Peer Review
HW 10

Q.1) **Summary**: The problem statement was given and six experiments were performed on different data sets. The purpose of the experiments was to compare the error rate obtained by maximizing the two cost functions. Three different data sets were taken for that purpose and separation hyperplane for each data set was obtained by maximizing each of the two cost functions. The first data set was taken where the classes were well separated but did not have normal distribution. The second data set was taken where the classes were not well separated and did not have normal distribution. The final data set was taken with classes having uncorrelated data and normally distributed. For each data set the classification error was calculated using hyperplane obtained from both the cost functions. The plots showing the data point and the separation hyperplane were shown and it was concluded from the experiments that the selection of the cost function depends on the data set. If the data set is clustered then $J(w) = w^T S_B w$ should be used as cost function and if the data set has normal distributions then $J(w) = w^T S_B w / w^T S_W w$ should be used as cost function.

Good points: Three different data sets are taken to perform the experiment and compare the results obtained by them. The results are showed with plots of data and separation hyperplane and classification error in each case is calculated.

Can Improve: The two cost functions are referred as simple and complex cost functions; without explicitly stating, which one is simple and which one is complex. Instead of that numbering the cost functions and referring to the equation numbers would make it easier to understand. Nothing is mentioned about the within class and between class variance of the data and how it affects the selection of cost function.

Q.2) **Summary:** The purpose of this experiment was to compare the two classifiers – Support Vector Machine (SVM) and Neural Network (NN). The data taken for this experiment was same as in previous question. The data was three dimensional with 2 classes. For NN experiment all three features were taken into account. The entire data set was divided into training set and testing set. The NN consisted of three layers – input, hidden and output layer. The input layer had three neurons corresponding to each input feature. The experiment consisted of training the network by adjusting the weights using back propagation and optimizing by using gradient descent method. The sigmoid function was used in the neurons of hidden layer. Number of neurons in the hidden layer was 1, 3 or 10 and for each network configuration, the network was first trained and then tested using test data set. Each time the training and testing errors were calculated. A Gaussian function was also experimented with number of neurons in the hidden layer as 3 but had higher testing error. From the results it was concluded that as the number of neurons in the hidden layer increased, the training error decreased but the testing error increased. This was due to overfitting the weights to the training data set.

For SVM, two features were considered in a single experiment. Two such experiments were performed. The performance of SVM is shown by plotting a graph of

false positives vs. true positives. Comparing SVM and NN, it was concluded that when the classes were well separated then SVM gave better results and when the classes were not well separated then NN gave better results. Thus there is no definite conclusion drawn and which classifier must be used is dependent on the type of data set.

Good Points: Same data set was used to compare SVM and NN. Number of neurons in the hidden layer was varied and for each case error was calculated.

Can Improve: The feature vector for NN and SVM should not be changed. NN had three features while SVM had two features in the experiment. The terms true positives and false positives should be explained as those are not commonly used terms and so reader may not be able to understand what the author is trying to say. Require more clarity on SVM plots and which software was used to perform to experiment.

Q.3) **Summary:** The three classifiers are described and compared in this experiment. Given a test sample, the distance between test sample and training data is calculated and arranged in ascending order. Different distances were considered and error in each case was compared. The data set consisted of 6 classes. In Parzen window classifier, the distances obtained are weighted using Gaussian window and all the weighted distances of training data points within a specified window size, are added. The class corresponding to the maximum value of that sum was assigned to the new test point. The size of the window was varied. In k-nearest neighbor, uniform distribution window was used. The class corresponding to the maximum number of samples, among k training samples, was assigned to the new test sample. Here the value of k was varied. In nearest neighbor, the class of training sample having the smallest distance from the test sample was assigned to the test sample. The unequal priori probability was also taken into account while calculating the smallest distance. From the error results it was concluded that KNN had best accuracy while NN had best precision. But inspite of its high error value, Parzen window was still considered to be a better classifier.

Good Points: Parzen window was designed using various window sign. K-nearest neighbor was designed different values of K. Various distances were used to find distance between the test sample and training data set. Error for each case was calculated and compared for all three classifiers.

Can improve: N/A

Grade: OK