FIT5201 Assignment 2 Report

Name: Akshay Sapra
ID: 29858186

Question 1 Task I

Q1) Task 1

Notations

$X \to$ observation set (points)

$K \to$ set of Clusture labels

$Z \to$ set of latent Variables

$\theta \to$ set of unknown parameters

$$L(X|\theta_n) = \log p(X|\theta) \quad -①$$

log likelihood of incomplete data.

$$\text{where } p(X|\theta_n) = \frac{p(Z,X|\theta_n)}{P(Z|X,\theta_n)} \quad \text{(Baye's Theor)} \quad -②$$

using ① & ②

$$L(X|\theta_n) = \log p(X|\theta_n)$$

$$= [\log p(X|\theta_n)] \sum_Z P(Z|X,\theta_n)$$

$$= \sum_Z P(Z|X,\theta_n) \log p(X|\theta_n) \quad \left(\text{since } \sum P(Z|X,\theta_n) = 1\right)$$

$$= \sum_Z P(Z|X,\theta_n) \log \frac{p(Z,X|\theta_n)}{P(Z|X,\theta_n)} \quad (\text{from } ②)$$

(M.W. Mak, 2005)

$$\sum_{z} P(Z|X,\theta_n) \log p(Z,X|\theta_n) - \sum_{z} P(Z|X,\theta_n) \log P(Z|X,\theta_n).$$

$$= E_Z\{\log p(Z,X|\theta_n) | X,\theta_n\} - E_Z\{\log P(Z|X,\theta_n) | X,\theta_n\}$$

$$= Q(\theta_n|\theta_n) + R(\theta_n|\theta_n)$$

where $E_Z \rightarrow$ Expectation wrt Z

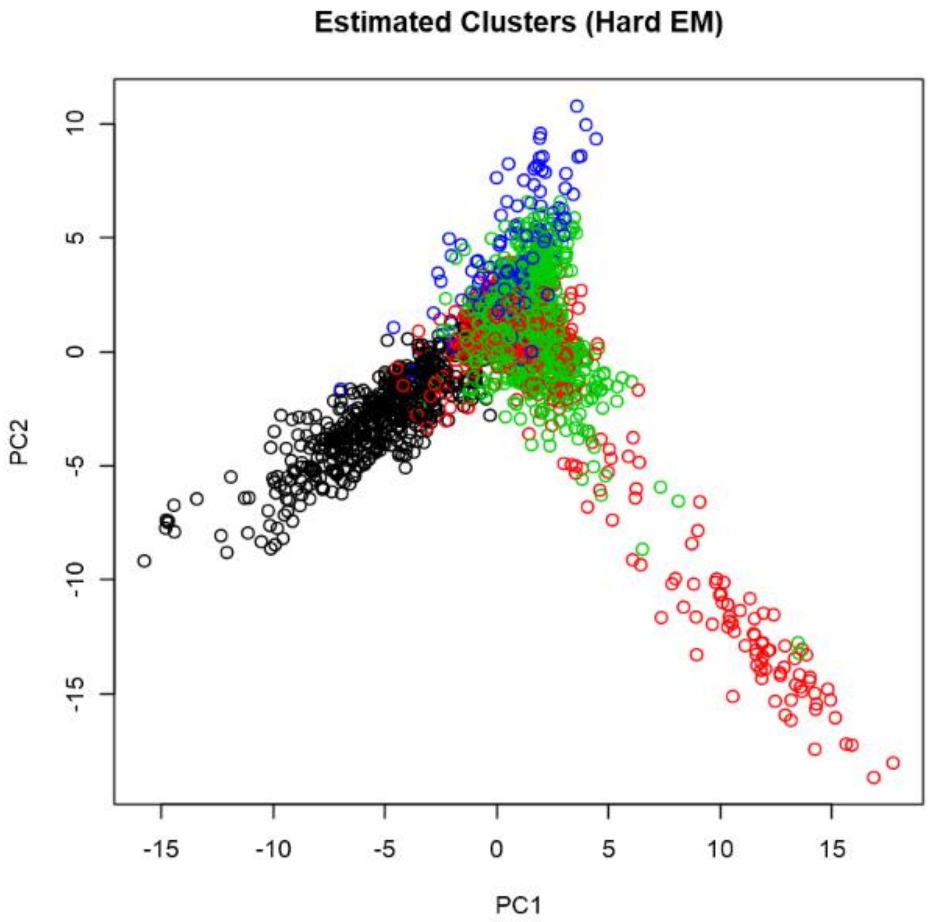
$$Q(\theta|\theta_n) = E_Z\{\log p(Z,X|\theta) | X,\theta_n\}$$

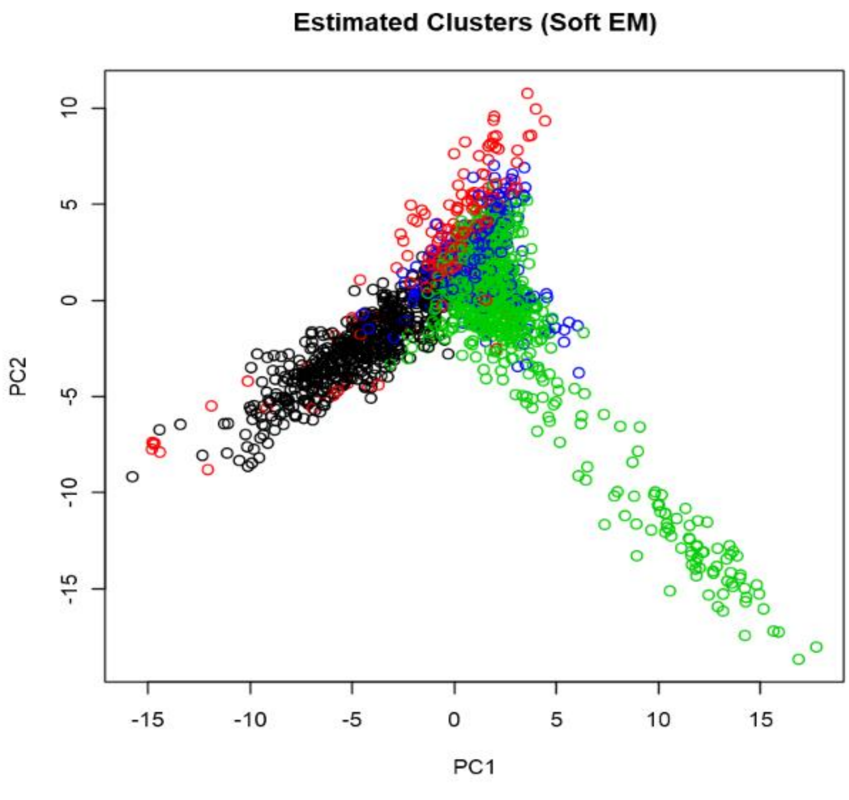
$$= \sum_{i=1}^{T} \sum_{j=1}^{J} E\{\delta_t^{(j)} | x_i,\theta_n\} \log\left[p(x_i|s_t^{(j)}=1, \phi^{(j)}) \pi^{(j)}\right]$$

Reference :- informit.com/articles/articles-aspx?p= ~~3027~~ 363730& seqNum=2

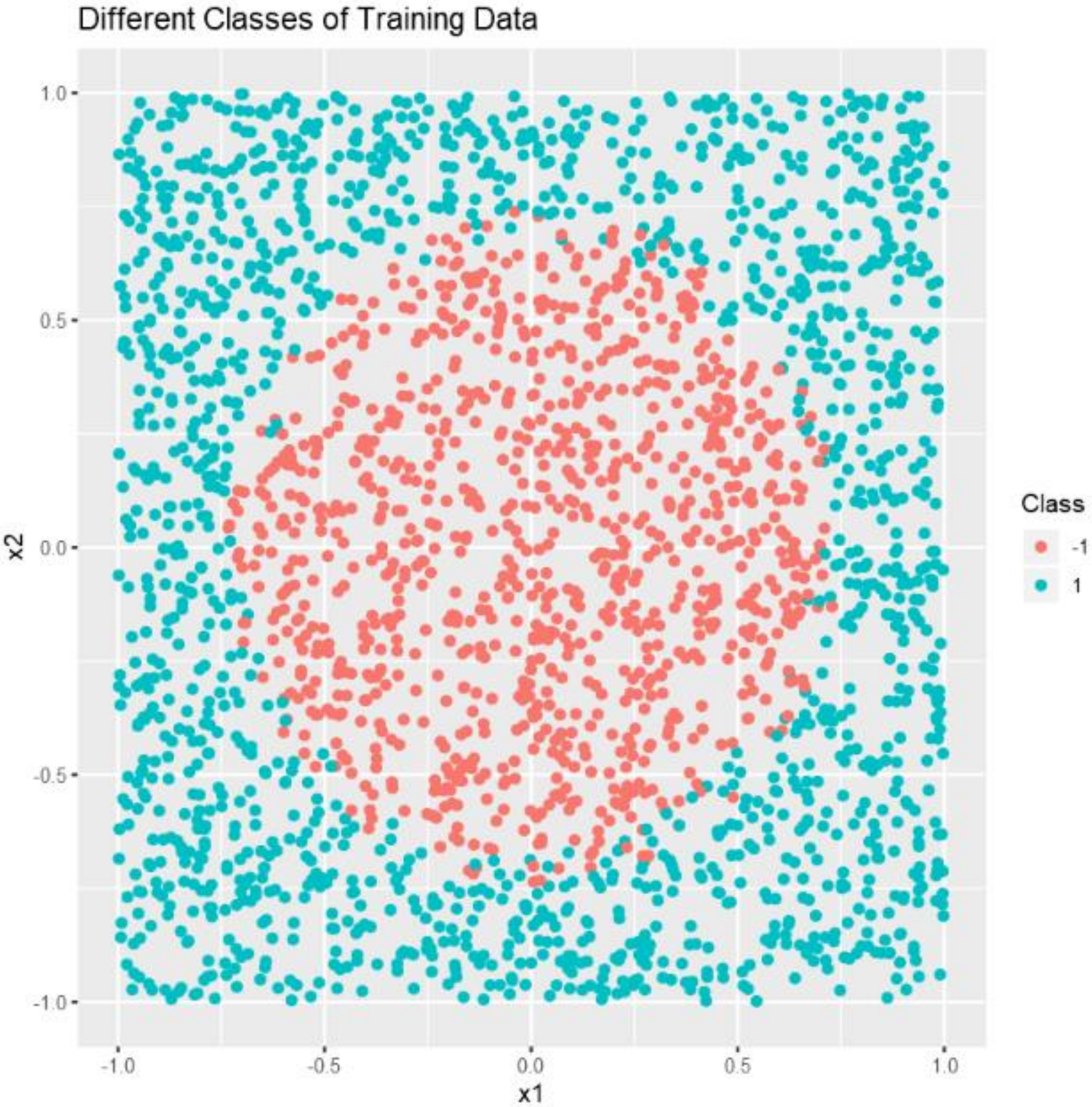
(M.W. Mak, 2005)

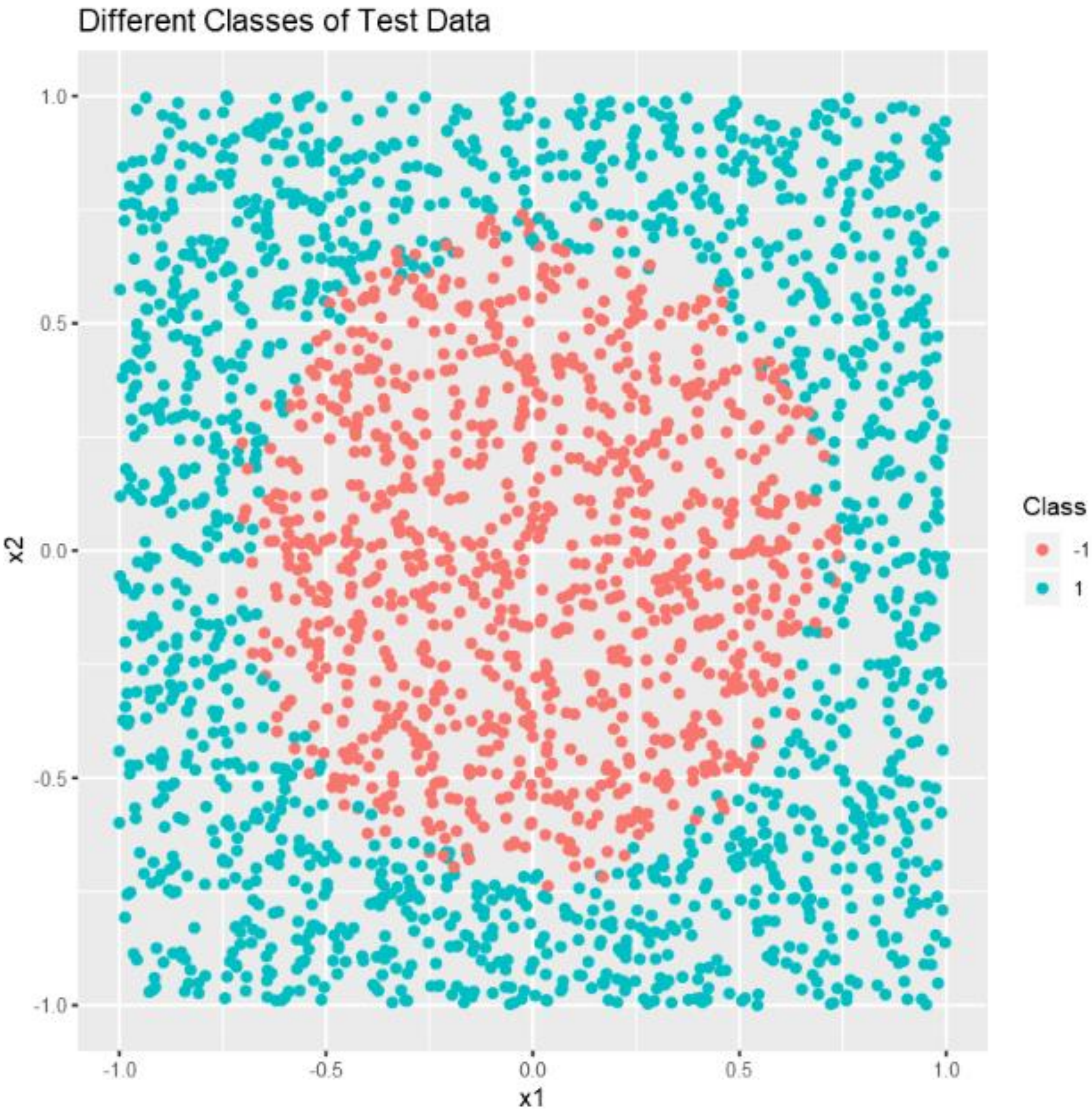
Question 1 Task IV





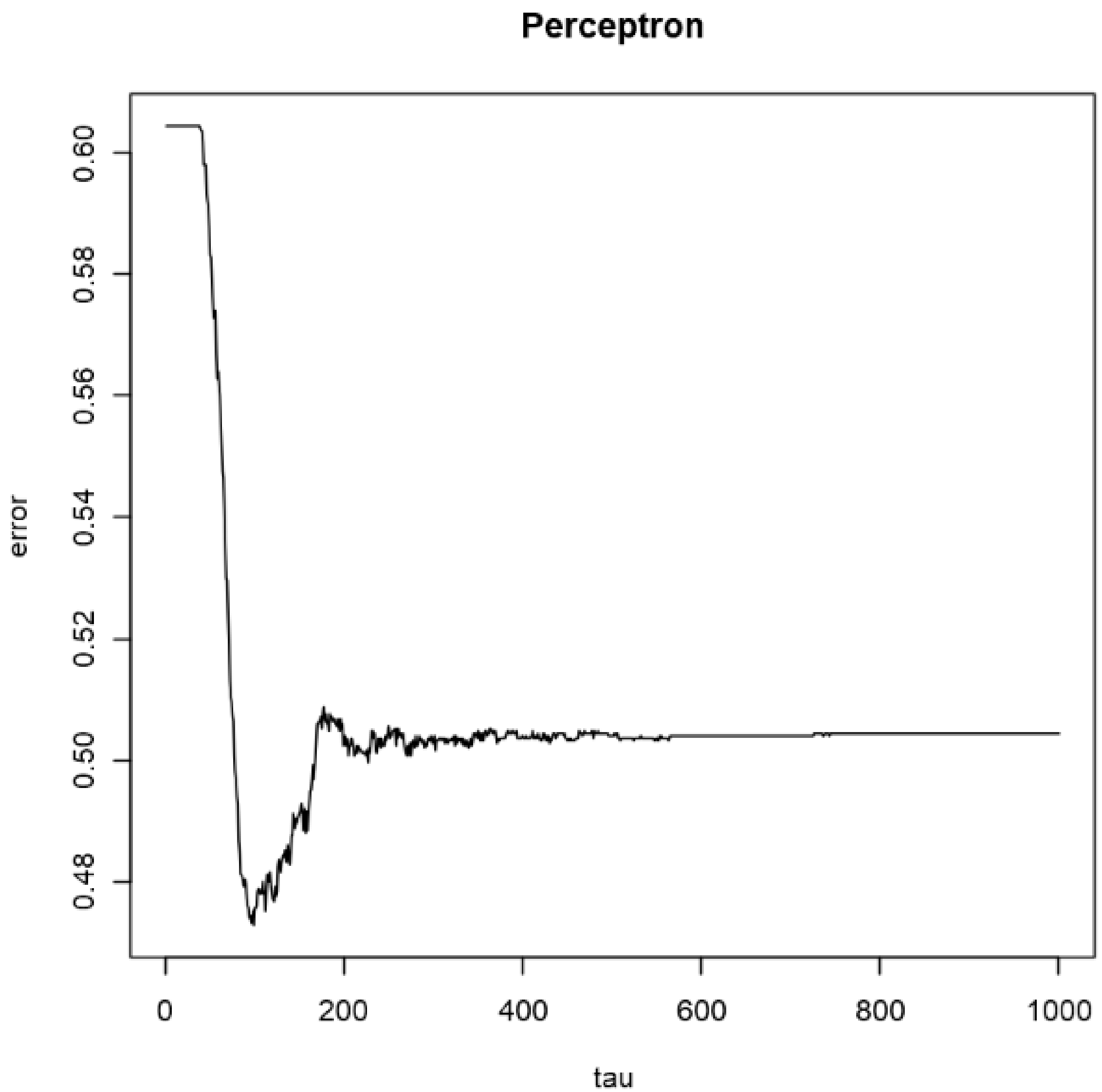
Question 2 Task I

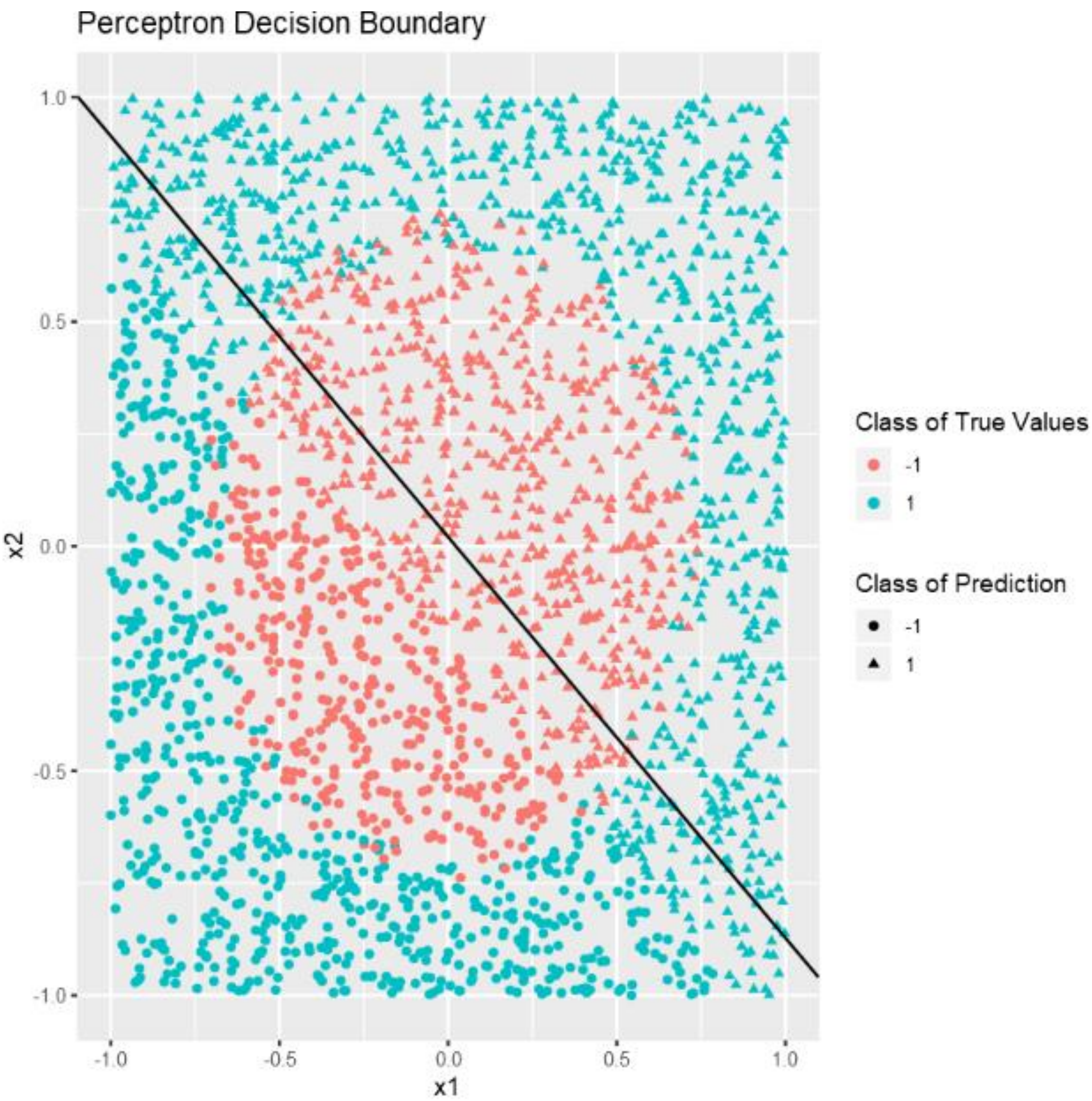




Different Classes of Test Data
x2
x1
1.0
0.5
0.0
-0.5
-1.0
-1.0
-0.5
0.0
0.5
1.0
Class
-1
1

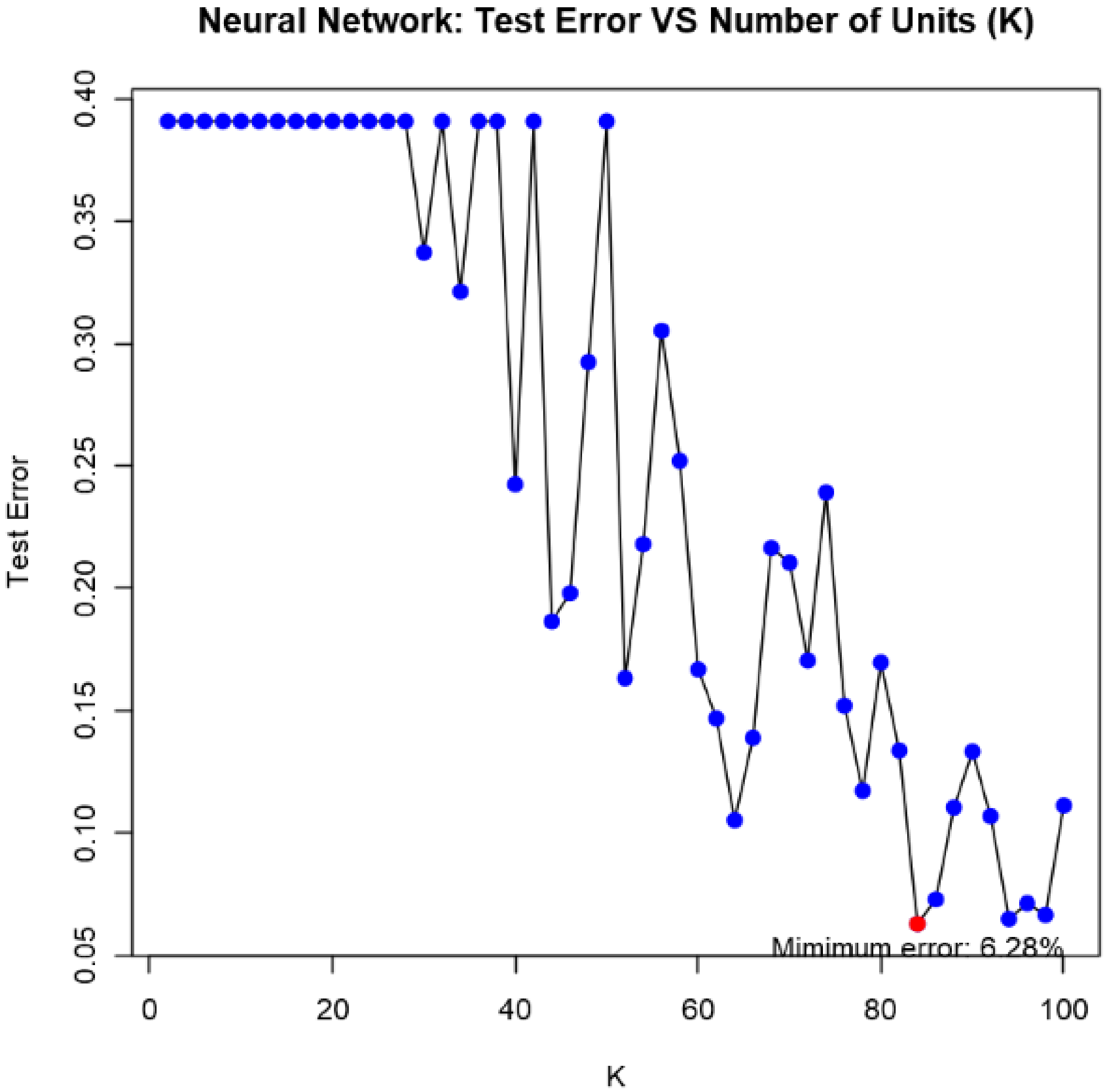
Question 2 Task II

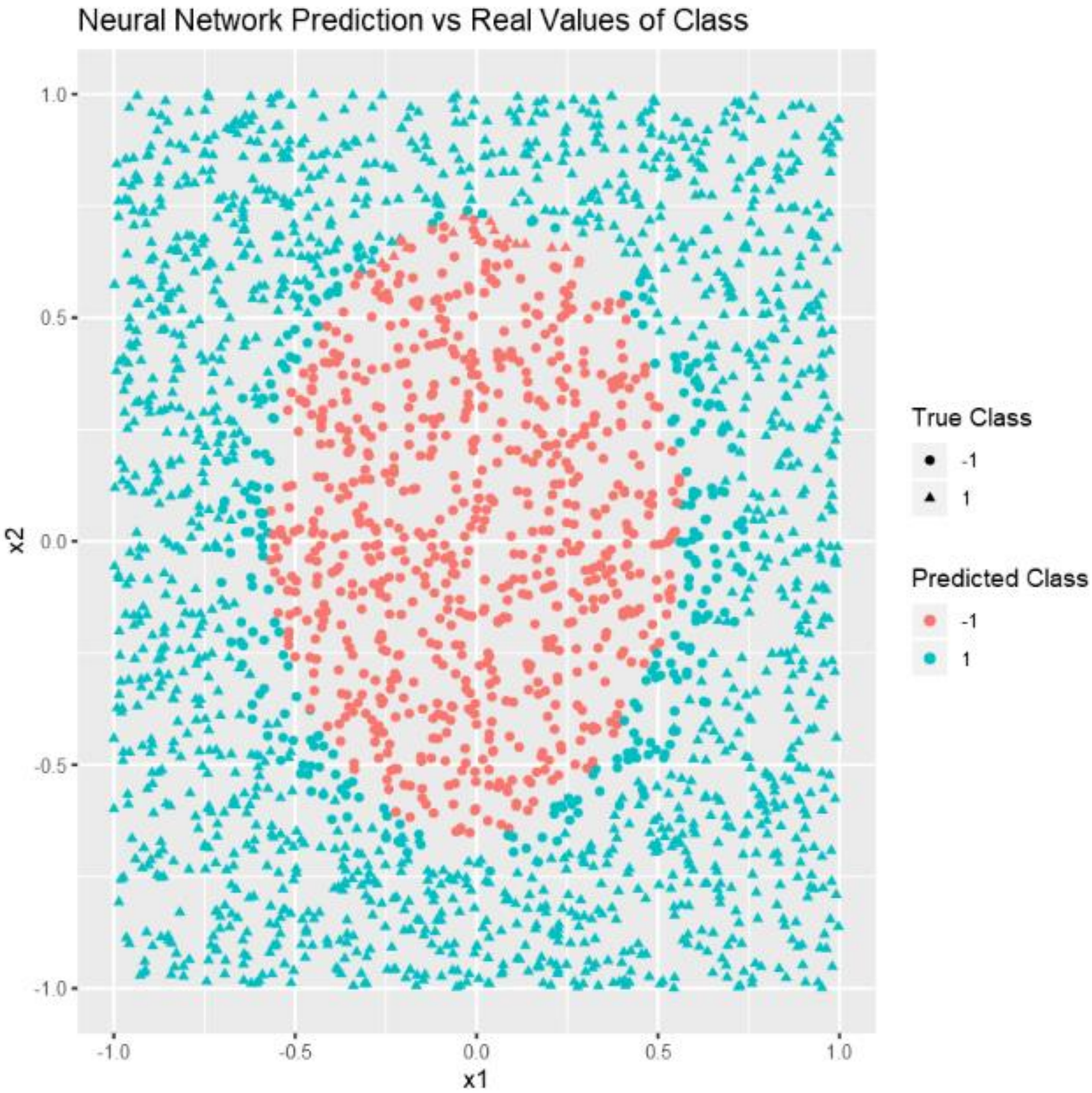




Perceptron Decision Boundary
1.0
0.5
0.0
-0.5
-1.0
x2
-1.0
-0.5
0.0
0.5
1.0
x1
Class of True Values
-1
1
Class of Prediction
-1
1

Question 2 Task III





Neural Network Prediction vs Real Values of Class
x2
1.0
0.5
0.0
-0.5
-1.0
x1
-1.0
-0.5
0.0
0.5
1.0
True Class
-1
1
Predicted Class
-1
1

Question 2 Task IV

K	Error_Percentage
2	39.08
4	39.08
6	39.08
8	39.08
10	39.08
12	39.08
14	39.08
16	39.08
18	39.08
20	39.08
22	39.08
24	39.08
26	39.08
28	39.08
30	33.72
32	39.08
34	32.12
36	39.08
38	39.08
40	24.24
42	39.08
44	18.64
46	19.80
48	29.24
50	39.08
52	16.32
54	21.80
56	30.52
58	25.20
60	16.68
62	14.68
64	10.52
66	13.88
68	21.64
70	21.04
72	17.04
74	23.92
76	15.20
78	11.72
80	16.96
82	13.36
84	6.28
86	7.28
88	11.04
90	13.32
92	10.68
94	6.48
96	7.12
98	6.64
100	11.12

```
print(paste0("Error for Perceptron is ", Perceptron.error*100," %"))
```

```
[1] "Error for Perceptron is 51.68 %"
```

```
print(paste0("In Neural Network, For K = ", min_x," units, error was minimum at ",min_y*100, " %"))
```

```
[1] "In Neural Network, For K = 84 units, error was minimum at 6.28 %"
```

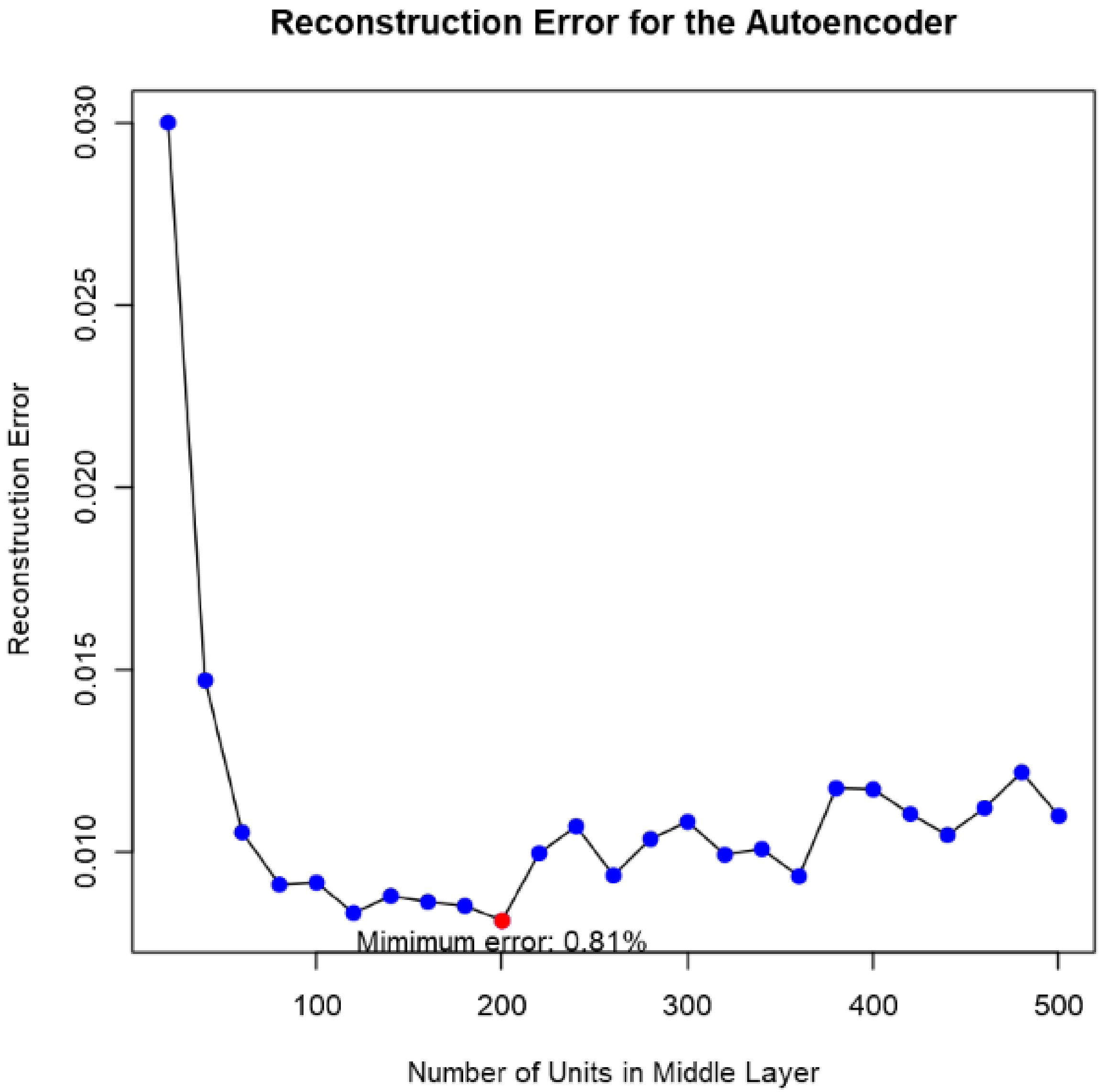
Question 2 Task V

Task V

In your PDF report explain the reason(s) responsible for such difference between perceptron and a 3-layer NN

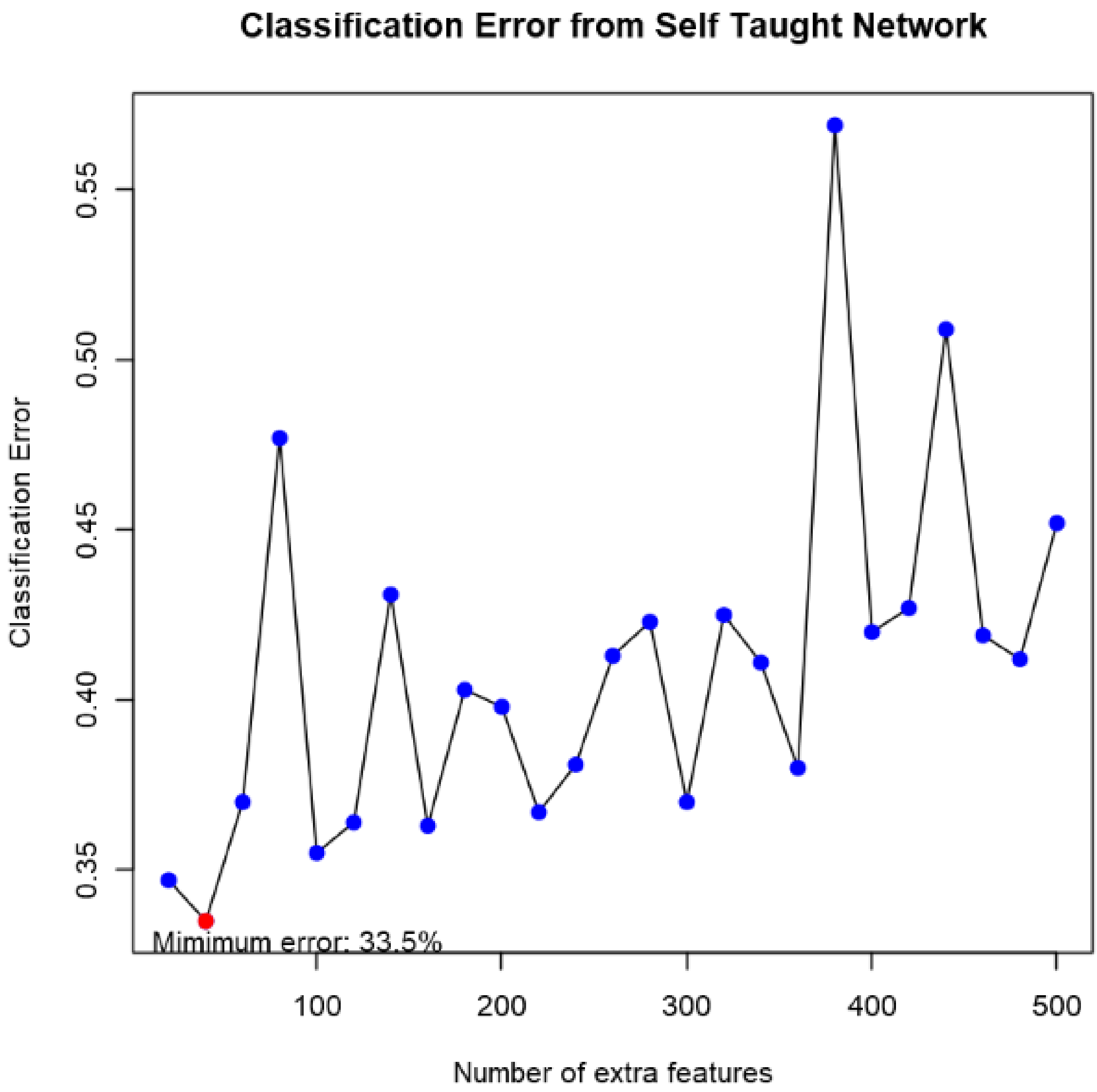
> -In this task I have explored both Perceptron and Neural Network for binary classification. It is evident from the graphs and errors that Perceptron model didn't performed well. It is primarily because Perceptron work well when data is linearly separable. While on the same data Neural Network performed with very low misclassification error because it is non linear classifier.

Question 3 Task III



The graph shows that there is sudden drop of error reaching global minimum of 0.81% where number of units of neurons are 200. After which there is gradual increase in the error with increase of number of neurons which could be possibly due to overfitting.

Question 3 Task VI



Question 3 Task VIII

Task VIII

Comparing the plot from Step III and VI, do you observe any relation between the reconstruction error and misclassification error? Explain your finding and add them to your PDF report

- It is evident from both the geaphs that classification error (Task VI) which was initially less but started increasing with minor fluctuations and it is safe to say it is highly unstable learning process where as Reconstruciton error (Task III) was initially very high but reduced to the global minimum and fluctuated very little when compared with the fluctuations of classification error

References

M.W. Mak, S. K. (2005, Jan 3). *Expectation-Maximization Theory*. Retrieved from informit.com: http://www.informit.com/articles/article.aspx?p=363730&seqNum=2