Class 7 Cody Xu

5 In-class exercises

Answer the following questions, and upload your results to your github repo. Remember, your answers do not have to be correct to earn participation points!

1.Bagging is a special case of random forests under which case?

When the only 1 predictors for data. Even if we choose from a random subset of predictors it will not change the result.

2.What are the hyperparameters we can control for random forests?

The sequence of p predictors of the trees. The depth of the trees.

3.Suppose you have the following paired data of (x,y): (1,2), (1,5), (2,0). Which of the following are valid bootstrapped data sets? Why/why not?

(1,0), (1,2), (1,5)

(1,2), (2,0)

(1,2), (1,2), (1,5)

The first and the last are the valid bootsrapped datasets. Data set 2 has too less data to separate to training and test set.

4.For each of the above valid bootstapped data sets, which observations are out-of-bag (OOB)?

Data set one no OOB. For data set 3 (2,0) is OOB

5.You make a random forest consisting of four trees. You obtain a new observation of predictors, and would like to predict the response. What would your prediction be in the following cases?

Regression: your trees make the following four predictions: 1,1,3,3.

The prediction will be the average of these result (1+1+3+3)/4=2. The prediction will be 2.

Classification: your trees make the following four predictions: “A”, “A”, “B”, “C”.

The prediction will be ‘A’, since it has the most votes.