Assignment 6

Introduction to Data Analytics

Prof. Nandan Sudarsanam & Prof. B. Ravindran

1. On a particular data set, we use the ensemble method approach to building a predictor and achieve state of the art performance. Is it possible for some of the individual models in this ensemble to have poor performance as measured on the training data?

   (a) no

   (b) yes

2. With regards to bagging and boosting, which among the following are true?

   (a) the different learners in bagging can be trained in parallel

   (b) the different learners in boosting can be trained in parallel

   (c) individual classifiers in bagging are trained with all data points in the training set

   (d) individual classifiers in boosting are trained with all data points in the training set

3. Can the boosting technique be applied on regression problems? Can bagging be applied on regression problems?

   (a) no, no

   (b) no, yes

   (c) yes, no

   (d) yes, yes

4. In the general context of classification, re-weighting the data points (relative to an original training data set where the points are un-weighted) can lead to

   (a) change in the underlying optimisation problem that is solved

   (b) change in the positions of data points in the feature space

   (c) change in the decision surface generated by the classifier

   (d) change in the nature of the data set from being linearly separable to becoming linearly non-separable (in case the original data was linearly separable)

5. If one feature (compared to all others) is a very strong predictor of the class label of the output variable, then all of the trees in a random forest will have this feature as the root node.

   (a) false

(b) true

6. Suppose we have 2-class training data which is linearly separable. We use the perceptron training algorithm to build a classifier. What is the number of possible solutions that can be obtained through this method?

   (a) 1
   (b) depends on the size of the margin separating the data points belonging to the two classes
   (c) infinite
   (d) depends on the size of the margin separating the data points belonging to the two classes and the learning rate parameter

7. By using a linear activation function in the output layer of a neural network for solving regression tasks, are we constraining the resultant model to be linear in the input features?

   (a) no
   (b) yes

8. In the backpropogation algorithm, the gradient of the error with respect to the weight vector is itself a vector. What does the direction of this vector indicate?

   (a) it points in the direction of steepest decrease in the error
   (b) it points in the direction of steepest increase in the error
   (c) it indicates that component of the weight vector that results in maximum error
   (d) it indicates that component of the weight vector that results in minimum error

9. Given a multi-class data set, you choose to use an artificial neural network to build a classification model for this data. How would you determine the number of hidden layer nodes to use for this task?

   (a) same as the number of input layer nodes
   (b) same as the number of output layer nodes
   (c) through cross validation
   (d) maximum of the number of input and output layer nodes

10. In the lectures, we saw how to train a 7 layer auto encoder network. In case we wanted to perform classification on the data used for training this network, while making use of the trained network, a suitable approach would be to

    (a) add an additional eighth layer on top of the 7 layers as the output layer and train the entire network for the classification task
    (b) add an additional eighth layer on top of the 7 layers as the output layer and only modify the weights between layers 7 and 8 in training for the classification task
    (c) discard the top 3 layers, add an additional layer on top of the 4^{th} layer as the output layer and train the entire network for the classification task
    (d) discard the top 3 layers, add an additional layer on top of the 4^{th} layer as the output layer and only modify the weights between layers 4 and 5 in training for the classification task