

CS423 Data Mining: Assignment 3

Due date: 15 Nov 2016

Instruction

In this work you will implement a variety of Bayesian classifiers for classifying the data you have collected . It is possible that the data can be divided into more than two classes. However, it is highly recommended that you first work on a simpler binary classification problem before attempting to do multi-class ones.

There are two variations of Bayesian classifiers that you will investigate. Assuming that your data is normally distributed (Gaussian) implement:

1. Normal Discriminant Analysis with full covariance matrix.
2. Normal Discriminant Analysis using Naive assumption.

Measure the performance of the two classifiers by cross-validation technique. That is, to partition the data into two sets: training set and testing set. Train your classifier on the training set and test the learned classifier on the testing set. Record the classification errors and repeat the process 10 times. Finally, report your classification errors together with their standard deviations from the 10 runs. Discuss (in length) about your results and the choice of classifier used.

Submission

Email your code and a report to jakramate.b@cmu.ac.th by 15 November 2016. Prefixing the subject of your email with '[423HW3]'.