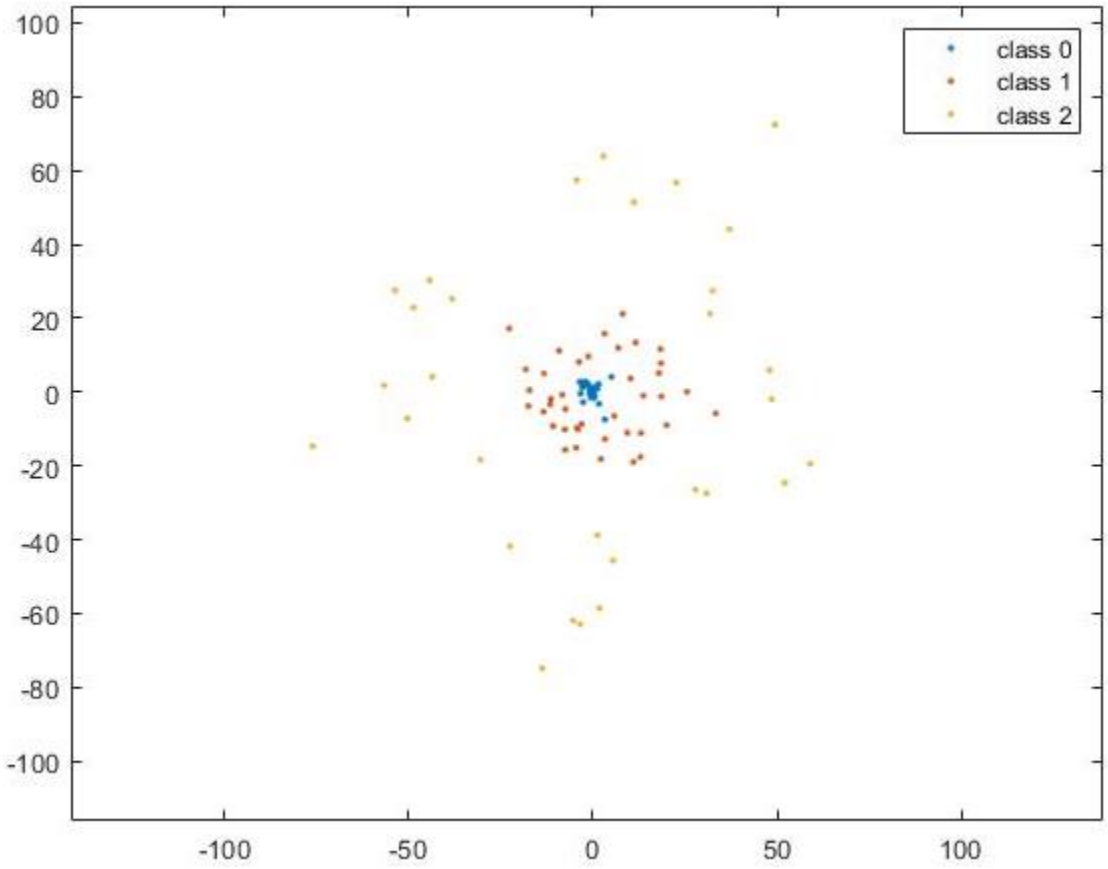
Question 1:

The problem is to classify 2D real vectors into 3 classes using Multi-layer Perceptrons with two fully connected layers with adaptive wights followed by a softmax layer. The outputs(3 in this case) approximate the class posteriors. As per MAP classification rule, the class of input sample is decided by choosing the class corresponding to the greatest class posterior given by the model.

There are 24 competing models (1-12 perceptrons in first layer and, sigmoid and elu activation function). Each model is trained with an objective to minimize the cross-entropy loss. 10-fold cross validation is performed for model selection. In every fold, the dataset is split for training and validation, the models are trained to minimize cross entropy loss and the ratio of number_of_correct_classifications/total_no_samples is calculated and this value is averaged for all 10 folds. This gives the performance measure for every competing model. The model with greatest performance measure is selected and again trained with entire training data samples.(If many models have same maximum performance measure, the model with a smaller number of perceptrons is elected in order to reduce the model parameters)

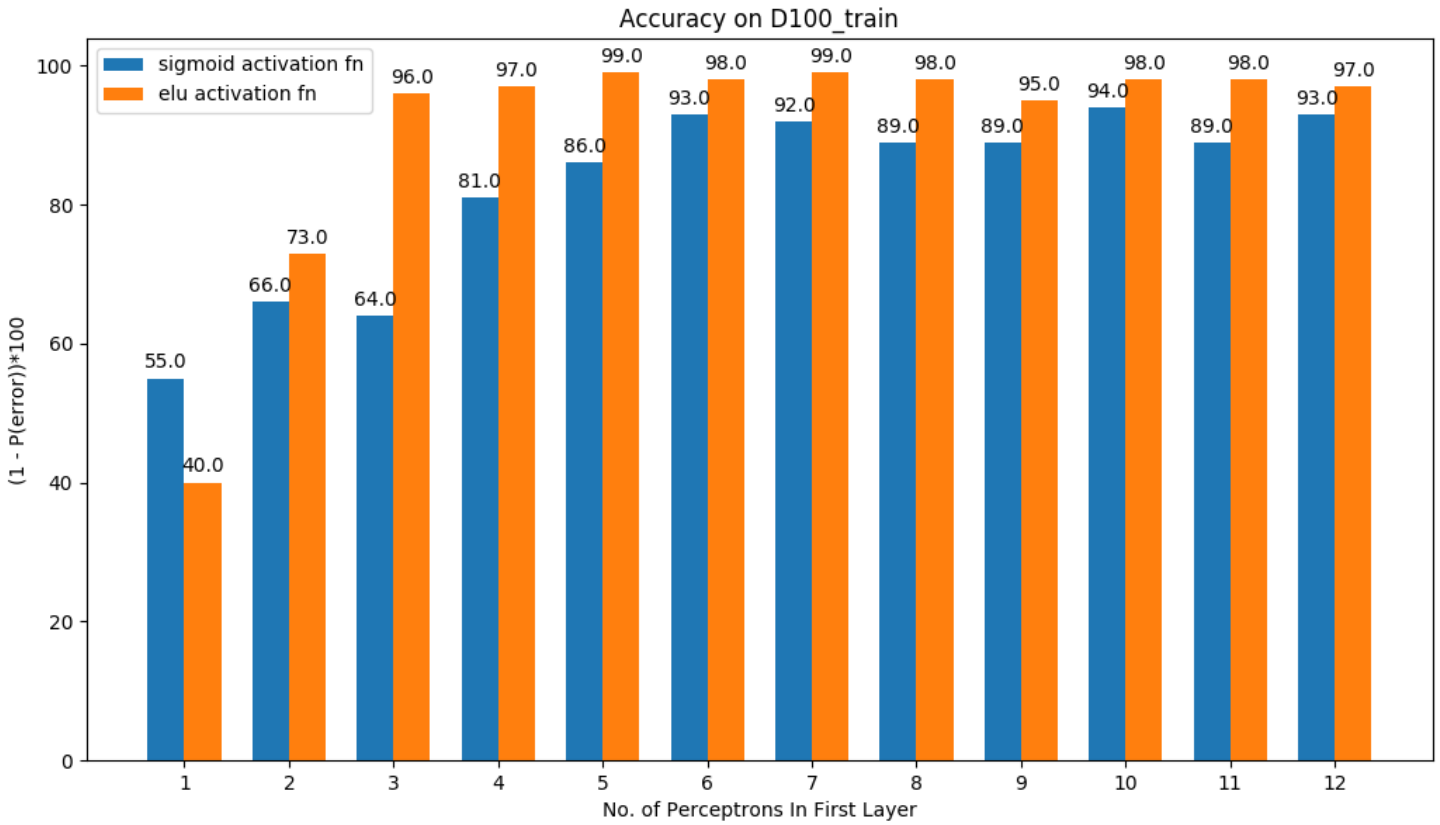
We have three training datasets with 100, 500 and 1000 samples each. 10-fold cross validation is performed for these three datasets. Now we have three models which are evaluated with 10,000 validation samples. For every model, the ratio of number_of_correct_classifications/total_no_samples is calculated.

a) FOR D100_train:



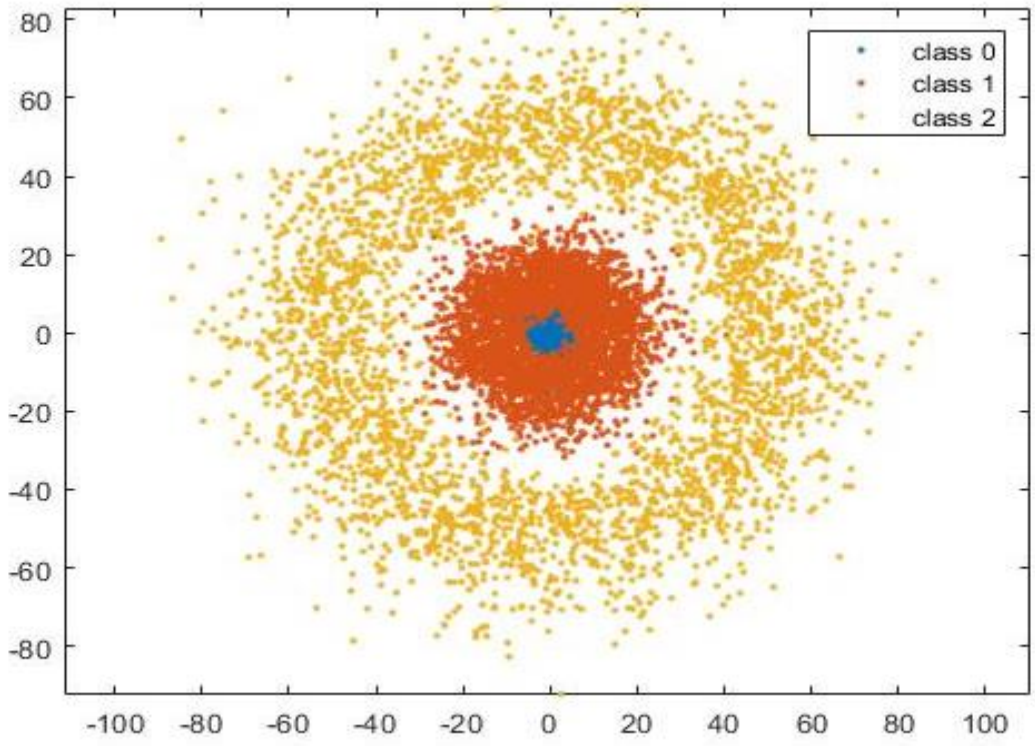
The above image shows the training dataset.

10-fold cross-validation results:



Selected Model: elu activation function with 5 perceptrons in first layer. **Corresponding Performance Measure:** 0.99

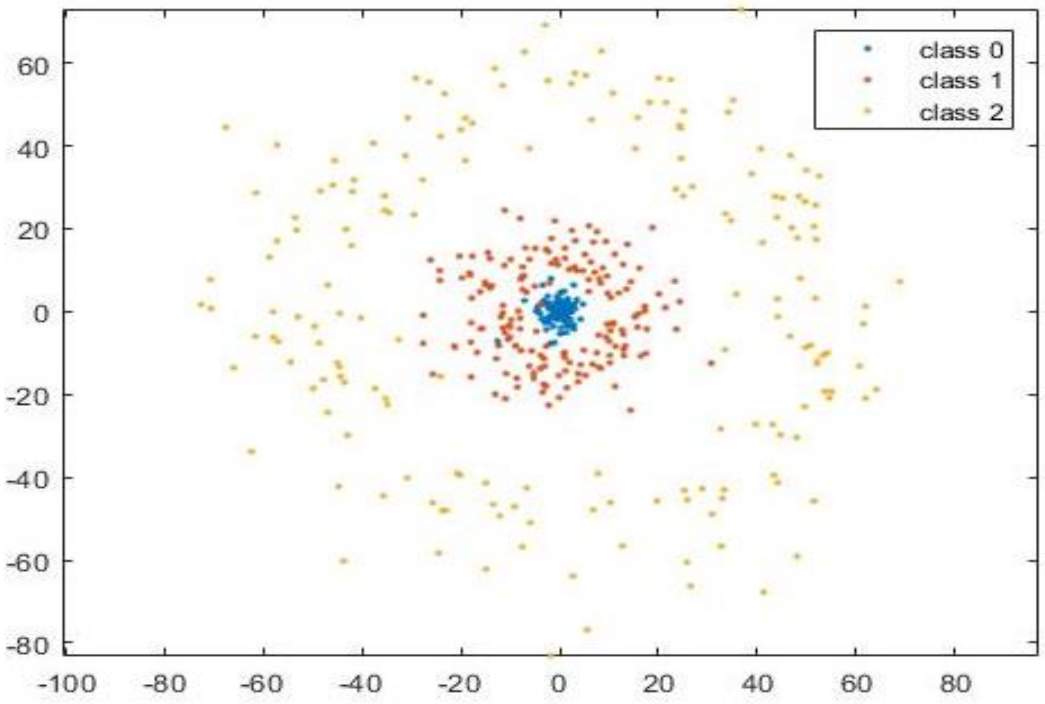
The following image shows the 10,000 validation samples used.



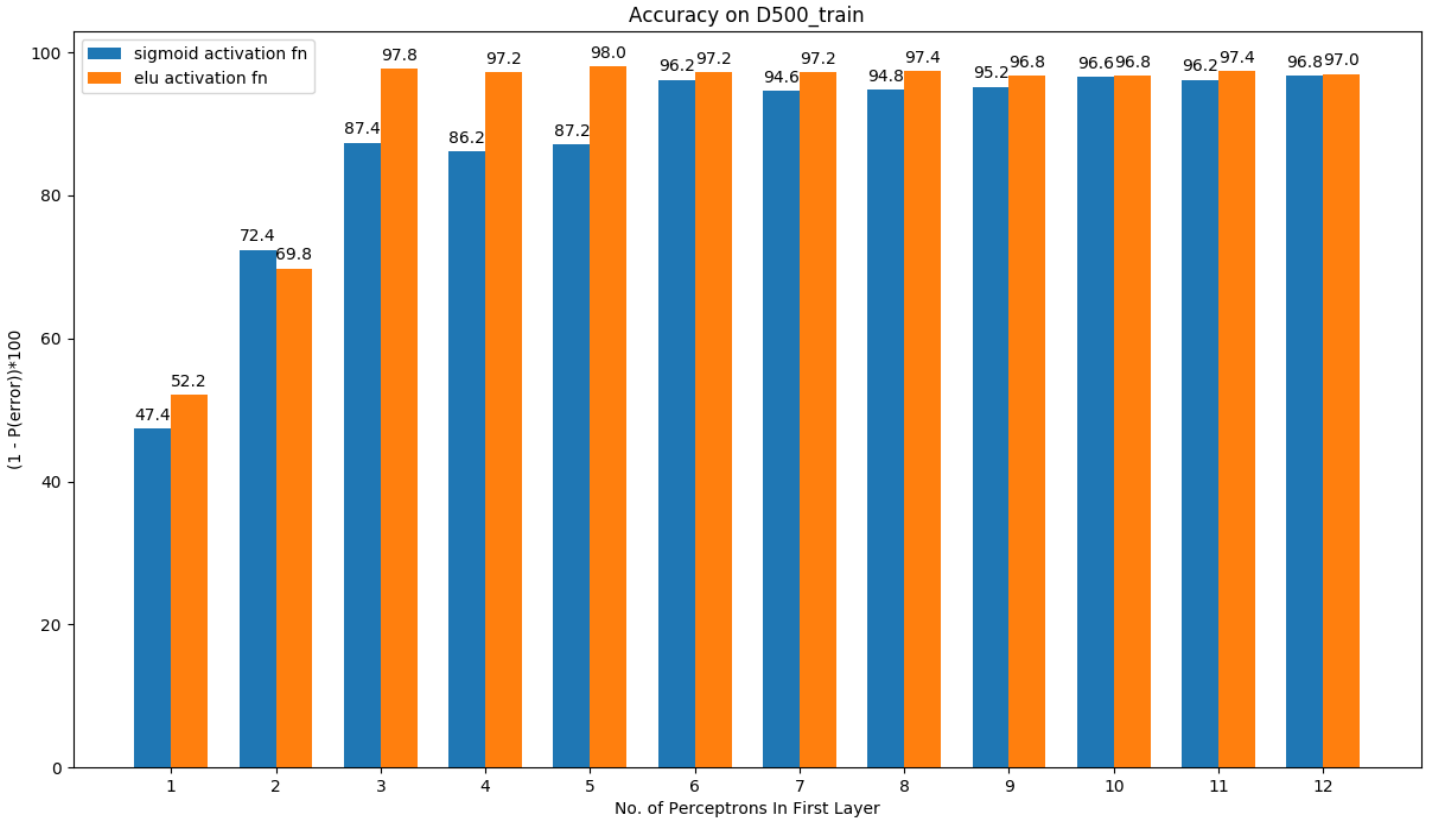
The selected model is trained using the entire 100 samples in the training set. This trained model is evaluated using the above shown validation data. The **Probability of Error** is **0.0367.**

b) FOR D500_train:

The below image shows the training dataset.



10-fold cross-validation results:

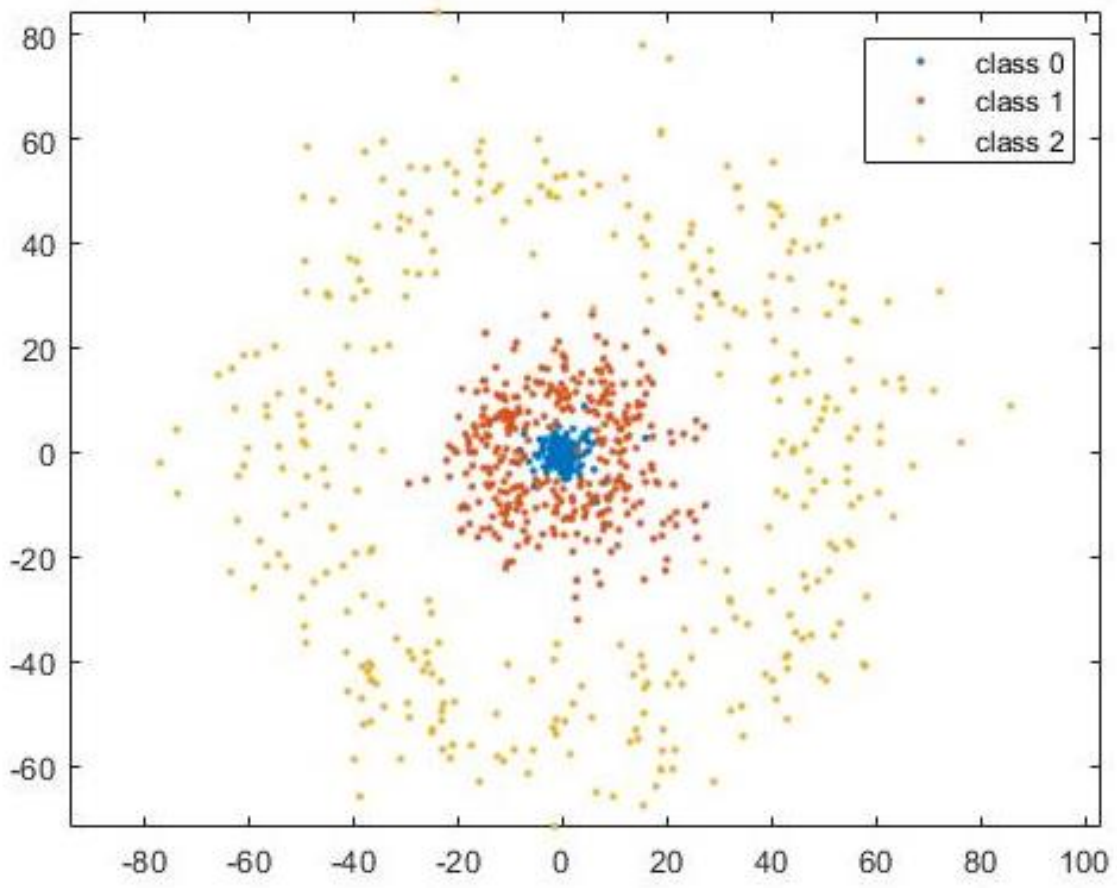


Selected Model: elu activation function with 5 perceptrons in first layer. **Corresponding Performance Measure:** 0.98

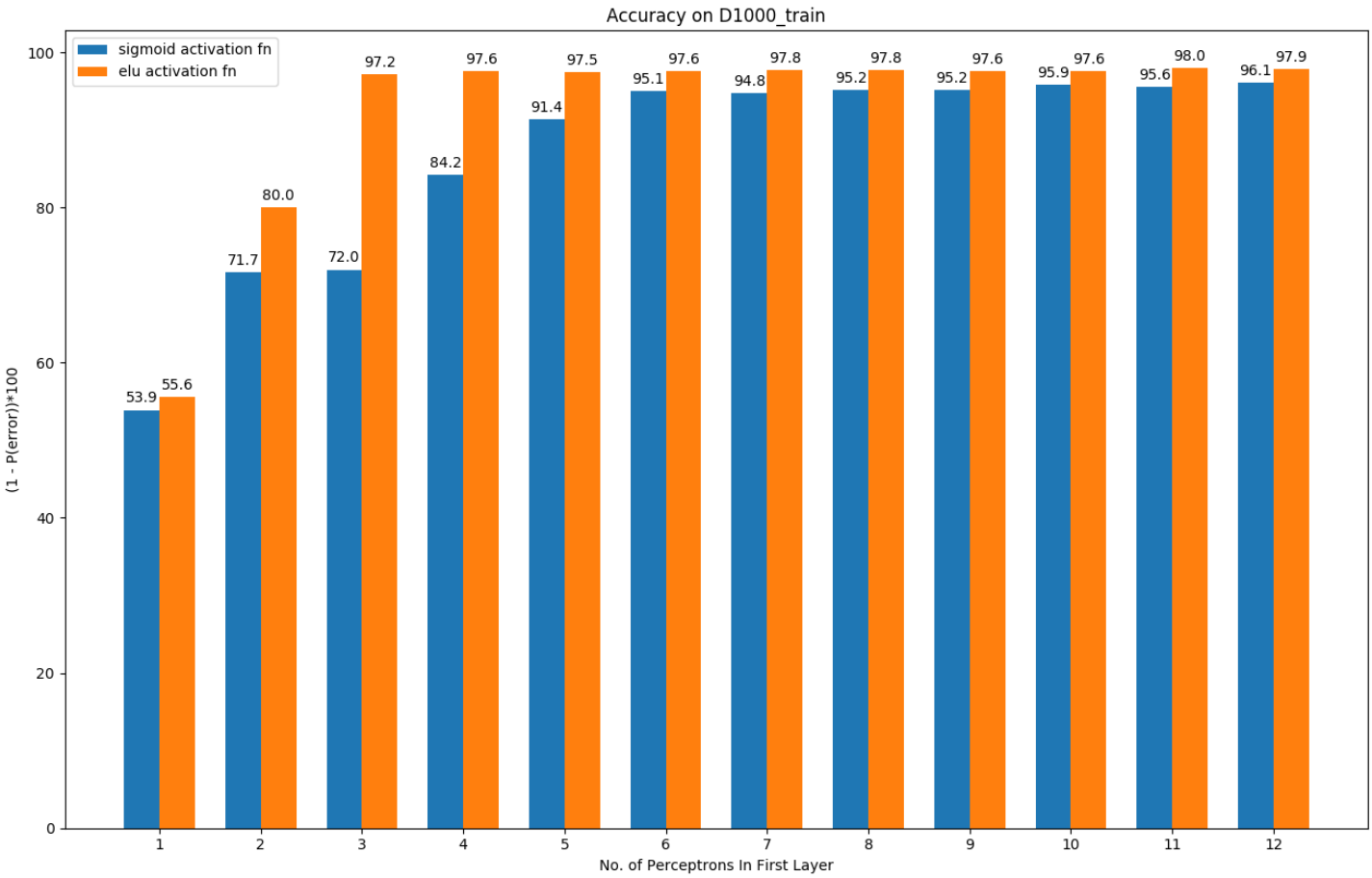
The selected model is trained using the entire 500 samples in the training set. This trained model is evaluated using the above shown validation data. The **Probability of Error** is **0.0266.**

c) FOR D1000_train:

The below image shows the training dataset.



10-fold cross-validation results:



Selected Model: elu activation function with 7 perceptrons in first layer. **Corresponding Performance Measure:** 0.978

The selected model is trained using the entire 1000 samples in the training set. This trained model is evaluated using the above shown validation data. The **Probability of Error** is **0.028.**

Summary:

	Selected Activation fn	**Selected no. of perceptrons**	**P(error) on Dval**
D100_train	elu	5	0.0367
D500_train	elu	5	0.0266
D1000_train	elu	7	0.028

Question 2:

The problem is to classify 2D real vectors into 3 classes using MAP classifier where the class conditional PDFs are modeled as Gaussian Mixture Models.

10-fold cross validation is performed to select an optimal model order for the class conditional PDFs. The GMM parameters are estimated using Expectation-Maximization. The model(for class conditional PDFs) which has maximum average log-likelihood for validation data is selected in the K-fold step. Then the selected model's parameters are estimated using EM using the entire training data.

The class prior is calculated by counting the number of samples from the training data. The class posterior is proportional to the product of class conditional PDF and corresponding class prior. As per MAP rule, the class corresponding to the greatest product is selected.

a) For D100_train

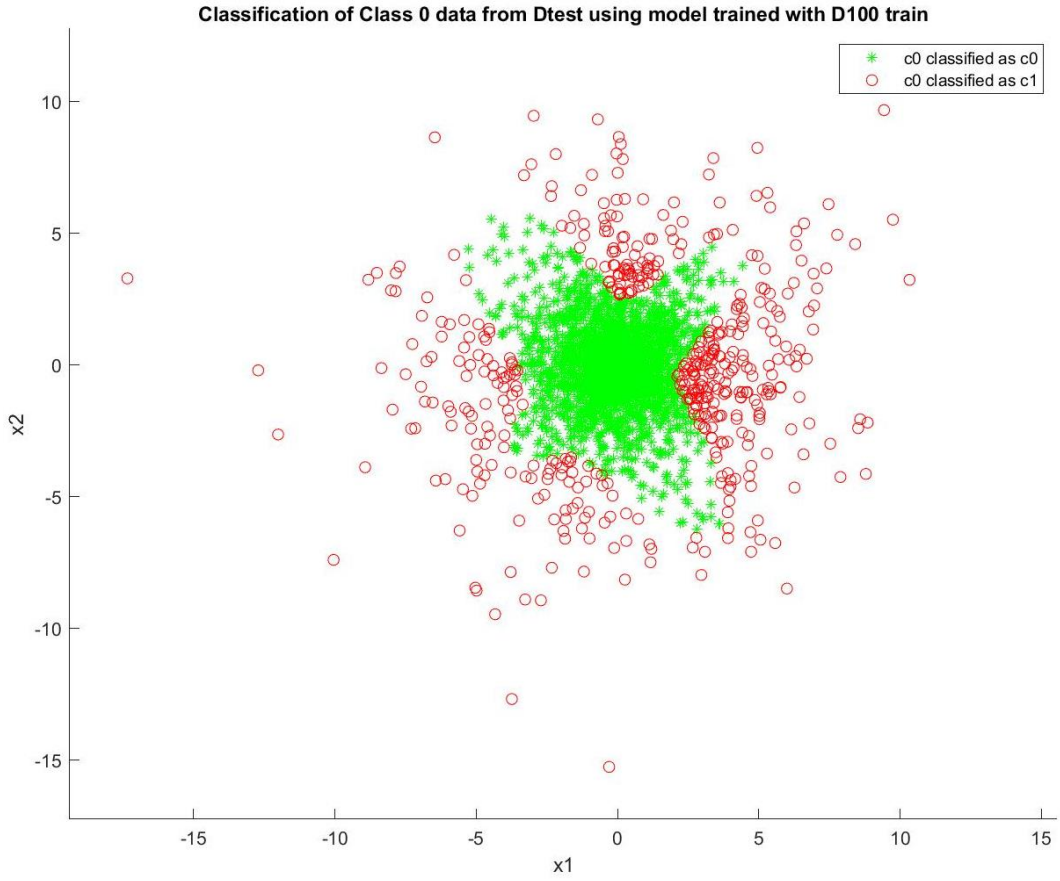
The 10-fold cross-validates selected the following model orders for class conditional PDFs.

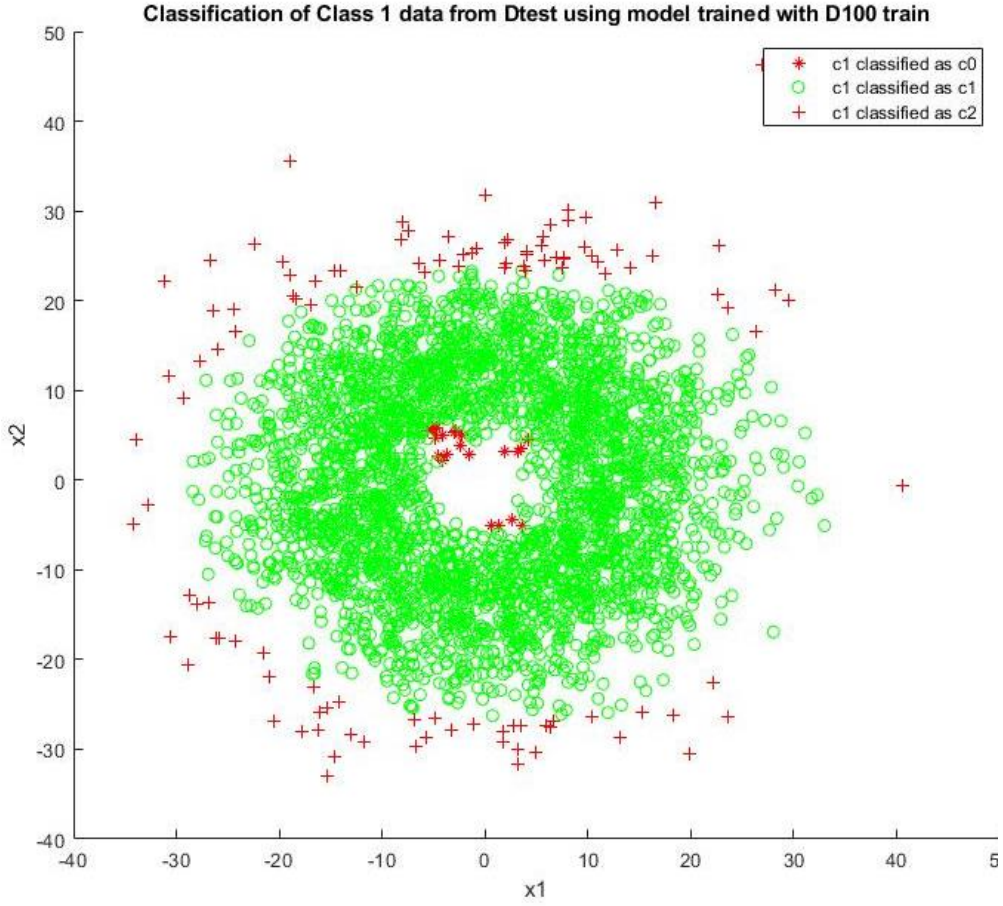
	Selected Model Order
Class 0	2
Class 1	1
Class 2	1

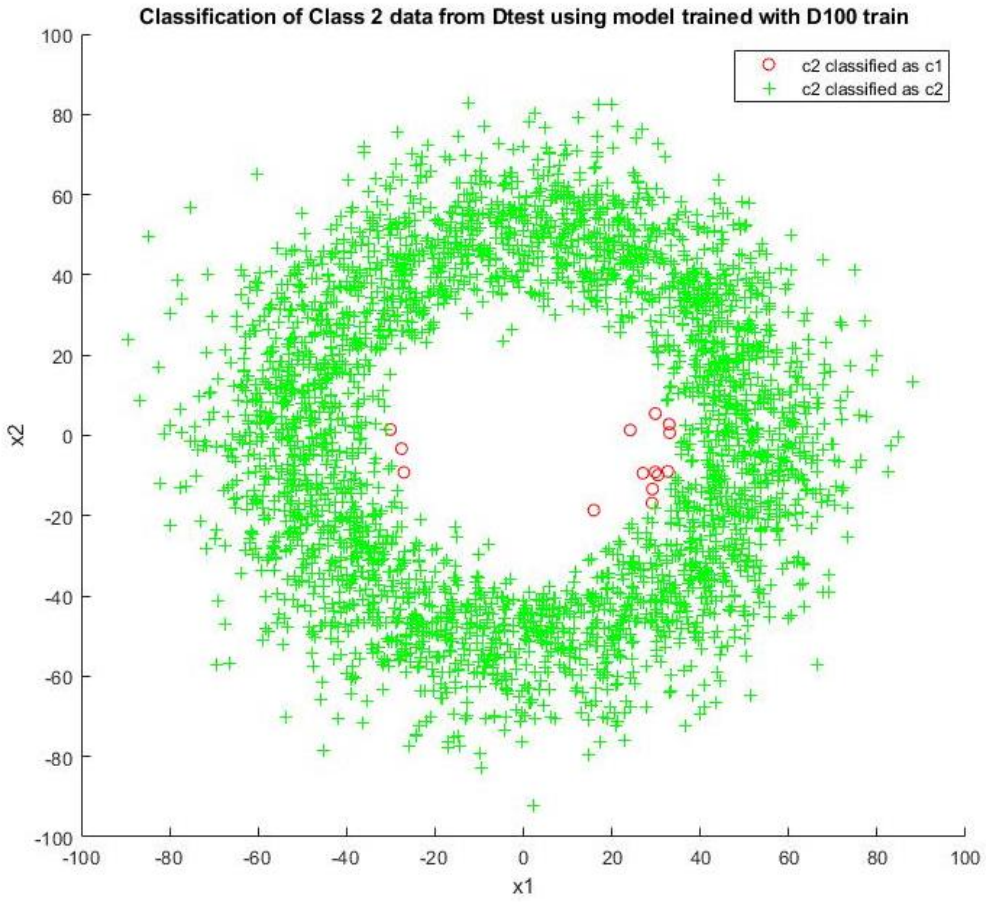
The 10,000 validation samples are classified using our selected models and MAP rule.

		Decision		
		0	1	2
True Label	0	2918	473	0
	1	22	3158	114
	2	0	14	3301

Probability of error = 0.0623


Classification of Class 0 data from Dtest using model trained with D100 train
c0 classified as c0
c0 classified as c1
x1
x2


Classification of Class 1 data from Dtest using model trained with D100 train
c1 classified as c0
c1 classified as c1
c1 classified as c2
x1
x2



b) For D500_train:

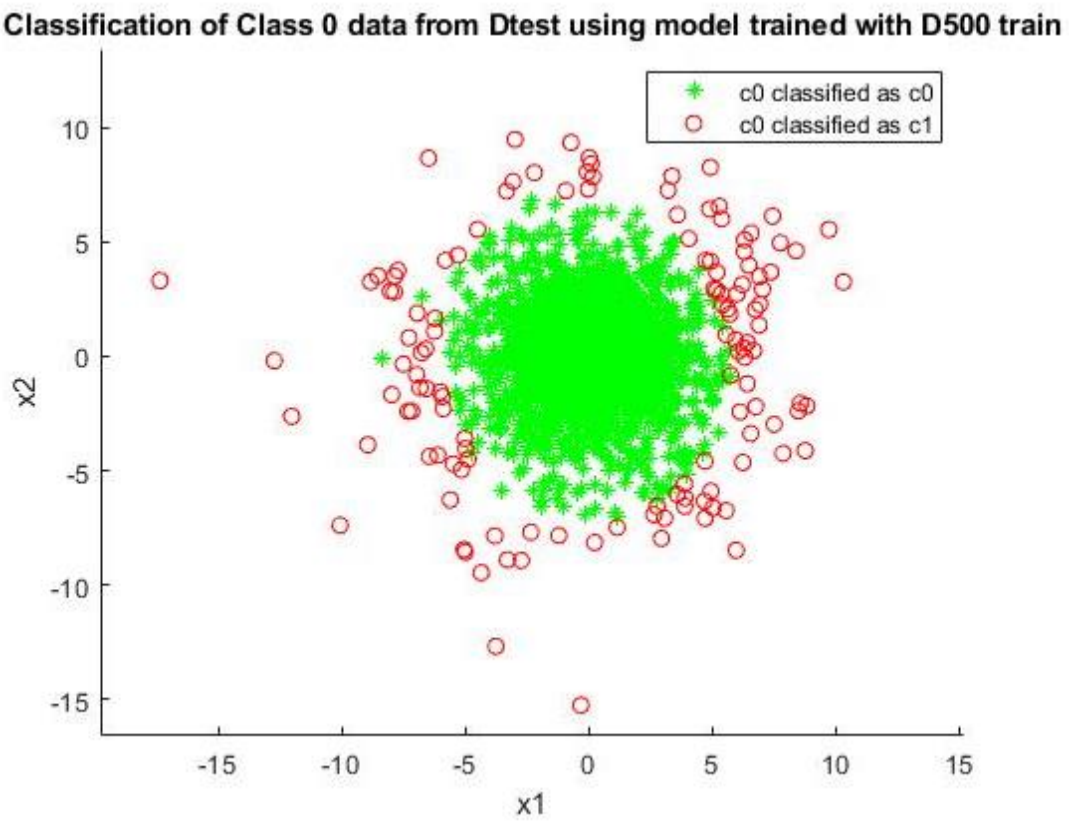
The 10-fold cross-validates selected the following model orders for class conditional PDFs.

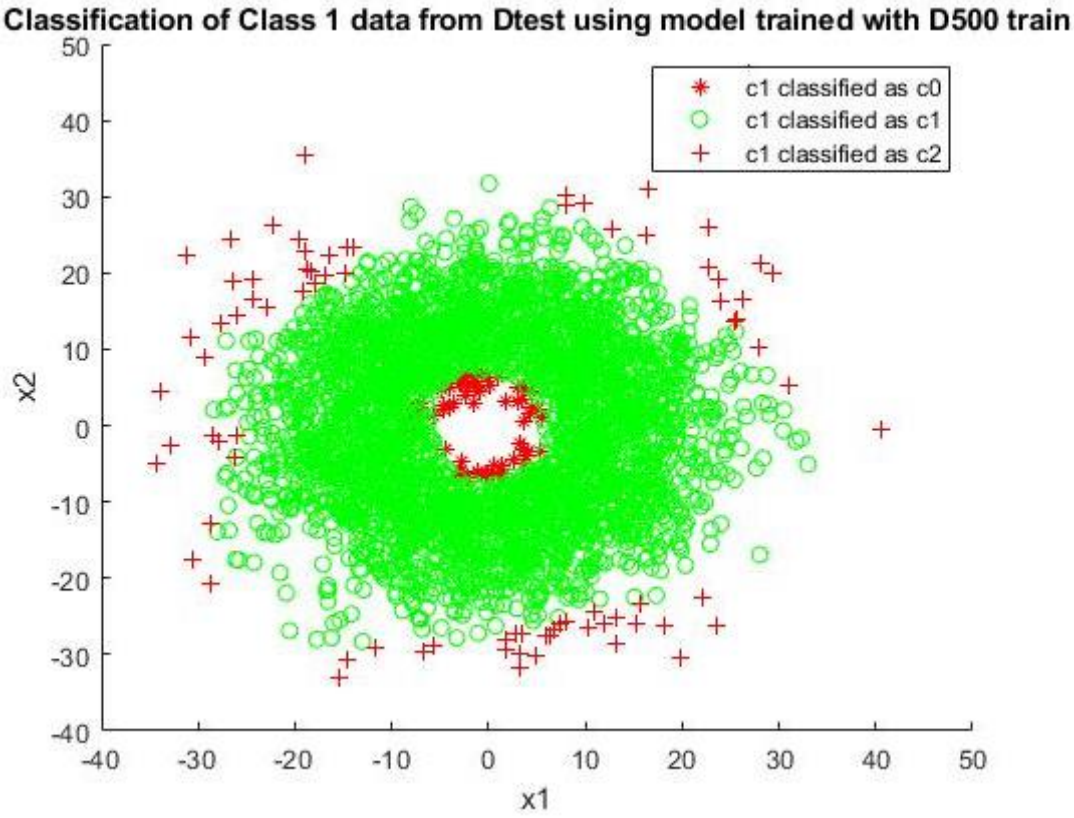
	Selected Model Order
Class 0	3
Class 1	4
Class 2	4

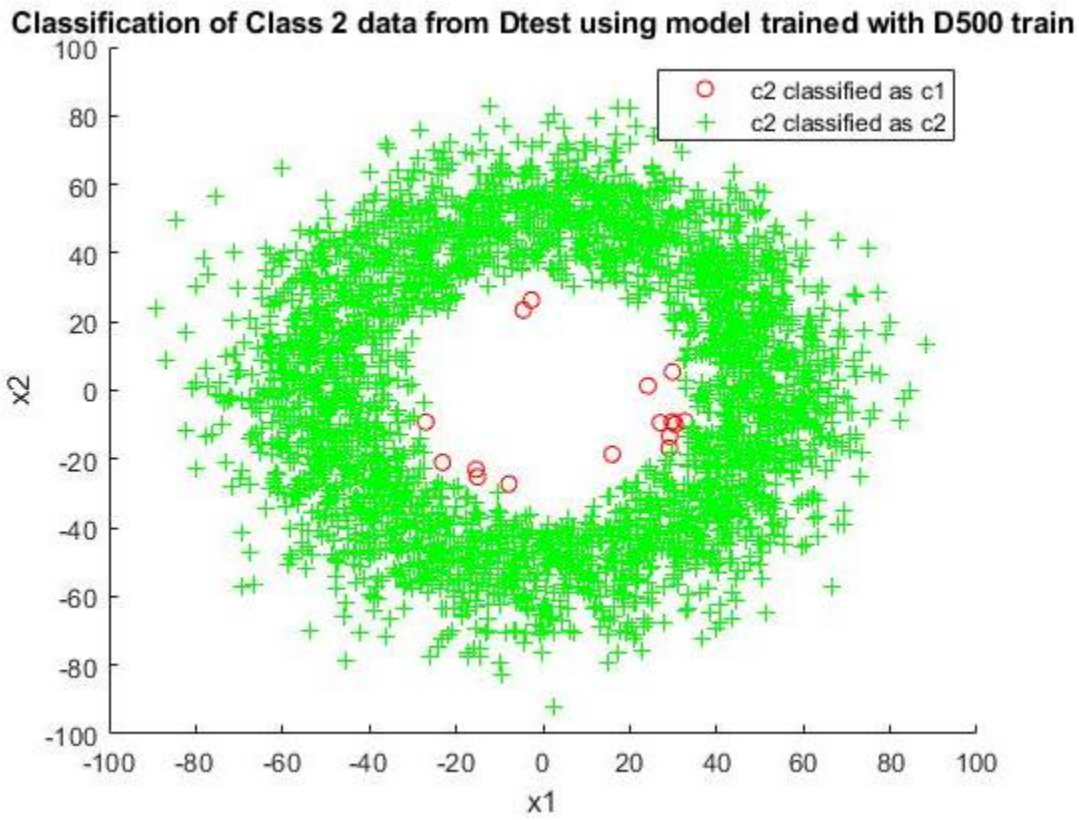
The 10,000 validation samples are classified using our selected models and MAP rule.

		Decision		
		0	1	2
True Label	0	3260	131	0
	1	66	3148	80
	2	0	16	3299

Probability of error = 0.0293


Classification of Class 0 data from Dtest using model trained with D500 train
c0 classified as c0
c0 classified as c1
x1
x2


Classification of Class 1 data from Dtest using model trained with D500 train
c1 classified as c0
c1 classified as c1
c1 classified as c2
x1
x2



c) For D1000_train:

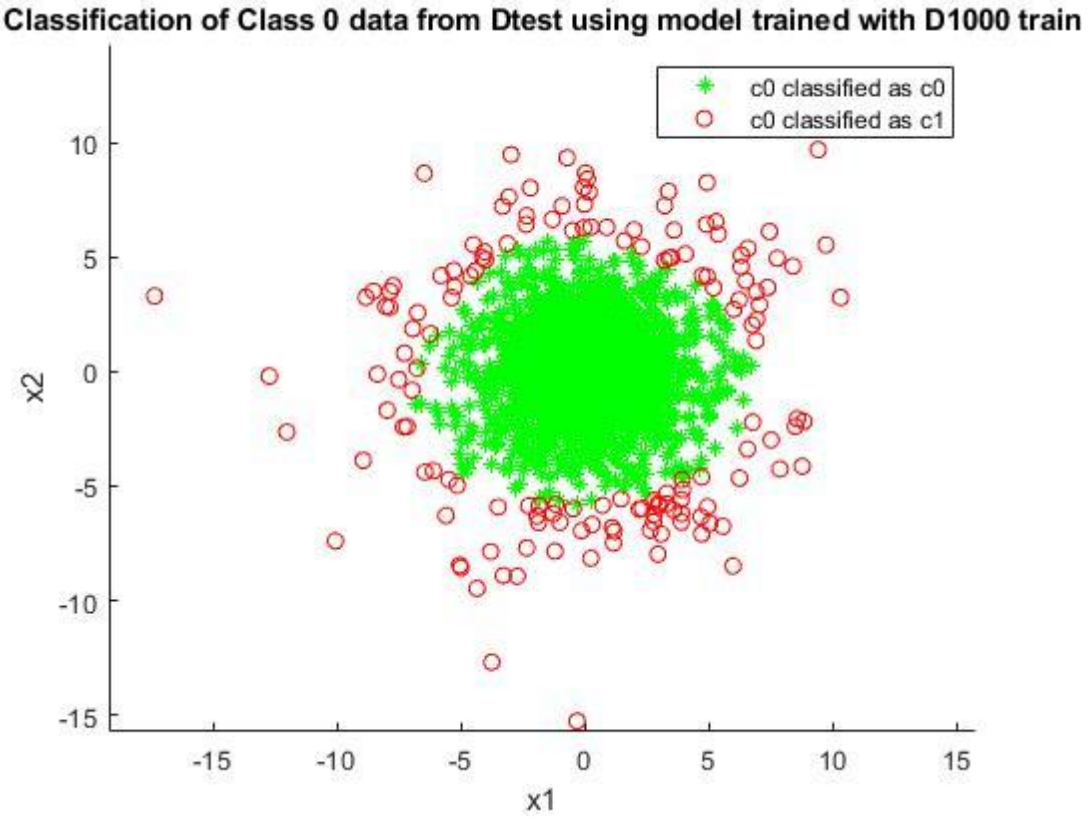
The 10-fold cross-validates selected the following model orders for class conditional PDFs.

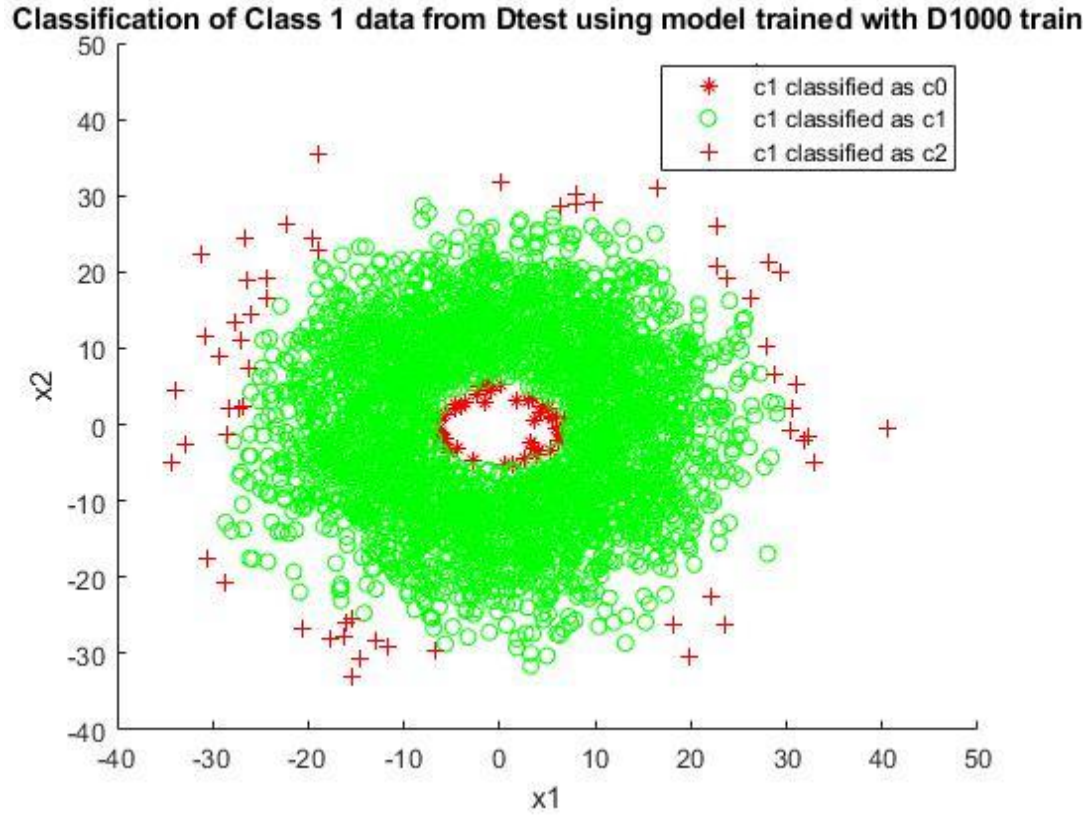
	Selected Model Order
Class 0	4
Class 1	4
Class 2	5

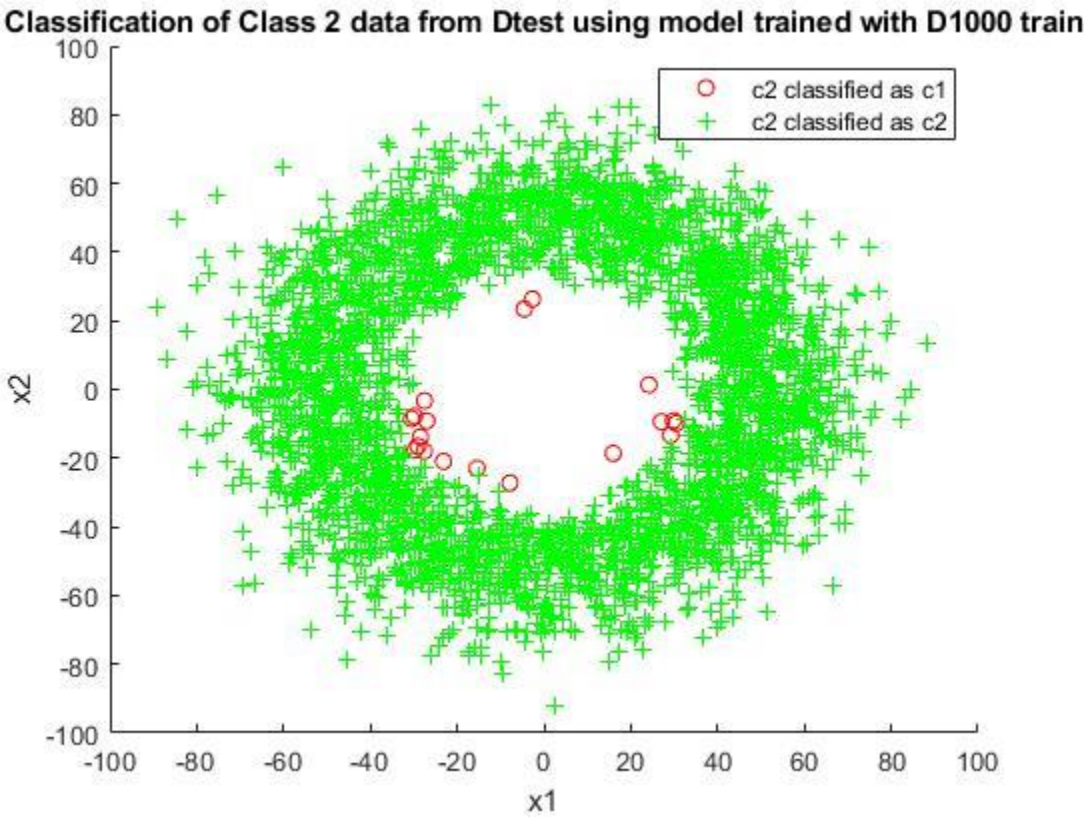
The 10,000 validation samples are classified using our selected models and MAP rule.

		Decision		
		0	1	2
True Label	0	3238	153	0
	1	46	3188	60
	2	0	19	3296

Probability of error = 0.0278


Classification of Class 0 data from Dtest using model trained with D1000 train
c0 classified as c0
c0 classified as c1
x1
x2


Classification of Class 1 data from Dtest using model trained with D1000 train
c1 classified as c0
c1 classified as c1
c1 classified as c2
x1
x2



All code can be found in https://github.com/veeraragav/EECE5644_Into_To_ML/tree/master/Assignment_3