SENG 474
Assignment 2

Zhengtang Wang
V00802086

Question 1.

1.1 Error rate $= 1 - \left((1)(\frac{1}{3}) + (0)(\frac{1}{3}) + (0)(\frac{1}{3}) \right) = \frac{2}{3} \approx 0.667$ (□ △ ○)

1.2 Error rate $= 1 - \left((0.7)(\frac{1}{3}) + (0.3)(\frac{1}{3}) + (0)(\frac{1}{3}) \right) = \frac{2}{3} \approx 0.667$ (□ △ ○)

1.3 Error rate $= 1 - \left((1)(\frac{1}{2}) + (0)(\frac{1}{4}) + (0)(\frac{1}{4}) \right) = \frac{1}{2} = 0.5$ (□ △ ○)

1.4 Error rate $= 1 - \left((0.7)(\frac{1}{2}) + (0.3)(\frac{1}{4}) + (0)(\frac{1}{4}) \right) = \frac{23}{40} = 0.575$ (□ △ ○)

Question 2.

2.1 Descending order:

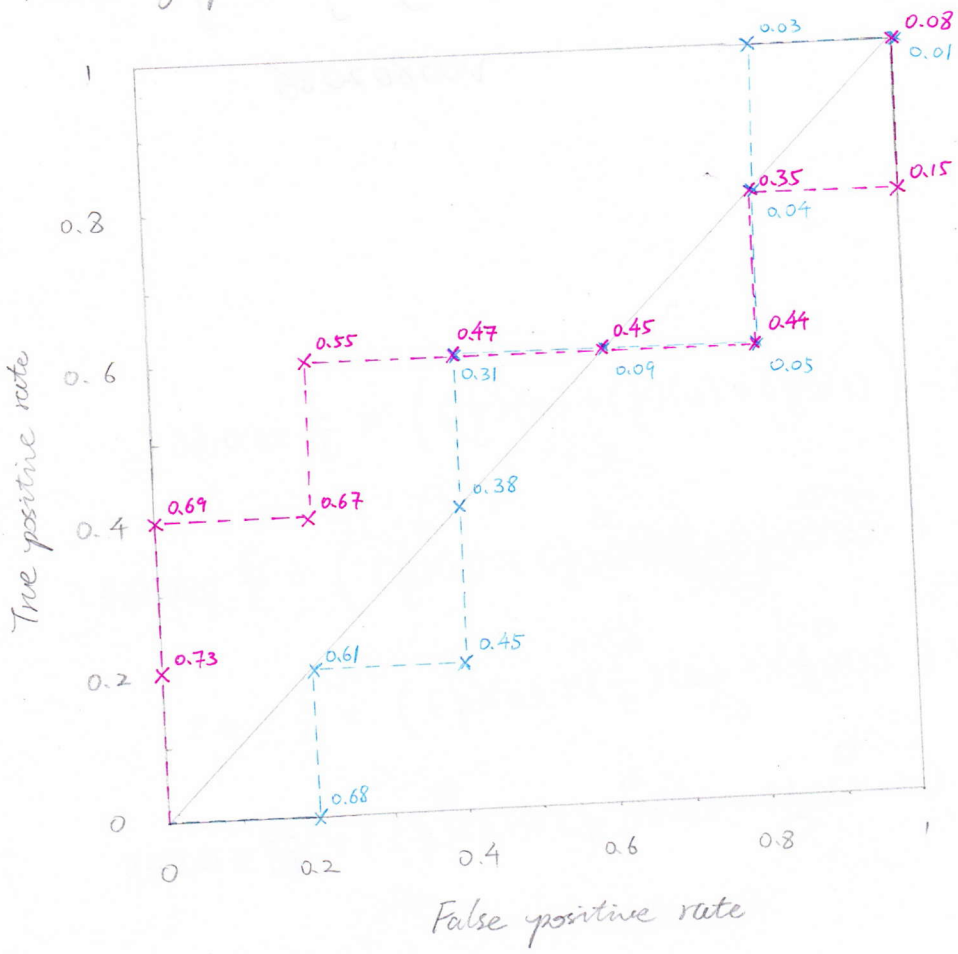
A

1	P	0.73
2	P	0.69
5	N	0.67
4	P	0.55
6	N	0.47
9	N	0.45
3	N	0.44
10	P	0.35
8	N	0.15
7	P	0.08

B

3	N	0.68
1	P	0.61
5	N	0.45
7	P	0.38
4	P	0.31
6	N	0.09
8	N	0.05
10	P	0.04
2	P	0.03
9	N	0.01

ROC graph:



2.2 When threshold $t = 0.5$

Instances 1, 2, 5 and 4 greater than 0.5 (classifier A)
(P, P, N, P)

$$\text{precision} = \frac{tp}{tp + fp} = \frac{3}{3+1} = \frac{3}{4} = 0.75$$

$$\text{recall} = \frac{tp}{tp + fn} = \frac{3}{3+2} = \frac{3}{5} = 0.6$$

$$F\text{-measure} = 2 \cdot \frac{\text{precision} \cdot \text{recall}}{\text{precision} + \text{recall}} = 2 \times \frac{0.75 \times 0.6}{0.75 + 0.6} = \frac{2}{3} \approx 0.667$$

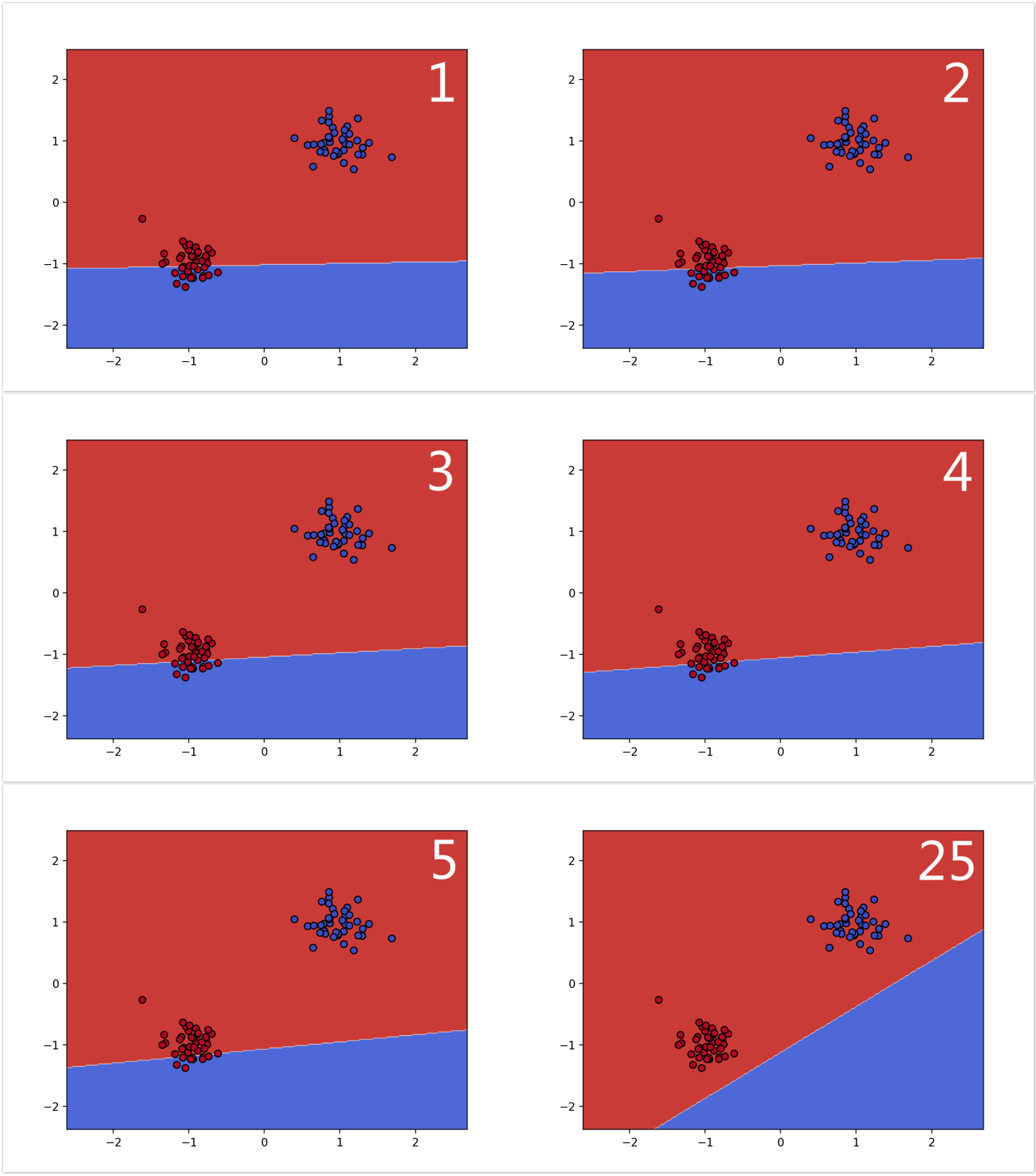
2.3 The curve of unbiased random classifier is plotted above.

When threshold between $t=1$ and $t=0.45$, classifier A better than random classifier.

When threshold from $t=0.38$ to $t=0.09$, and from $t=0.04$ to $t=0.01$, classifier B better than random classifier.

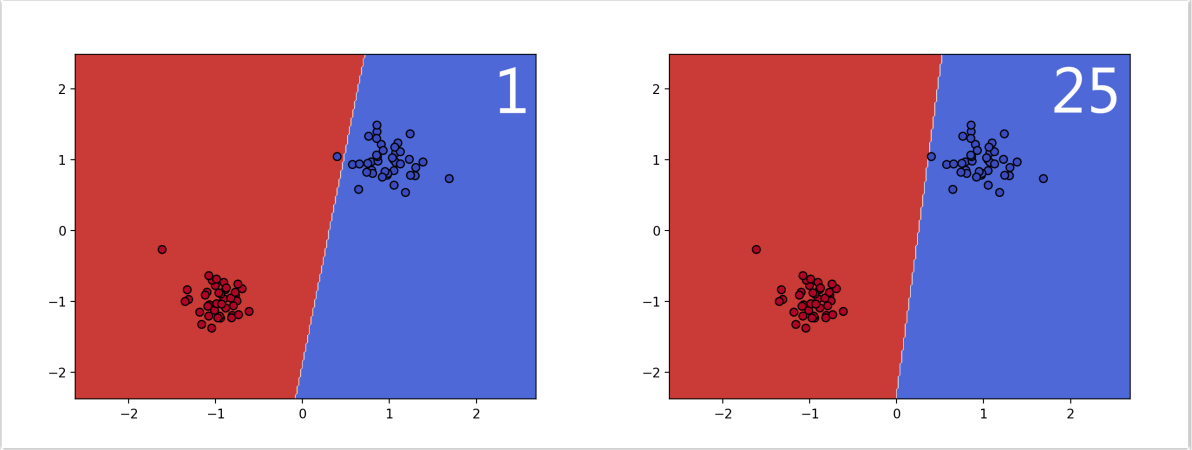
Question 3.

3.1 n_epochs = 25, learning_rate = 0.001



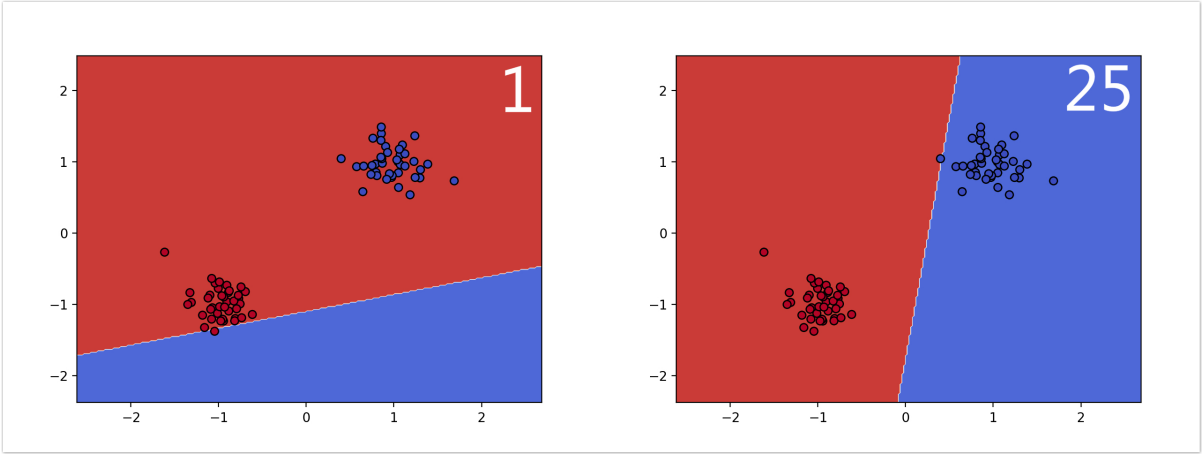
The decision boundary is moving upward and trying to separate two classes of data. But unable to do so due to the limited number of epochs, and a low learning rate.

3.2 n_epochs = 25, learning_rate = 0.1



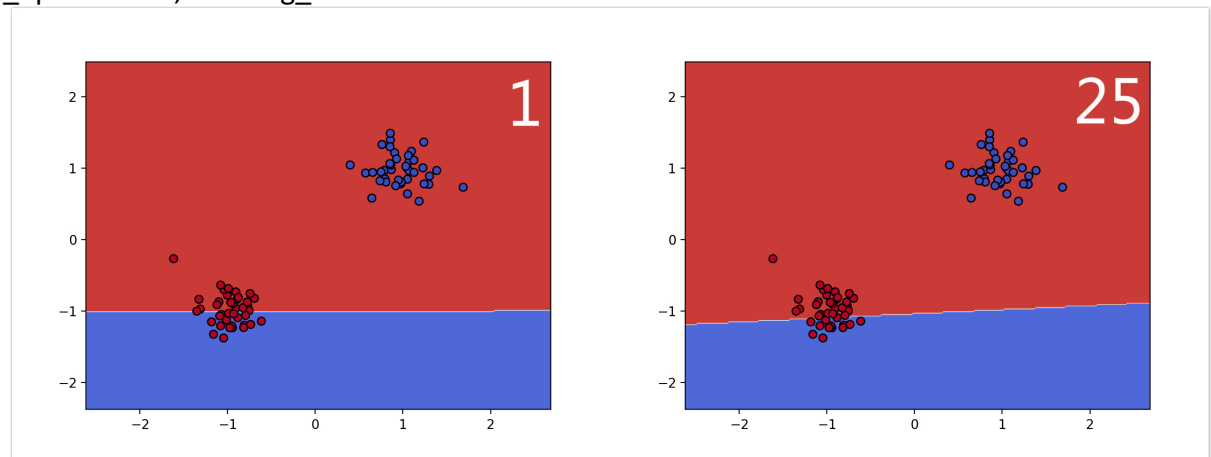
The initial plot is almost correctly classified two classes of data, with just one blue dot on the red area. On the second plot, the decision boundary successfully separate two classes, and the perceptron stop since all data are classified correctly. So, no changes on the rest plots.

n_epochs = 25, learning_rate = 0.01



The decision boundary is moving upward and trying to separate two classes of data, the decision boundary shifts a lot at first, then slow down when it get closer to the group of blue data. Almost correctly classified all the data, but not finished due to limited number of epochs.

n_epochs = 25, learning_rate = 0.0001



The decision boundary is slightly shifts a little bit through the 25 plots, due to the slow learning speed, perceptron cannot classify these data in 25 epochs. Need either higher learning rate or more epochs to do this classification.

Question 4.

4.1 $P(D|\theta) = \prod_i P(x_i|\theta) = \prod_i \theta^{x_i}(1-\theta)^{1-x_i}$

$\because \theta = P(x=T)$ & $D = \{T \times 7, F \times 3\}$

$\therefore x_i = 7, \ 1-x_i = 3$

$P(D|\theta) = \theta^7(1-\theta)^3$

Finding the maximum of $P(D|\theta)$ by solve the derivative $= 0$

$$\frac{d}{d\theta}\left(\theta^7(1-\theta)^3\right) = 7\theta^6(1-\theta)^3 + \theta^7 \cdot \frac{d}{d\theta}\left((1-\theta)^3\right)$$

$$0 = 7\theta^6(1-\theta)^3 + \theta^7 \cdot 3(1-\theta)^2 \cdot \frac{d}{d\theta}(1-\theta)$$

$$0 = 7\theta^6(1-\theta)^3 + 3\theta^7(1-\theta)^2 \cdot (-1)$$

$$0 = 7\theta^6(1-\theta)^3 - 3\theta^7(1-\theta)^2$$

$$3\theta^7(1-\theta)^2 = 7\theta^6(1-\theta)^3$$

$$3\theta = 7(1-\theta)$$

$$3\theta = 7 - 7\theta$$

$$10\theta = 7$$

$$\theta = \frac{7}{10} = 0.7$$

4.2 $D = \{0.1, 0.1, 0.12, 0.15, 0.15, 0.5, 0.55, 0.6, 0.68\}$

Fit in histogram of A: 2(0.1) ✓, (0.12) ✓, 2(0.15) ✓, (0.5) ×, (0.55) ×, (0.6) ×, (0.68) $\frac{5}{9} \approx 0.556$

Fit in histogram of B: 2(0.1) ✓, (0.12) ✓, 2(0.15) ✓, (0.5) ×, (0.55) ✓, (0.6) ✓, (0.68) ✓ $\frac{8}{9} \approx 0.889$

mode(A) = 0.05, mode(B) = 0.1, 0.2

mean(A) = $[(14)(0.05)+(4)(0.1)+(3)(0.15)]/(14+4+3) = 0.07$, mean(B) = ... = 0.24

mode(D) = 0.1, 0.15, mean(D) = 0.33

∵ B has similar mode, mean, and higher ratio to represent D's data.

∴ Overall, B is more likely to explain data D.