Question 1

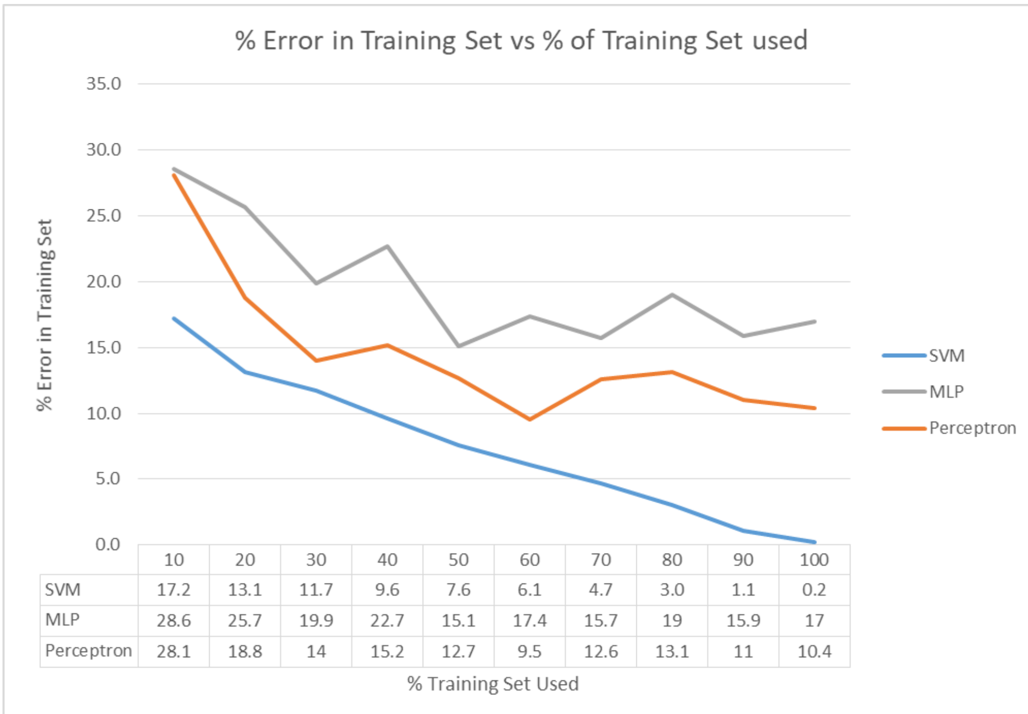
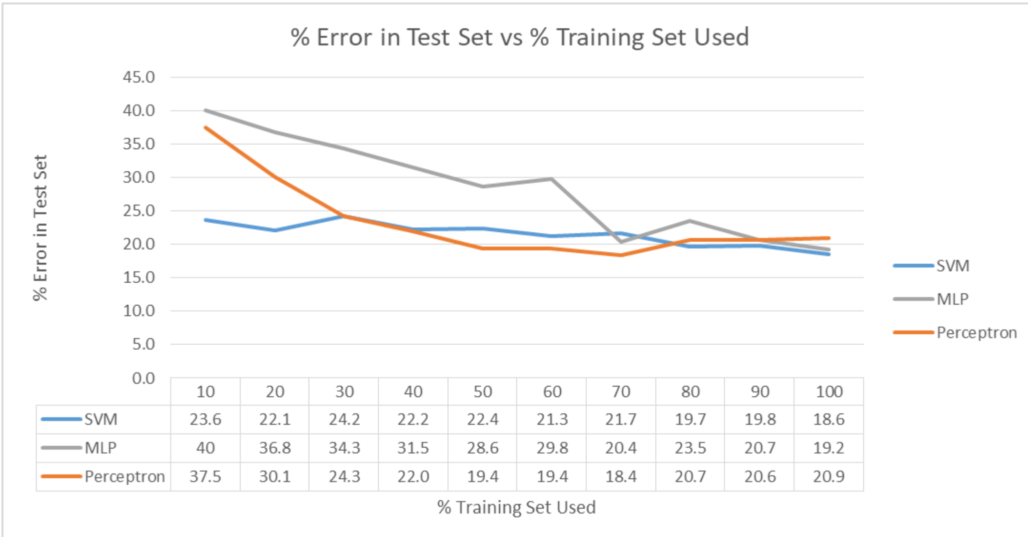


Figure 1: A graph showing percent correct of the training set as a function of the percentage of the training set used.



	10	20	30	40	50	60	70	80	90	100
SVM	23.6	22.1	24.2	22.2	22.4	21.3	21.7	19.7	19.8	18.6
MLP	40	36.8	34.3	31.5	28.6	29.8	20.4	23.5	20.7	19.2
Perceptron	37.5	30.1	24.3	22.0	19.4	19.4	18.4	20.7	20.6	20.9

Figure 2: A graph showing percent correct of the test set as a function of the percentage of the training set used.

Of the three algorithms tested, SVM appears to perform the best on both the training and test set, while MLP performs the worst. For MLP, we used a softmax algorithm which takes the dot product of the weights times the training data and then does this:

$$softmax(label = x) = \frac{e^{w_x t}}{\sum_{n=0}^{9} e^{w_n t}} \tag{1}$$

Where t is the training data and there are 10 labels, 0-9.

Question 2

Part a

1) 4.0 GPA - P, the data match the decision tree
2) 3.9 GPA - P, the data match the decision tree
3) 3.9 GPA - P, the data match the decision tree
4) 3.8 GPA - yes publications - P, the data match the decision tree
5) 3.6 GPA - no publications - rank 2 university - P, the data match the decision tree

6) 3.6 GPA - yes publications - P, the data match the decision tree
7) 3.4 GPA - no publications - rank 3 university - N, the data match the decision tree
8) GPA 3.4 - No publication - Rank 1 University - N, data match the tree
9) GPA 3.2 - N, data match the tree
10) GPA 3.1 - N, data match the tree
11) GPA 3.1 - N, data match the tree
12) GPA 3.0 - N, data match the tree

Part b

For the entire set, there are 6 positives and 6 negatives. Hence, the entropy of the set is:

$$E(S) = -\frac{1}{2}\log_2\frac{1}{2} - \frac{1}{2}\log_2\frac{1}{2} = 1 \tag{2}$$

For GPA, the information gained is:
GPA ¿= 3.9: 3 P, 0 N

$$E(GPA >= 3.9) = -1\log_2 1 - 0\log_2 0 = 0 \tag{3}$$

3.2 ¡ GPA ¡ 3.9: 3 P, 2 N

$$E(3.2 < GPA < 3.9) = -\frac{3}{5}\log_2\frac{3}{5} - \frac{2}{5}\log_2\frac{2}{5} = 0.9710 \tag{4}$$

GPA ¡= 3.2: 0 P, 4 N

$$E(GPA <= 3.2) = -0\log_2 0 - 1\log_2 1 = 0 \tag{5}$$

$$I(GPA) = \frac{1}{4} * 0 + \frac{5}{12} * 0.9710 + \frac{1}{3} * 0 = 0.4046 \tag{6}$$

$$Gain(GPA) = E(S) - I(GPA) = 1 - 0.4046 = 0.5954 \tag{7}$$

For university rank, the information gained is:
University rank = 1: 3 P, 2 N

$$E(rank = 1) = -\frac{3}{5}\log_2\frac{3}{5} - \frac{2}{5}\log_2\frac{2}{5} = 0.9710 \tag{8}$$

University rank = 2: 2 P, 1 N

$$E(rank = 2) = -\frac{2}{3}\log_2\frac{2}{3} - \frac{1}{3}\log_2\frac{1}{3} = 0.9183 \tag{9}$$

University rank = 3: 1 P, 3 N

$$E(rank = 3) = -\frac{1}{4}\log_2\frac{1}{4} - \frac{3}{4}\log_2\frac{3}{4} = 0.8113 \tag{10}$$

$$I(rank) = \frac{5}{12} * 0.9710 + \frac{1}{4} * 0.9183 + \frac{1}{3} * 0.8113 = 0.9046 \tag{11}$$

$$Gain(rank) = E(S) - I(rank) = 1 - 0.9046 = 0.0954 \tag{12}$$

For whether the student has publications, the information gained is:
Published = Yes : 3 P, 2 N

$$E(Published = Yes) = -\frac{3}{5}\log_2\frac{3}{5}) - \frac{2}{5}\log_2\frac{2}{5} = .9710 \tag{13}$$

Published = No : 3 P, 4 N

$$E(Published = No) = -\frac{3}{7}\log_2\frac{3}{7} - \frac{4}{7}\log_2\frac{4}{7} = .9852 \tag{14}$$

$$I(Published) = \frac{5}{12}(.9710) + \frac{7}{12}(.9852) = 0.9792 \tag{15}$$

$$Gain(Published) = E(S) - I(Published) = 1 - 0.9792 = 0.0208 \tag{16}$$

For the quality of the student's recommendations, the information gained is:
Recommendations = Good : 5 P, 3 N

$$E(Recommendations = Good) = -\frac{5}{8}\log_2\frac{5}{8} - \frac{3}{8}\log_2\frac{3}{8} = .9544 \tag{17}$$

Recommendations = Normal : 1 P, 3 N

$$E(Recommendations = Normal) = -\frac{1}{4}\log_2\frac{1}{4} - \frac{3}{4}\log_2\frac{3}{4} = .8113 \tag{18}$$

$$I(Recommendations) = \frac{8}{12}(.9544) + \frac{4}{12}(.8113) = 0.9067 \tag{19}$$

$$Gain(Recommendations) = E(S) - I(Recommendations) = 1 - 0.9067 = 0.0933 \tag{20}$$

For the root of the decision tree, the best attribute to use is GPA, since it has the highest information gain of all the attributes. Because a GPA of 4.0 always leads to P and a GPA of 3.3 always leads to N, we only need to investigate the next node for when GPA = 3.6.
For the subset, there are 3 positives and 2 negatives. Hence, the entropy of the subset is:

$$E(S) = -\frac{3}{5}\log_2\frac{3}{5} - \frac{2}{5}\log_2\frac{2}{5} = 0.9710 \tag{21}$$

For university rank, the information gained is:
University rank = 1 : 1 P, 1 N

$$E(rank = 1) = -\frac{1}{2}\log_2\frac{1}{2} - \frac{1}{2}\log_2\frac{1}{2} = 1 \tag{22}$$

University rank = 2 : 1 P, 0 N

$$E(rank = 2) = -1\log_2 1 - 0\log_2 0 = 0 \tag{23}$$

University rank = 3 : 1 P, 1 N

$$E(rank = 3) = -\frac{1}{2}\log_2\frac{1}{2} - \frac{1}{2}\log_2\frac{1}{2} = 1 \tag{24}$$

$$I(rank) = \frac{2}{5} * 1 + \frac{1}{5} * 0 + \frac{2}{5} * 1 = 0.8 \tag{25}$$

$$Gain(rank) = E(S) - I(rank) = 0.9710 - 0.8 = 0.1710 \tag{26}$$

For whether the student has publications, the information gained is:
Published = Yes : 2 P, 0 N

$$E(Published = Yes) = -1\log_2 1 - 0\log_2 0 = 0 \tag{27}$$

Published = No : 2 P, 1 N

$$E(Published = No) = -\frac{2}{3}\log_2\frac{2}{3} - \frac{1}{3}\log_2\frac{1}{3} = 0.9183 \tag{28}$$

$$I(Published) = \frac{2}{5}(0) + \frac{3}{5}(.9183) = 0.5510 \tag{29}$$

$$Gain(Published) = E(S) - I(Published) = 0.9710 - 0.5510 = 0.42 \tag{30}$$

For the quality of the student's recommendations, the information gained is:
Recommendations = Good : 3 P, 2 N

$$E(Recommendations = Good) = -\frac{3}{5}\log_2\frac{3}{5}) - \frac{2}{5}\log_2\frac{2}{5} = .9710 \tag{31}$$

Recommendations = Normal : 0 P, 0 N

$$E(Recommendations = Normal) = -0\log_2 0 - 0\log_2 0 = 0 \tag{32}$$

$$I(Recommendations) = 1(.9710) + 0(0) = 0.9710 \tag{33}$$

$$Gain(Recommendations) = E(S) - I(Recommendations) = 0.9710 - 0.9710 = 0 \tag{34}$$

The best attribute to use next would be whether the student has publications. Because if the student has publications the answer is always positive, then we only need to examine when the student doesn't have publications to create the next child node in the decision tree.
For the subset, there is 1 positive and 2 negatives. Hence, the entropy of the subset is:

$$E(S) = -\frac{1}{3}\log_2\frac{1}{3} - \frac{2}{3}\log_2\frac{2}{3} = 0.9183 \tag{35}$$

For university rank, the information gained is:
University rank = 1 : 0 P, 1 N

$$E(rank = 1) = -0\log_2 0 - 1\log_2 1 = 0 \tag{36}$$

University rank = 2 : 1 P, 0 N

$$E(rank = 2) = -1 \log_2 1 - 0 \log_2 0 = 0 \tag{37}$$

University rank = 3 : 0 P, 1 N

$$E(rank = 3) = -0 \log_2 0 - 1 \log_2 1 = 0 \tag{38}$$

$$I(rank) = \frac{1}{3} * 0 + \frac{1}{3} * 0 + \frac{1}{3} * 0 = 0 \tag{39}$$

$$Gain(rank) = E(S) - I(rank) = 0.9710 - 0 = 0.9710 \tag{40}$$

For the quality of the student's recommendations, the information gained is:
Recommendations = Good : 1 P, 2 N

$$E(Recommendations = Good) = -\frac{1}{3} \log_2 \frac{1}{3} - \frac{2}{3} \log_2 \frac{2}{3} = 0.9183 \tag{41}$$

Recommendations = Normal : 0 P, 0 N

$$E(Recommendations = Normal) = -0 \log_2 0 - 0 \log_2 0 = 0 \tag{42}$$

$$I(Recommendations) = 1(.9183) + 0(0) = 0.9183 \tag{43}$$

$$Gain(Recommendations) = E(S) - I(Recommendations) = 0.9183 - 0.9183 = 0 \tag{44}$$

The best attribute to use next would be the university rank. Because there is no further information necessary to categorize positives and negatives, recommendation quality is not needed to determine the final outcome. Thus, the decision tree is complete.

Part 3

The decision tree we obtained in part b is identical to the decision tree provided in part a. Because all of the data match the decision tree, this is not surprising.

Question 3

Part a

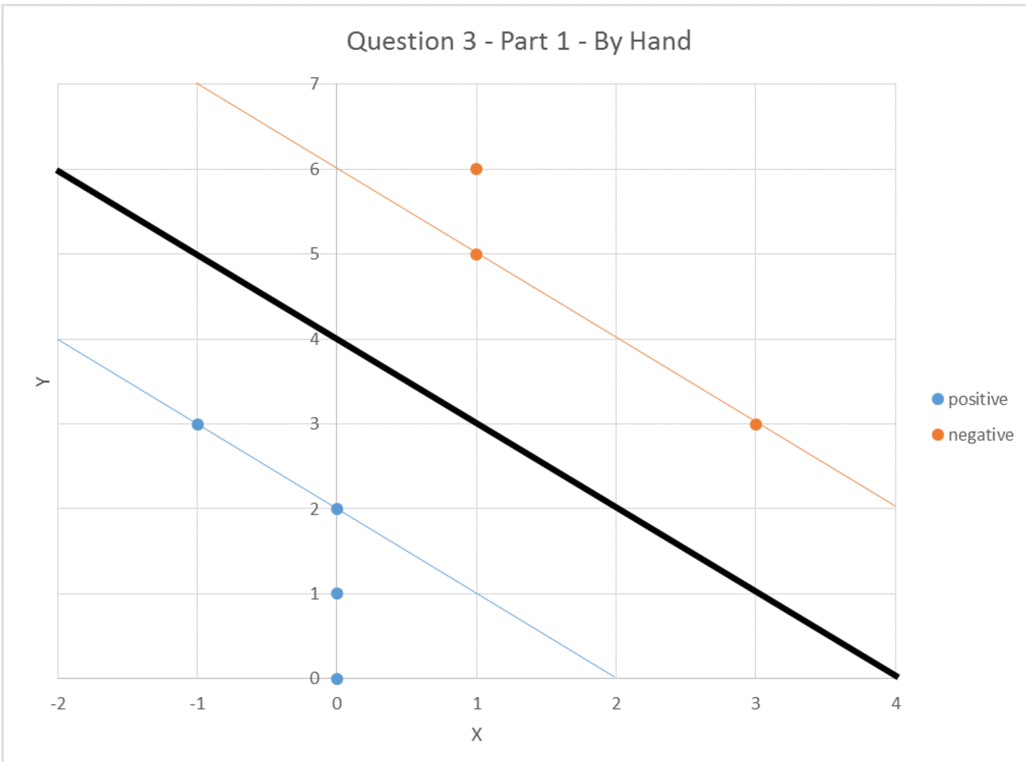


Figure 3: Positive data set is blue. Negative data set is orange. Hand drawn linear classifier in black. Support vectors for corresponding data sets also included

Part b

The parameters w and b for this line are -1 and 4, respectively. The equation below represents the drawn classifer.

$$h(x) = -x + 4 \tag{45}$$

Part c

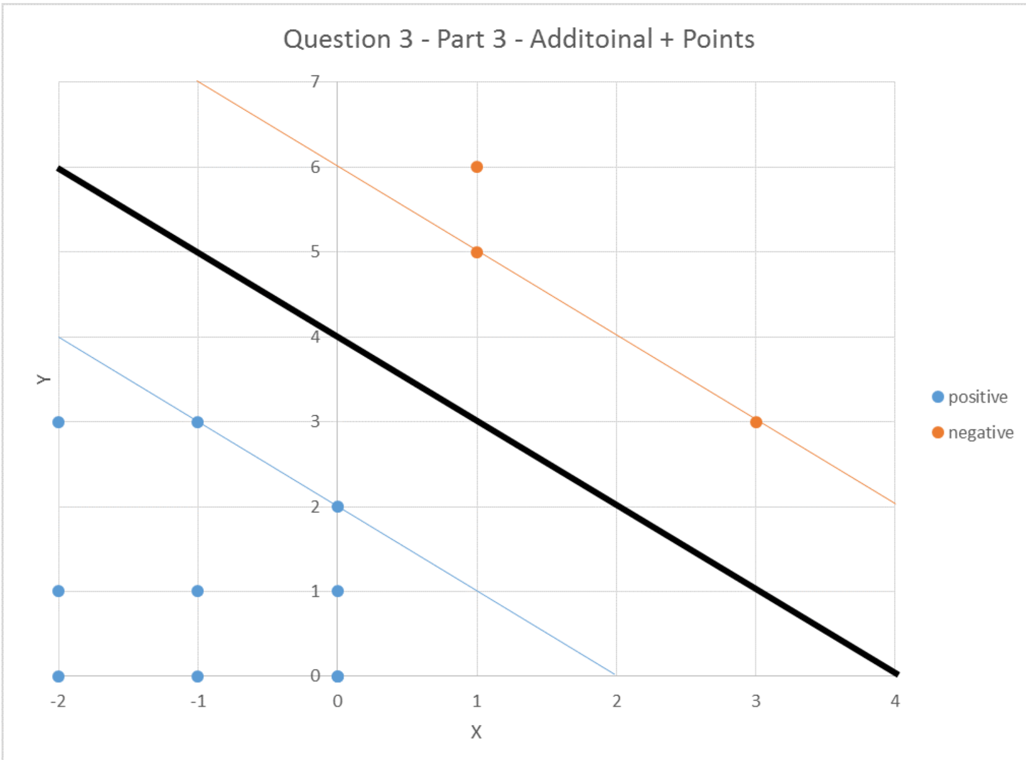


Figure 4: Positive data set is blue. Negative data set is orange. Hand drawn linear classifier in black. Support vectors for corresponding data sets also included

The addition of more data to the positive data set did not change the linear SVM because the values did not change the position of the corresponding support vectors. As a result, the values of w and b are unchanged and the equation below still represents the classifer in the graph.

$$h(x) = -x + 4 \tag{46}$$

Question 4

Part a

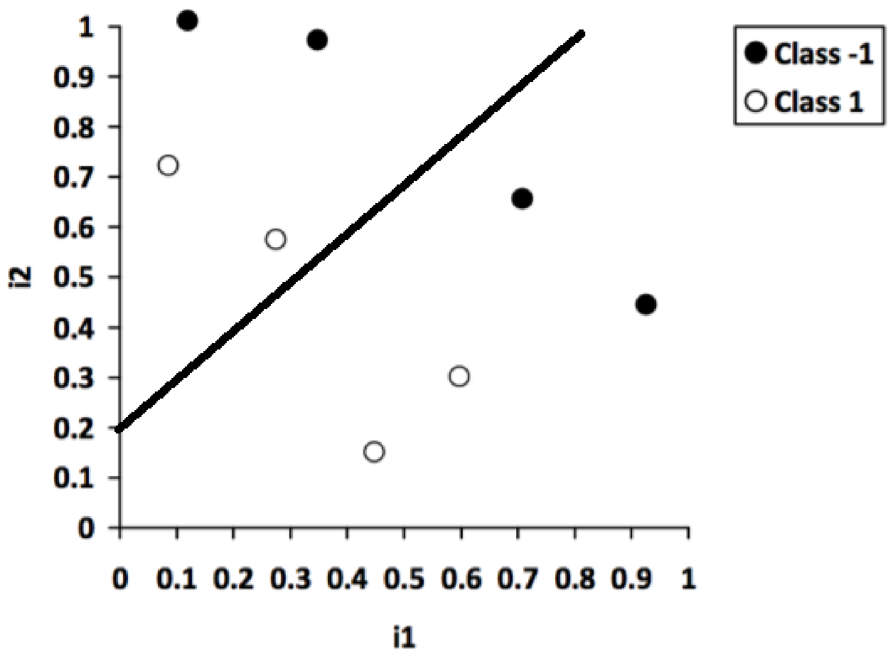


Figure 5: The plot with

Question 5

Part a

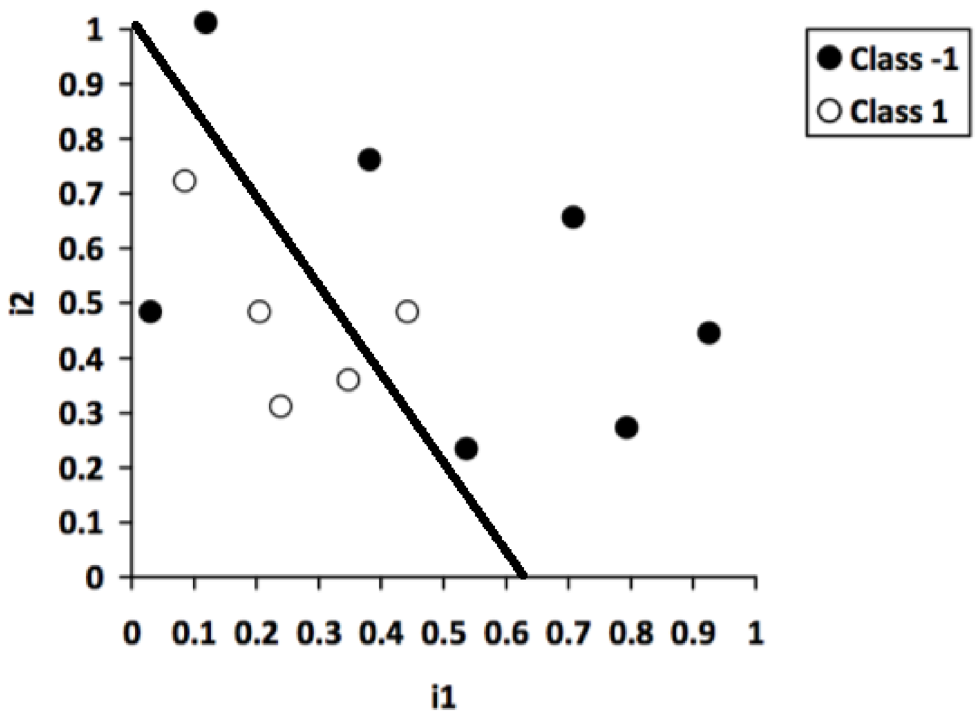


Figure 6: The minimum error possible is 2.

Part b

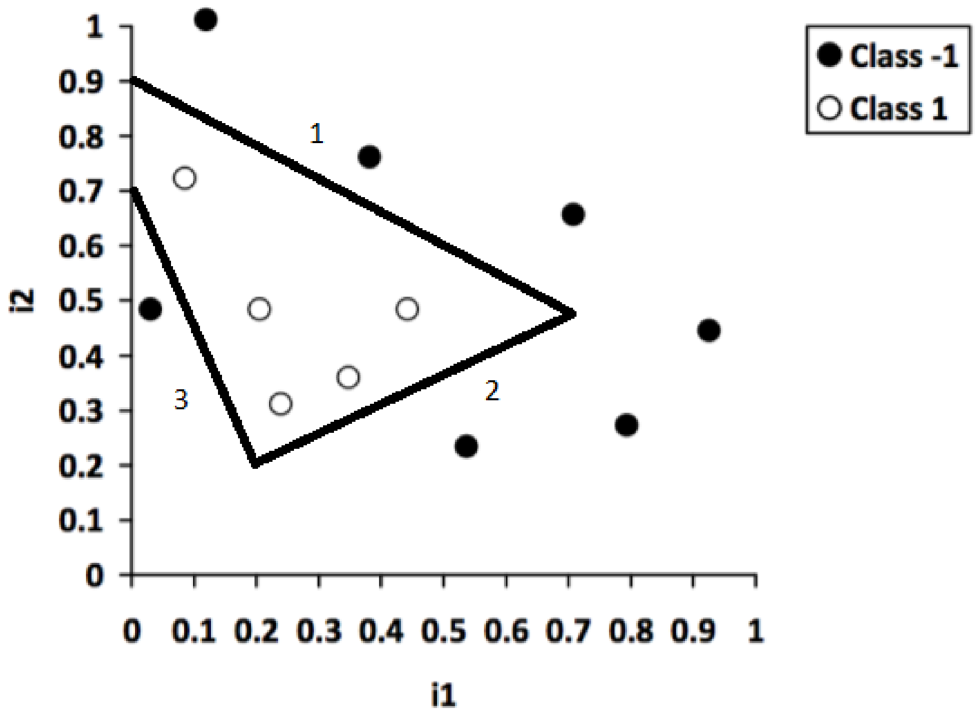


Figure 7: A multilayer perceptron with 3 layers separates the two categories.

The weights for the lines are:
$w_0 = -0.9, w_1 = \frac{4}{7}, w_2 = -1$ for line 1
$w_0 = -0.08, w_1 = -0.6, w_2 = 1$ for line 2
$w_0 = -0.7, w_1 = 2.5, w_2 = 1$ for line 3