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

An ensemble bagging is a random forest. A random forest is an average method.

Random forest is decorrelated bagging.

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

Hyperparameters are the number of decision trees in the forest and the number of features considered by each tree when splitting a node. (The parameters of a random forest are the variables and thresholds used to split each node learned during training).

* 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)

We could only take paired data and cannot change the predicted value. 2 and 3 could be chosen.

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

For the second dataset, (1,5) is the OOB, and for the third dataset, (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.

Regression takes the average (1+1+3+3)/4=2

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

Categorical classification takes the mode of the predictors, so it is A.