**Answer the following questions, and upload your results to your github repo. Remember, your answers do not have to be correct to earn participation points!**

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

Bagging is a special case of random forests under the situation where instead of a subset of m predictors are used in making stumps for growing the trees, in bagging all predictors are being chosen to grow the tree

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

We can control the number of m predictors being chosen each time for making the stump. Thus, the value of ‘m’ is the hyperparameter; And the 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)**

The iii set is a valid bootstrapped data set. For set i, it has (1,0) which is not from the original data set; in set ii, while both of the elements are from the original dataset, the number of elements does not equal to the original data set; For set iii, it has same number of elements and all the elements are copies of the original data set.

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

(2,0) is the OOB for dataset iii, the other two data sets are not valid bootstrapped data sets.

* **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".**

Answer:

i: The prediction will be ‘2’

ii: The prediction will be ‘A’