Week 6 - AYU - Individual

**Problem** (Sample - Question 10)

Determine which of the following statements about random forests is/are true?

I. If the number of predictors used at each split is equal to the total number of available predictors, the result is the same as using bagging.

1. When building a specific tree, the same subset of predictor variables is used at each split.
2. Random forests are an improvement over bagging because the trees are decorrelated.
3. None
4. I and II only
5. I and III only
6. II and III only
7. The correct answer is not given by (A), (B), (C), or (D).

**Problem** (Sample - Question 41)

For a random forest, let p be the total number of features and m be the number of features selected at each split.

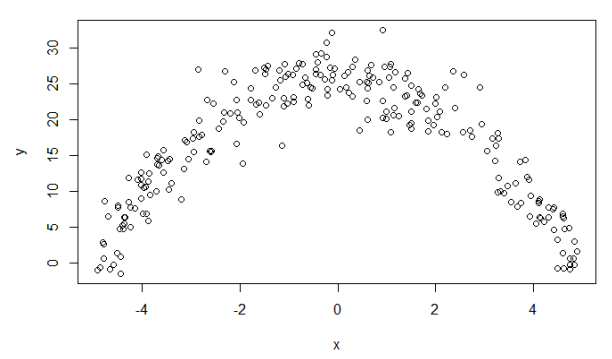
Determine which of the following statements is/are true.

I. When , random forest and bagging are the same procedure.

1. is the probability a split will not consider the strongest predictor.
2. The typical choice of is .
3. None
4. I and II only
5. I and III only
6. II and III only
7. The correct answer is not given by (A), (B), (C), or (D)

**Problem** (Sample - Question 39)

You are given a dataset with two variables, which is graphed below. You want to predict y using .



Determine which statement regarding using a generalized linear model (GLM) or a random forest is true.

1. A random forest is appropriate because the dataset contains only quantitative variables.
2. A random forest is appropriate because the data does not follow a straight line.
3. A GLM is not appropriate because the variance of y given x is not constant.
4. A random forest is appropriate because there is a clear relationship between y and
5. A GLM is appropriate because it can accommodate polynomial relationships.