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

   The estimated depth of a Decision Tree trained on a one million instance training set can vary depending on the complexity of the data and the specific algorithm implementation. In general, Decision Trees have an **average depth of log2(N),** where N is the number of instances in the training set. So, for a one million instance training set, the estimated depth would be around log2(1,000,000) ≈ 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?**

   The Gini impurity of a node is usually lower than that of its parent. The Gini impurity is a measure of the node's impurity or the degree of class mixture in the node. Splitting a node based on a feature reduces the impurity, and thus **the Gini impurity of the resulting child nodes is typically lower than that of the parent node.**

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

   If a Decision Tree is overfitting the training set, reducing the maximum depth can be a good idea. **Overfitting occurs when the tree becomes too complex and starts to memorize the training examples instead of generalizing well to unseen data.** By reducing the maximum depth, the tree is constrained to be simpler and less likely to overfit. However, it's important to find the right balance and avoid underfitting, where the tree may not capture the underlying patterns in the data.

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

   Scaling the input features is generally not necessary or beneficial for Decision Trees. Decision Trees make splits based on feature thresholds, and the relative scales of the features do not affect the tree's structure or performance. However, if the features have different units or scales, it might be a good idea to normalize or standardize them to ensure fair comparisons during the splitting process.

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?**

   Estimating the training time for a Decision Tree on a larger dataset depends on various factors such as the complexity of the data, the algorithm implementation, the computational resources, and any optimizations used. However, as a rough estimate, if it takes an hour to train a Decision Tree on a 1 million instance training set, training another Decision Tree on a 10 million instance training set might take around 10 hours. **The training time tends to increase approximately linearly with the size of the training set.**

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

   The `presort=True` parameter in scikit-learn's DecisionTreeClassifier speeds up training when the training set is small or when the maximum depth is limited. However, if your training set has 100,000 instances, enabling presorting is unlikely to provide a significant speedup. In fact, it may even slow down the training process due to the overhead of sorting the data. It's generally recommended to leave `presort` as False unless you have a specific reason to enable it.