1. What is the estimated depth of a Decision Tree trained (unrestricted) on a one million instance training set?

Ans>> The general formula is for depth of decision tree is `log2(m)^3`, where `m` is the number of instances, and here with 1 million data set the depth will be log2(10^6) = 20

2. Is the Gini impurity of a node usually lower or higher than that of its parent? Is it always lower/greater, or is it usually lower/greater?

Ans>> It is lower than it’s parents, if one child is smaller than the other, it is possible for it to have a higher gini score than its parrent

3. Explain if its a good idea to reduce max depth if a Decision Tree is overfitting the training set?

Ans>> If the our decision tree is overfitting then we should reduce the max-depth for the algorithm. Because by reducing the max-depth it will be less variance and regulrarize the model.

4. Explain if its a good idea to try scaling the input features if a Decision Tree underfits the training set?

Ans>> The tree based model is not required scaling of the input features, so there will be no effect on model underfitting or overfitting.

5. How much time will it take to train another Decision Tree on a training set of 10 million instances if it takes an hour to train a Decision Tree on a training set with 1 million instances?

Ans>> We calculate the traning time of decision tree algorithms is by using this formula.

`Traning time = n \* 10m \* log(10m) / n \* m \* log(m)`, where `n` is the time and `m` is the instance number. In this case after calculation it will be 11.677 hours.

6. Will setting presort=True speed up training if your training set has 100,000 instances?

Ans>> The `presort = True` is work when the dataset is smaller a like under few thousand instance, and this case the data set size is 100,000, s here it will make the algorithms slow.

7. Follow these steps to train and fine-tune a Decision Tree for the moons dataset:

a. To build a moons dataset, use make moons(n samples=10000, noise=0.4).

b. Divide the dataset into a training and a test collection with train test split().

c. To find good hyperparameters values for a DecisionTreeClassifier, use grid search with cross-validation (with the GridSearchCV class). Try different values for max leaf nodes.

d. Use these hyperparameters to train the model on the entire training set, and then assess its output on the test set. You can achieve an accuracy of 85 to 87 percent.

8. Follow these steps to grow a forest:

a. Using the same method as before, create 1,000 subsets of the training set, each containing 100 instances chosen at random. You can do this with Scikit-ShuffleSplit Learn's class.

b. Using the best hyperparameter values found in the previous exercise, train one Decision Tree on each subset. On the test collection, evaluate these 1,000 Decision Trees. These Decision Trees would likely perform worse than the first Decision Tree, achieving only around 80% accuracy, since they were trained on smaller sets.

c. Now the magic begins. Create 1,000 Decision Tree predictions for each test set case, and keep only the most common prediction (you can do this with SciPy's mode() function). Over the test collection, this method gives you majority-vote predictions.

d. On the test range, evaluate these predictions: you should achieve a slightly higher accuracy than the first model (approx 0.5 to 1.5 percent higher). You've successfully learned a Random Forest classifier!