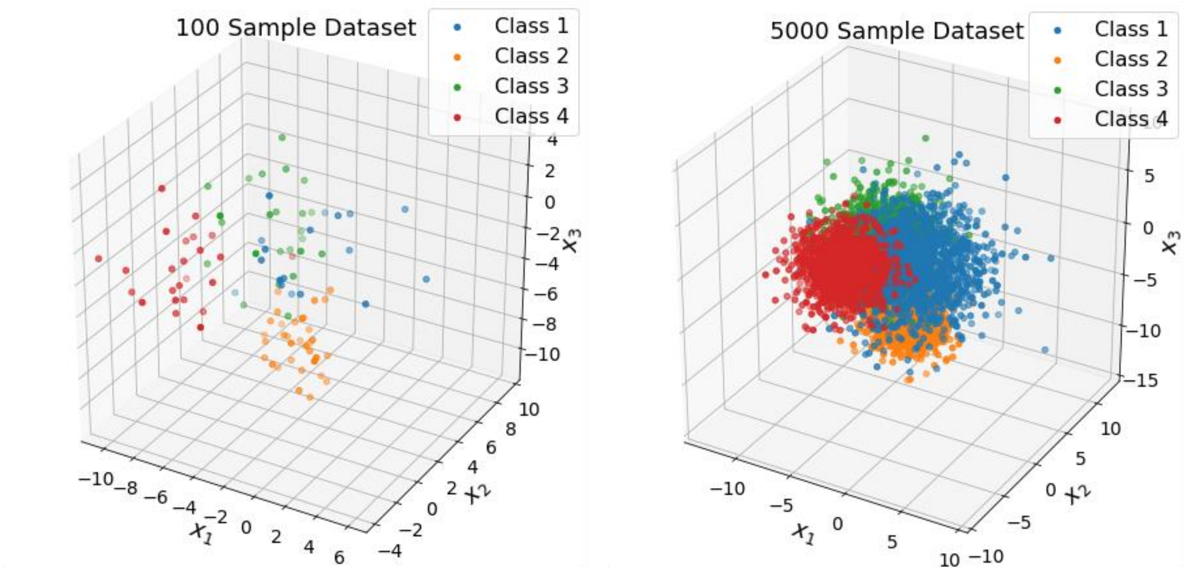
Tyler Cerilli
EECE 5644

Homework 2
https://github.com/tpcerilli/EECE-5644-Homeworks/tree/master/HW3

Question 1

Visualization of the data:



Confusion Matrix (rows: Predicted class, columns: True class):

```
[[22520     1   468  1025]
 [   28 21874   588  3999]
 [  631   345 22225  1359]
 [ 1611  2888  1621 18817]]
```

Total Number of Misclassified Samples: 14564
Empirically Estimated Probability of Error (MAP): **0.1456**

Training

For the multilayer perceptron model (MLP), we test from 5 to 19 neurons

100 Samples...
Best no. of neurons: 12
Probability of error (Training): 0.24

200 Samples...
Best no. of neurons: 13
Probability of error (Training): 0.205

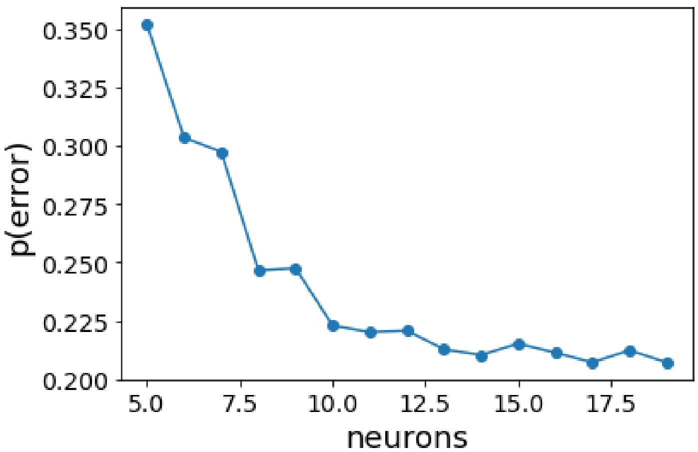
500 Samples...
Best no. of neurons: 16
Probability of error (Training): 0.196

<u>1000 Samples...</u>
Best no. of neurons: 15
Probability of error (Training): 0.19499999999999998

<u>2000 Samples...</u>
Best no. of neurons: 14
Probability of error (Training): 0.193

<u>5000 Samples...</u>
Best no. of neurons: 17
Probability of error (Training): 0.20719999999999997

After about 10 neurons, the probability of error begins to plateau on the training data (for the 5000 sample case):



Now we will test the optimal number of neurons (calculated on the training data) on the test data.

Testing

<u>100 Samples Trained evaluated on 100000 Sample Test...</u>
Number of neurons: 12
Probability of error (Test): 0.24431000000000003

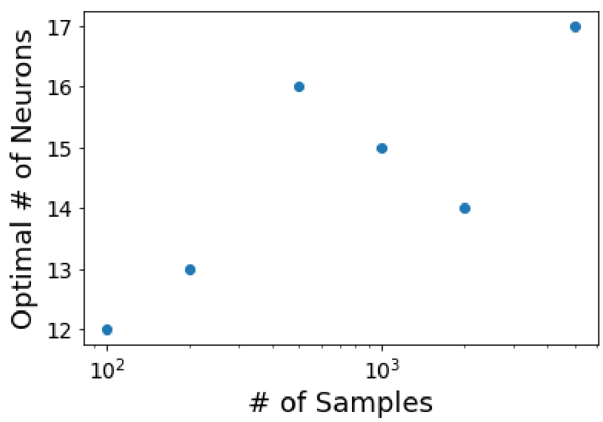
<u>200 Samples Trained evaluated on 100000 Sample Test...</u>
Number of neurons: 13
Probability of error (Test): 0.20889000000000002

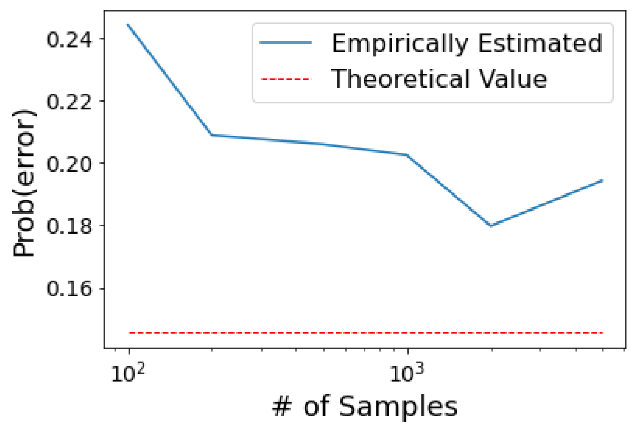
<u>500 Samples Trained evaluated on 100000 Sample Test...</u>
Number of neurons: 16
Probability of error (Test): 0.20604

<u>1000 Samples Trained evaluated on 100000 Sample Test...</u>
Number of neurons: 15
Probability of error (Test): 0.20252000000000003

<u>2000 Samples Trained evaluated on 100000 Sample Test...</u>
Number of neurons: 14
Probability of error (Test): 0.17979

<u>5000 Samples Trained evaluated on 100000 Sample Test...</u>
Number of neurons: 17
Probability of error (Test): 0.19428999999999996



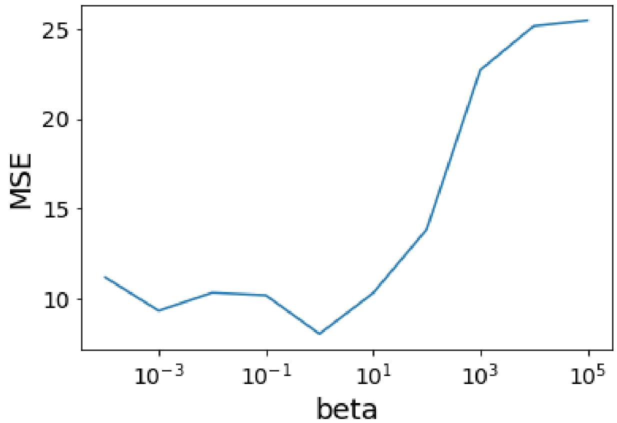


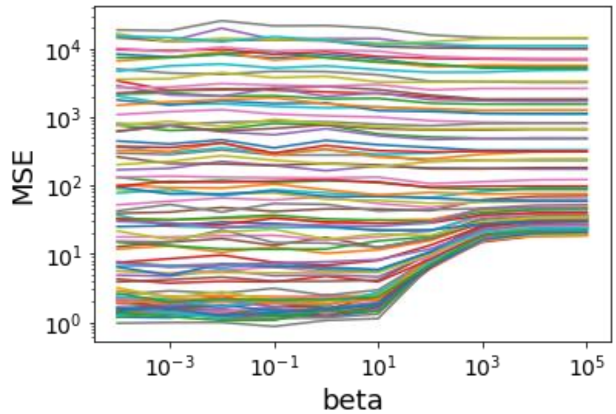
The optimal number of neurons does not appear to closely follow a particular trend, except that possibly a higher number of samples increases the optimal number of neurons in this case. Meanwhile, as expected, a higher number of training samples increases the performance (accuracy) of the model.
Also, the MLP never achieves the theoretical value of accuracy and misses by about 5%.

Question 2

mse of beta: 7.987717400527491

optimal beta: 1





As alpha increases, so does the MSE as seen on the figure on the right. The lower the line, the lower the alpha. However, it seems when alpha is very high (10^4), the increase in beta accounts for the noise and produces a better mean squared error.

Meanwhile an increase in beta with a fixed alpha, will increase the MSE with beta > ~10.

MAP ; adds prior

$$\hat{\theta}_{MAP} = \underset{\theta}{\arg\max} \sum \left[\log(p(x^{(i)}|\theta)) + \log(p(\theta))\right] \rightarrow p(\theta) = \mathcal{N}(\theta \,|\, \overset{\mu}{0}, \overset{\sigma}{\lambda^{-1} I})$$

$$PNLL(\theta) = NLL + \lambda \sum_{j=1}^{n} \theta_j^2$$
$$= NLL + \lambda \theta^T \theta$$
$$+ \lambda \|\theta\|_2^2$$

$$\hat{\theta}_{MAP} = \underset{\theta}{\arg\max} \Sigma\left[\log(p(x^{(i)}|\theta)) + \log(p(\theta))\right]$$
$$(X\theta - y)^T (X\theta - y)$$
$$\log(p(\theta)) = \mathcal{N}(\theta | 0, \beta I)$$
$$= \beta I \tilde{x}^T y$$

$$\boxed{\hat{\theta}_{MAP} = (\tilde{X}^T \tilde{X} + \beta I)\, \tilde{x}^T y}$$