

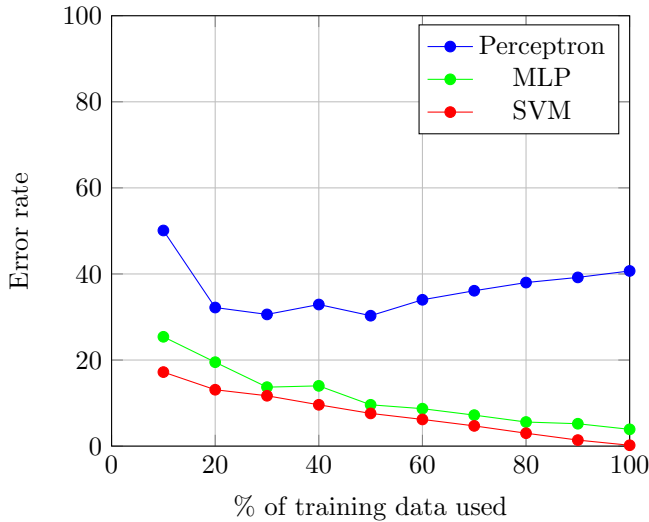
Assignment 2

Kyle Reagle and Yigit Gungor

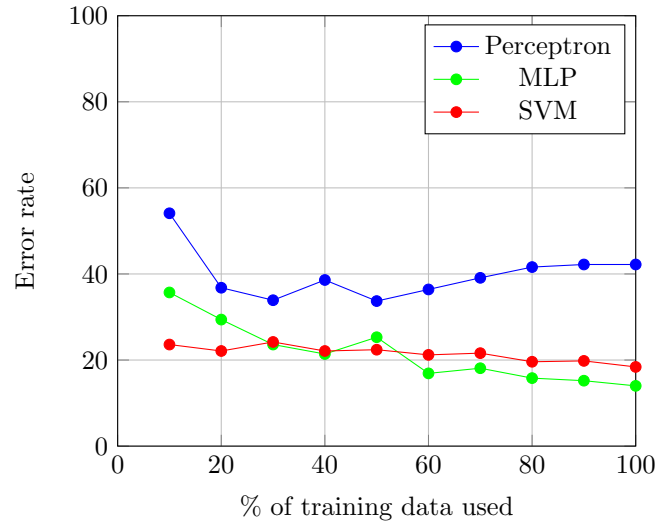
October 24, 2017

Question 1:

Predictions performed on the complete training data set



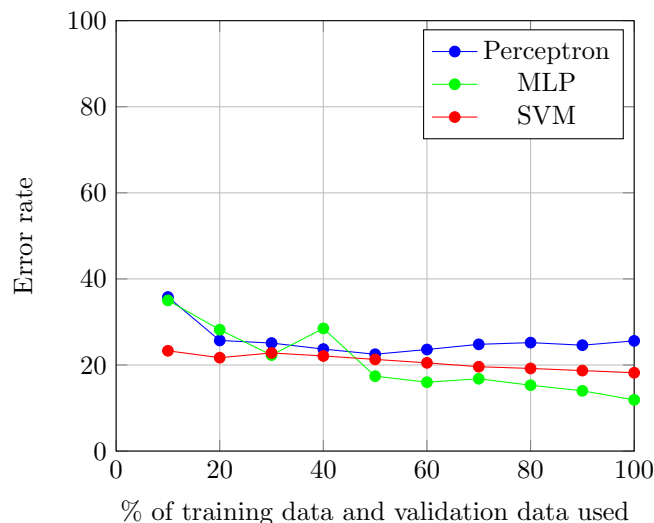
Predictions performed on the complete testing data set



Based on the graphs, it's obvious that the perceptron performed the worst of all the algorithms, especially since it was the slowest. Initially, the perceptron's prediction rate improved, but slowly got worse as more training data was used. The MLP and the SVM all improved their prediction rate as more training data was given. Since the perceptron and SVM were relatively easy to program, there wasn't much fine tuning to be done, and there didn't seem to be any failure conditions. The MLP was very inaccurate with high learning rates (alpha) around 1, so the learning rate had to be set around 0.001 for the best results. The MLP also seemed to perform the best with 50 hidden nodes. Anything higher or lower than 50 seemed to decrease the prediction performance. All the algorithms had lower error rates as the number of iterations increased, but it obviously slowed down run time. For example, the perceptron got to around 15% error rate at 25 iterations, but that required around half an hour of run time.

Extra Credit:

Predictions performed on the complete testing data set



Surprisingly, the perceptron's prediction performance improved a lot after doing training with the addition of validation data. The MLP and the SVM improved slightly, decreasing the error rate around 1% to 2%.

Question 2:

- (a) The provided tree correctly categorizes the provided examples.

Below is a table showing the output of the tree corresponding to the attributes given in the table.

No.	Table	Tree
1	P	P
2	P	P
3	P	P
4	P	P
5	P	P
6	P	P
7	N	N
8	N	N
9	N	N
10	N	N
11	N	N
12	N	N

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

GPA

$$E(GPA \geq 3.9) = -1\log_2(1) - 0\log_2(0) = 0$$

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

$$E(GPA \leq 3.2) = -0\log_2(0) - 1\log_2(1) = 0$$

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

$$Gain(GPA) = E(S) - I(GPA) = 1 - 0.4046 = 0.5954$$

Published

$$E(Published = Yes) = -\frac{3}{5}\log_2(\frac{3}{5}) - \frac{2}{5}\log_2(\frac{2}{5}) = 0.9710$$

$$E(Published = No) = -\frac{3}{7}\log_2(\frac{3}{7}) - \frac{4}{7}\log_2(\frac{4}{7}) = 0.9852$$

$$I(Published) = \frac{5}{12} * 0.9710 + \frac{7}{12} * 0.9792 = 0.9792$$

$$Gain(Published) = E(S) - I(Published) = 1 - 0.9792 = 0.0954$$

Rank

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

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

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

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

$$Gain(rank) = E(S) - I(rank) = 1 - 0.9046 = 0.0954$$

Recommendation

$$E(Recommendation = Good) = -\frac{5}{8}\log_2(\frac{5}{8}) - \frac{3}{8}\log_2(\frac{3}{8}) = 0.9544$$

$$E(Recommendation = Normal) = -\frac{1}{4}\log_2(\frac{1}{4}) - \frac{3}{4}\log_2(\frac{3}{4}) = 0.8113$$

$$I(Recommendation) = \frac{8}{12} * 0.9544 + \frac{4}{12} * 0.8113 = 0.9067$$

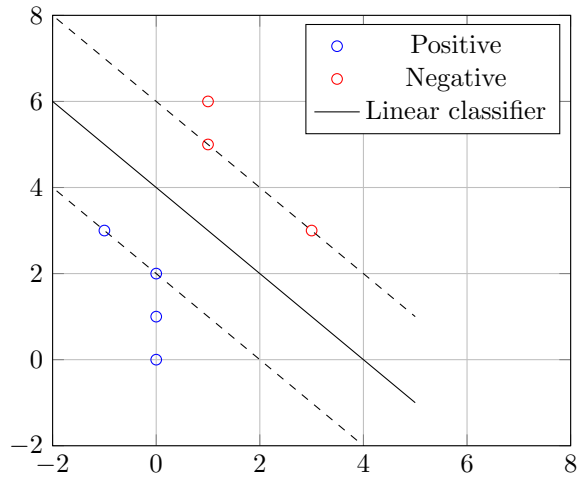
$$Gain(Recommendation) = E(S) - I(Recommendation) = 1 - 0.9067 = 0.0933$$

After selecting GPA as the first attribute, we will see that $(GPA \leq 3.2)$ and $(GPA \geq 3.9)$ reaches to consistent answers of P and N in that order. For the middle group we find that having published prior research reaches to a P answer with 0.42 gain. Lastly, ranking will identify the remainder data with N,P,N as in the order of the ranks.

- (c) Decision tree generated in part b is identical to the given decision tree. Since we used the most optimized way to set the tree in part b, we get the same result as the given example.

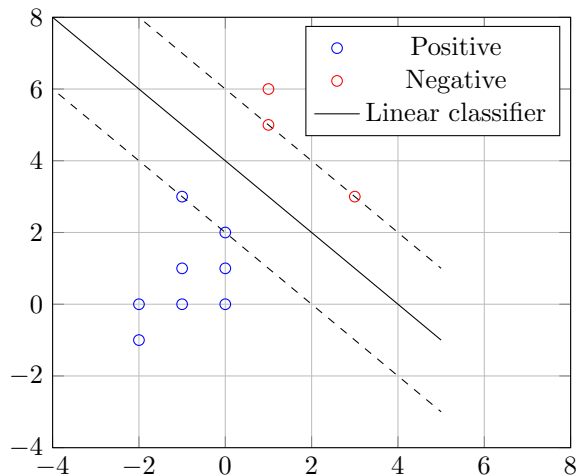
Question 3:

- (a) Linear classifier by inspection:



- (b) $w = -1, b = 4$

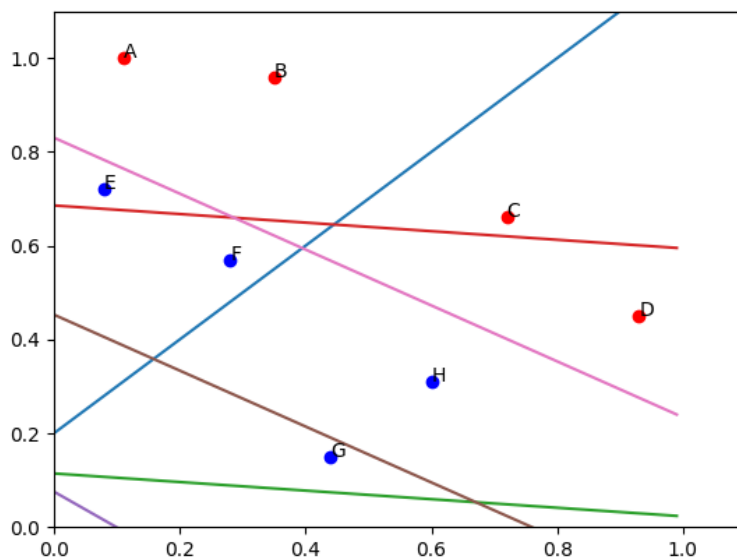
- (c) The linear SVM is still $h(x) = -x + 4$



Question 4:

- (a) Below are two examples of output from a program we wrote to classify the data points.

Figure 1: Final linear separator is the pink line

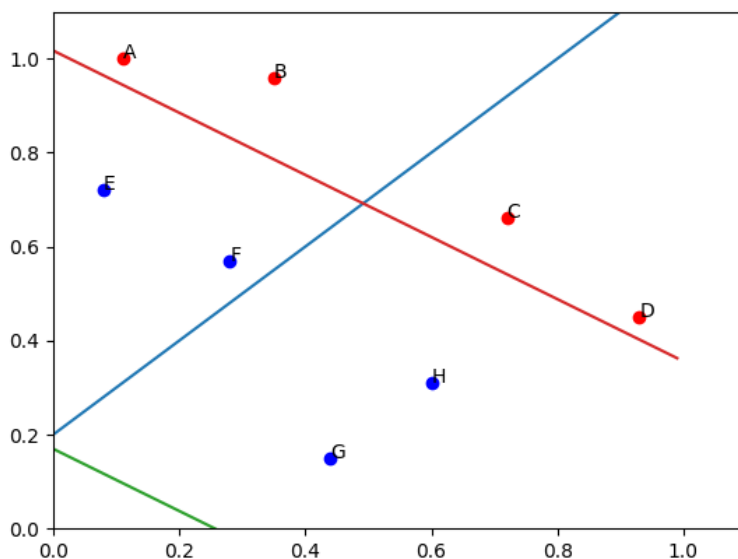


```

Iteration 1
Missclassified points: ['C', 'D', 'E', 'F']
New weight vector: [-0.8, -0.43999999999999995, -2.3200000000000003]
Iteration 2
Missclassified points: ['E', 'F', 'G', 'H']
New weight vector: [0.19999999999999996, -0.15999999999999992, -1.7500000000000004]
Iteration 3
Missclassified points: ['E', 'F', 'G', 'H']
New weight vector: [1.2, -0.15999999999999992, -1.7500000000000004]
Iteration 4
Missclassified points: ['D', 'E']
New weight vector: [0.19999999999999996, -2.02, -2.6500000000000004]
Iteration 5
Missclassified points: ['E', 'F', 'G', 'H']
New weight vector: [1.2, -1.58, -2.6500000000000004]
Iteration 6
Missclassified points: ['E', 'F', 'H']
New weight vector: [2.2, -1.58, -2.6500000000000004]

```

Figure 2: Final linear separator is the red line

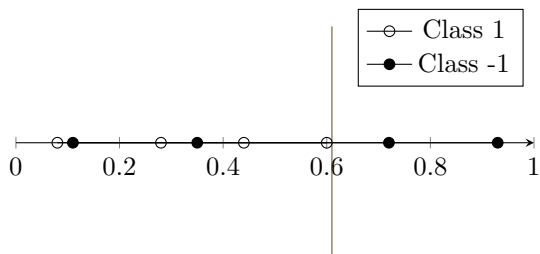


```

Iteration 1
Missclassified points: ['C', 'D', 'E', 'F']
New weight vector: [-0.8, -0.8600000000000001, -1.9]
Iteration 2
Missclassified points: ['E', 'F', 'G', 'H']
New weight vector: [0.19999999999999996, -0.7800000000000001, -1.18]
Iteration 3
Missclassified points: ['E', 'F', 'G', 'H']
New weight vector: [1.2, -0.7800000000000001, -1.18]

```

- (c) The plot below shows the line that misclassifies the least number of data points, which are the two black dots to the left of the linear separator. This line can be represented with a weight ratio of $-w_0/w_1 = 0.6$



$$MSE = ((0.6 - 0.11)^2 + (0.6 - 0.35)^2)/8 = 0.038$$

Question 5:

- (a) By observation we can see that a line with the following formula will separate most of the data point.

$$i_2 = -i_1 + 1$$

Using this line, the minimum error will be 2/12 due to two Class -1 points below the line.

- (b) To separte the data point provided, we decided that we would need three lines. These would be:

$$i_2 = -i_1 + 1$$

$$1 - 1i_1 - 1i_2 = 0$$

$$w_{0,0} = 1, w_{0,1} = -1, w_{0,2} = -1$$

$$i_2 = i_1 - 0.05$$

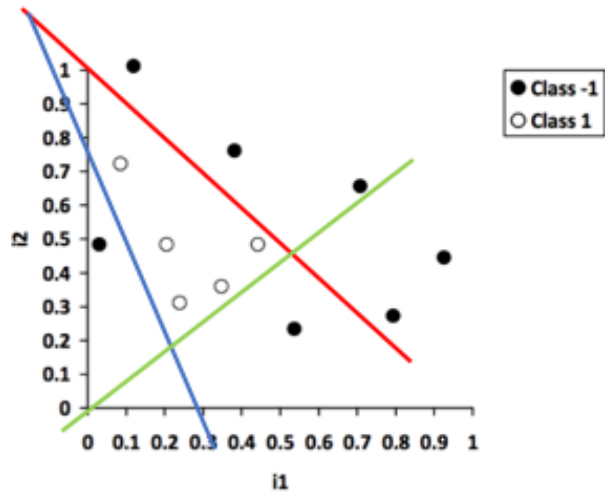
$$-0.05 + 1i_1 - 1i_2 = 0$$

$$w_{1,0} = -0.05, w_{1,1} = 1, w_{1,2} = -1$$

$$i_2 = -0.02i_1 + 0.08$$

$$0.08 - 0.02i_1 - 1i_2 = 0$$

$$w_{2,0} = 0.08, w_{2,1} = -0.02, w_{2,2} = -1$$



The data points that are within these lines will be classified as Class 1 and outside as Class -1.