

Problem 1

- i. Pruning is a data compression technique in machine learning and search algorithms that reduces the size of decision trees by removing sections of the tree that are non-critical and redundant to classify instances. Pruning reduces the complexity of the final classifier, and hence improves predictive accuracy by the reduction of overfitting.
- ii. Pruning of the decision tree is accomplished by replacing a whole subtree by a leaf node. this measurement is based on validation data. in the other hand, the pruning is usually done when the model is over fitted.
- iii.
 - 1. Stop growing when data split is not statistically significant
 - 2. Acquire more training data
 - 3. Remove irrelevant attributes (manual process – not always possible)
 - 4. Grow full tree, then post-prune