**1. Which of the following statement is true ?**  
***A. In Gradient Boosting (GD) and Stocharstic Gradient Boosting (SGD), we update set of parameters in an iterative manner.***  
B. In Stocharstic Gradient Boosting (SGD), we have to run for all the sample for a single update of parameter in each iteration.  
C. In Gradient Boosting (GD), we either use entire data or a subject of training data to update a parameter in each iteration.  
  
**2. Which of the following hyper parameter of Random Forest increase, causes overfit the data ?**  
***Answer : depth of tree***  
  
**3. Imagine you are working with analytics vidya and you want to develop a machine learning algorithm which predict no. of views on the article. Your analysis is based on teachers name, author name, no. of articles written by same author on the analytics vidya platform in past etc. Which of the following evalution matrix would you choose in that case ?**  
***Answer : Min square error***  
  
**4. Lets say that you are using action funX in hidden layer of neural network at a particular neuron for given input you get the output -0.001. Which of the following activation function should X represent.**  
***Answer : fun 8***  
  
**5. Which of the following are one of the important step to preprocess the text in NLP based project.**  
A. scemming  
B. topward removal  
C. object standardization  
***D. All of these***  
  
**6. Adding a anon important feature to a linear regression method may result in**  
***Answer : increase in R^2 (square of R)***  
  
**7. In KNN model, it is very likely to overfit due to cause of dimensionality. Which of the following option would you consider to handle this problem.**  
***Answer : dimensionality reduction & feature selection***  
  
**8. Which of the following is true about the gradient boosting tree ?**  
A. in each stage introduce a regression tree to compensate the shortcoming of the existing model.  
B. we can use gradient distance (GD) method to minimze the loss finction  
***C. Both of these***  
  
**9. To apply bagging to regression trees, which of the following are true in that case ?**  
A. we build the n regression with n bootstrap sample  
B. we take the average of n regression tree  
C. each tree has high varience with low biased.  
***D. All of the above***  
  
**10. When you find noise in data, which of the following option will you considered in KNN ?**  
***Answer : I will increase the value of k.***  
  
**11. Suppose you want to predict the class of the data point, x=1 and y=1 using eucleadian distance in 3NN in which class these data points belongs to ?**  
***Answer : positive (+) class***  
  
**12. Which of the following will be eucleadian distance between the two data points A(1,3) and (2,3)**  
***Answer : 1***  
  
**13. Suppose you are working on a binary classification problems with three input features and you choose to apply a bagging algorithm X. On this data, you choose max\_features = 2 and the n\_estimators = 3, assume that each estimation has 70% accuracy. Note that algorithm X is aggregating the result of individual estimates based on maximum voting. What will be the maximum accuracy you can get ?**  
***Answer : 100%***  
  
**14. In random forest or gradient boosting algorithm, features can be of any type, for example it can be a continuous features or categorical features. Which of the following option is true when you consider this type of feature.**  
***Answer : Both the algorithm can handle real valued attributes by discretizing them.***  
  
**15. Which of the following is true about training and testing error in the case described below. Suppose you want to apply Adaboost algorithm on data D which has 'T' observation. You have set half of the data for training and half for testing initially. Now you want to increase the no. of data points for training. T1, T2, ------- Tn where T1 > T2 < T3 < ---- < Tn**  
***Answer : the difference between training error and testing error decreases as the no. of observation increases.***  
  
**16. Suppose you are given 3 variables x, y, z, the pearson corelation coefficient for (x,y), (y,z) & (x,z), c1, c2 and c3 respectively. Now you have added 2 in all the values of X, and substract 2 from all the values of Y and Z remains the same. The new coeeficient (x,y), (y,z) & (x,z) are given by D1, D2 respectively. How do the values of D1, D2, D3 relates to c1, c2, c3**  
***Answer : D1=c1, D2=c2, D3=c3.***