In this assignment you are supposed to build the bayesian classifiers and mixture models for datasets assigned to your group. This assignment is more focused on analysis of classification techniques and getting used to handling data in Machine Learning instead of getting classification accuracy (as you will be getting pretty well accuracy in most of the cases).

Task 1:

Bayesian Classifiers: Builds Bayesian model using the given dataset for the given case numbers.

CaseNumber: 1 -- Bayes with Covariance same for all classes

2 -- Bayes with Covariance different for all classes

3 -- Naive Bayes with C = \sigma^2*I.

4 -- Naive Bayes with C same for all classes.

5 -- Naive Bayes with C different for all classes.

Datasets Details:

Dataset for each group is specified here.

1. Linearly separable data:

2-dimensional artificial data of 3 or 4 classes.Some datasets have four classes and some have three classes. Work with the one that you have got.

Each class has 500 data points (First 500 is class1, next 500 is class 2 like that).

2. Non Linearly seperable data:

Each group has three classes. It is in the fromat "groupNo_classNumber.txt" (i.e. 1_c1.txt).

3. OverLapping data:

Each class has 500 data points (First 500 is class1, next 500 is class 2 like that).

4. Real Data:

Real world data of 3 classes. The real world data sets correspond to the formant frequencies F1 and F2 for vowel utterances.

Each file contains 1500 points, (First 500 is class1, next 500 is class 2 like that). The File name in format "group_groupNumber.txt".

Divide the dataset into training, validation and testing. For better tuning of the parameters, you can perform k-cross validation.

Note: Each batch of students must use the datasets identified for that batch only.

You are supposed to make two functions for training and testing. The declaration of both the functions are given here. Please follow it "strictly".