yxu66

a. I think the file breast-cancer is more challenging. Because the data is more complicated, so it is hard to predict very correctly. And the precision and recall is more meaningful to this data sets than nursery, since the precision tells the how accurate the diagnosis is and the recall tells how many patient that will get recurrence that you can pick out.

b.The accuracy of training set will higher than that of test set, since the decision tree is built based on the training set. The result of training set of nursery, lymphography, tennis and restaurant was overfitting.

c. In breast-cancer test, the accuracy of train is not 100%, there is noisy here.

d.In test case that have enough information, the tree will not look very different. Since the tree is based on training set, which is 80% of the data sets, but maybe slightly different. If the size of data set is small, then the decision tree may very different.