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

When we use all the predictors or when do not restrict the predictors on which the model should be built

* What are the hyperparameters we can control for random forests?
  1. Bagging or Bootstrap aggregation parameter to control the number of trees
  2. M, maximum number of predictors, out of our parameter set
  3. Depth of trees
  4. Minimum number of samples required in leaf nodes(also a restriction on tree growth)
  5. Minimum samples split, minimum number of samples placed in a node before it is split
* 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 a valid dataset, there is no (1,0) in the dataset

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

invalid dataset, as both observations are present in the dataset but sample size is 2

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

Valid dataset, as both observations are present in the dataset

* For each of the above valid bootstapped data sets, which observations are out-of-bag (OOB)?
  1. Not a valid dataset
  2. (1,5)
  3. (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.

We will predict the response to be 2, average for (2\*1 + 2\*3)/4

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

A, we will take the mode of predictions.