

Joe Czepil
V00774878
SENG 474
Assignment 2

Question 1)

1.1)

Rectangle with a percentage of 33% will be predicted correctly every time. Every other prediction will be wrong. Therefore, the error rate is 66%.

1.2)

Predicting rectangle 70% of the time would result in rectangle being correctly predicted $0.7 \times 0.333 = 23.3\%$ of the time. Predicting triangle 30% of the time would result in triangle being correctly predicted $0.3 \times 0.333 = 9.9\%$ of the time. Therefore, the error rate would be $1 - 0.233 - 0.0999 = 66\%$.

1.3)

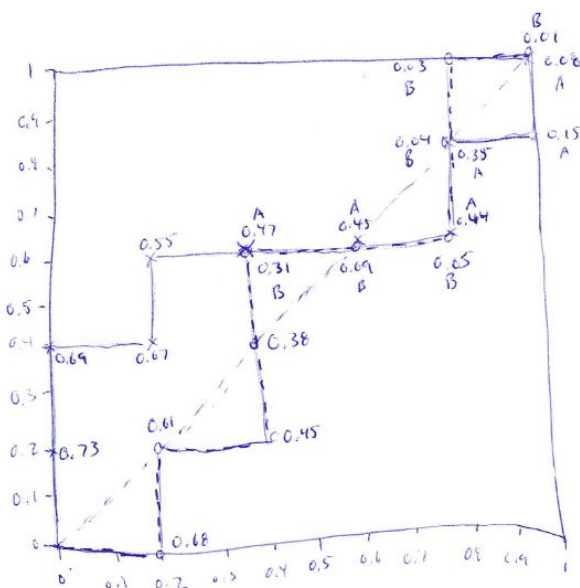
Predicting rectangle 100% of the time would result in 50% of predictions being correct as 50% of the data is rectangles. Therefore, the error rate would be 50%.

1.4)

Predicting rectangle 70% of the time would result in rectangle being correctly predicted $0.7 \times 0.5 = 35\%$ of the time. Predicting triangle 30% of the time would result in triangle being correctly predicted $0.3 \times 0.25 = 7.5\%$. Therefore, the error rate would be $1 - 0.35 - 0.075 = 57.5\%$.

Question 2)

2.1)



2.2)

Classifier A

Precision: 3/4

Recall: 3/5

$$\text{F-measure} = 2 * \left(\frac{\frac{3}{4} * \frac{3}{5}}{\frac{3}{4} + \frac{3}{5}} \right) = 0.667$$

2.3)

See graph in question 2.1

Thresholds when classifier A did better than random: 0.45 - 1

Thresholds when classifier B did better than random: 0.09 - 0.38, 0.01 - 0.04

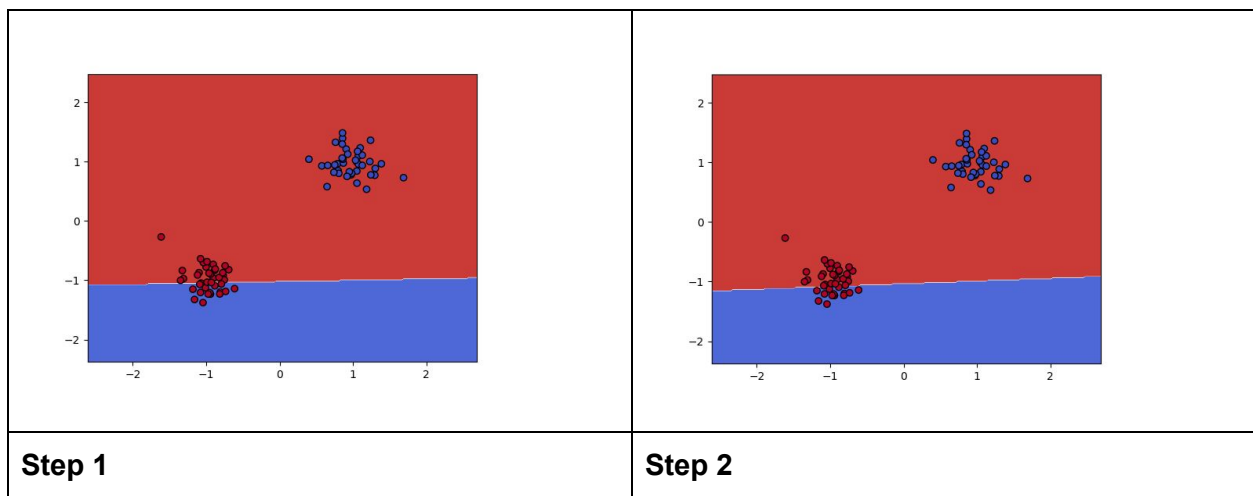
2.4)

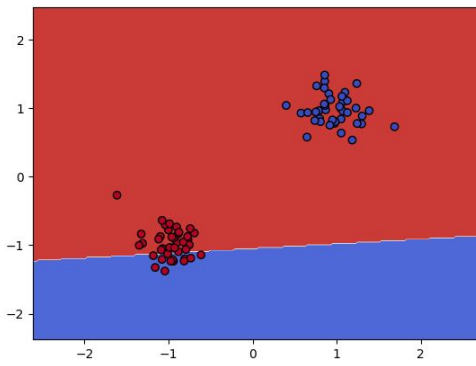
Not a graduate student.

Question 3)

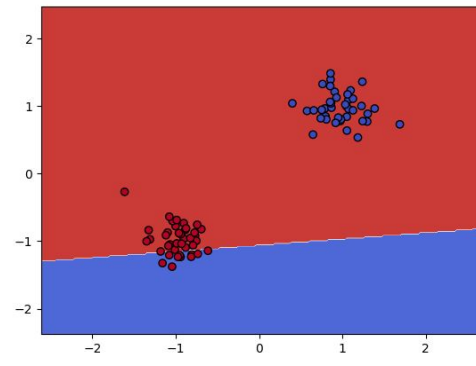
3.1)

With each iteration of fitting the line, the line moves slightly closer to the best solution. As can be seen in the diagrams below, the line slowly moves counter clockwise until neither cluster is inside the blue area. Obviously, this is not a complete solution but if more iterations occurred, eventually the blue cluster would end up in the blue area and the red cluster would end up in the red area.

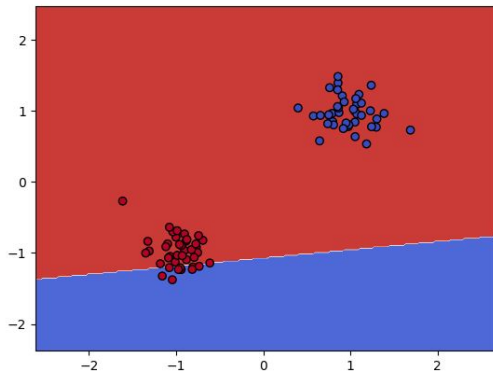




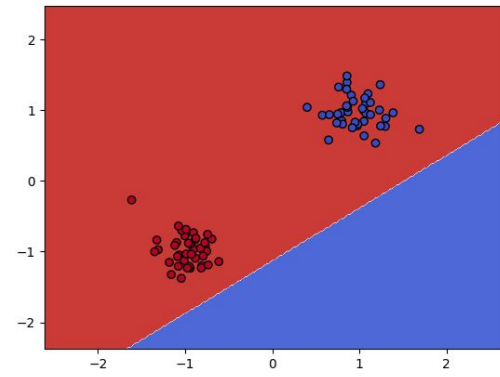
Step 3



Step 4



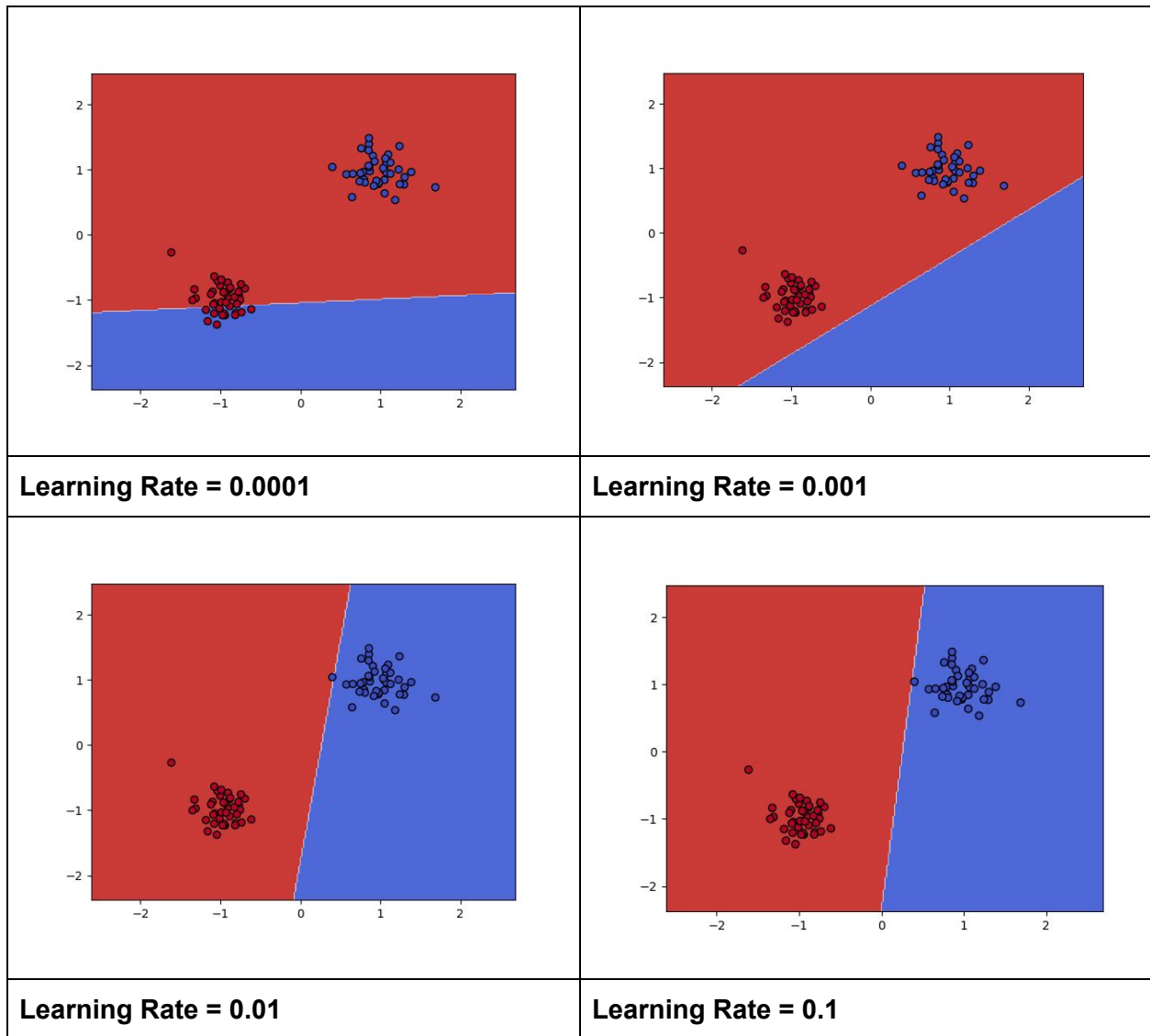
Step 5



Final Step

3.2)

As the learning rate is increased, the classifying line gets a more and more accurate result. With a learning rate of 0.0001 the initial line barely changes. A learning rate of 0.1 yields an acceptable but far from optimal result. Curiously, the difference between 0.1 and 0.01 is quite small.



3.3)

Not a graduate student.

Question 4)

4.1)

π	$\pi^7(1 - \pi)^3$
0.1	0.000000073
0.2	0.0000066
0.3	0.000075
0.4	0.00035
0.5	0.00098
0.6	0.0018
0.7	0.0022
0.8	0.0017
0.9	0.00048
1	0

The MLE is 0.0022.

4.2)

Range	A (Red)	B (Blue)	D
0.05	14	3	
0.1	4	4	2
0.15	3	2	3
0.2		4	
0.25			
0.3		3	
0.35		2	
0.4			
0.45			

0.5			1
0.55		1	1
0.6		1	1
0.65			
0.7		1	1

Model B (Blue) is more likely to describe the data from D. D has no values within the 0.05 range of model A which has 14 values out of 21 and three values match at 0.55, 0.5, and 0.7 with model B (Blue).