Indian Institute of Technology Tirupati

Course: EE5106: Pattern Recognition & Machine Learning

Programming Assignment_3

Q.1

Choose one dataset from the datasets that have been uploaded. Now Using ML estimation estimate the parameters μ and ∑ for each class and then classify the test data by using likelihoods. Do cross Validation on the dataset and report accuracy and show the confusion matrix for each case.

Q.2

Repeat Q.1 by using posteriors instead of likelihoods.

****

What is Cross Validation?

Assume that you are dealing with a 2 class problem and the dataset has 1000 data points (500 for each class). Now divide these 1000 data points into 5 groups (200 in each group). Make sure that each group has equal data points from both the classes. Now select 4 groups and separate class_1 and class_2 data points and perform ML estimation for the parameters μ and ∑. Finally classify those data points which are present in the group (Test data) that you have not selected earlier and report accuracy. Repeat this process by changing , other 4 parts for training data and different left out part for Test data. In this example we will end up with 5 cases. This process is known as cross validation.

What is Confusion Matrix?

Confusion Matrix for a two class problem looks like this:

Confusion Matrix	Predicted As Class_1	Predicted As Class_2
Test Data Point Belong to Class_1	TP	FP
Test Data Point Belong to Class_2	FP	TP

TP: True Positive FP: False Positive

For each test data point find whether it is TP or FP and finally calculate accuracy as

$$\text{Accuracy} = \frac{\text{Total True positives}}{\text{No. of test data points}} \text{X } 100$$