In-class exercises

**Submitted by:**

Pallavi Agarwal - 86917465

1. Bagging is a special case of random forests under which case?

Random forests is a special case of Bagging in that while the idea is to use bagging, the way it is implemented is only a subset of features (usually given by sqrt(n) where n is the number of features) is selected to split each node in a tree, unlike in bagging where we consider all features for splitting.

1. What are the hyperparameters we can control for random forests?
   1. The number of predictors per split, though mostly taken as sqrt(n)
   2. Tree size: number of trees, maximum depth, number of leaves per branch
2. 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)

a and c are the valid set because we are looking for multisets of at least the same size as original data. b is a subset which would be boosting not bagging

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

a has out of bag observation (1,0)

1. 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”.
2. 2
3. A