CARLETON UNIVERSITY
SCHOOL OF COMPUTER SCIENCE
WINTER 2020

COMP 5107 Assignment IV Due March 26, 2020

Consider the two-class pattern recognition problem in which the class conditional distributions are both normally distributed with arbitrary means M_1 and M_2, and covariance matrices Σ_1 and Σ_2 respectively. Assume that you are working in a 3-D space (for example, as in Assignment II) and that the covariance matrices are not equal.

(a) Generate 200 *training* points of each distribution before diagonalization and plot them in the (x_1–x_2) and (x_2–x_3) domains.

(b) Using these *training* points estimate the parameters of each distribution using a maximum likelihood and a Bayesian methodology. In the latter, assume that you know the covariances. Plot the convergence of the parameters with the number of samples in each case.

(c) Using these same *training* points estimate each univariate distribution using a Parzen Window approach. In this case, work with the features in each dimension separately, and with an appropriate Gaussian kernel. For the output, you must plot the final learned *distribution* of the features in each dimension, and print out their "sample" mean and variance in each dimension.

(d) Using the *estimated* distributions, compute the optimal Bayes discriminant function (for the ML, Bayes and Parzen schemes) and plot it in the (x_1–x_2) and (x_2–x_3) domains.

(e) Generate 200 new points for each class for **testing** purposes, classify them and report the classification accuracy. *Do this (i.e., using all 400 points) using a ten-fold cross validation and a leave-one-out method*.

(f) Repeat (a)-(d) for the same data after you have diagonalized it.