Jiashu Liao Ziru Liu
of Rutgers

Assignment 2

CS 440

Intro to Artificial Intelligence

Submitted to:
Pro. Kostas Bekris
Rahul Shome
Chaitanya Mitash

Contents

1 Task 1 **2**

2 Task 2 **5**

3 Task 3 **7**

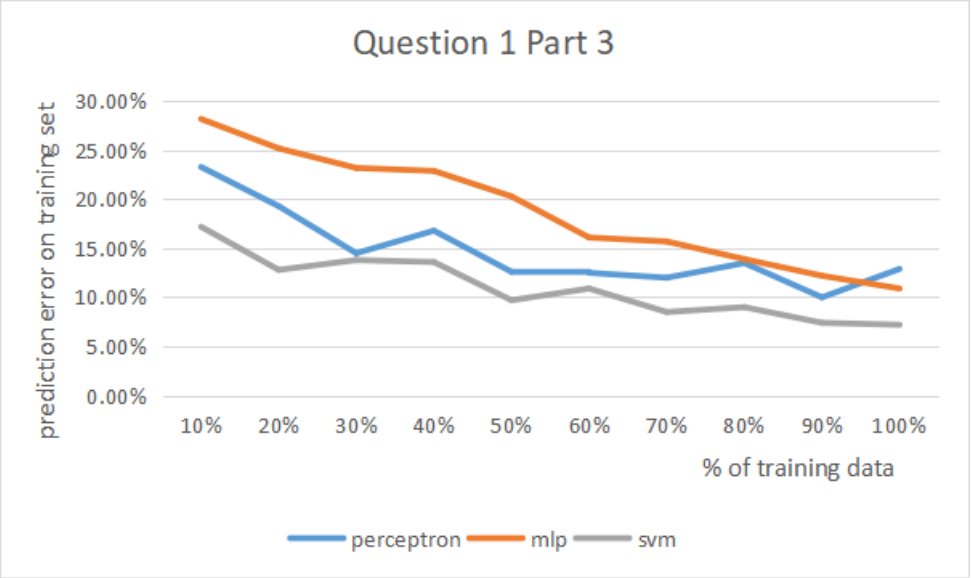
4 Task 4 **9**

5 Task 5 **18**

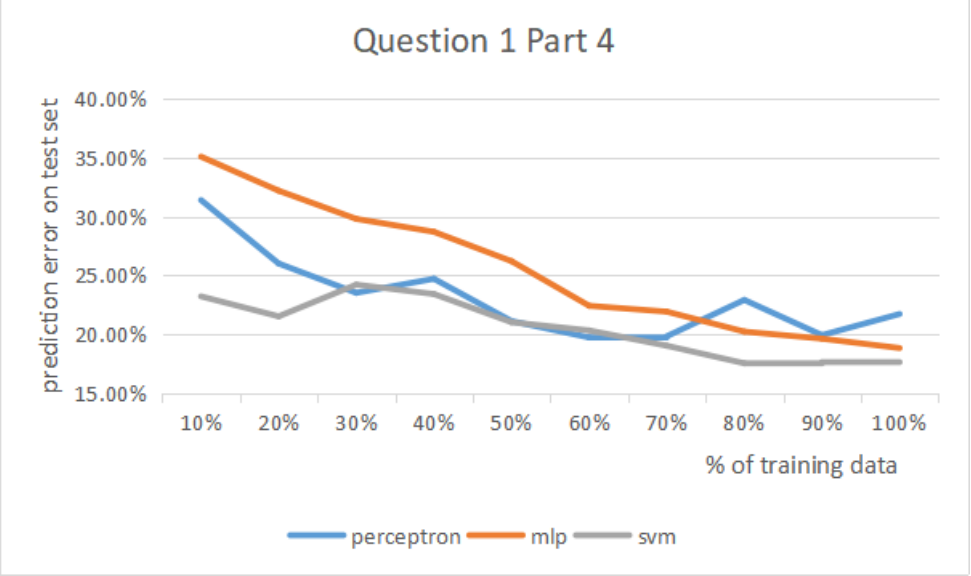
6 Extra **21**

1 Task 1

For question 1 and 2, please see the attachment files.
Question 3:



Question 4:

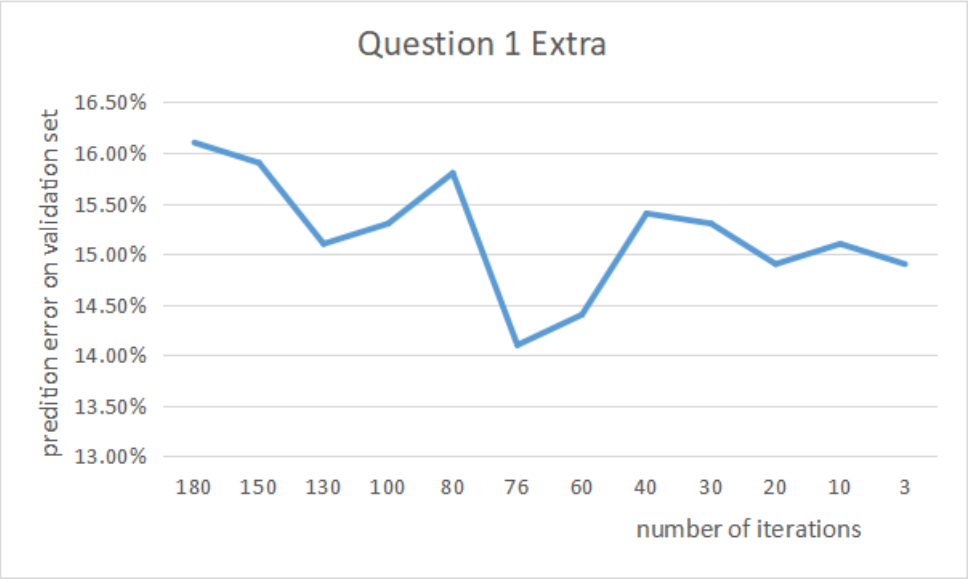


Question 5:
As the size of the training set increases, the three algorithms have different performance behaviors, respectively. From the plots provided above, it is easily to see that MLP is mostly influenced by the size of the training set; it has the highest prediction error for 10% training set, and as the size increases, the prediction error drops dramatically. For the Perceptron, it is much less influenced by the training set size; according to the graph, as the size increases, the prediction errors for part 3 are fluctuating between $20\% - 25\%$, and the change is not constant. For SVM, it has similar behaviors compared to Perceptron, except for that its prediction error constantly drops as the training set size increases.

When comparing the plots of part 3 and part 4, it is easily to find that plot 3 has better performance than plot 4. This is necessary because plot 3 is the result that is generated from applying predictions on the completed training set, whose data are more related to the real training data. However for plot 4, the behavior is a little bit worse because the data from the testing set is less related to the real training data.

If we have more time and more computational resources, we will prefer to develop more hidden layer and more units per layer. The MLP will be able to fit with the data better since it will have better classification ability. Of course, we need to avoid the over-fitting problem at the same time by introducing k-fold cross validation method.

Question 6:
We pick the number of iteration as the hyper-parameters. Since the SVM has more distinguished performance than other algorithm on plot 3 and plot 4, we decided to conduct our experiments on it. After using validation set to tune our hyper-parameter on the SVM, we found out that the SVM will get a "local"(because we cannot ensure it is global minimum) best/minimum performance at iteration = 76. The image below illustrates it.



2 Task 2

Question a
Yes, it is correct.
Question b
First we compute the highest information gain attribute.

$$\begin{aligned} Gain(GPA) &= I(\tfrac{6}{12}, \tfrac{6}{12}) - Remainder(GPA) \\ &= 1 - (\tfrac{3}{12}I(1,0) + \tfrac{5}{12}I(\tfrac{3}{5}, \tfrac{2}{5}) + \tfrac{4}{12}I(0,1)) \\ &= 1 - (0 + 0.4041 + 0) \\ &= 0.596 \end{aligned}$$

$$\begin{aligned} Gain(University) &= I(\tfrac{6}{12}, \tfrac{6}{12}) - Remainder(University) \\ &= 1 - (\tfrac{5}{12}I(\tfrac{3}{5}, \tfrac{2}{5}) + \tfrac{3}{12}I(\tfrac{2}{3}, \tfrac{1}{3}) + \tfrac{4}{12}I(\tfrac{1}{4}, \tfrac{3}{4})) \\ &= 1 - (0.404 + 0.225 + 0.27) \\ &= 0.0207 \end{aligned}$$

$$\begin{aligned} Gain(Published) &= I(\tfrac{6}{12}, \tfrac{6}{12}) - Remainder(Published) \\ &= 1 - (\tfrac{5}{12}I(\tfrac{3}{5}, \tfrac{2}{5}) + \tfrac{7}{12}I(\tfrac{3}{7}, \tfrac{4}{7})) \\ &= 1 - (0.4046 + 0.59) \\ &= 0.0954 \end{aligned}$$

$$\begin{aligned} Gain(Recommendation) &= I(\tfrac{6}{12}, \tfrac{6}{12}) - Remainder(Recommendation) \\ &= 1 - (\tfrac{8}{12}I(\tfrac{5}{8}, \tfrac{3}{8}) + \tfrac{4}{12}I(\tfrac{1}{4}, \tfrac{3}{4})) \\ &= 1 - (0.6363 + 0.27) \\ &= 0.0933 \end{aligned}$$

Second Gain(GPA) > Gain(University) > Gain(Published) > Gain(Recommendation).
We pick GPA as the first attribute in the decision tree.

$$\begin{aligned} Remainder(University) &= \tfrac{5}{12}(\tfrac{2}{5}I(\tfrac{1}{2}, \tfrac{1}{2}) + \tfrac{1}{5}I(1,0) + \tfrac{2}{5}I(\tfrac{1}{2}, \tfrac{1}{2})) \\ &= 0.8 \end{aligned}$$

$$\begin{aligned} Remainder(Published) &= \tfrac{5}{12}(\tfrac{2}{5}I(1,0) + \tfrac{3}{5}I(\tfrac{1}{3}, \tfrac{2}{3})) \\ &= 0.5510 \end{aligned}$$

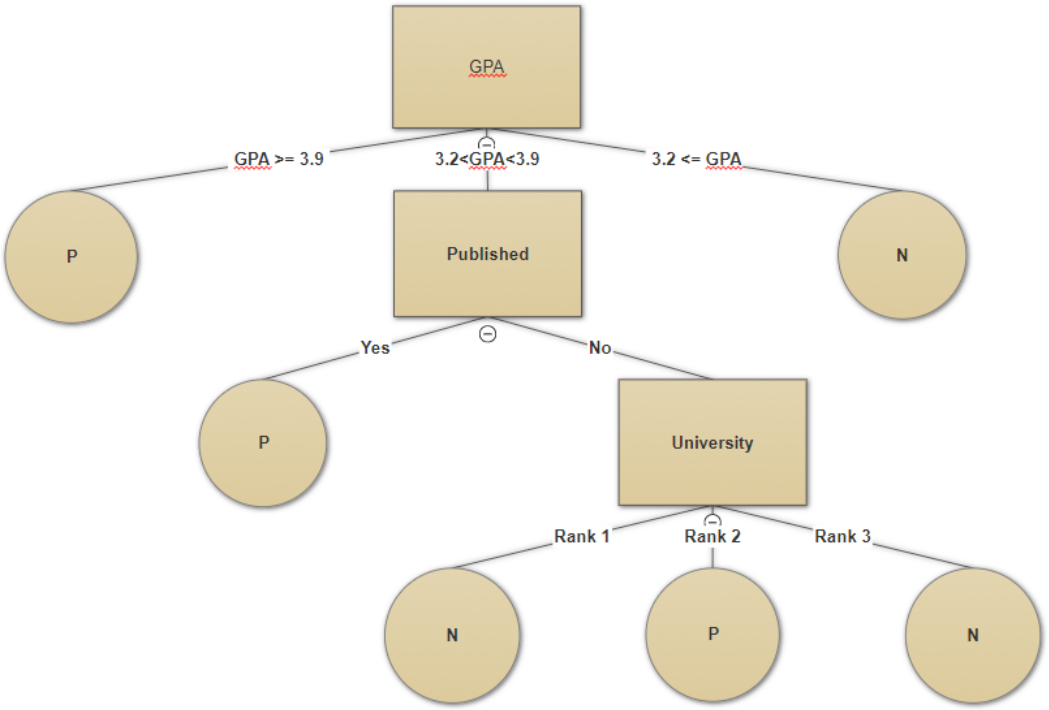
$$\begin{aligned} Remainder(Recommendation) &= (\tfrac{5}{5}I(\tfrac{3}{5}, \tfrac{2}{5})) \\ &= 0.9710 \end{aligned}$$

Remainder(Published) < Remainder(University) < Remainder(Recommendation), we pick Published as the attribute for case 3.2 < GPA ≤3.6
Then for Yes case, only P. For No case:

$$\begin{aligned} Remainder(University) &= (\tfrac{1}{2}I(1,0) + \tfrac{1}{2}I(1,0)) + (\tfrac{1}{2}I(0,1) \\ &= 0 \end{aligned}$$

$$\begin{aligned} Remainder(Recommendation) &= (\tfrac{3}{3}I(\tfrac{1}{3}, \tfrac{2}{3})) \\ &= 0.918 \end{aligned}$$

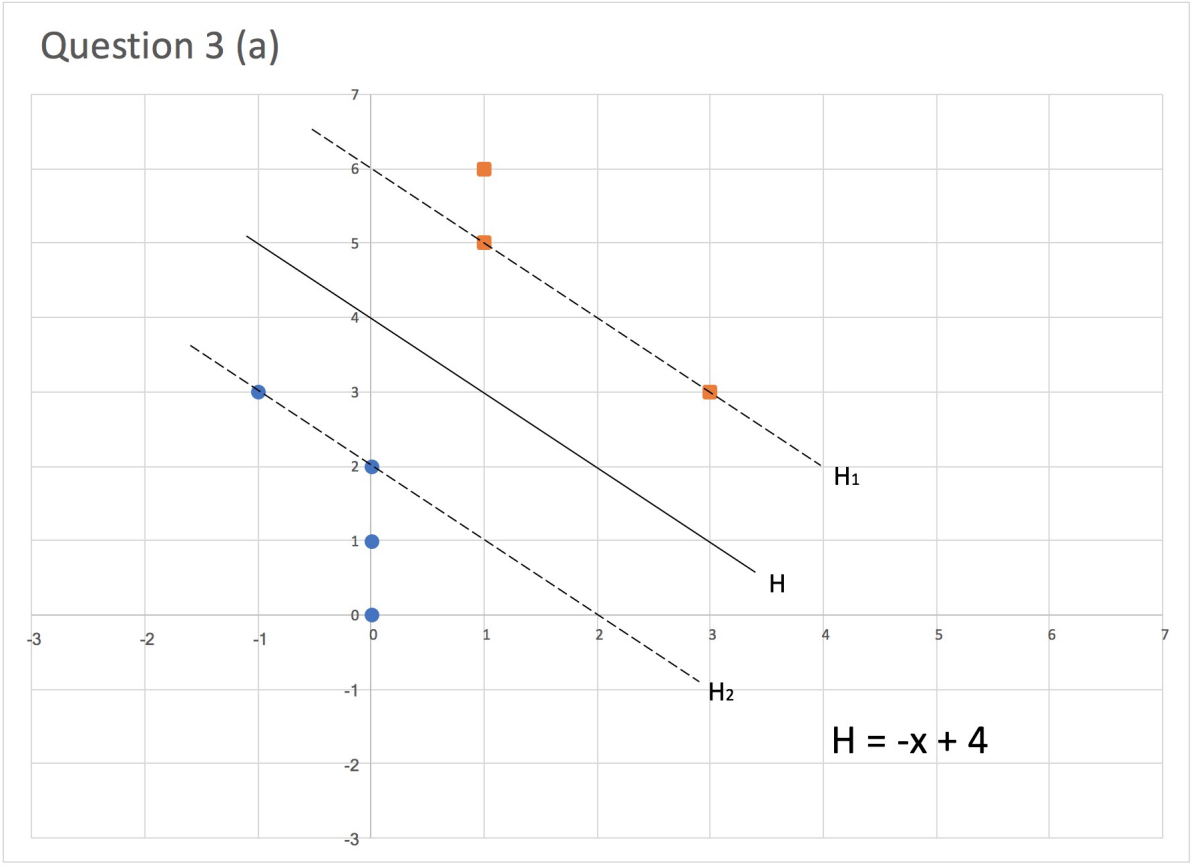
Therefore, under No case, we will pick University as the attribute for case No
Then the final tree will be as below:



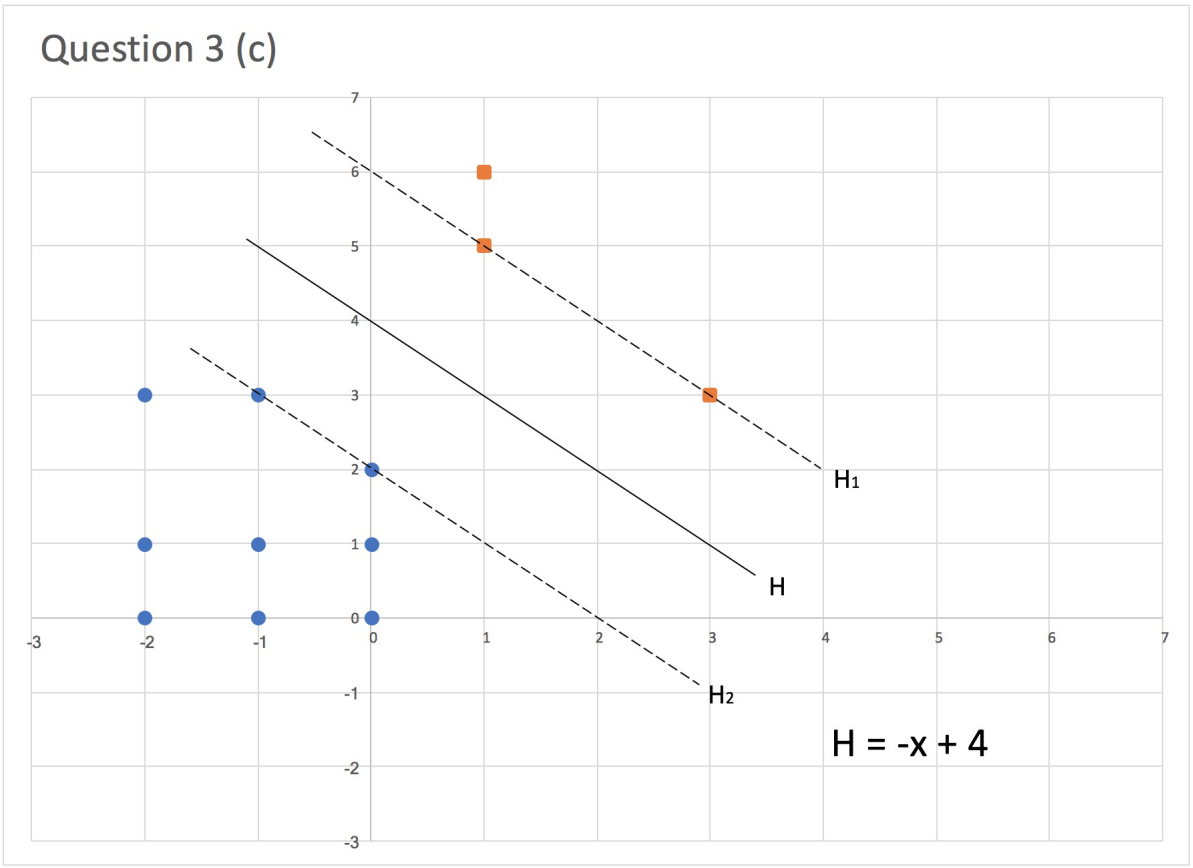
Yes, this is a coincidence because the given examples are very limited, so the tress classify every attribute in the same way.

3 Task 3

Question A:



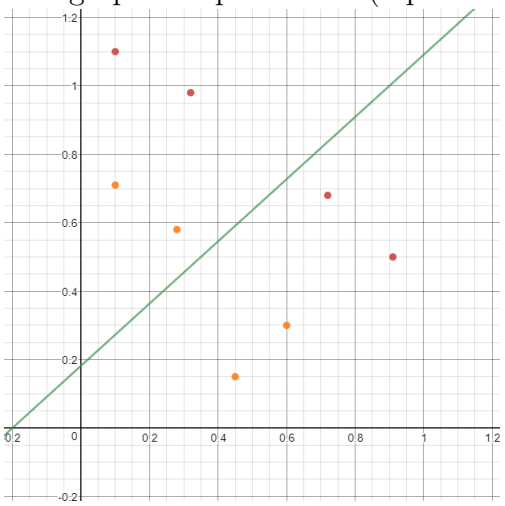
Question B:
$h(x) = w^t x + b$ as the equation of the separating hyper lane. By the graph, the separating lane is on the point (0, 4) and (1,3). Then $3 = w + b$ and $4 = b$. We get $b = 4$ and $w = -1$
Question C:



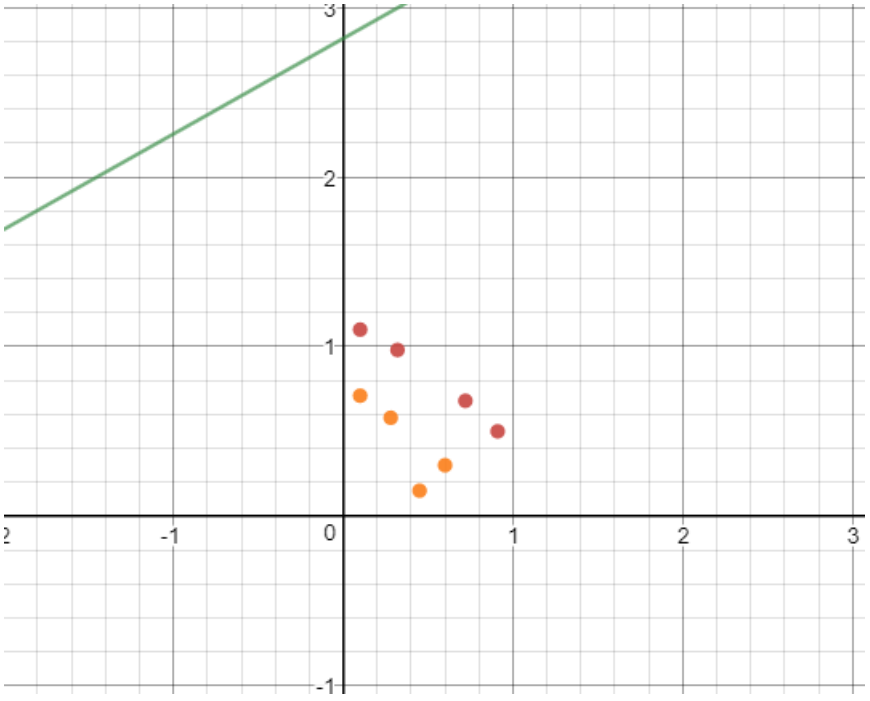
By the graph the separating lane is on the same point as the Question a. Then the w = -1 and b = 4.

4 Task 4

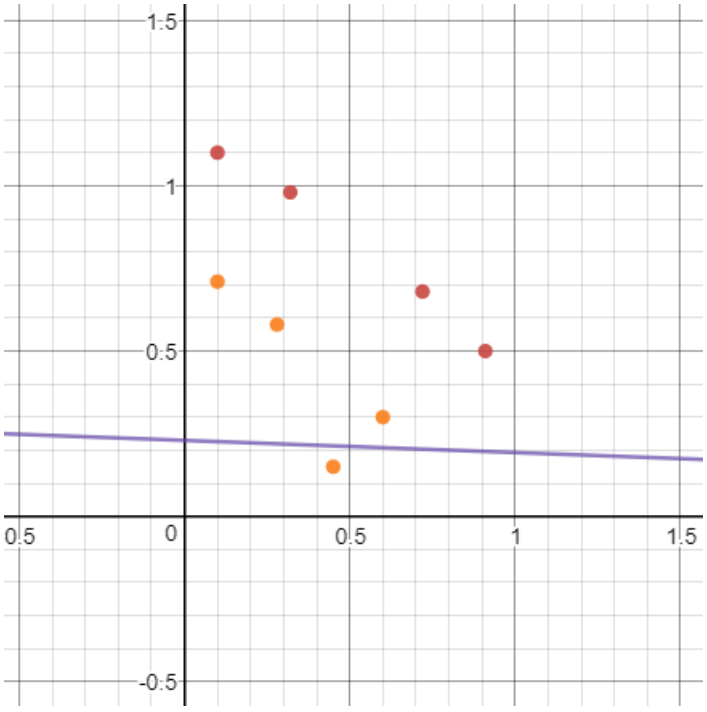
Question a:
The initial $h(x) = w_0 b + w_1 x_1 + w_2 x_2$ for $b = 1, w_0 = 0.2, w_1 = 0.1, w_1 = -1$. Learning rate is 0.3.
We have two classified data:
Positive $(0.1, 0.71)(0.28, 0.58)(0.45, 0.15)(0.6, 0.3)$, while $forg(x) \geq 1, h(x) = 1$, the class is positive. Marked as orange in graphs below.
Negative $(0.1, 1.1)(0.32, 0.98)(0.72, 0.68)(0.91, 0.5)$, while $g(x) < 0, h(x) = -1$, the class is negative. Marked as red in red below.
The graph is separated as (3 points are misclassified):



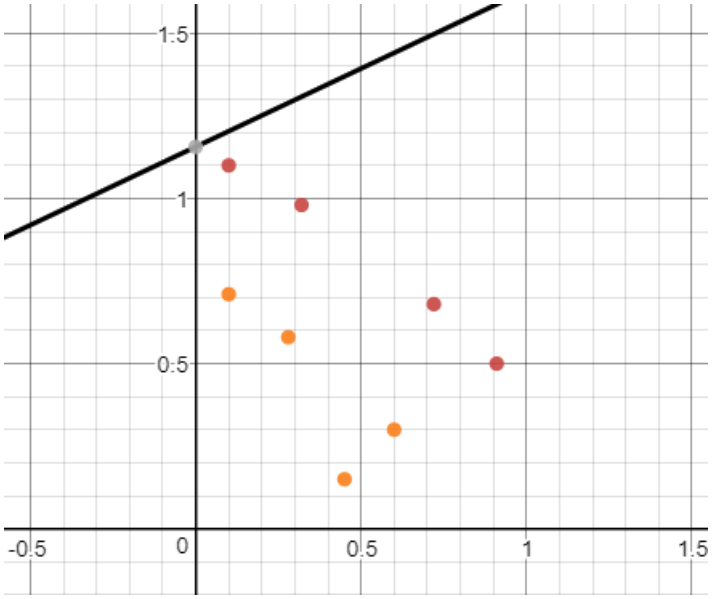
For point (0.1, 0.71), which right class is 1.
$w_0 b = 0.2 * 1 = 0.2$
$w_1 x_1 = 1 * 0.1 = 0.1$
$w_2 x_2 = (-1) * 0.71 = -0.71$
h(x) = -1, which is wrong, then update the weight:
$w_0 = 0.2 + 0.3 * 2 * 1 = 0.8$
$w_1 = 0.1 + 0.3 * 2 * 0.1 = 0.16$
$w_2 = -0.71 + 0.3 * 2 * 0.71 = -0.284$
The graph is separated as (4 points are misclassified):



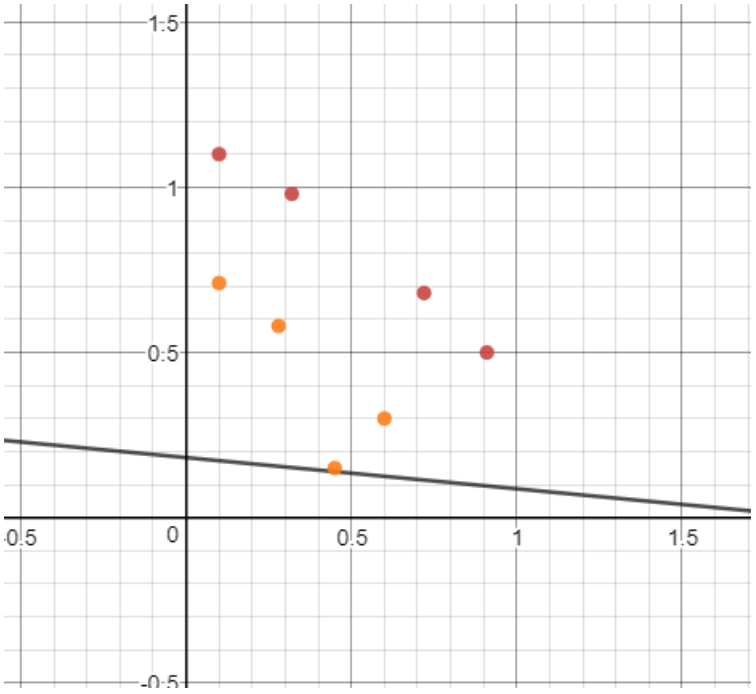
For point (0.32, 0.98), which right class is -1.
$w_0 b = 0.8 * 1 = 0.8$
$w_1 x_1 = 0.16 * 0.32 = 0.0516$
$w_2 x_2 = -0.284 * 0.98 = -0.27832$
h(x) = 1, which is wrong, then update the weight:
$w_0 = 0.8 - 0.3 * 2 * 1 = 0.2$
$w_1 = 0.16 - 0.3 * 2 * 0.32 = -0.032$
$w_2 = (-0.284) - 0.3 * 2 * 0.98 = -0.872$
The graph is separated as (3 points are misclassified):



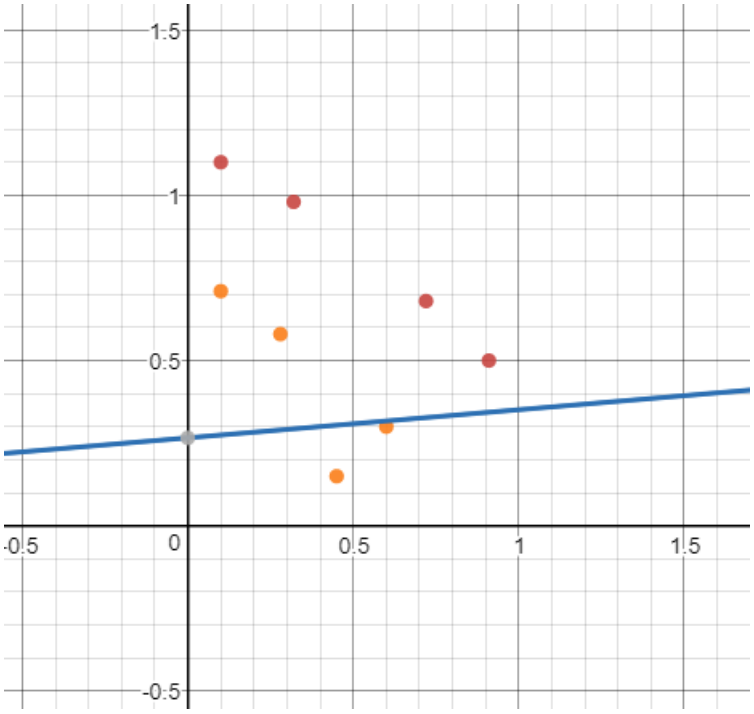
For point (0.6, 0.3), which right class is 1.
$w_0 b = 0.2 * 1 = 0.2$
$w_1 x_1 = -0.032 * 0.6 = -0.0192$
$w_2 x_2 = -0.872 * 0.3 = -0.2616$
h(x) = -1, which is wrong, then update the weight:
$w_0 = 0.2 + 0.3 * 2 * 1 = 0.8$
$w_1 = -0.032 + 0.3 * 2 * 0.6 = 0.328$
$w_2 = -0.872 + 0.3 * 2 * 0.3 = -0.692$
The graph is separated as (4 points are misclassified):



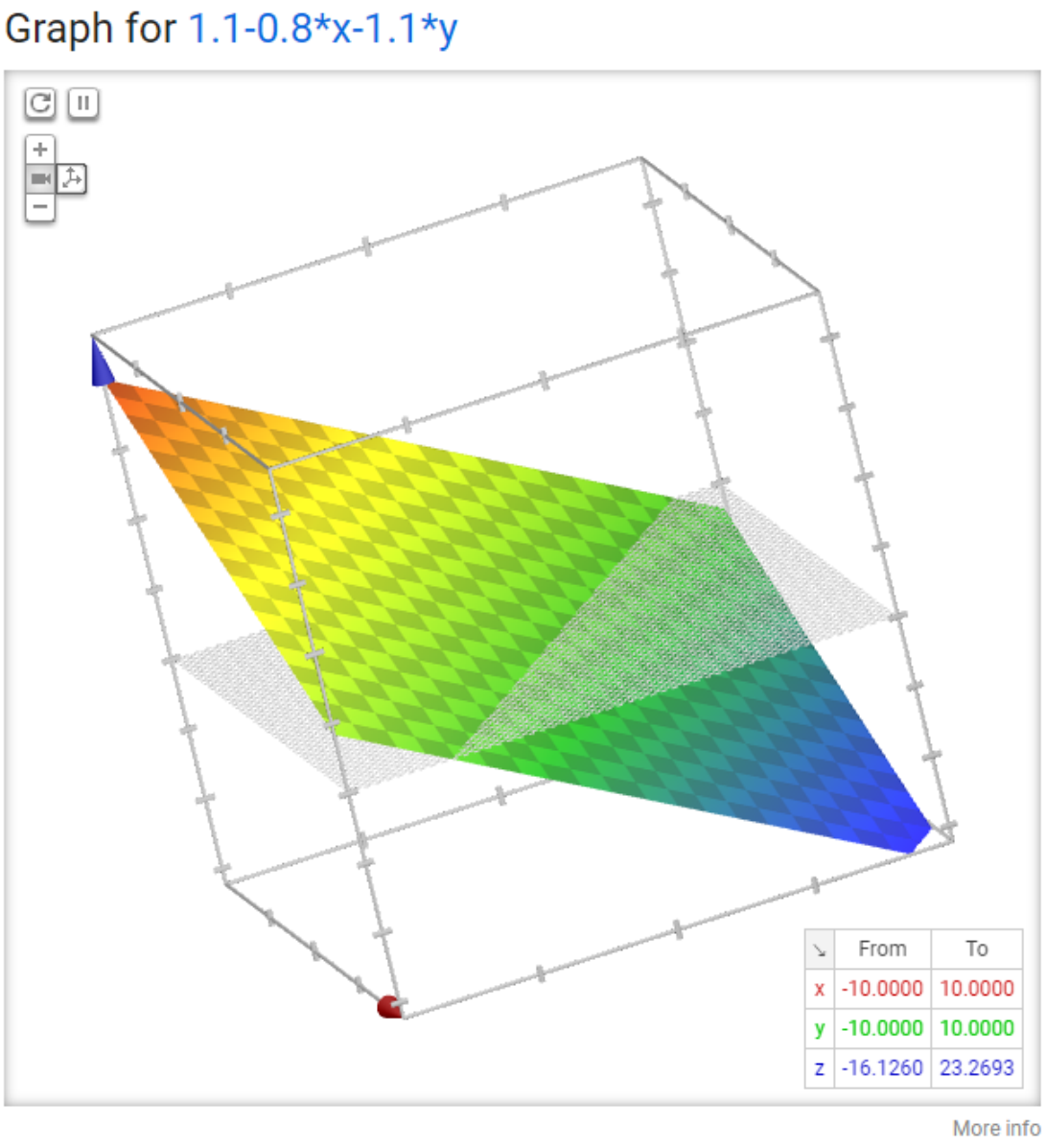
For point (0.72, 0.68), which right class is -1.
$w_0 b = 0.8 * 1 = 0.8$
$w_1 x_1 = 0.328 * 0.28 = 0.23616$
$w_2 x_2 = -0.692 * 0.58 = -0.40136$
h(x) = 1, which is wrong, then update the weight:
$w_0 = 0.8 - 0.3 * 2 * 1 = 0.2$
$w_1 = 0.328 - 0.3 * 2 * 0.72 = -0.104$
$w_2 = -0.692 - 0.3 * 2 * 0.68 = -1.1$
The graph is separated as (3 points are misclassified):



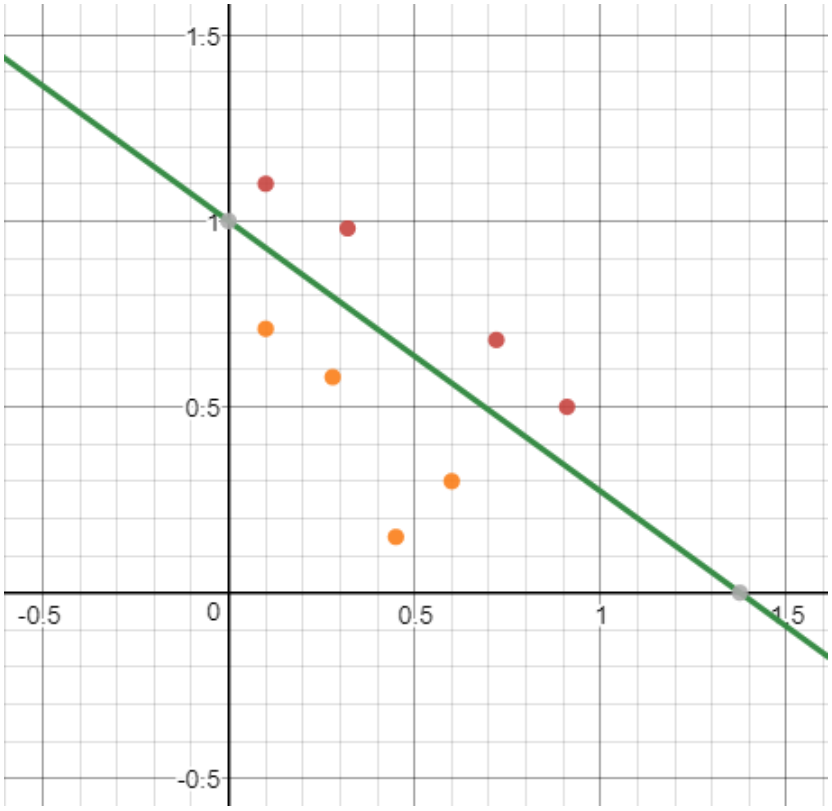
For point (0.28, 0.58), which right class is -1.
$w_0 b = 0.2 * 1 = 0.2$
$w_1 x_1 = -0.104 * 0.28 = -0.02912$
$w_2 x_2 = -1.1 * 0.58 = -0.638$
h(x) = -1, which is wrong, then update the weight:
$w_0 = 0.2 + 0.3 * 2 * 1 = 0.8$
$w_1 = -0.104 + 0.3 * 2 * 0.28 = 0.064$
$w_2 = -1.1 + 0.3 * 2 * 0.58 = -0.752$
The graph is separated as (2 points are misclassified):



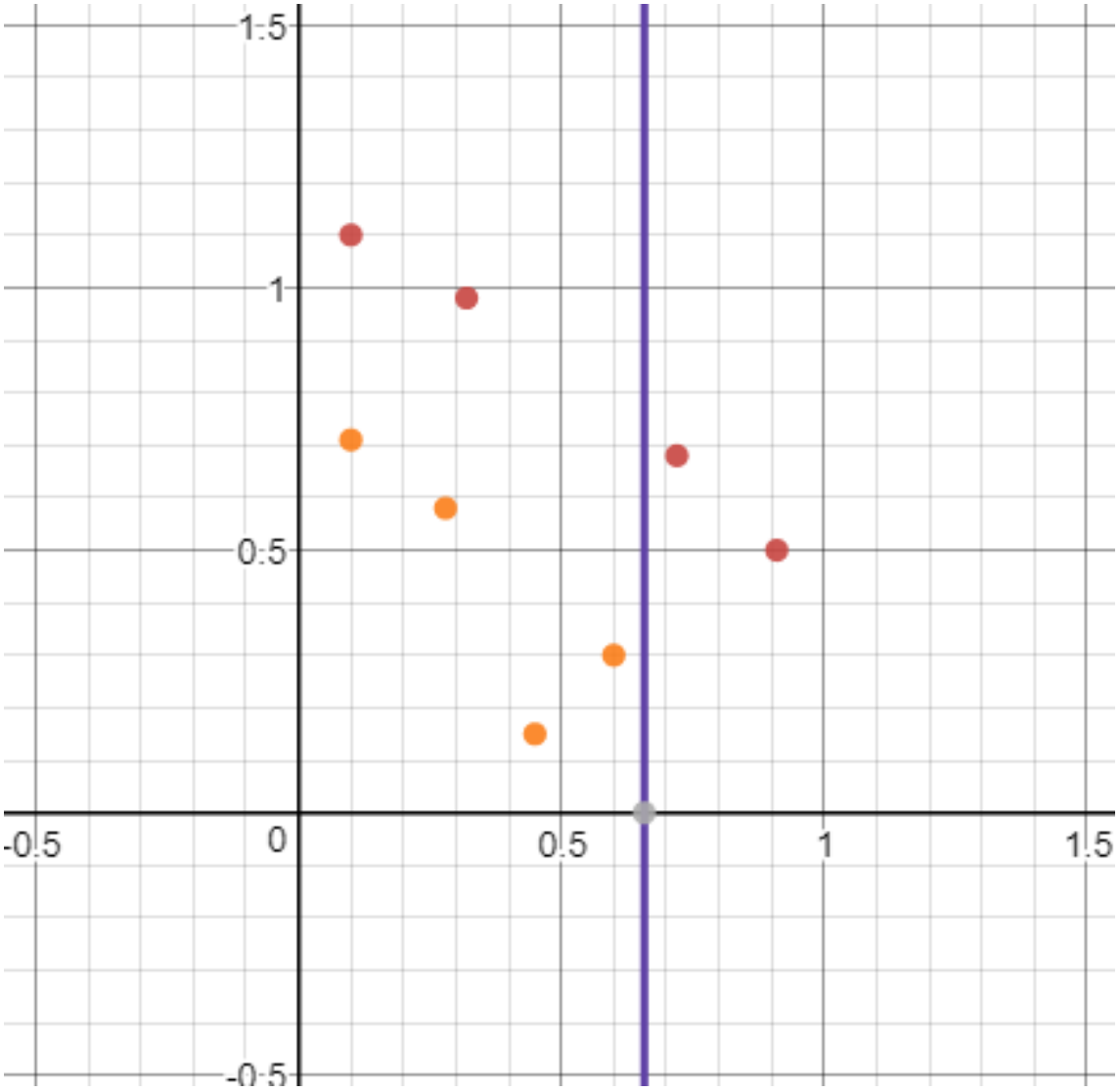
Question b: Since it is not perfect classification line yet for $h(x) = g(0.8 + 0.064 * x_1 - 0.752 * x_2)$. The perfect classification line should be $h(x) = g(1.1 - 0.8 * x_1 - 1.1 * x_2)$. In 3 dimension space as:



The picture as below showed the perfect classification line in 2 dimension space:



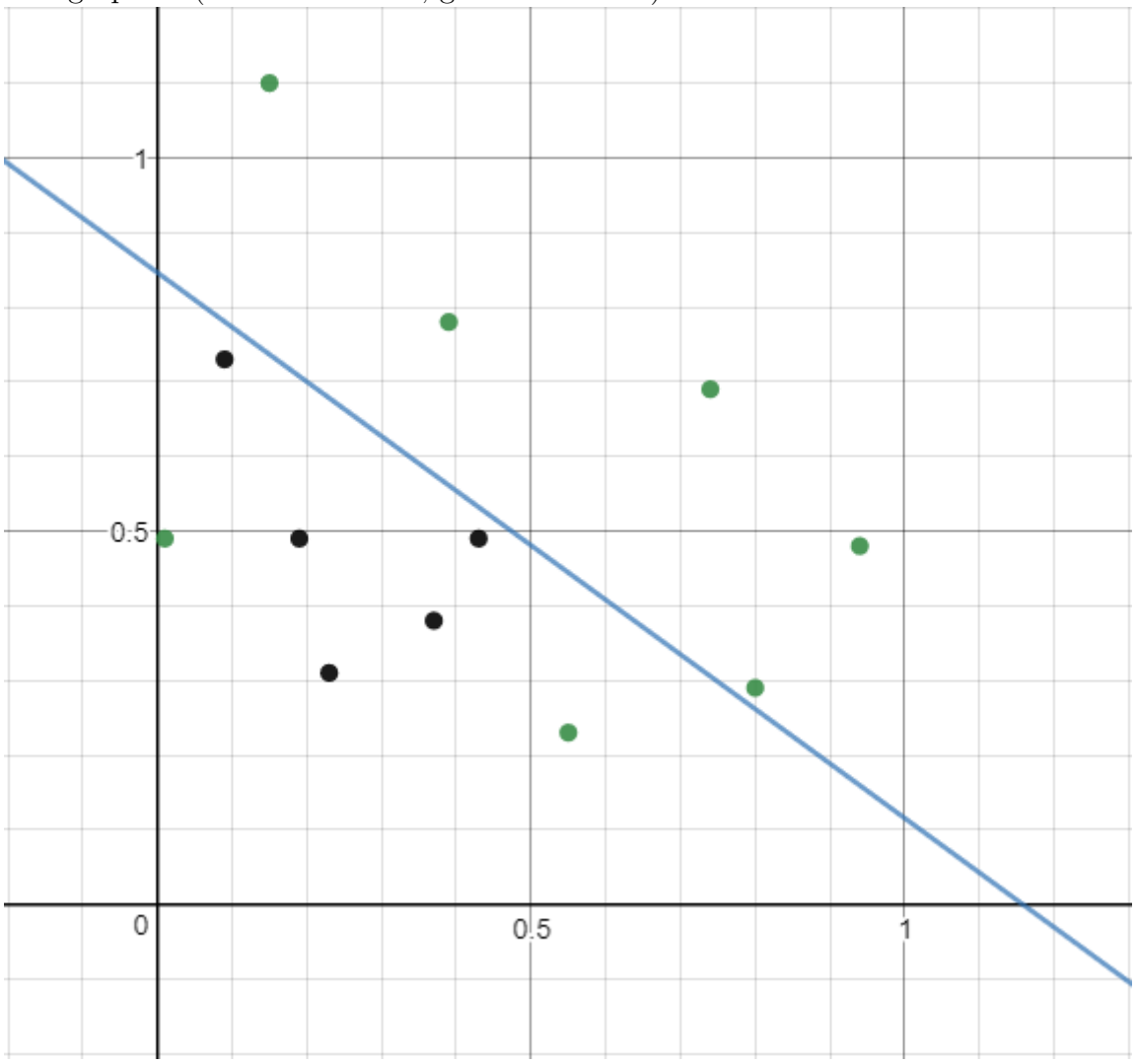
Question c:
Because we only have w_0 and w_1, so the line separates input space must be perpendicular to x-axis. $w_1 = -0.75$ and $w_0 = 0.5$
The graph is separated as (2 points are misclassified):



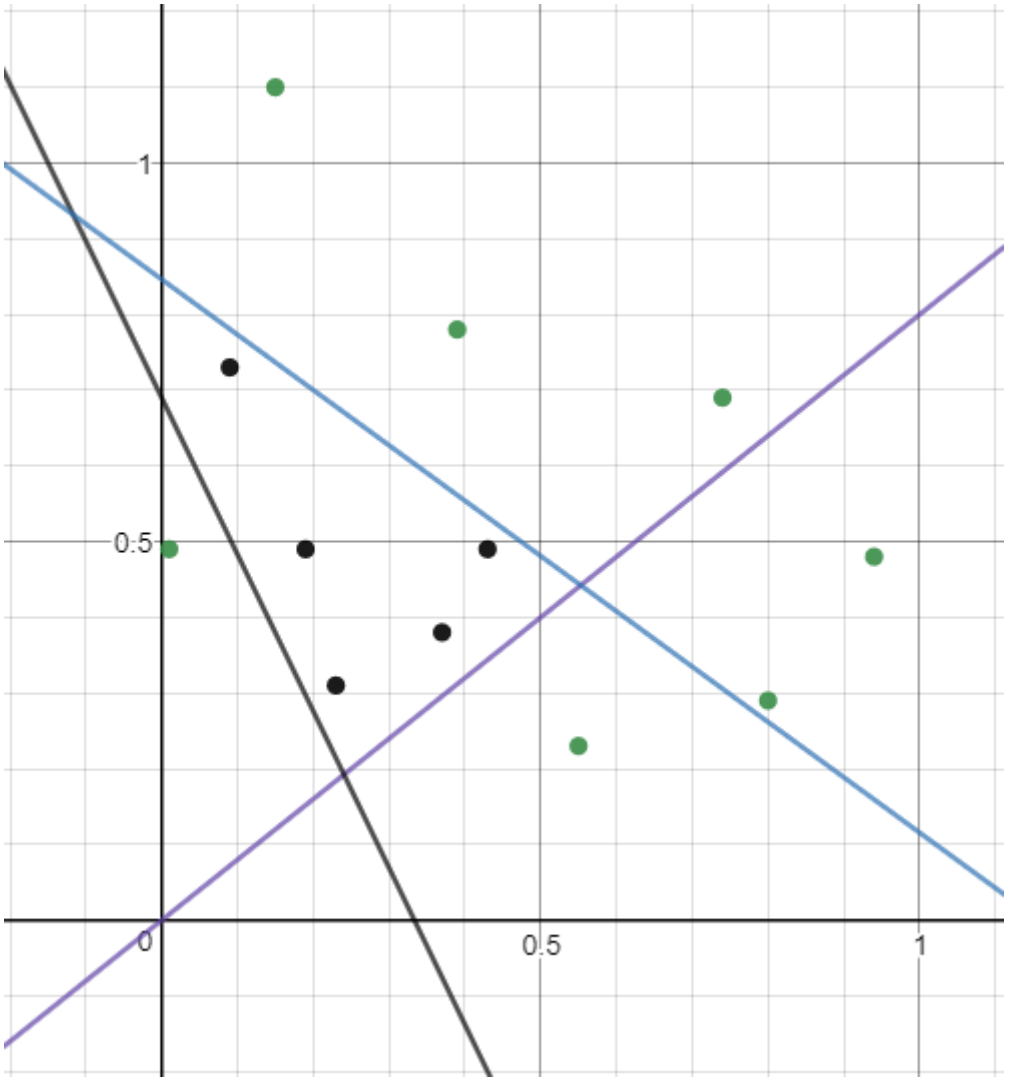
$h(w) = -0.76x_1 + 0.5$
Error $= \frac{1}{2}Err^2$
Error $= 2 * \frac{1}{2}(-1-1)^2 = 4$

5 Task 5

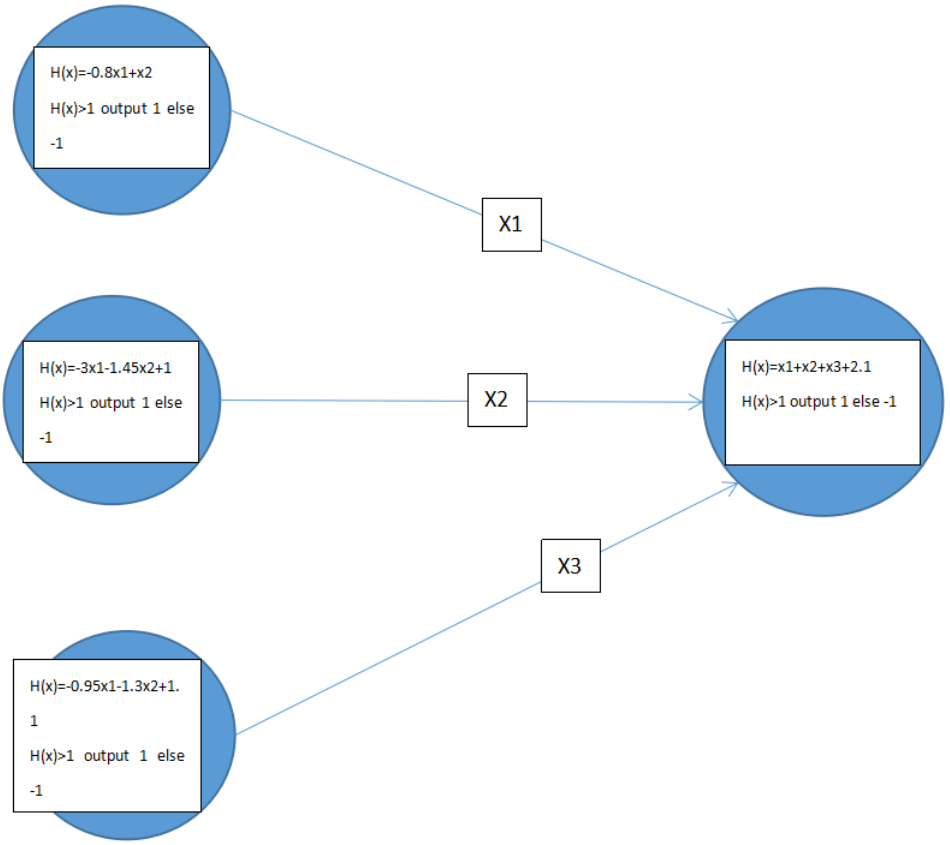
Question a: For $w_0 = 1.1$, $w_1 = -0.95$, $w_2 = -1.3$
Error $= \frac{1}{2}Err^2$
Error $= 2 * \frac{1}{2}(-1-1)^2 = 4$, because two point are misclassified.
The graph as (Black is classs 1, green is class -1) :



Question b:
The graph as (Black is classs 1, green is class -1) :



These points will be classified correctly by the three lines, the perceptrons will be as below:


H(x)=-0.8x1+x2
H(x)>1 output 1 else
-1
X1
H(x)=-3x1-1.45x2+1
H(x)>1 output 1 else
-1
X2
H(x)=x1+x2+x3+2.1
H(x)>1 output 1 else -1
X3
H(x)=-0.95x1-1.3x2+1.
1
H(x)>1 output 1 else
-1

6 Extra