In-class exercises

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

When the random subset of m is equal to the total predictor p.

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

The number of predictors in the random subset—m.

The depth of the tree

The number of trees

**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)

Among all these three data sets, data set 3 is valid bootstrapped data set, since all the elements in this data set are belong to the original data set.

For 2, this data set only has 2 number

For 1, (1, 0) is not in the original data

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

For 2. (1,2), (2,0), observation (1,5) is OOB

For 3. (1,2), (1,2), (1,5), observation (2,0) is OOB

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. 2 –the average

2. The prediction should be ‘A’