Vibudh Singh – 98716913  
MBAN - BAIT 509 Class Meeting 07 – Class activity

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!

* Bagging is a special case of random forests under which case?

When the total number of randomly selected predictor variables would be equal to the total number of predictor variables. Both will give very similar results every time we include most of the significant variables in the random forest trees.

* What are the hyperparameters we can control for random forests?

The hyperparameters we can control for random forests are:

* The number of predictor variables that we what to choose
* The number of samples we want to choose in our bagged data sample
* 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. (1,0), (1,2), (1,5)
  2. (1,2), (2,0)
  3. (1,2), (1,2), (1,5)

1 -> Not. Because it is not a subset of the original dataset

2-> Yes. Because it is a subset of the original dataset

3-> Yes. Because it is a subset with repeated elements

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

2-> Out of bag is (1,5)

3-> Out of bag is (2,0)

* 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?
  1. Regression: your trees make the following four predictions: 1,1,3,3.
  2. Classification: your trees make the following four predictions: “A”, “A”, “B”, “C”.

1). We will take the mean value, i.e., 2  
2). We will take the mode value, i.e., A