* Bagging is a special case of random forests under which case?
* When Bagging can generate different trees, it is a special case of random forests. Decorrelated ensemble bagging is the random forest. When M(all the dataset)=p(a new subset the predict)
* What are the hyperparameters we can control for random forests?
* How many predictors we will choose every time as m
* Different predictors we will choose as p
* Number of trees
* 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)
* Number 2 and 3 are the correct since one x only corresponds to one y. We can not change the original value.
* The number of the sample equals the number of data in dataset, so number 2 is not a valid bootstrapped data sets
* For each of the above valid bootstapped data sets, which observations are out-of-bag (OOB)?

For number 3, the out-of-bag 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.
  2. Classification: your trees make the following four predictions: “A”, “A”, “B”, “C”.

1. Take the average of the result in regression: (1+1+3+3)/4=2
2. Take the mode of result in classification: “A”