.5$ the loss is 1.25
$C = 0.05$ " " " 0.575

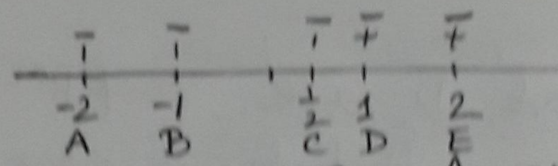
<u>case 3:</u> C and E are the support vectors.
D is violating the margin.

in point C and E, we get $\omega = \frac{4}{3}$ and
$b = -\frac{5}{3}$

$\sum \max(0, 1 - y(\omega x + b)) = 1.33$

For $C = 5$, the loss $= 7.538$
$C = 0.5$, the loss is 1.545
$C = 0.05$, loss $=$ 0.9465

1.(c)



Case 1: Lets assume A and D are support vectors and B and C are ~~the~~ violating the margin constraints.

in A:

$wx + b = -1$

$-2w + b = -1$

in D:

$wx + b = +1$

$w + b = 1$.

we get $w = \frac{2}{3}$ and $b = \frac{1}{3}$

From the formula:

after calculating $\sum \max(0, 1 - y(wx + b))$

we get $\frac{7}{3} = 2.33$

For $C = 5$, the loss is $\frac{1}{2}\|w\|^2 + C \sum \max(0, 1 - y(wx + b))$

$= 11.87$

For $C = 0.5$, the loss is 1.385

For $C = 0.05$, " " " 0.339

* Case 4 : support vectors are A and E
violation points : B, C, D.

here $w = \frac{1}{2}$, $b = 0$.

$\sum \max(0, 1 - y(wx+b)) = 2.25$.

for $C = 5$, loss is 11.375

$C = 0.5$, loss is 1.25

$C = 0.05$, loss is 0.2375 .

* Case 5 : support vectors are B and E
violation points : C and D.

here $w = \frac{2}{3}$ and $b = -\frac{1}{3}$

$\sum \max(0, 1 - y(wx+b)) = 1.67$

For $C = 5$, Loss $= 8.57$

$C = 0.5$, Loss $= 1.055$

$C = 0.05$, loss $= 0.3035$.

o, optimal parameters for $C = 5$ are $w = \frac{4}{3}$ & $b = -\frac{5}{3}$

" " for $C = 0.5$, $w = \frac{2}{3}$ & $b = -\frac{1}{3}$

" " for $C = 0.05$, $w = \frac{1}{2}$ & $b = 0$.

1.3 Part c 3 / 3

✓ - 0 pts Correct

$$a_3(x) = \begin{cases} +1 & \text{when } wx+b>0 \\ -1 & \text{otherwise} \end{cases}$$

Let's find out the sets of w_1, w_2, w_3 and b for hidden node a_1 which satisfies example 1.

$$w_1 x_1 + w_2 x_2 + w_3 x_3 + b > 0$$
$$(-1)\cdot 0 + (-1)\cdot 0 + (-1)\cdot 0 + 1 > 0$$

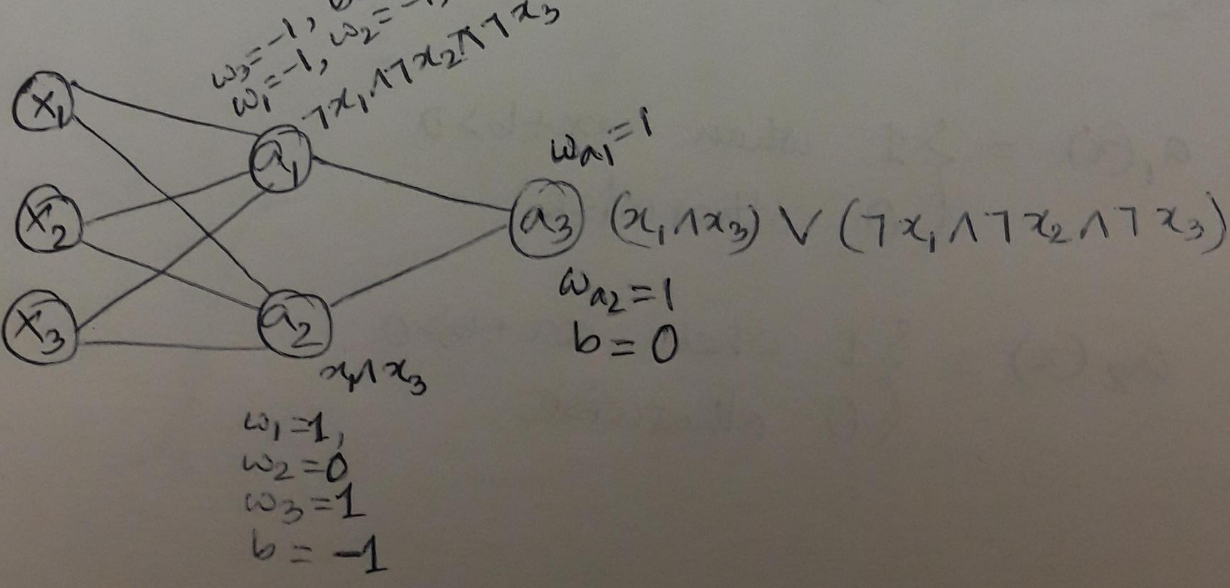
So, a_1 : $w_1 = -1$
$w_2 = -1$
$w_3 = -1$
$b = 1$

Now pick the values of w_1, w_2, w_3 and b for hidden node a_2 which satisfies example 6 and 8

$$w_1 x_1 + w_3 x_3 + b > 0$$
$$1 \cdot 1 + 1 \cdot 1 + (-1) > 0$$

Finally for the output node a_3,

lets, $w_{a_1} = 1$, $w_{a_2} = 1$ and $b = 0$



2(a) The data set of the given decision tree:

Example	x_1	x_2	x_3	y
1:	0	0	0	+1
2:	0	0	1	-1
3:	0	1	0	-1
4:	0	1	1	-1
5:	1	0	0	-1
6:	1	0	1	+1
7:	1	1	0	-1
8:	1	1	1	+1

The possible logical function of the output node a_3 is $(x_1 \land x_3) \lor (\neg x_1 \land \neg x_2 \land \neg x_3)$

Let's put the function $\neg x_1 \land \neg x_2 \land \neg x_3$ for hidden node a_1 and $x_1 \land x_3$ in hidden node a_2.

Then decide the activation function of a_1 and a_2 and a_3.

$$a_1(x) = \begin{cases} 1 & \text{when } wx+b>0 \\ 0 & \text{otherwise} \end{cases}$$

$$a_2(x) = \begin{cases} 1 & \text{when } wx+b>0 \\ 0 & \text{otherwise} \end{cases}$$

2.1 Part a 3 / 3

✓ - 0 pts Correct

2.(b) A general method to convert a decision tree to a 2-layer neural network:

As we transformed the decision tree of problem 1(a) into a 2-layer neural network:

→ we can formulate the output of the decision tree as a combination of different logical functions in hidden nodes, as neural network can be represented in any logical function.

→ At each hidden node, individual logical function will be calculated. And every node has different weights and bias according to their activation function.

→ All the outputs of hidden nodes will go to the output node as inputs. In output node, the final output is calculated according to ~~these~~ final function on those inputs. Here, the weights and bias for inputs of the output node will be according to the activation function of output node.

2.2 Part b 2 / 3

✓ **- 1 pts** **Not a clear construction method.**

2(c) → Ensemble of k decision trees can be represented as a single 2-layer neural network with k hidden units. Activation function of each hidden unit will be logical function of each decision tree.

→ The output node will take outputs of all the k ~~di~~ hidden nodes with activation function = Sign (sum of all inputs)

→ for output node, all weights of inputs will be same and that is 1 and bias is 0.

→ If output ~~is~~ comes positive, majority of the k hidden node predicted +1. If the output comes negative, this means, the majority of the k hidden node predict −1.

2.3 Part c 3 / 3

✓ - 0 pts Correct

According to the chain rule:

$$\frac{\partial}{\partial x} \log e^{exp} = \frac{1}{e^{exp}} \cdot e^{exp} \cdot \frac{\partial exp}{\partial x}$$

So, we can write

$$\sum_{n=1}^{N} \left[x - \frac{1 \cdot \sum_{k=1}^{C} e^{w_k^T x + b} \cdot x}{\sum_{k=1}^{C} e^{w_k^T x + b}} \right]$$

$$= \sum_{n=1}^{N} \left[x - \sum_{k=1}^{C} \frac{e^{w_k^T x + b} \cdot x}{\sum_{k=1}^{C} e^{w_k^T x + b}} \right]$$

$$= \sum_{n=1}^{N} \left[x - \sum_{k=1}^{C} P(Y=k|x)\, x \right]$$

$$= \sum_{n=1}^{N} \left[x - x \sum_{k=1}^{C} P(Y=k|x) \right]$$

$$= \sum_{n=1}^{N} x \left[1 - \sum_{k=1}^{C} P(Y=k|x) \right]$$

3. For each example, the probability:

$$P(Y=i|K) = \frac{e^{w_i^T x+b}}{\sum_{k=1}^{c} e^{w_x^T +b}}$$

So, total probability:

$$P(Y|K) = \prod_{n=1}^{N} \frac{e^{w_i^T x+b}}{\sum_{k=1}^{c} e^{w_k^T x+b}}$$

Taking conditional log likelihood of the cost function:

$$\log P(Y/K) = \sum_{n=1}^{N} \log \frac{e^{w_i^T x+b}}{\sum_{k=1}^{c} w_k^T x+b}$$

$$= \sum_{n=1}^{N} \left[\log e^{w_i^T x+b} - \log \sum_{k=1}^{c} e^{w_k^T x+b}\right]$$

$$\boxed{\text{Gradient} = \frac{\partial L}{\partial w}} \qquad = \sum_{n=1}^{N} \left[\cancel{\log e} \; w_i^T x+b - \log \sum_{k=1}^{c} e^{w_k^T x+b}\right]$$

So,

$$\frac{\partial}{\partial w}\left[\sum_{n=1}^{N} [w_i^T x+b - \log \sum_{k=1}^{c} e^{w_k^T x+b}\right]$$

$$= \sum_{n=1}^{N} \frac{\partial}{\partial w}\left[w_i^T x+b - \log \sum_{k=1}^{c} e^{w_k^T x+b}\right]$$

$$= \sum_{n=1}^{N}\left[x - \frac{\partial}{\partial w}\left(\log \sum_{k=1}^{c} e^{w_k^T x+b}\right)\right] \qquad \left[\text{As } b \text{ is constant } \& \frac{\partial w}{\partial w} = 1\right.$$

$$\cancel{= \sum_{n=1}^{N}}$$

3 Q3 (your answer should say 472 if you are enrolling in 472) **3 / 5**

- ✓ **- 1 pts** **Gradient for bias not calculated.**
- ✓ **- 1 pts** **Partially correct.**

You have forgotten that the The log-likelihood function is the sum of what you have here. (Before applying the partial derivation)