CSE 465- COMPUTER ASSIGNMENT 3
SUPPORT VECTOR MACHINES
(DUE DATE: MAY 15'17)

1. Randomly generate 5,000 examples from each of two Gaussian distributions with the same diagonal covariance matrix $\alpha I_{50}, \alpha = 50$, centered one $(-1, \ldots, -1)$ and $(1, \ldots, 1)$ in R^{50}. Examples from the first distribution are labeled with 1, others with +1. Split the data into two equal sets of size 5,000, training and test [hint: the examples have been already generated for you. You may use the files negative.dat and positive.dat].

2. Download and install any SVM software library

3. Select a family of kernels, polynomial or Gaussian. Use SVM classification for training set sizes 1,000, 2,000, 3,000, 4,000 and 5,000, and find the kernel parameters that minimize the test error. Plot the test error as a function of the parameters and mark the minimum.

4. For a fixed parameter setting, plot the training and test errors as a function of the training set size and estimate the asymptotic test error.

5. Compute the Bayes error, i.e., the error of the best possible clas- sifier by using the density functions of the two classes. Compare with the results obtained using SVMs.