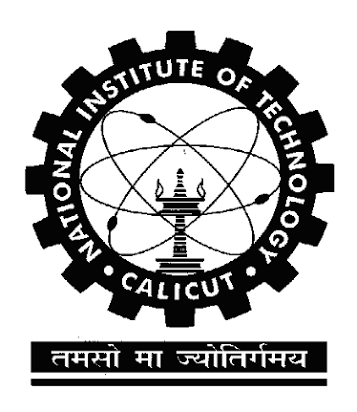
**National Institute of Technology**

**Calicut**



**PATTERN RECOGNITION AND ANALYSIS**

**REPORT**

**ASSIGNMENT NO-5**

**SUBMITTED BY:**

Bojanapally Veda Sree B150734EC

**Assignment 5a:** Analysis on MLP

Fit using MLP, with varying number of hidden-units.

(a) Define three 2 dimensional guassians: G1, G2 and G3 (use matlab function: mvnrnd)

(b) G1 has μ1 = 0.25, 0.3 and Σ1 = 0.2, 0.25; 0.25, 0.4

(c) G2 has μ2 = 0.5, 0.6 and Σ2 = 0.3, 0.1; 0.1, 0.4

(d) G3 has μ2 = 0.7, 0.75 and Σ2 = 0.3, 0.1; 0.1, 0.4

(e) Generate 100 data points from all three.

(f) Use first 80 data points from each as training set and the next 20 from each as your test set.

(g) So you now have 2-dimensional inputs, with 3 classes. Train a MLP network with 3 hidden

units.

(h) What is the training accuracy? What is the testing accuracy?

How does variation in the number of hidden units affect accuracy?

(a) Vary the number of hidden-units to 2 and 4.

(b) How does the accuracy value change? Is it more, less?

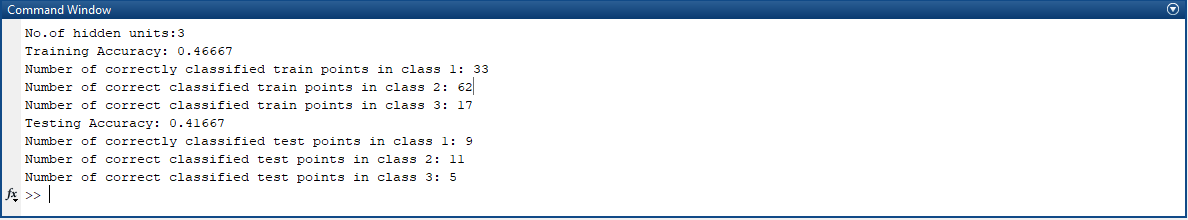
(c) Vary it to 10, how does the testing accuracy change?

**Procedure:**

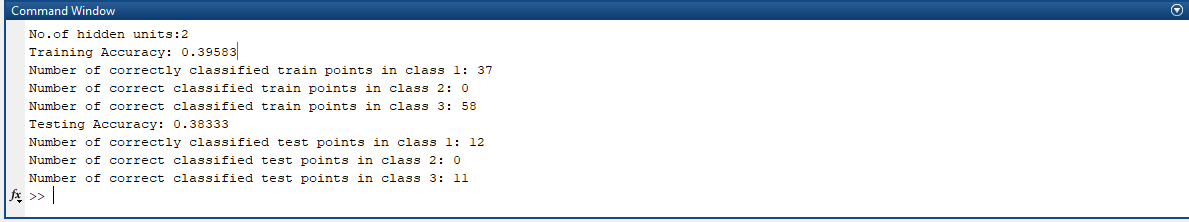
* Three 2D Gaussians of given mean and variance were defined and 100 data points are generated from each of them using mvnrnd() function.
* First 80 data points of each Gaussian is concatenated and taken as the training set and the remaining 20 is concatenated and taken as test set.
* Number of inputs, hidden units and outputs are given as n\_ip, n\_hu and n\_op respectively.
* The weight matrix of first layer(w1) of dimension (n\_ip x n\_hu) is initialized randomly using rand() function. Similarly the weight matrix of output layer of dimension (n\_hu x n\_op) is initialized.
* All the training points are passed through the network for forward pass and then back propagation is done and this is repeated for 100 times for more accuracy.
* For the training points belonging to the first class, target vector is taken as [1 0 0] similarly [0 1 0] for class 2 and [0 0 1] for class3.
* A sigmoid function is defined as the activation function.
* The weights are then updated by finding the derivatives of the error at the output with respect to the corresponding weight and thus back propagation is carried out.
* For each of the training point, outputs are calculated and if maximum output is for the class to which it actually belongs to, count for that class accuracy is increased by 1.
* Training accuracy is calculated as total number of correct classifications divided by total number of train points.
* For each of the test point, outputs are calculated and if maximum output is for the class to which it actually belongs to, count for that class accuracy is increased by 1.
* Testing accuracy is calculated as total number of correct classifications divided by total number of test points.
* The above procedure is repeated for different number of hidden units and accuracies are calculated.

**Observations and results:**

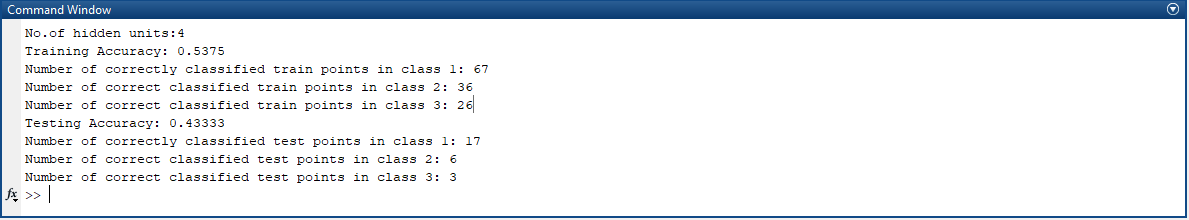
Taking 3 hidden units



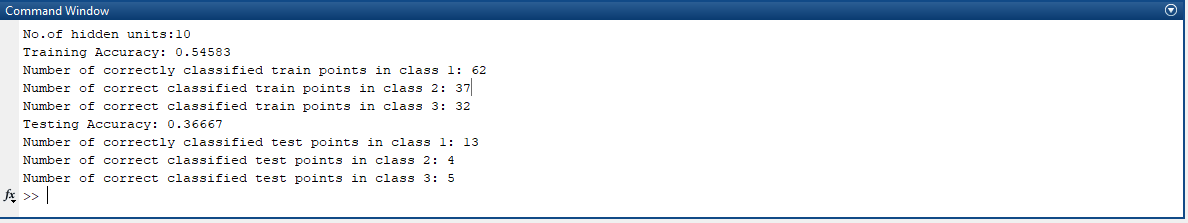
Taking 2 hidden units



Taking 4 hidden units



Taking 10 hidden units



We observe that as the no. of hidden units are increased from 2 to 3 to 4, both the training and testing accuracy increases. But, when the no. of hidden units are increased to 10, the training accuracy increases but the testing accuracy is reduced because of overfitting.

**Assignment 5b:** Application of MLP for classification.

(a) Define three 4 dimensional gaussians.

(b) Generate 100 data points for each class using an appropriate co-variance matrix and means.

(c) Use FLD to reduce to 3 dimensions.

(d) Classify using a 2-layer MLP with 3 hidden units.

(e) What is the training and testing accuracies?

(f) How does the accuracy values compare with Euclidean classification you did before?

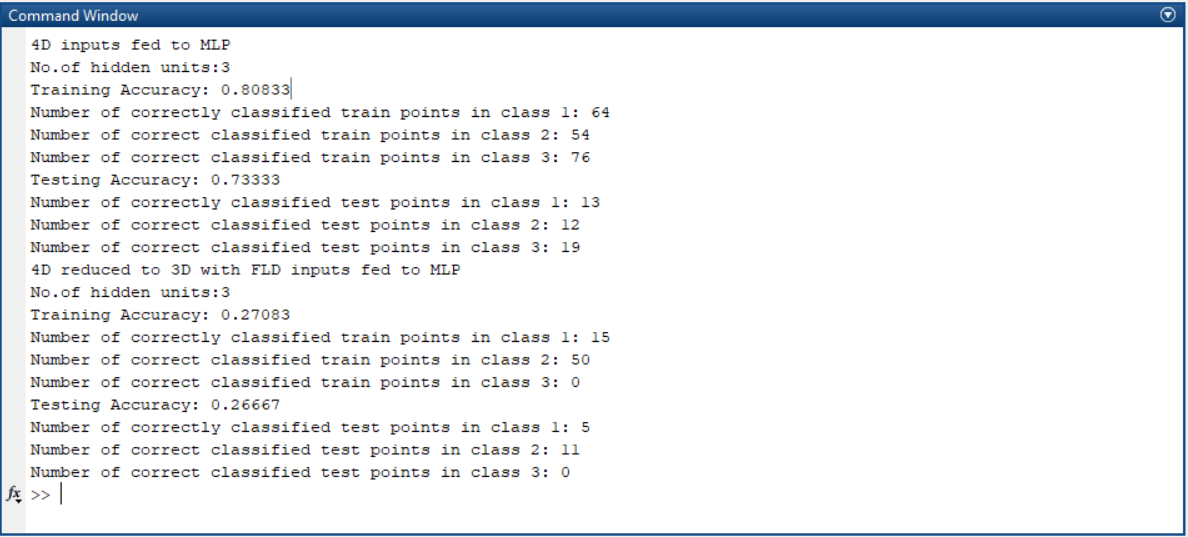
**Procedure:**

* Three 4D Gaussians were defined and 100 data points generated from them using mvnrnd() function.
* Procedure for previous question is repeated for getting accuracy in 4D.
* FLD was used to reduce the dimension from 4 to 3 similar to the procedure in assignment 4.
* For this 3D data, the above procedure is repeated(forward pass and back propagation through MLP) and accuracies computed.
* A comparison of accuracy obtained using Euclidean distance (as in previous assignment) and MLP is done.

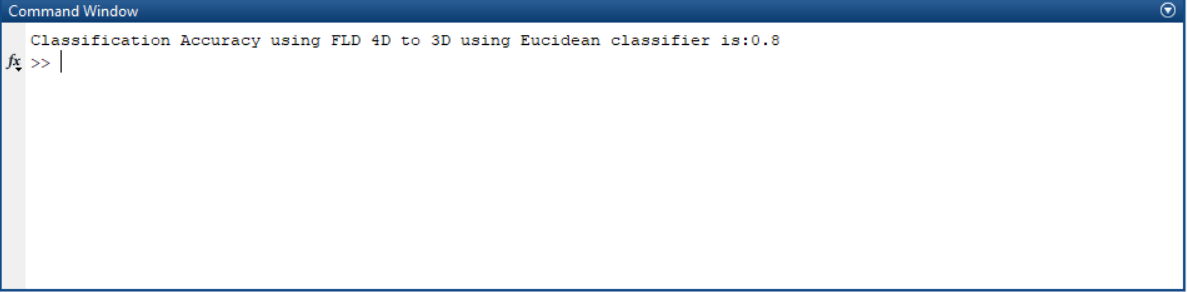
**Observations and results:**

Training and testing accuracies for

* 4D Gaussians fed to MLP and
* Reduced from 4D to 3D using FLD and fed to MLP

****

Training and testing accuracies with Euclidean classification



The accuracies are found to be higher when 4D data is directly fed to MLP when compared to Euclidean or when reduced to 3D.