IE412 HW#5
Tianqi Wu

Problem 2:
Since age is a continuous numeric value between 0 and 120, we can discretize it by partitioning the range to two subsets. We can set age 0-50 as young and 40-120 as old. Then, we can continue with the basic binary decision tree algorithm.

Problem 3:
It depends on different situations. If each classifier is given the same whole dataset, ensemble of multiple such classifiers would not make a difference since all the classifiers learn the same dataset and would make the same error. If each classifier is able to learn different subsets of the whole dataset, combination of multiple such classifiers would lead to a nontrivial increase of the classification accuracy since the error made by one classifier may be corrected by others.

Problem 4:
SVM will work while Decision-tree induction and Naive Bayes algorithm may not work. Since we have a relatively large dataset containing 100 tissue samples with 10000 genes of binary attribute. For SVM, the complexity of trained classifier is characterized by the number of support vectors rather than the dimensionality of the data. Hence, SVM is able to work with the dataset with large dimensions. Decision tree induction may work since the attribute is binary. However, an induced tree may grow deep and it would be slow with high computation cost and. Also, it may overfit the training data, leading to poor accuracy of unseen samples. For Naive Bayes classification, it assumes the class conditional independence and it will lose the accuracy for the micro-array data.