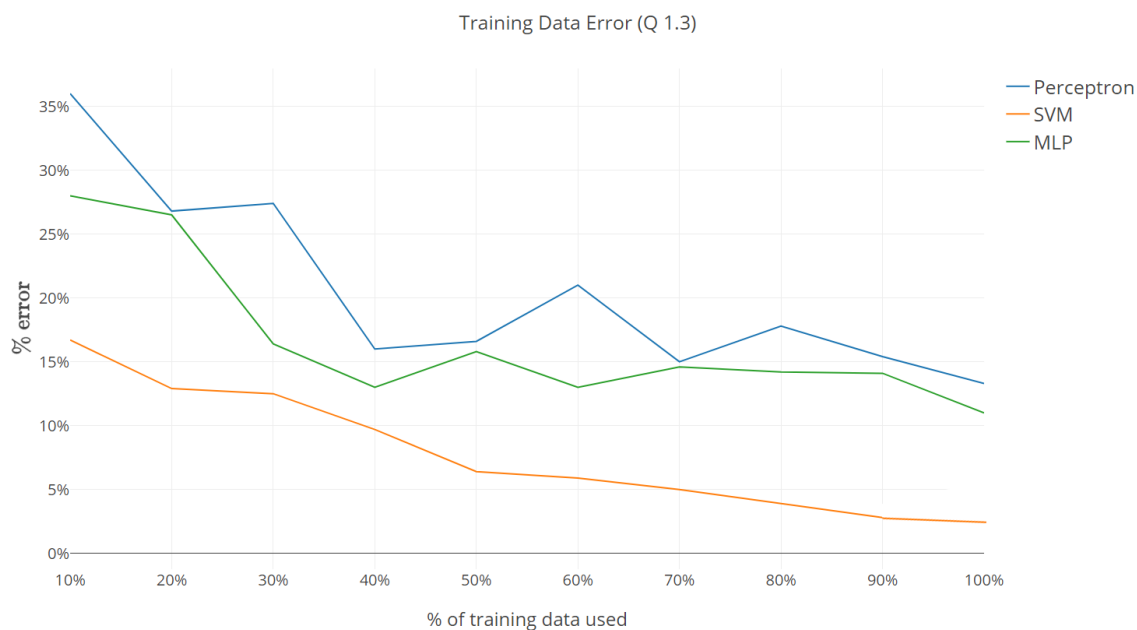


# Assignment 2

by Tomer Levy and Dhruv Bhardwaj

October 25, 2017

## 1 Question 1



### Perceptron Training

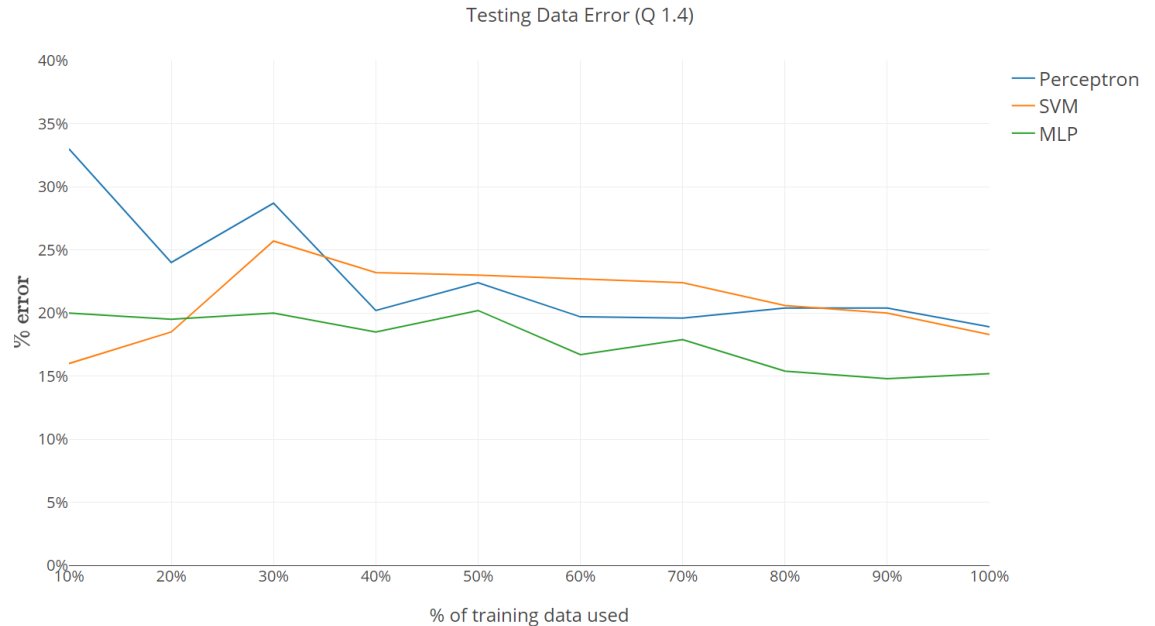
10%: 64% on training. 20% : 83.5% on training. 30% : 82.6% on training. 40% : 84% on training. 50% : 83.4% on training. 60% : 79% on training. 70% : 85% on training. 80% : 82.2% on training. 90% : 84.6% training. 100% : 86.7% on training.

### SVM Training

10%: 83% on training. 20% : 87.5% on training. 30% : 87.6% on training. 40% : 90.3% on training. 50% : 93.6% on training. 60% : 94.1% on training. 70% : 95% on training. 80% : 96.2% on training. 90% : 97.3% on training. 100%: 97.5% on training.

## MLP Training

10%: 72% on training. 20% : 73.5% on training. 30% : 83.6% on training. 40% : 87% on training. 50% : 84.2% on training. 60% :87% on training. 70% : 86.4% on training. 80% : 86.8% on training. 90% : 86.9% training. 100% : 89% on training.



(Percent error is found doing 100% - the value recieved on testing)

## Perceptron Testing

10%: 67% on testing. 20% : 76% on testing. 30% : 71.3% on testing. 40% : 79.8% on testing. 50% : 77.6% on testing. 60% :80.3% on testing. 70% : 80.4% on testing. 80% : 79.6% on testing. 90% : 79.7% on testing. 100% : 81.1% on testing.

## SVM Testing

10%: 84% on testing. 20% : 81.5% on testing. 30% : 74.3% on testing. 40% : 76.8% on testing. 50% : 77% on testing. 60% :77.3% on testing. 70% : 77.6% on testing. 80% : 79.4% on testing. 90% : 80% on testing. 100%: 81.7% on testing.

## MLP Testing

10%: 80% on testing. 20% : 80.5% on testing. 30% : 80% on testing. 40% : 81.5% on testing. 50% : 79.8% on testing. 60% :83.2% on testing. 70% : 82.1% on testing. 80% : 84.6% on testing. 90% : 85.2% on testing. 100%: 84.8% on testing.

## Q1.5 (summary)

According to the plotted values we've recieved from our testing set, our classification methods have some differences between them. One of the most flagrant differences is that perceptron does the worst out of all of them, more generally with less training data percentage used. For example, with 10% training data, perceptron hit a 67% accuracy, while the other two were around 80-85%. Towards the end of using all the training data, perceptron and SVM seemed to converge at 81.5% whereas MLP got 85% accuracy. Given more time and resources, we could have ran tons of iterations on our classification methods; this takes a ton of time. We could also apply a learning rate if we knew how, which would also give us better results and longer runtime. Further, if we had a larger training data size we may have also been able to get past those pesky 7, and 9 errors that our program continuously gets wrong.

The runtimes of these algorithms are extremely varying, SVM runs extremely fast, perceptron runs pretty slow, and MLP takes a very long time. With these points in mind, we'd have to say that we like SVM the best out of all these classification methods, due to its indistinguishable accuracy from compared methods, and its extremely fast runtime.

## 2 Question 2

A) We will say that the provided tree correctly categorizes the provided examples. Every example given by the table can be proven by the given tree. Every GPA  $\geq 3.9$  is a pass, whereas anything less than 3.2 is a fail. If GPA is between 3.2 and 3.9, then publications is checked. If publications is yes, then yes, otherwise, check university ranking. Recommendation is irrelevant. These given examples on the table all fall properly within the tree, thus it is a valid decision tree.

B)  $I(6/12, 6/12) = 1$

$$Gain(A) = I\left(\frac{p}{p+n}, \frac{n}{p+n}\right) - \sum_{i=1}^n \frac{p_i + n_i}{p+n} I\left(\frac{p_i}{p_i + n_i}, \frac{n_i}{p_i + n_i}\right)$$

### Step 1

GPA: [Possible: 3.9-4.0] : 3 [PPP]////[Possible 3.2-3.9] : 5 [PPPNN]////[Possible 0-3.2] : 4[NNNN] Gain[GPA]= $1 - \left[\frac{3}{12} B\left(\frac{3}{3}\right) + \frac{5}{12} B\left(\frac{3}{5}\right) + \frac{4}{12} B\left(\frac{0}{4}\right)\right] = 1 - [0 + 0.40456 + 0] = 0.59544$

University Ranking: [Possible: 1] : 5 [PPPNN]////[Possible 2] : 3 [PPN]////[Possible 3] : 4[NNNN] Gain[University Ranking]= $1 - \left[\frac{5}{12} B\left(\frac{3}{5}\right) + \frac{3}{12} B\left(\frac{2}{3}\right) + \frac{4}{12} B\left(\frac{1}{4}\right)\right] = 1 - [.40456 + 0.22957 + 0.27043] = 0.09544$

Publication: [Possible: Yes] : 5 [PPPNN]////[Possible No] : 7 [PPPNNNN] Gain[Publication]= $1 - \left[\frac{5}{12} B\left(\frac{3}{5}\right) + \frac{7}{12} B\left(\frac{3}{7}\right)\right] = 1 - [.40456 + 0.57472] = 0.02072$

Recommendation: [Possible: Good] : 8 [PPPPPPPN]////[Possible Normal] : 4[PNNN] Gain[Recommendation]= $1 - [\frac{8}{12} B(\frac{5}{8}) + \frac{4}{12} B(\frac{1}{4})] = 1 - [0.39445 + 0.27043] = 0.33512$

We take GPA as it provides the best gain

## Step 2

$$I(2/5, 3/5) = 0.971$$

University Ranking: [Possible: 1] : 2 [PN]////[Possible 2] : 1[P]////[Possible 3] : 2 [PN] Gain[University Ranking]= $0.971 - [\frac{2}{5} B(\frac{1}{2}) + \frac{1}{5} B(\frac{1}{1}) + \frac{2}{5} B(\frac{1}{2})] = 0.971 - [0.4 + 0 + 0.4] = 0.171$

Publication: [Possible: Yes] : 2 [PP]////[Possible No] : 3 [PNN] Gain[Publication]= $0.971 - [\frac{3}{5} B(\frac{1}{3}) + \frac{2}{5} B(\frac{2}{2})] = 0.971 - [0 + 0.5510] = 0.42$

Recommendation: [Possible: Good] : 5 [PPPNN] Gain[Recommendation]= $0.971 - [\frac{5}{5} B(\frac{3}{5})] = 0.971 - [0.971] = 0$

We take Publication as it provides the best gain

## Step 3

$$I(2/3, 1/3) = 0.9183$$

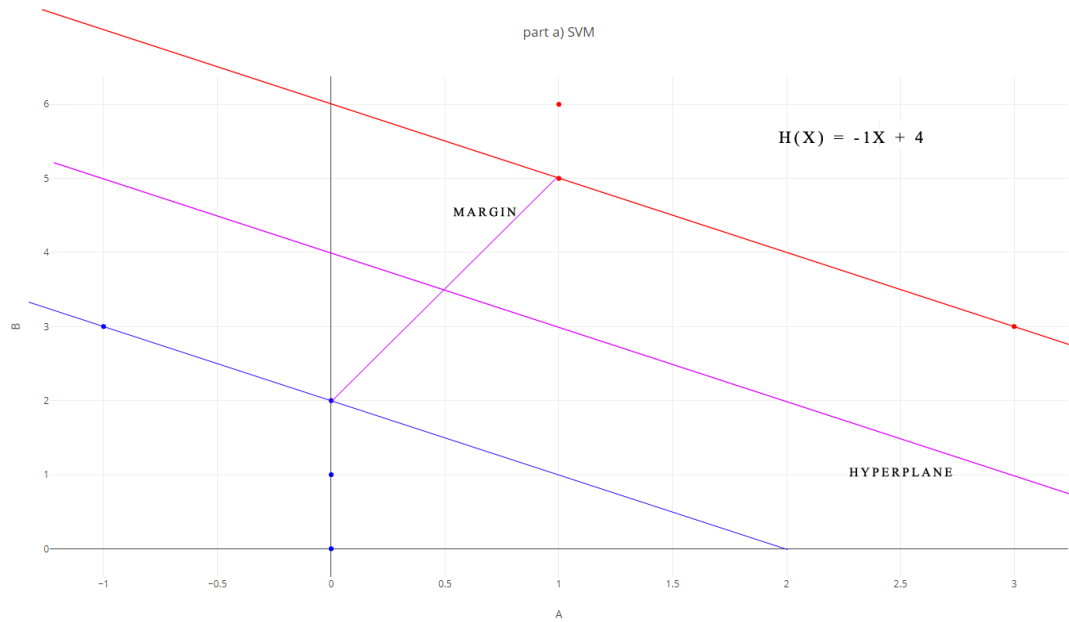
University Ranking: [Possible: 1] : 1 [N]////[Possible 2] : 1[P]////[Possible 3] : 1 [N] Gain[University Ranking]= $0.9183 - [\frac{1}{3} B(\frac{0}{1}) + \frac{1}{3} B(\frac{1}{1}) + \frac{1}{3} B(\frac{0}{1})] = 0.9183 - [0] = 0.9183$

Recommendation: Nothing Gain[Recommendation]=0 We take University Ranking as it provides the best gain

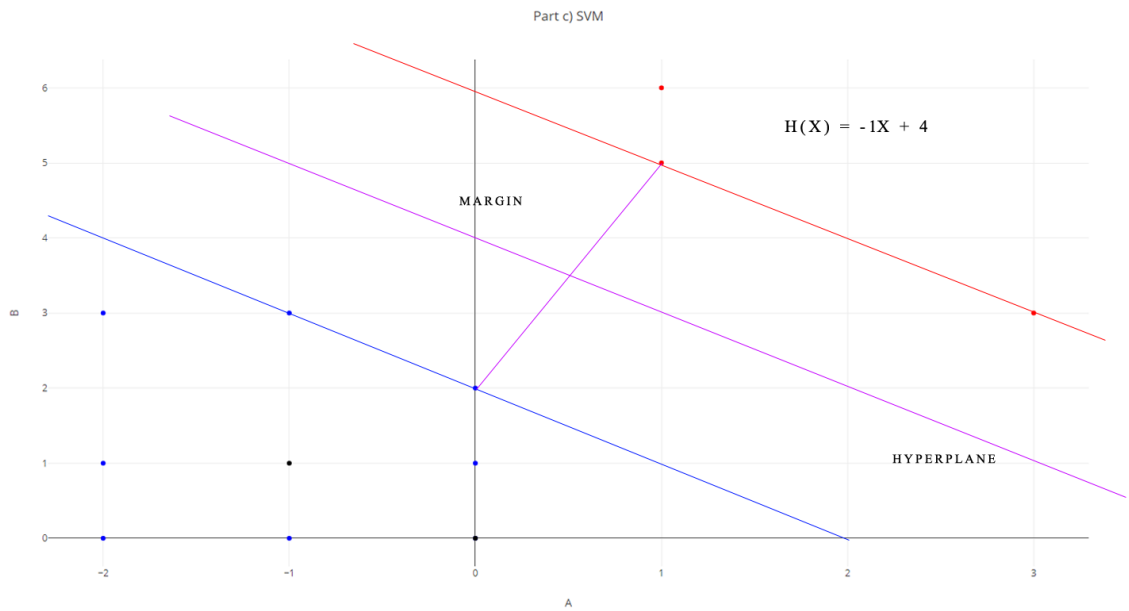
### Part C:

The decision tree we get in return of this long process turns out to be the exact same as the decision tree that was originally given to us. Following a certain path will always give you a specific solution, and since we have not observed an anomaly, this will always be true.

### 3 Question 3



This plot is for part a and b for Question 3. As you can see, our hypothesis function is  $h(x) = -x + 4$ . The margin is maximized and the lines of the positive and negative points are parallel



The added positive points have no affect on the graph / hyperplane / hypothesis.  
Our hypothesis function remains the same.

## 4 Question 4

a) There are 8 points in the graph (in order of increasing x values):

(0.075, 0.725) (0.125, 1.125) (0.25, 0.55) (0.35, 0.95)

(0.45, 0.15) (0.6, 0.3) (0.7, 0.65) (0.925, 0.45)

Initial weights:

$w_0 = 0.2$

$w_1 = 1$

$w_2 = -1$

constant offset  $i_0 = 1$

$f(x) = w_0 * i_0 + w_1 * i_1 + w_2 * i_2$

$= 0.2 + i_1 - i_2$

Line on graph:  $i_2 = 0.2 + i_1$

Let  $h(x) = 1$  when  $f(x) > 0$ , and  $h(x) = 0$  otherwise

$f(0.075, 0.725) = 0.2 + 0.075 - 0.725 < 0$

$h(x) = 0$

$f(0.125, 1.125) = 0.2 + 0.125 - 1.125 < 0$

$h(x) = 0$ , wrong

$f(0.25, 0.55) = 0.2 + 0.25 - 0.55 < 0$

$h(x) = 0$

$f(0.35, 0.95) = 0.2 + 0.35 - 0.95 < 0$

$h(x) = 0$ , wrong

$f(0.45, 0.15) = 0.2 + 0.45 - 0.15 > 0$

$h(x) = 1$ , wrong

$f(0.6, 0.3) = 0.2 + 0.6 - 0.3 > 0$

$h(x) = 1$ , wrong

$f(0.7, 0.65) = 0.2 + 0.7 - 0.65 > 0$

$h(x) = 1$

$f(0.925, 0.45) = 0.2 + 0.925 - 0.45 > 0$

$h(x) = 1$

4 misclassified points using  $w_1 = 1$  and  $w_2 = -1$

Getting line #2: Use a learning rate of 0.75, point (0.125, 1.125)

Error = expected value - calculated value

$w_1 = w_1 + (\text{learning rate}) * (\text{error}) * (i_1)$

$= 1 + (0.75) * (1 - 0) * (0.125)$

$= 1.094$

$w_2 = w_2 + (\text{learning rate}) * (\text{error}) * (i_2)$

$= -1 + (0.75) * (1 - 0) * (1.125)$

$= -0.156$

$f(i_1, i_2) = 1.094 * i_1 - 0.156 * i_2 + 0.2$

$i_2 = 7.013 * i_1 + 1.282$

$f(0.075, 0.725) = 0.2 + 0.075(1.094) - 0.725(0.156) > 0$   
 $h(x) = 1$ , wrong  
 $f(0.125, 1.125) = 0.2 + 0.125(1.094) - 1.125(0.156) > 0$   
 $h(x) = 1$   
 $f(0.25, 0.55) = 0.2 + 0.25(1.094) - 0.55(0.156) > 0$   
 $h(x) = 1$ , wrong  
 $f(0.35, 0.95) = 0.2 + 0.35(1.094) - 0.95(0.156) > 0$   
 $h(x) = 1$   
 $f(0.45, 0.15) = 0.2 + 0.45(1.094) - 0.15(0.156) > 0$   
 $h(x) = 1$ , wrong  
 $f(0.6, 0.3) = 0.2 + 0.6(1.094) - 0.3(0.156) > 0$   
 $h(x) = 1$ , wrong  
 $f(0.7, 0.65) = 0.2 + 0.7(1.094) - 0.65(0.156) > 0$   
 $h(x) = 1$   
 $f(0.925, 0.45) = 0.2 + 0.925(1.094) - 0.45(0.156) > 0$   
 $h(x) = 1$   
 4 misclassified points using  $w_1 = 1.094$  and  $w_2 = -0.156$

Getting line #3: Updating with point (0.6, 0.3):

$w_1 = 1.094 + 0.75 * (0 - 1) * 0.6$   
 $= 0.644$   
 $w_2 = -0.156 + 0.75 * (0 - 1) * 0.3$   
 $= -0.381$   
 $f(i_1, i_2) = 0.2 + 0.644*i_1 - 0.381*i_2$   
 $i_2 = 1.69*i_1 + 0.525$   
 $f(0.075, 0.725) < 0$   
 $f(0.125, 1.125) < 0$ , wrong  
 $f(0.25, 0.55) > 0$ , wrong  
 $f(0.35, 0.95) > 0$   
 $f(0.45, 0.15) > 0$ , wrong  
 $f(0.6, 0.3) > 0$ , wrong  
 $f(0.7, 0.65) > 0$   
 $f(0.925, 0.45) > 0$   
 4 misclassified values using  $w_1 = 0.644$ ,  $w_2 = -0.381$

Getting line #4: Updating weights with point (0.45, 0.15):

$w_1 = 0.644 + 0.75 * (0 - 1) * 0.45$   
 $= 0.307$   
 $w_2 = -0.381 + 0.75 * (0 - 1) * 0.15$   
 $= -0.494$   
 $f(i_1, i_2) = 0.2 + 0.307*i_1 - 0.494*i_2$   
 $i_2 = 0.621*i_1 + 0.405$   
 $f(0.075, 0.725) < 0$   
 $f(0.125, 1.125) < 0$ , wrong  
 $f(0.25, 0.55) > 0$ , wrong  
 $f(0.35, 0.95) > 0$

$f(0.45, 0.15) > 0$ , wrong  
 $f(0.6, 0.3) > 0$ , wrong  
 $f(0.7, 0.65) > 0$   
 $f(0.925, 0.45) > 0$   
 4 misclassified points with  $w1 = 0.307$ ,  $w2 = -0.494$

Getting line #5: Updating weights using (06, 0.3):

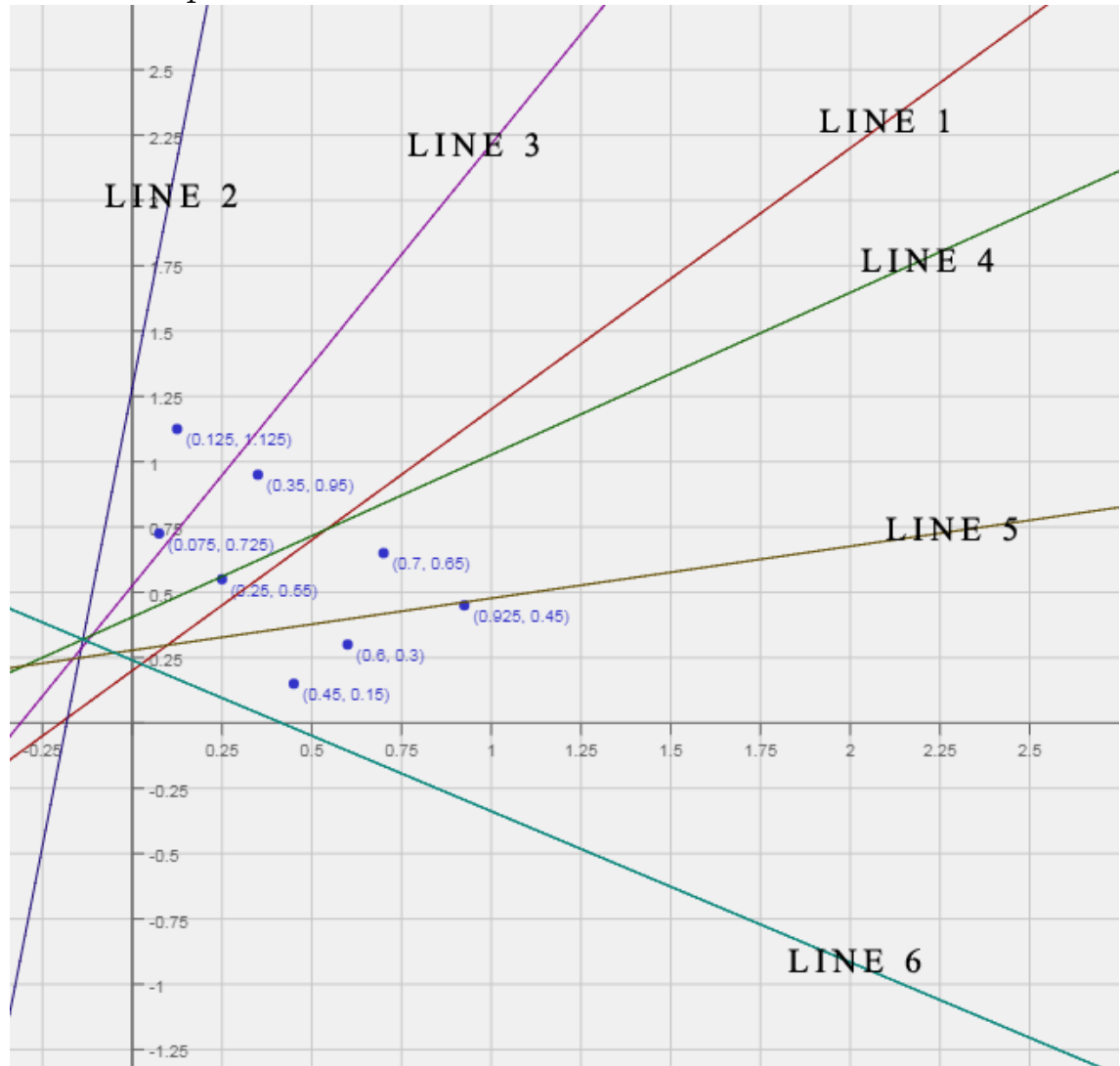
$w1 = 0.307 + 0.75 * (0 \ 1) * 0.6$   
 $= -0.143$   
 $w2 = -0.494 + 0.75 * (0 \ 1) * 0.3$   
 $= -0.719$   
 $f(i1, i2) = 0.2 - 0.143*i1 - 0.719*i2$   
 $i2 = -0.199*i1 + 0.278$   
 $f(0.075, 0.725) < 0$   
 $f(0.125, 1.125) < 0$ , wrong  
 $f(0.25, 0.55) < 0$   
 $f(0.35, 0.95) < 0$ , wrong  
 $f(0.45, 0.15) > 0$ , wrong  
 $f(0.6, 0.3) < 0$   
 $f(0.7, 0.65) < 0$ , wrong  
 $f(0.925, 0.45) < 0$ , wrong  
 5 misclassified points with  $w1 = -0.143$ ,  $w2 = -0.719$

Getting line #6: Updating weights using point (0.45, 0.15)

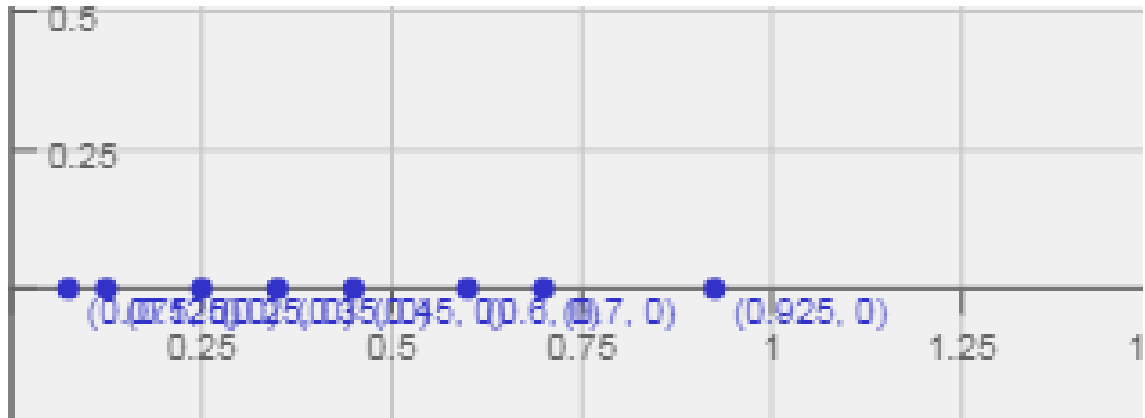
$w1 = -0.143 + 0.75 * (0 \ 1) * 0.45$   
 $= -0.481$   
 $w2 = -0.719 + 0.75 * (0 \ 1) * 0.15$   
 $= -0.832$   
 $f(i1, i2) = 0.2 - 0.481*i1 - 0.832*i2$   
 $i2 = -0.578*i1 + .24$   
 $f(0.075, 0.725) < 0$   
 $f(0.125, 1.125) < 0$ , wrong  
 $f(0.25, 0.55) < 0$   
 $f(0.35, 0.95) < 0$ , wrong  
 $f(0.45, 0.15) < 0$   
 $f(0.6, 0.3) < 0$   
 $f(0.7, 0.65) < 0$ , wrong  
 $f(0.925, 0.45) < 0$ , wrong  
 4 misclassified points



## Final Graph With Lines



Part B) separating line on graph:  $i_2 = 1 - i_1$   
 $f(x) = 0.2 - 0.2 \cdot i_1 - 0.2 \cdot i_2$   
 $w_1 = -0.2, w_2 = -0.2$



Part C) The diagram would be the points all long the x-axis, with a y-value of 0. Everything else is the same.

Best separation would be at  $i1 = 0.65$ , with 2 points to the right, and 6 points to the left. With this separation, only 2 points are misclassified, leaving an error of 25%.

$$F(x) = w_0 * i_0 + w_1 * i_1$$

Assuming  $w_0 = 0.2$ , given  $i_0 = 1$ , and using  $i_1 = 0.65$ :

$$0 = 0.2 + 0.65 * w_1 \quad -0.2/0.65 = w_1 \quad w_1 = -0.307$$

## 5 Question 5

### 5.1 Part A

Negative Points:

(0.05,0.5), (0.12,1.1), (0.4,0.75), (0.54,0.25), (0.71,0.65), (0.79,0.26), (0.93, 0.45)

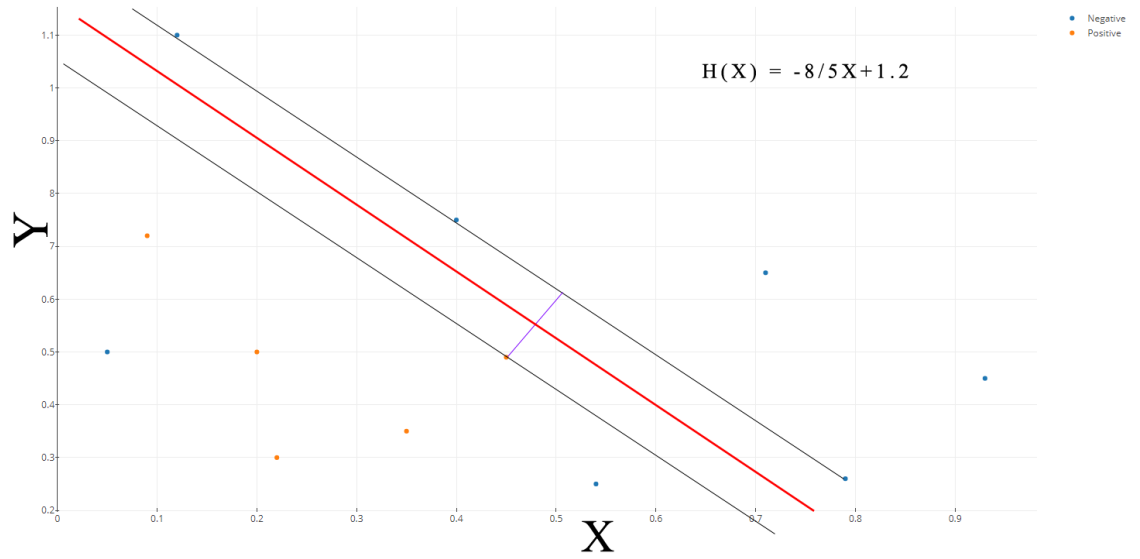
Positive Points:

(0.09,0.72), (0.2,0.5), (0.22,0.3), (0.35,0.35), (0.45,0.49)

After some calculation, we find the hypothesis for our perceptron to be  $h(x) = 8/5x + 1.2$ .

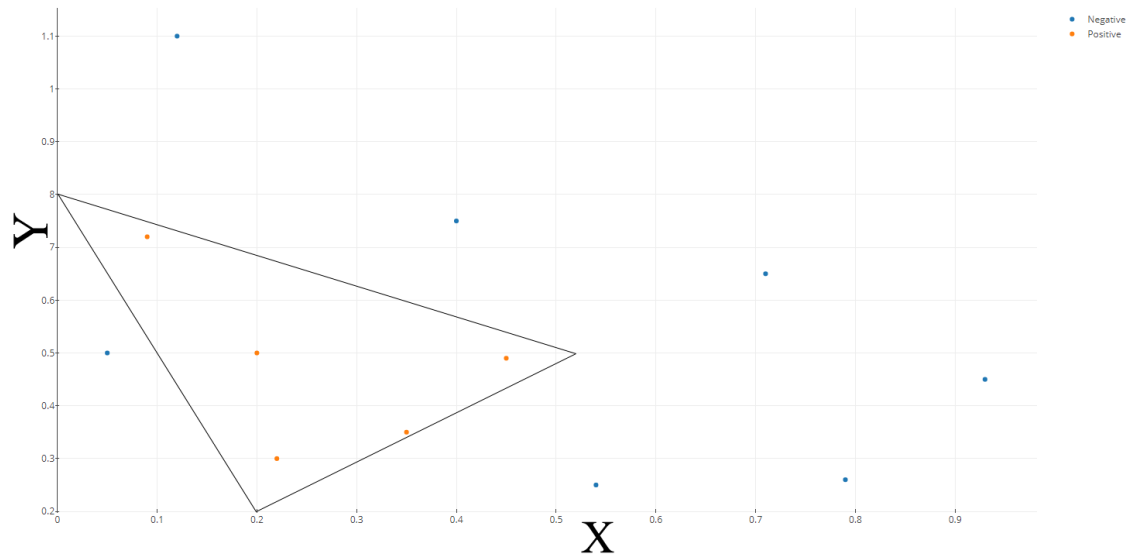
Our calculated minimum error ( Distance to hyperplane was  $13/200$ ) = 0.065

Question 5



## 5.2 Part B

Question 5



$$0 = 0.8 - 4x_1 - x_2 = \text{line 1}$$

$$W_{1,0} = 0.8 \text{ --- } W_{1,1} = -4 \text{ --- } W_{1,2} = -1$$

$$0 = -0.3125 + 1.5625x_1 - x_2 = \text{line 2}$$

$$W_{1,0} = 0.3125 \text{ --- } W_{1,1} = 1.5625 \text{ --- } W_{1,2} = -1$$

$$0 = 0.8 - 0.58x_1 - x_2 = \text{line 3}$$

$$W_{1,0} = 0.8 \text{ --- } W_{1,1} = -0.58 \text{ --- } W_{1,2} = -1$$