**In Class Exercise 3**

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

* When the trees in the ensemble tend to be correlated i.e., they share similarities.

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

* Number of branches in the decision tree
* Subset of predictors or number of predictors

**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)**

i and iii are valid data sets since they are related in some sense with the paired data.

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

The following observations are OOB in each data set

1. (1,0)
2. None
3. (1,2)

**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. –** Prediction 2

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