

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).

Build Gaussian Mixture Models.

The task is to perform classification by modeling the the class conditional density as Gaussian Mixture Model for the following datasets:

- a) Image dataset
- b) Digit dataset
- c) Handwritten dataset
- d) Spiral dataset

Image Data set : 23-dimensional image data corresponding to different scenes. Every group is provided with the data for three scenes. Please check [here](#) to find the mapping for your group. The image folders are provided so that you can see the images on which you are working. You are required to work on the features extracted (which can be found in the corresponding feature files). The 23 dimensional features include color histogram, edge directed histograms and Entropy of wavelet coefficients extracted for local blocks of an image for a particular scene. Multiple images for a particular scene are provided in separate files inside the respective folders corresponding to different classes. Dataset can be downloaded from [here](#).

Digit Dataset : 39-dimensional speech data corresponding to isolated digit utterances. Every group is provided with the data for five digits. The information about the digits provided for each group can be obtained in: [group_digit_mapping](#). The 39 dimensional MFCC features extracted frame by frame from utterances for a particular digit by multiple people are provided in separate files inside the respective folders corresponding to classes. Data for digits can be downloaded from [here](#).

Handwritten Characters

The dataset for handwritten characters can be found [here](#). The features are the pixels of the stroke.

Spiral Dataset

The spiral dataset can be found here: [Spiral Data](#). The first 500 points are one class and the rest are other class.

Report Submission - Minimum Requirements:

The report should include the following things for appropriate tasks.

Experiment for both Diagonal and Full covariance matrices.

Write your own Bayesian classifier and GMM codes.

Decision Boundaries and Decision Surfaces (2D and 3D).

Plot ROC, DET curves, Confusion Matrices with accuracies.

Plot of GMM means with data points for Spiral Data.