1. (1 point) Which of the following is / are true about weak learners used in ensemble model?

   1. They have low variance and they don't usually overfit
   2. They have high bias, so they can not solve hard learning problems
   3. They have high variance and they don't usually overfit

      A. 1 and 2
      B. 1 and 3
      C. 2 and 3
      D. none of the above

   Answer: A

   Explanation
   Weak learners are sure about particular part of a problem. So they usually don't overfit which means that weak learners have low variance and high bias.

2. (1 point) In an election, N candidates are competing against each other and people are voting for either of the candidates. Voters don't communicate with each other while casting their votes. Which of the following ensemble method works similar to above-discussed election procedure?

      A. bagging
      B. boosting
      C. a and b
      D. none

   Answer: A

   Explanation
   In bagged ensemble, the predictions of the individual models won't depend on each other.

3. (1 point) Which of the following is/are true about Random Forest and Gradient Boosting ensemble methods?

   1. Both methods can be used for classification task
   2. Random Forest is used for classification whereas Gradient Boosting is used for regression task
   3. Random Forest is used for regression whereas Gradient Boosting is used for Classification task
   4. Both methods can be used for regression task

   A. 1 and 2

   B. 2 and 3

   C. 1 and 4

   D. 3 and 4

Answer: C

Explanation
Self Explanatory

4. (1 point) To apply bagging to regression trees which of the following is/are true?

   1. We build the N regression with N bootstrap sample

   2. We take the average the of N regression tree

   3. Each tree has a high variance with low bias

     A. 1 and 2

     B. 2 and 3

     C. 1 only

     D. 1,2 and 3

Answer: D

Explanation
We apply all the steps.

5. (1 point) (Multiple select) Considering the AdaBoost algorithm, which among the following statements is true?

     A. In each stage, we try to train a classifier which makes accurate predictions on any subset of the data points where the subset size is at least half the size of the data set

     B. In each stage, we try to train a classifier which makes accurate predictions on a subset of the data points where the subset contains more of the data points which were miscalssified in earlier stages

     C. The weight assigned to an individual classifier depends upon the number of data points correctly classified by the classifier

     D. The weight assigned to an individual classifier depends upon the weighted sum error of misclassified points for that classifier

Answer: B,D

Explanation
The classifier chosen at each stage is the one that minimises the weighted error at that stage.The weight of a point is high if it has been misclassified more number of times in the previous iterations. Thus, maximum error minimisation is performed by trying to

correctly predict the points which were misclassified in earlier iterations. Also, weights are assigned to the classifiers depending upon their accuracy which again depends upon the weighted error (for that classifier).

6. (1 point) (Multiple select) Which among the following are some of the differences between bagging and boosting?

   A. In bagging we use the same classification algorithm for training on each sample of the data, whereas in boosting, we use different classification algorithms on the different training data samples

   B. Bagging is easy to parallelise whereas boosting is inherently a sequential process

   C. In bagging we typically use sampling with replacement whereas in boosting, we typically use weighted sampling techniques

   D. In comparison with the performance of a base classifier on a particular data set, bagging will generally not increase the error whereas boosting may lead to an increase in the error

   Answer: B,C,D

   Explanation
   With regards to the last option, boosting can result in an increase in error over a base classifier due to over-emphasis on existing noise data points in later iterations.

   Connected questions
   Consider the following data set with x feature and y target variable We want to use

x	y
12	130
18	146
22	174

   gradboost method to build a regression model on the base model.

7. (1 point) What is the base model?

   A. $y = 80$

   B. $y = 164$

   C. $y = 150$

   D. $y = 146$

   Answer: C

   Explanation
   $$\bar{y} = \frac{130 + 146 + 174}{3} = 150$$

8. (1 point) Residuals for each data point

    A. $-30, -10, 20$

    B. $-20, -4, 24$

    C. $20, 4, -24$

    D. $30, 10, -20$

Answer: B

Explanation

Residuals $= y_i - \bar{y}$