William Shu & Zhiqian Ma

1. Remove the least information gain feature of all features because it provides the least information gain.
2. Unprune: Training Set Accuracy: 0.99650 Test Set Accuracy: 0.86538
   Pruned: Training Set Accuracy: 0.92995 Test Set Accuracy: 0.91827
   Error rate decrease: ((1-0.86538) - (1-0.91827))/(1-0.86538) = (0.13462 - 0.08173)/0.13462 = 39.29%
3. The more the nodes in decision tree, the more likely the decision tree memorize all the training data. So the decision tree tends to overfit the training data and cannot generalize. Because pruned tree has fewer nodes.