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

Random forest differs only in choosing a random subset of predictors for each tree, to decorrelate the trees.

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

The hyperparameters are B, the number of trees to be made, and m, the size of the subset of predictors to be randomly chosen for each 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)

Not valid because (1,0) is not drawn from the above paired data.

* 1. (1,2), (2,0)

Is valid because the data is drawn from the above paired data.

* 1. (1,2), (1,2), (1,5)

Is valid because the data is drawn from the above paired data and each observation drawn is replaced in the original data, meaning drawing the same observation twice is allowed.

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

In case 1, (2,0) is out of bag.

In case 2, (1,5) is out of bag.

In case 3, (2,0) is out of bag.

**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.

(1+1+3+3)/4 = 2

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

Mode = A, so prediction = A