Make reasonable assumptions as and whenever necessary. Answer the questions to any sequence but answers to all the parts of any question should appear together. Marks will be deducted if this is not followed properly.

1. Prove that Bayes Classifier is the minimum error rate based classifier.   5

2. Explain 10-fold cross validation. Under which circumstances, F-measure is more preferred than accuracy in order to evaluate the performance of a classifier? Define Recall and Precision. Are these two measures conflicting to each other?   3+3+3+3=12

3. Suppose there are C classifiers with individual error-rate ε. What is the probability that the ensemble classifier makes a wrong prediction? Explain the steps of Ada-Boost algorithm?
3+5=8

4. Suppose we have 5 1D data points: $x1=5$, $x2=6$, $x3=7$, $x4=8$, $x5=9$, with 1, 2, 3 as class 4 and 5 as class 2. Let we use polynomial kernel of degree 3 $K(x,y)=(xy+1)^3$. Let the values of Lagrangian multipliers are $\alpha1=2.1$, $\alpha2=0$, $\alpha3=1.2$, $\alpha4=0$, $\alpha5=4.8$. Then write down the equation for discriminant function.   5

6. Consider a dataset with a nominal target attribute (i.e., a nominal CLASS) and several predicting attributes. Suppose that the dataset contains 1000 instances and that the data instances in the dataset have been clustered into 10 clusters each one containing roughly 100 instances. Let c1, c2, c3, ..., c9, c10 be the cluster centroids. The clustering has been performed using Euclidean distance over the predicting attributes (without using the target attribute). Consider the following classification method: Given a test instance t and an integer k (k is much smaller than 100):

   1. Find the closest centroid to the test instance using Euclidean distance over the predicting attributes.
   2. Use Euclidean distance to select the k-nearest neighbours of t *among those instances that belong to the cluster represented by the closest centroid.*
   3. Use those k selected data instances to classify the test instance.

Would this classification algorithm behave as similar to the k-nearest neighbour classifier? If not what are the differences? Explain and if at all possible ILLUSTRATE your answer.
5

8. HMM was introduced in the class with the URN problem. There are 3 urns each with a number of red, green and blue balls. Balls are drawn WITH replacement from different urns.

How will the HMM formulation change if balls were drawn WITHOUT replacement. Give the mathematical formulation precisely. That is, derive expressions for probability of (a) state sequence given the observation sequence, and (b) the observation sequence
4+4=8

9. "Viterbi and Forward/Backward algorithms are one and the same with a mere difference in operator". Prove or disprove the statement with precise arguments.   7