Assignment 3 (Computer Exercise)

A First Simulation Example on Designing and Assessing Classifiers

- Generate two small (10 observations per class) training data sets both from binormal distribution with different mean vectors and identity covariance matrix (case I in text).
- Design a Linear Discriminant Classifier (LDA) using the training sets, which is another name for the estimated Bayes Classifier with equal covariance matrices.
- Generate two large (1000 observations per class) testing data sets to assess your classifier (in real life application we do not have this luxury).
- Assume equal priors and costs and calculate the two types of errors of your trained classifier. Compare its performance to the Bayes classifier, i.e., the actual Bayes classifier that results from using the mean vector and covariance matrix directly without estimation.
- Repeat the above 10 more times with different training-set sizes, e.g., 20, 40, 80, 100, 200, 300, 400, 500, 700, 1000. Plot the error vs. the training set size. What do you observe?