Assignment 2

By Brandon Young and Ruicheng Wu

1 problem 1

2 problem 2

2.1 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.

2.2 b.

Step I

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

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)

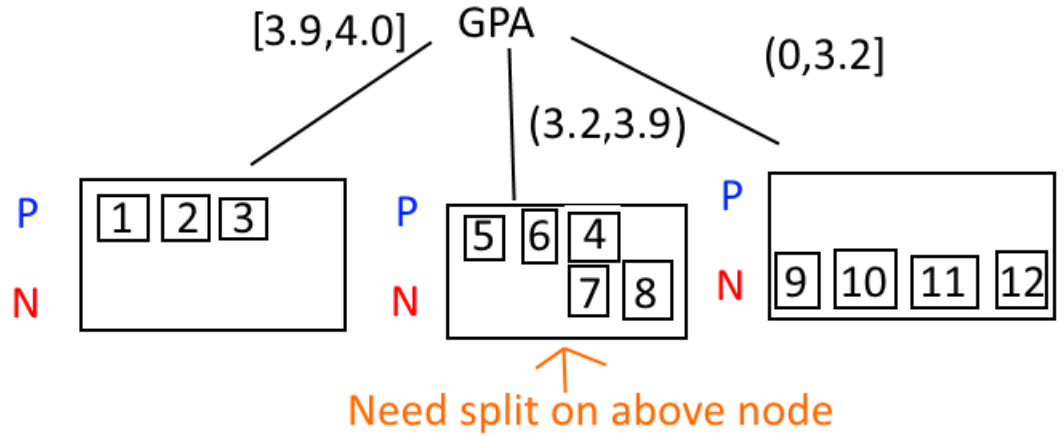Gain(GPA)=1-$[\frac{3}{12}$ B$(\frac{3}{3})+\frac{5}{12}$ B$(\frac{3}{5})+\frac{4}{12}$ B$(\frac{0}{4})]$=1-[0.0+0.404562747689+0.0] = 0.595437252311

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

[3.9,4.0] GPA (0,3.2]

(3.2,3.9)

P |1|2|3| P |5|6|4| P
                           |7|8|
N   N   N |9|10|11|12|

↑
Need split on above node

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

GPA ...

Published

Yes — No

P [6] [4]
N

P [5]
N [7] [8]

Need split on above node
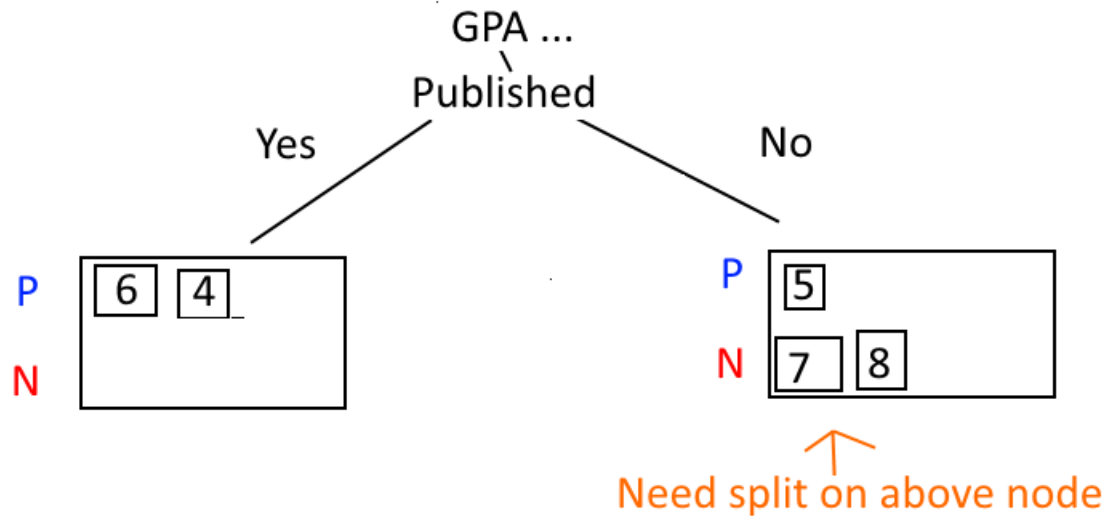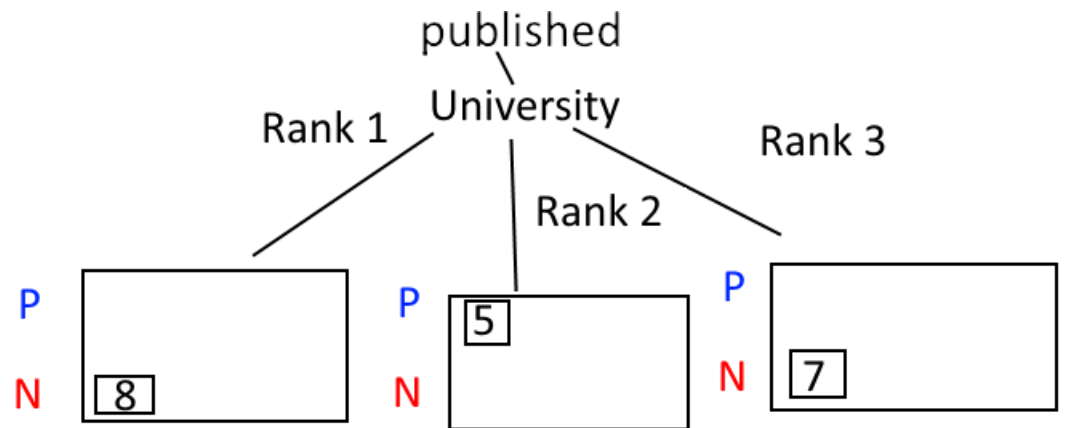
Step III
$$I(\tfrac{1}{3}, \tfrac{2}{3})=0.918295834054$$
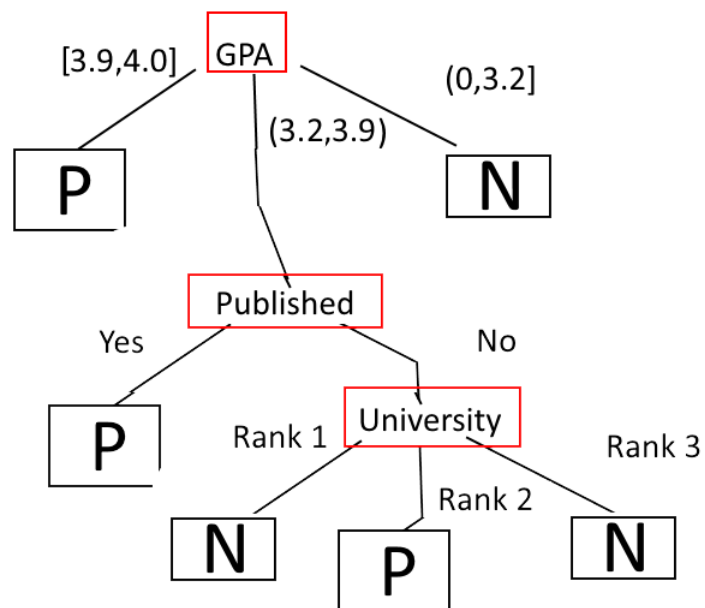University: Rank 1 — 1(N),Rank 2 — 1(P) , Rank 3 — 1(N)
Recommendation: good 3(PNN)
Gain(University)=0.918295834054-$[\tfrac{1}{3}$ B$(\tfrac{0}{1})+\tfrac{1}{3}$ B$(\tfrac{1}{1})+\tfrac{1}{3}$ B$(\tfrac{0}{1})]$=0.918295834054-[0.0+0.0+0.0]=0.918295834054

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

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

published

Rank 1   University   Rank 3

Rank 2

P   P 5   P

N 8   N   N 7

And the final tree to be returned is

[3.9,4.0] GPA   (0,3.2]

(3.2,3.9)

P   N

Published

Yes   No

P   Rank 1 University   Rank 3

Rank 2

N   P   N

5

2.3   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.

3 problem 3

3.1 a.

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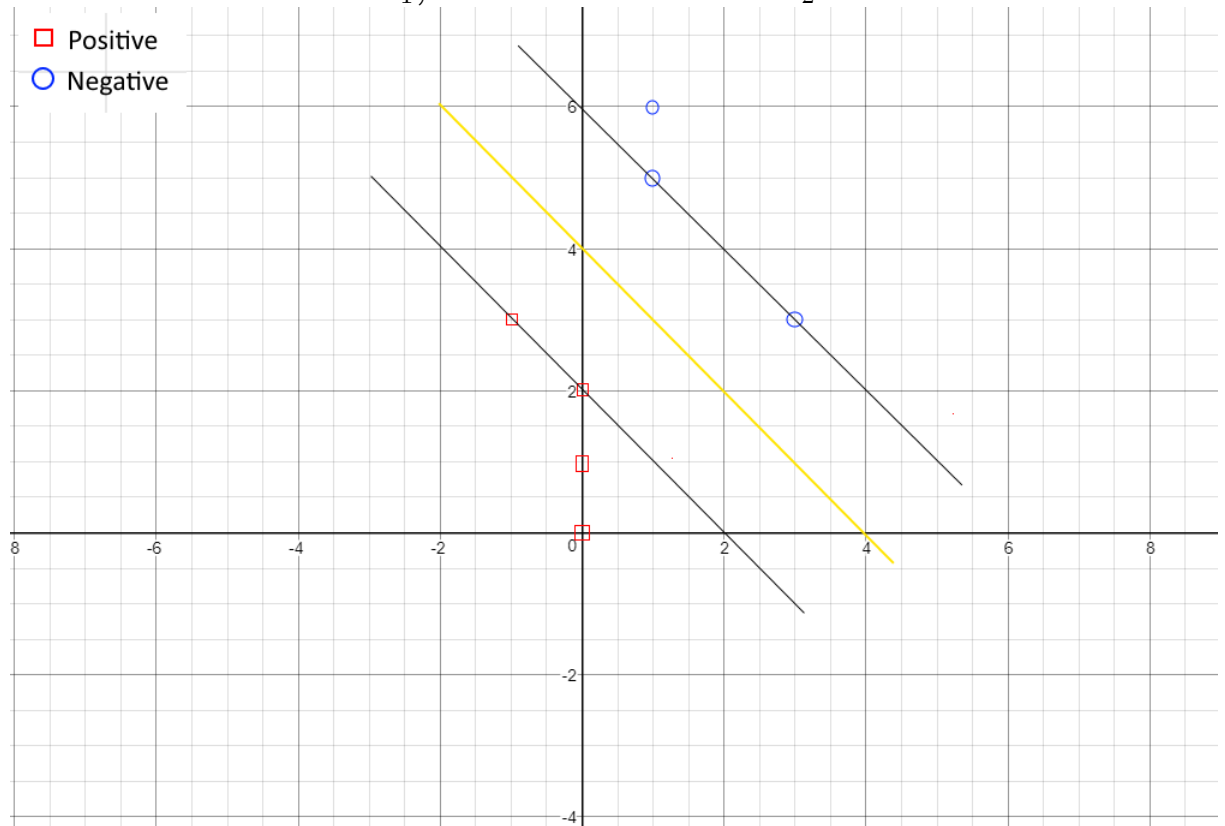
3.2 b.

$s_1 : (1,5)$, $s_2 : (-1,3)$, $w : (1,1)$, 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.

$[\text{-1 } 3]^T$ $[0 \text{ } 2]^T$ $[0 \text{ } 1]^T$ $[0 \text{ } 0]^T$, plug in: All positive , so those are 1;

$[1 \text{ } 5]^T$ $[1 \text{ } 6]^T$ $[3 \text{ } 3]^T$, plug in: All negative , so those are -1;

3.3   c.

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

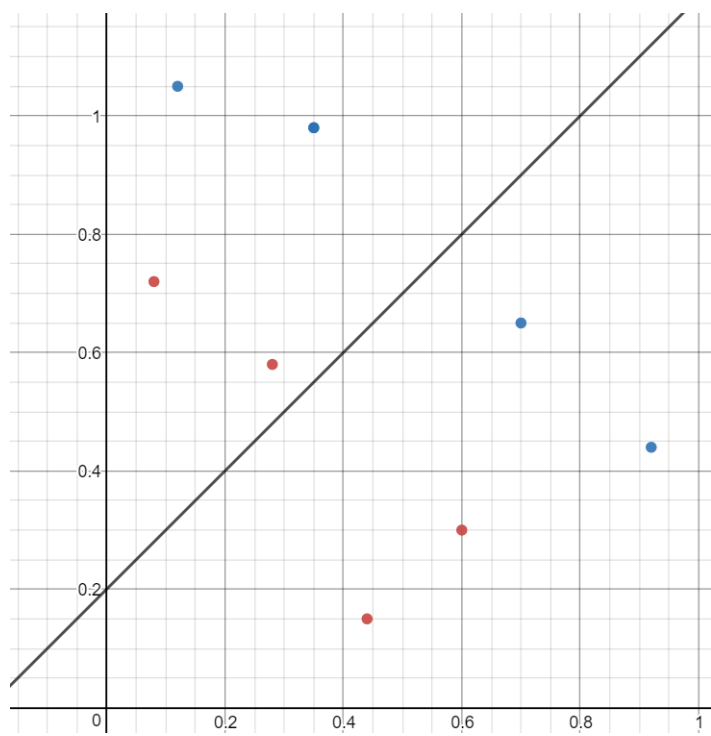We claim the $w^T : (-0.5, -0.5)$ and $b : 2$ are still same to parametrize with new data.

$[\text{-2 } 0]^T$ $[\text{-2 } 1]^T$ $[\text{-2 } 3]^T$ $[\text{-1 } 0]^T$ $[\text{-1 } 1]^T$ $[0 \text{ } 0]^T$, plug in: All positive , so those are 1;
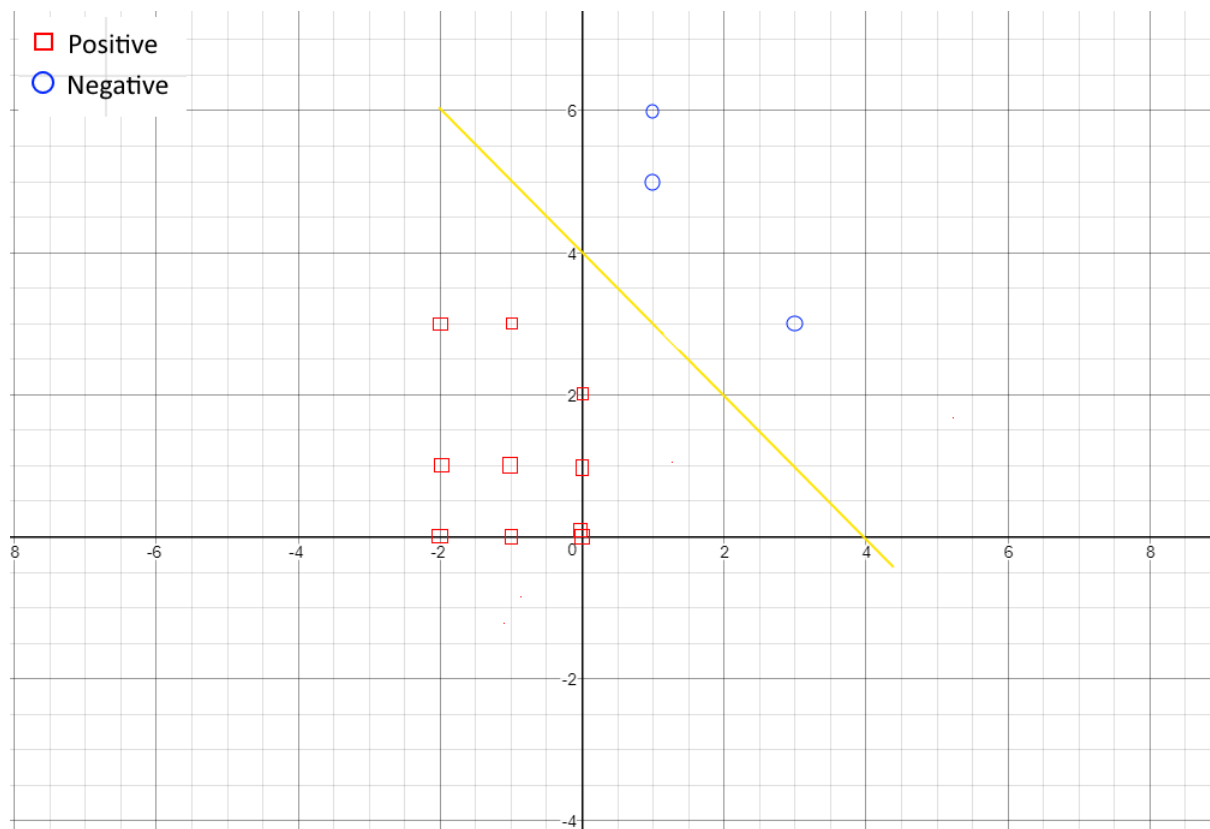
4   problem 4

4.1   a.

The initial linear separator (line 1) is $i_2 = -\frac{i_1}{i_2} - \frac{0.2}{i_2}$ or $i_2 = i_1 + 0.2$:

Assume that the points are:

- A: $(0.08, 0.72)$

- B: $(0.28, 0.58)$

- C: $(0.44, 0.15)$

- D: $(0.6, 0.3)$

- E: $(0.12, 1.05)$

- F: $(0.35, 0.98)$

- G: $(0.7, 0.65)$

- H: $(0.92, 0.44)$

where A, B, C and D are in class 1 and E, F, G and H are in class -1.

Use $f(x) = w_1 i_1 + w_2 i_2 + 0.2$ to classify the samples. If $f(x) < 0$ then x is in class -1 and $h_w(x) = 0$. If $f(x) \geq 0$ then x is in class 1 and $h_w(x) = 1$. By the initial linear separator, 4 samples are misclassified (A, B, G and H):

- A: $(0.08) - (0.72) + 0.2 = -0.44 < 0, h_w(A) = 0$

- B: $(0.28) - (0.58) + 0.2 = -0.1 < 0, h_w(B) = 0$

- C: $(0.44) - (0.15) + 0.2 = 0.49 > 0, h_w(C) = 1$

- D: $(0.6) - (0.3) + 0.2 = 0.5 > 0, h_w(D) = 1$

- E: $(0.12) - (1.05) + 0.2 = -0.73 < 0, h_w(E) = 0$

- F: $(0.35) - (0.98) + 0.2 = -0.43 < 0, h_w(F) = 0$

- G: $(0.7) - (0.65) + 0.2 = 0.25 > 0, h_w(G) = 1$

- H: $(0.92) - (0.44) + 0.2 = 0.68 > 0, h_w(H) = 1$

Let the learning rate, $\alpha = 0.5$ and use A to update the weights. Then $Err_A = y - h_w(A) = 1 - 0 = 1$.

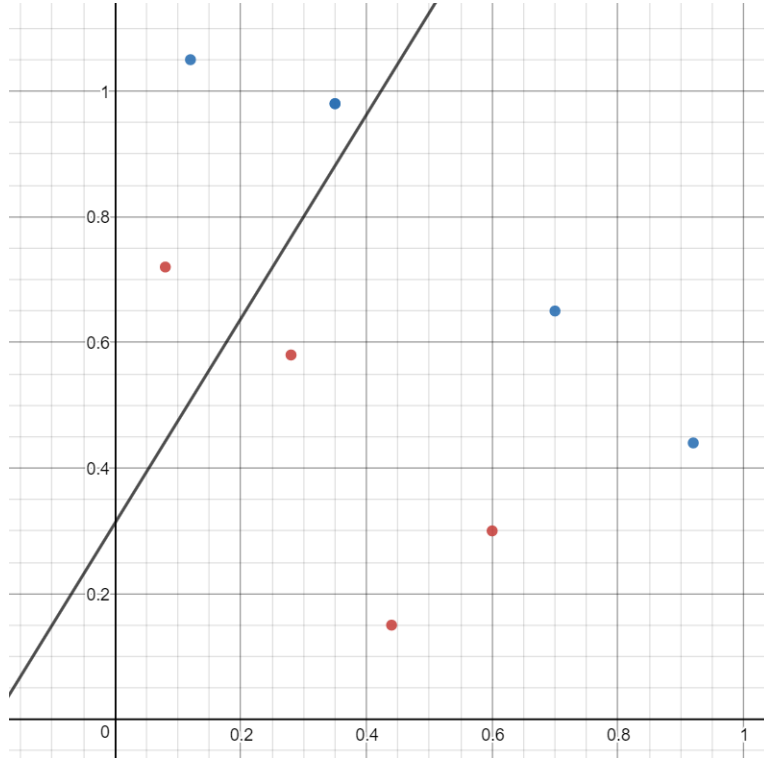$$w_1' \leftarrow w_1 + \alpha * Err_A * i_1$$

$$w_1' \leftarrow 1 + (0.5)(1)(0.08)$$

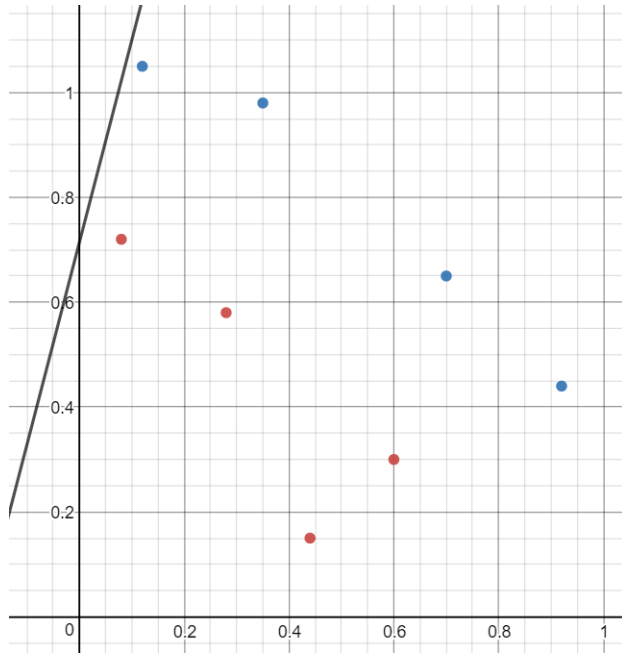$$w_1' = 1.04$$

$$w_2' \leftarrow (-1) + (0.5)(1)(0.72)$$
$$w_2' = -0.64$$

Hence the new weights are $w_1 = 1.04$ and $w_2 = -0.64$. Line 2, then, is: $i_2 = 1.625(i_1) + 0.3125$, with 3 misclassified points:
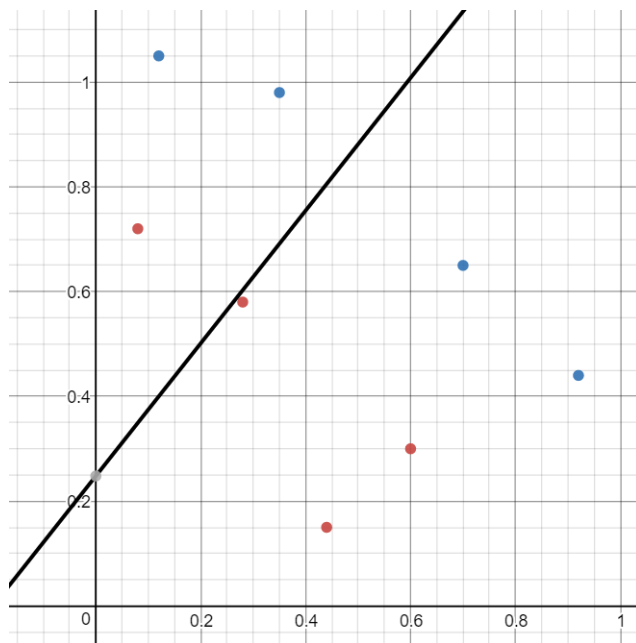

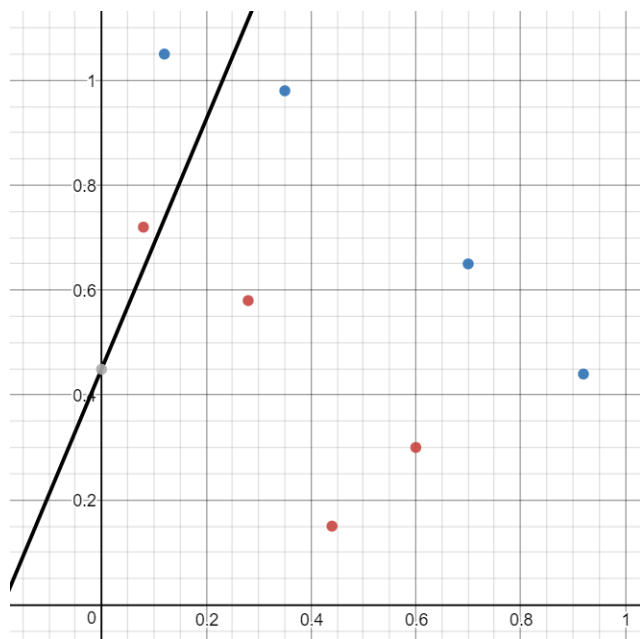Following a similar process as above, the next few iterations generate the following plots:

Update Line 2 with A to get $w_1 = 1.08$, $w_2 = -0.28$. Line 3: $i_2 = 3.857i_1 + 0.714$, with 4 misclassified points:
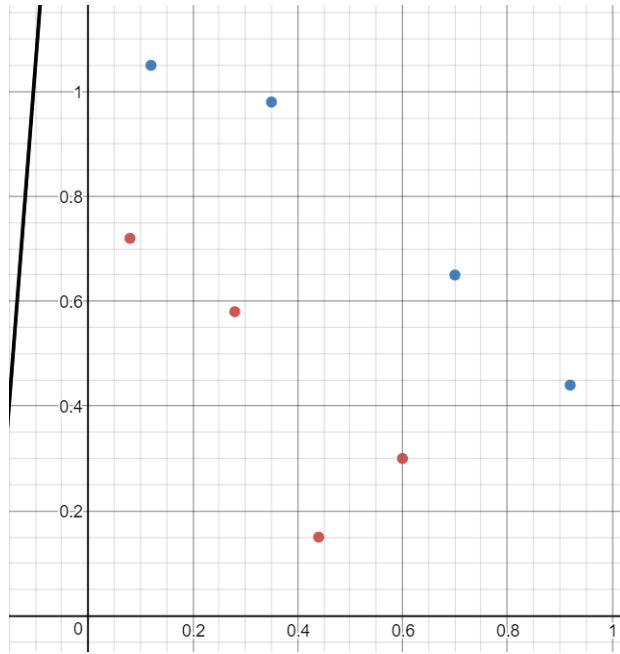
Update Line 3 with E to get $w_1 = 1.02$, $w_2 = -0.805$. Line 4, $i_2 = 1.267(i_1) + 0.248$, with 3 misclassified points:

Update Line 4 with A to get $w_1 = 1.06$, $w_2 = -0.445$. Line 5, $i_2 = 2.382(i_1) + 0.449$, with 4 misclassified points:

Update Line 5 with A to get $w_1 = 1.1$, $w_2 = -0.085$. Line 6, $i_2 = 12.941(i_1) + 2.35$, with 4 misclassified points:
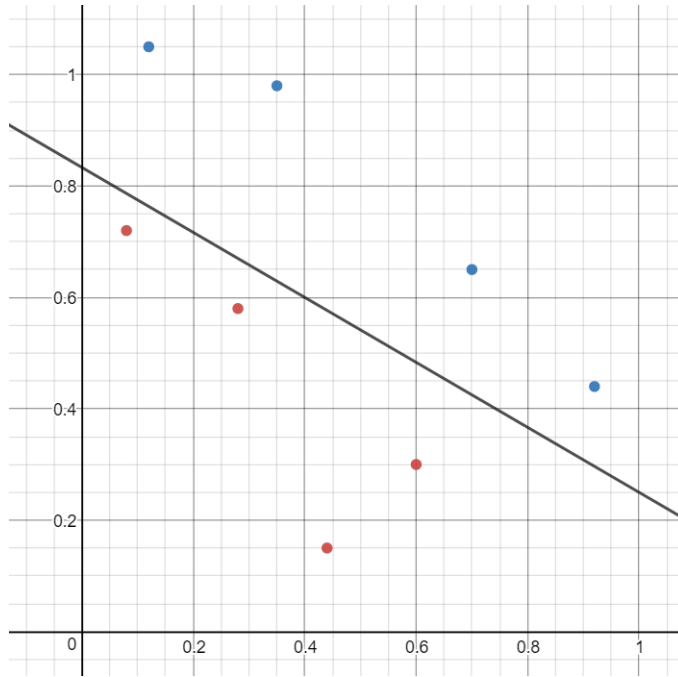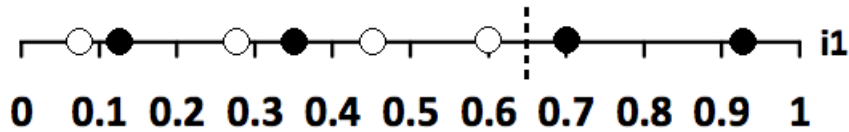
4.2   b.

The linear separator that achieves perfect classification (found after 172 iterations) is:

$$i_2 = -0.58333i_1 + 0.83333$$

where $w_1 = -0.14$ and $w_2 = -0.24$. The line in the graph:

4.3 c.

The diagram above shows only the i_1 value from each sample. The dotted line represents the location of the best possible split, which is at $i_1 = 0.65$, with only 2 misclassified points and a 25% error.

A separator in 1D is $w_1 i_1 + w_0 i_0$ or $w_1 i_1 + 0.2$ (assuming $w_0 = 0.2$) which can be used to solve for w_1:

$$w_1 = -\frac{0.2}{i_1} = -\frac{0.2}{0.65} = -0.3077$$

As a result, $w_1 = -0.3077$ is the weight that best classifies the

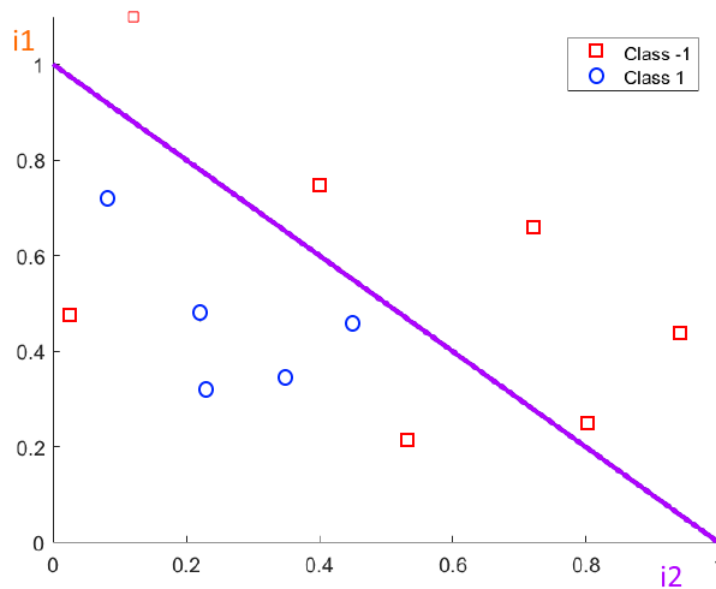samples.

5 problem 5

5.1 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 : $1 - 1 * x_1 - 1 * x_2 = 0$

5.2 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:
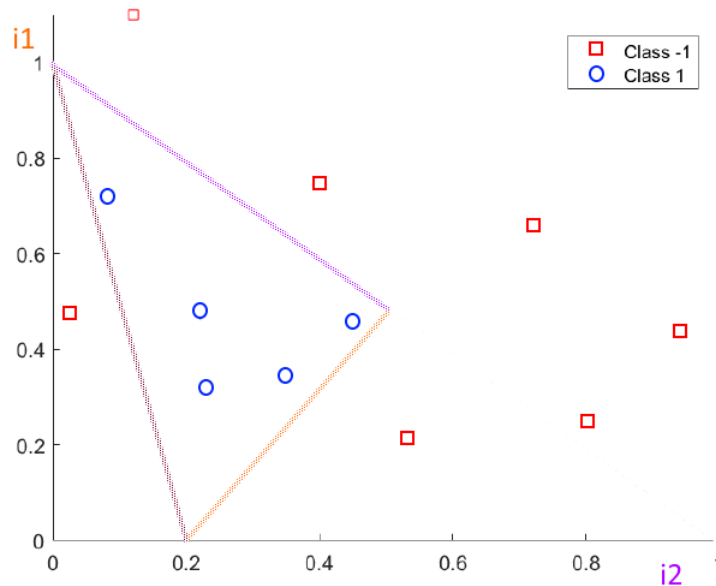
$0.2 - 1 * x_1 + 0.6 * x_2 = 0$

$W_{1,0} = 0.2, W_{1,1} = -1, W_{1,2} = 0.6$

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

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