Assignment 2

By Brandon Young and Ruicheng Wu

1 problem 2

1.1 a.

We claim the provided tree correctly categorize the provided example since every example can be inducted from this decision tree. Like GPA above 3.6 is P and below 3.3 is N, then with publication P,otherwise check University Rank, Only rank 2 will be P,other ranks are N. And recommendation doesn't matter.

1.2 b.

Step I

I($\frac{6}{12}, \frac{6}{12}$)=1

GPA: $[3.9, 4.0]$ $3(PPP)$, $(3.2, 3.9)$ $5(PPPNN)$, $[3.0, 3.2]$ $4(NNNN)$

University: Rank 1— 5(PPPNN),Rank 2— 3(PPN) , Rank 3— 4(PNNN)

Publication: Yes 5(PPPNN) , No 7(PPPNNNN)

Recommendation: good 8(PPPPPNNN) , normal 4(PNNN)

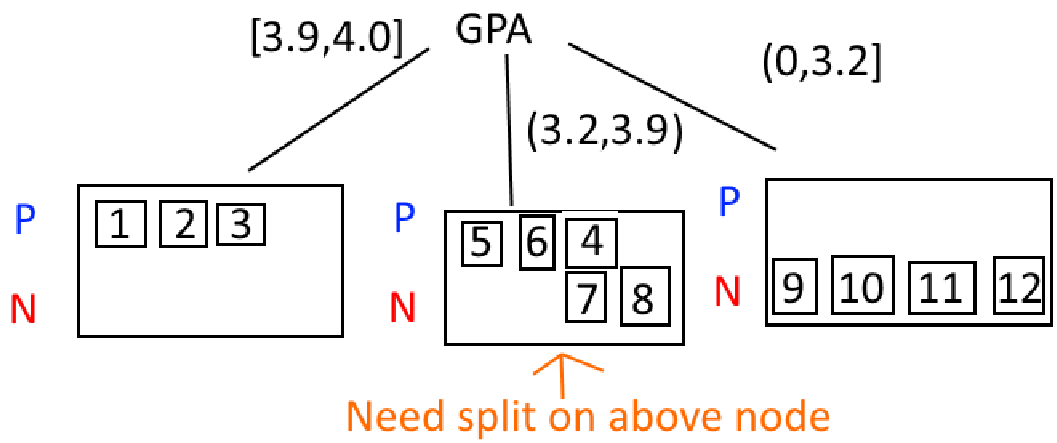
Gain(GPA)=1-[$\frac{3}{12}$ B($\frac{3}{3}$)+$\frac{5}{12}$ B($\frac{3}{5}$)+$\frac{4}{12}$ B($\frac{0}{4}$)]=1-[0.0+0.404562747689+0.0] = 0.595437252311

Gain(University)=1-[$\frac{5}{12}$ B($\frac{3}{5}$)+$\frac{3}{12}$ B($\frac{2}{3}$)+$\frac{4}{12}$ B($\frac{1}{4}$)]= 1-[0.404562747689+0.229573958514+0.270426041486] =0.095437252395

Gain(Publication)=1-[$\frac{5}{12}$ B($\frac{3}{5}$)+$\frac{7}{12}$ B($\frac{3}{7}$)]=1-[0.404562747689+0.574716412687]= 0.020720839624

Gain(Recommendation)=
1-[$\frac{8}{12}$ B($\frac{5}{8}$)+$\frac{4}{12}$ B($\frac{1}{4}$)]=1-[0.636289335283+0.270426041486]= 0.093284623231

So we pick GPA as the best Gain attribute in this level



Step II

I($\frac{2}{5}$, $\frac{3}{5}$)=0.970950594455

University: Rank 1— 2(PN),Rank 2— 1(P) , Rank 3— 2(PN)
Publication: Yes 2(PP) , No 3(PNN)
Recommendation: good 5(PPPNN)

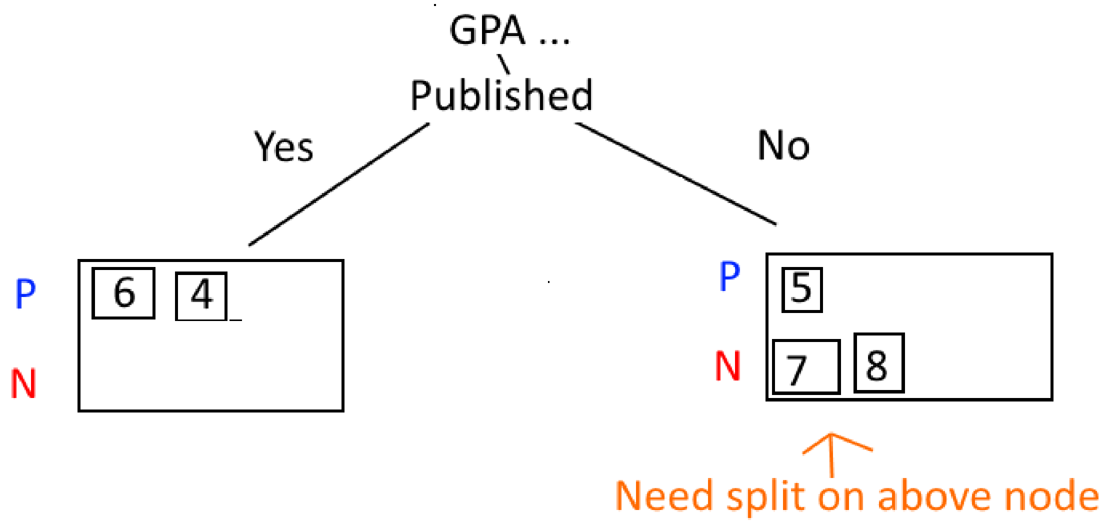
Gain(University)=0.970950594455-[$\frac{2}{5}$ B($\frac{1}{2}$)+$\frac{1}{5}$ B($\frac{1}{1}$)+$\frac{2}{5}$ B($\frac{1}{2}$)]=
0.970950594455-[0.4+0.0+0.4] = 0.170950594455

Gain(Publication)=0.970950594455-[$\frac{3}{5}$ B($\frac{1}{3}$)+$\frac{2}{5}$ B($\frac{2}{2}$)]=0.970950594455-
[0.550977500433+0.0]= 0.419973094022

Gain(Recommendation)=0.970950594455-[$\frac{5}{5}$ B($\frac{3}{5}$)]=0.970950594455-

[0.970950594455+0.0+0.0] = 0

So we pick Publication as the best Gain attribute in this level



Step III

I($\frac{1}{3}$, $\frac{2}{3}$)=0.918295834054

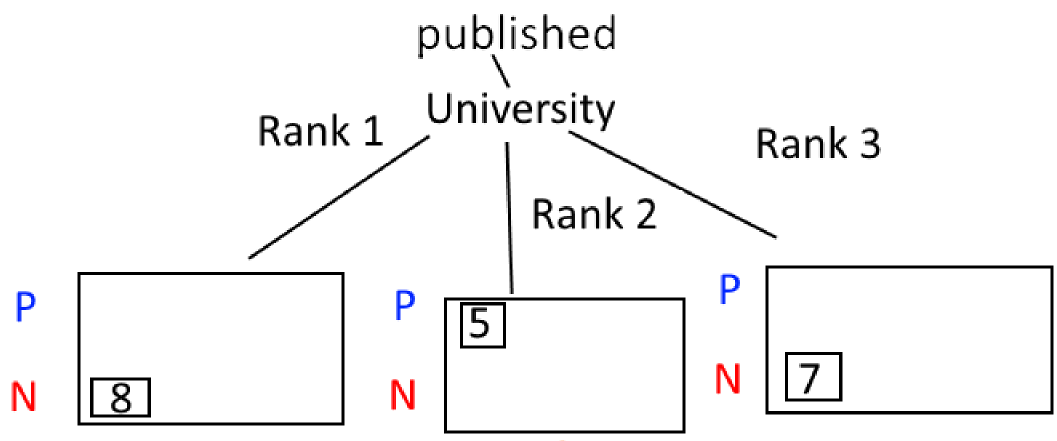
University: Rank 1 — 1(N),Rank 2 — 1(P) , Rank 3 — 1(N)

Recommendation: good 3(PNN)

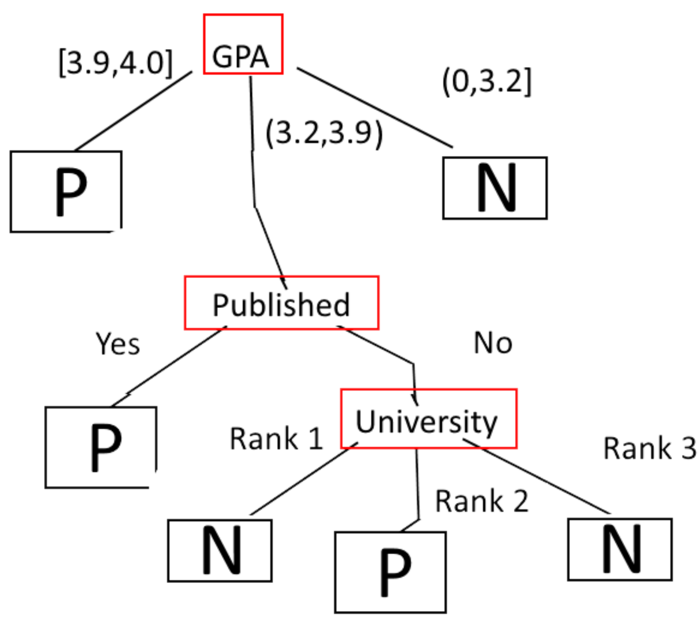
Gain(University)=0.918295834054-[$\frac{1}{3}$ B($\frac{0}{1}$)+$\frac{1}{3}$ B($\frac{1}{1}$)+$\frac{1}{3}$ B($\frac{0}{1}$)]=0.918295834054-[0.0+0.0+0.0]=0.918295834054

Gain(Recommendation)=0.918295834054-[$\frac{3}{3}$ B($\frac{1}{3}$)]=0.918295834054-[0.918295834054]= 0

So we pick University as the best Gain attribute in this level



And the final tree to be returned is



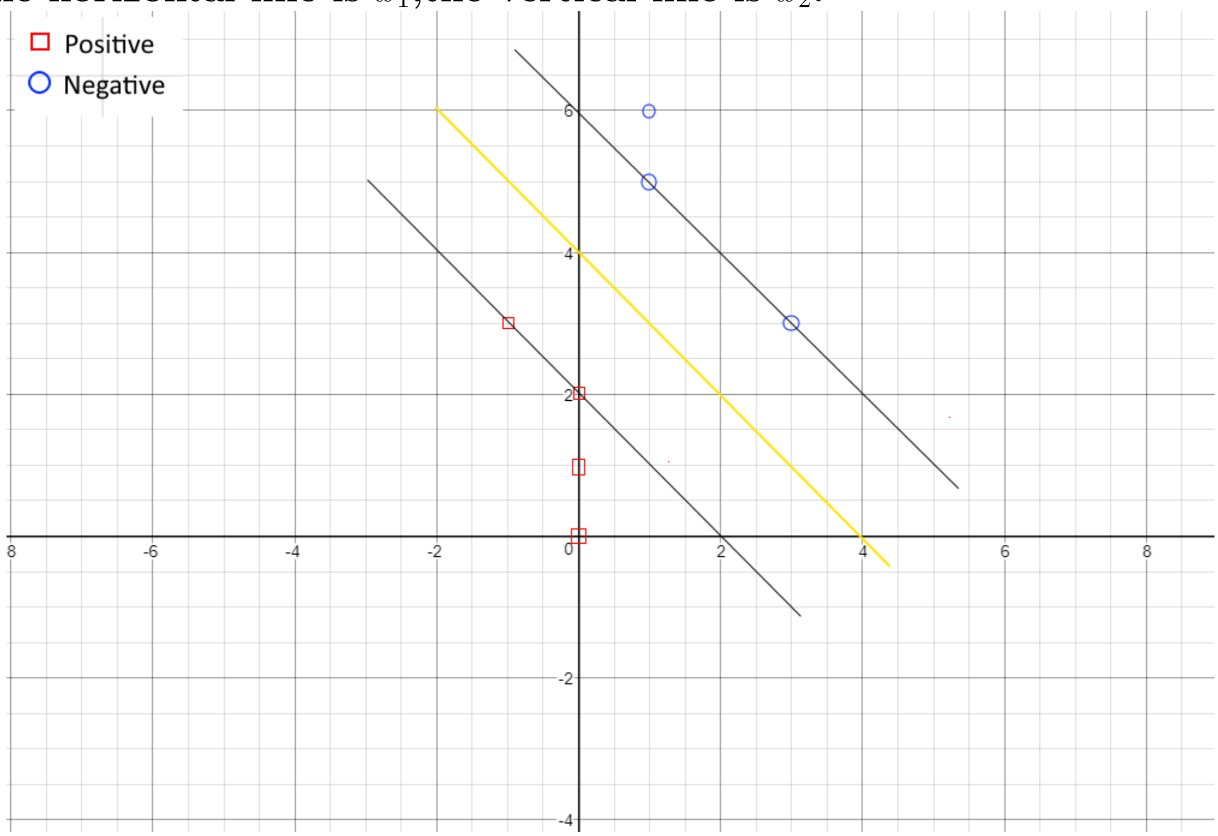
1.3 c.

The decision tree we got from b. is same from the provided one. This is not coincidence, it is believed both tree is generated by applying the decision tree algorithm.

2 problem 3

2.1 a.

The horizontal line is x_1,the vertical line is x_2.



2.2 b.

$s_1 : (1, 5)$, $s_2 : (-1, 3)$, $w : (1, 1)$, So decision boundary is: $y = (1, 1)x - 4$ Assume there is (a, a) for w^T, we will have

$$a + 5a + b = -1$$

$$-a + 3a + b = 1$$

Solve those equations for :

$$a = -0.5, b = 2$$

So we have $w^T : (-0.5, -0.5)$ and $b : 2$ for parameters. We can confirm the max Margin is $\frac{2}{||w^T||} = \frac{2}{(\frac{\sqrt{2}}{2})} = 2\sqrt{2}$, which verifies with the distance between $s_1 : (1, 5)$, $s_2 : (-1, 3)$, $\sqrt{(1 - (-1))^2 + (5 - 3)^2} = 2\sqrt{2}$.

And also we verify with given data.

[-1 3]^T [0 2]^T [0 1]^T [0 0]^T, plug in: All positive , so those are 1;

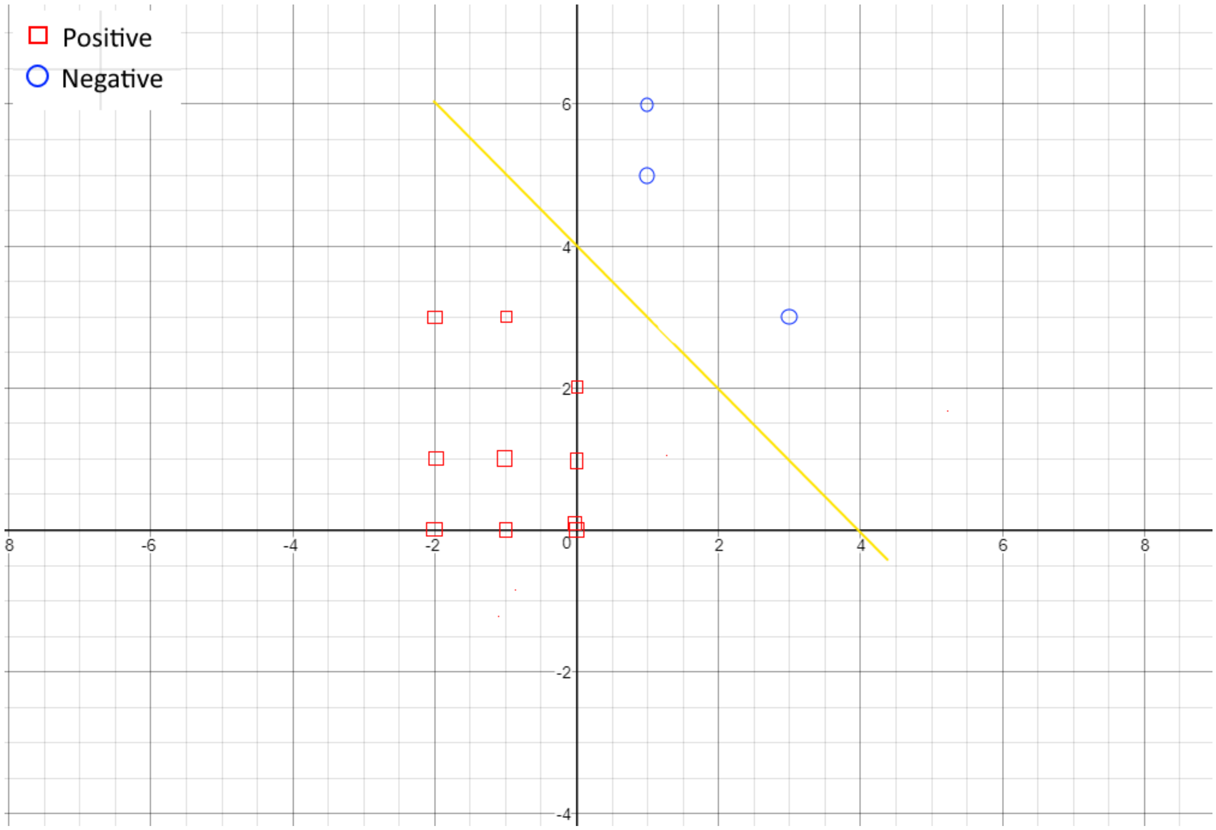
[1 5]^T [1 6]^T [3 3]^T , plug in: All negative , so those are -1;

2.3 c.

By inspection, the new added points are not support vectors(ie. the closest points to separating line)

We claim the $w^T : (-0.5, -0.5)$ and $b : 2$ are still same to parametrize with new data.

[-2 0]^T [-2 1]^T [-2 3]^T [-1 0]^T [-1 1]^T [0 0]^T, plug in: All positive , so those are 1;



3 problem 5

3.1 a.

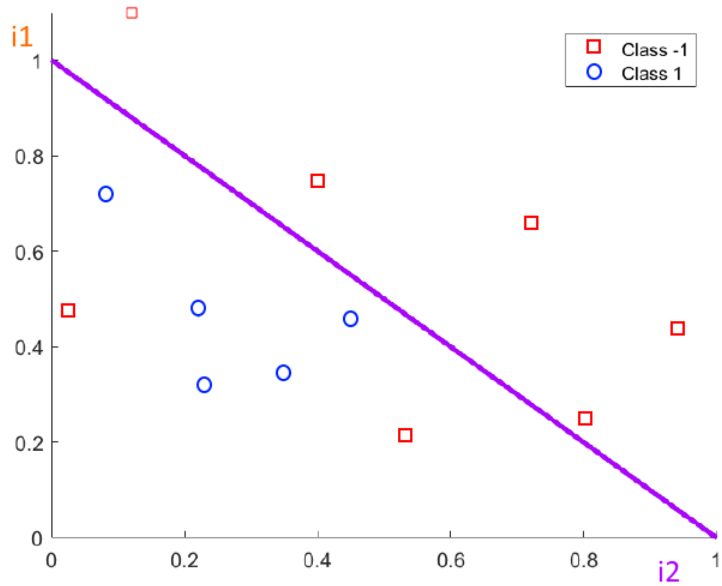
We define those points by inspection.

Class 1: (0.03,0.50),(0.11,1.2),(0.5,0.75),(0.54,0.23),(0.7,0.65),(0.8,0.25),(0.91,0.4

Class -1: (0.07,0.73),(0.23,0.49),(0.24,0.33),(0.35,0.35),(0.46,0.46);

The line can be observed from the graph, and we found the best single perceptron will have two unclassied points. So the minimum error will be $\frac{2}{12} = 0.167$;

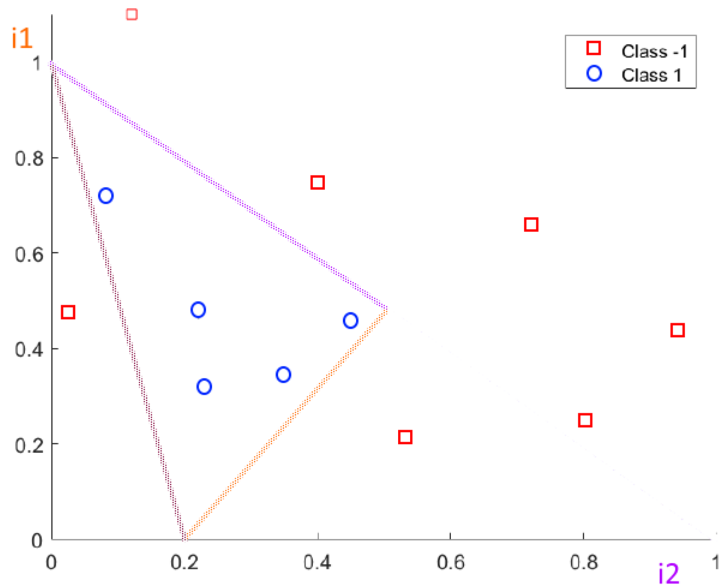
The formula for this dividing line is : $1 - 1 * x_1 - 1 * x_2 = 0$



3.2 b.

We claim that we need 3 separate lines to completely separate two classes. classes. So we need at least 3 perceptrons to compute classification functions. We will just use 3 perceptrons for simplicity.

Formula for line 1:
$1 - 1 * x_1 - 1 * x_2 = 0$
$W_{1,0} = 1, W_{1,1} = -1, W_{1,2} = -1$
Formula for line 2:
$0.2 - 1 * x_1 + 0.6 * x_2 = 0$
$W_{1,0} = 0.2, W_{1,1} = -1, W_{1,2} = 0.6$
Formula for line 3:
$-0.2 + 1 * x_1 + 0.2 * x_2 = 0$
$W_{1,0} = -0.2, W_{1,1} = 1, W_{1,2} = 0.2$

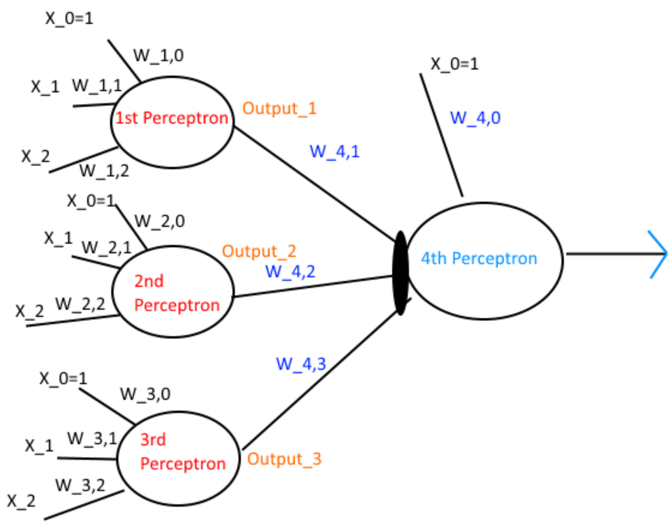


As we can see, a data point gives output as positive with all three perceptrons is class 1 and same applies to class -1. So in the next layer, we can simply apply "and" operation to the results of all thre eperceptrons. In that case, we only need one unit. And the output will be 1 if and only if all perceptrons in

1st layer output 1.

Hence we set

$$W_{4,0} = -2.5, W_{4,1} = 1, W_{4,2} = 1, W_{4,3} = 1$$


X_0=1
W_1,0
X_1
W_1,1
1st Perceptron
Output_1
X_2
W_1,2
X_0=1
W_2,0
X_1
W_2,1
2nd
Perceptron
Output_2
X_2
W_2,2
X_0=1
W_3,0
X_1
W_3,1
3rd
Perceptron
Output_3
X_2
W_3,2
W_4,1
W_4,2
W_4,3
X_0=1
W_4,0
4th Perceptron