**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!

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

When m =p, when data subset we choose equals to the total set.

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

The restricted number of predictors(m) used in the tree.

* 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)

The third one is the valid bootstrapped data set, because it has the same data length and are randomly drawn from the dataset while the first data set contains (1,0) which is out of the data set and the second one only contains two data.

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

For iii. OOB 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.

The average of 1,1,3,3, which is 2

* 1. Classification: your trees make the following four predictions: "A", "A", "B", "C".

The mode of "A", "A", "B", "C", which is A