

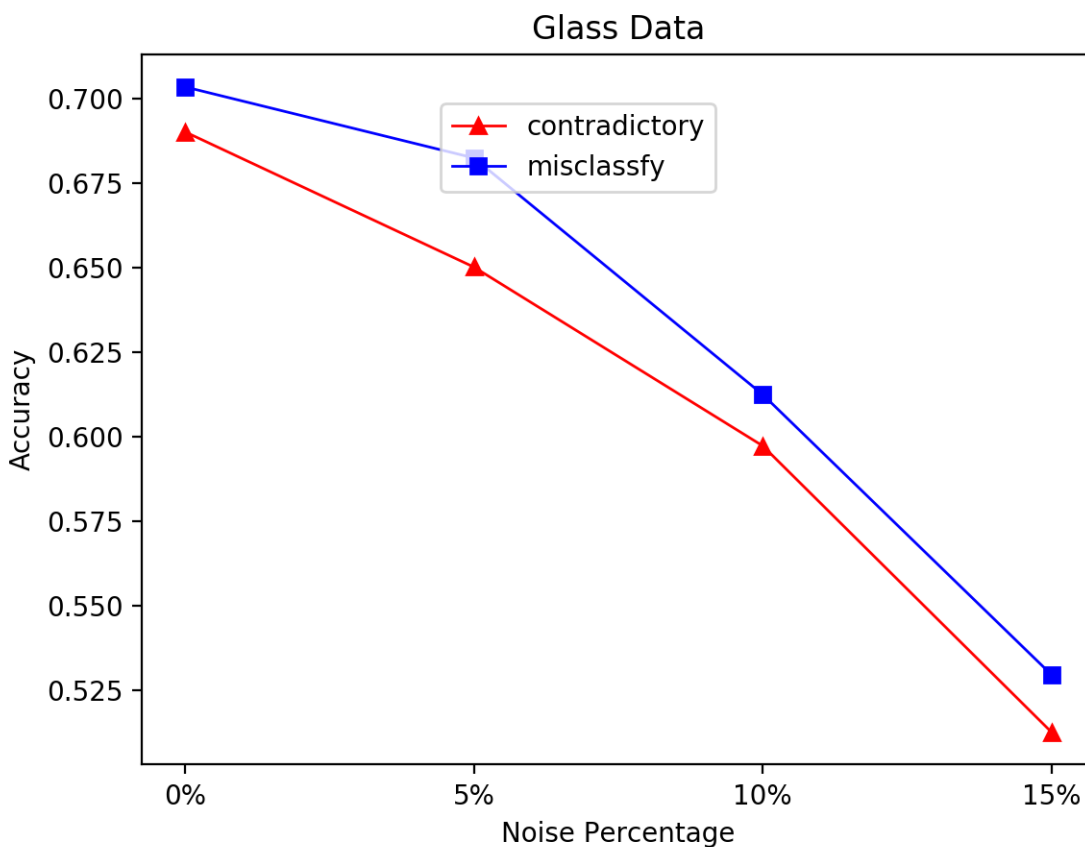
Question 1 :

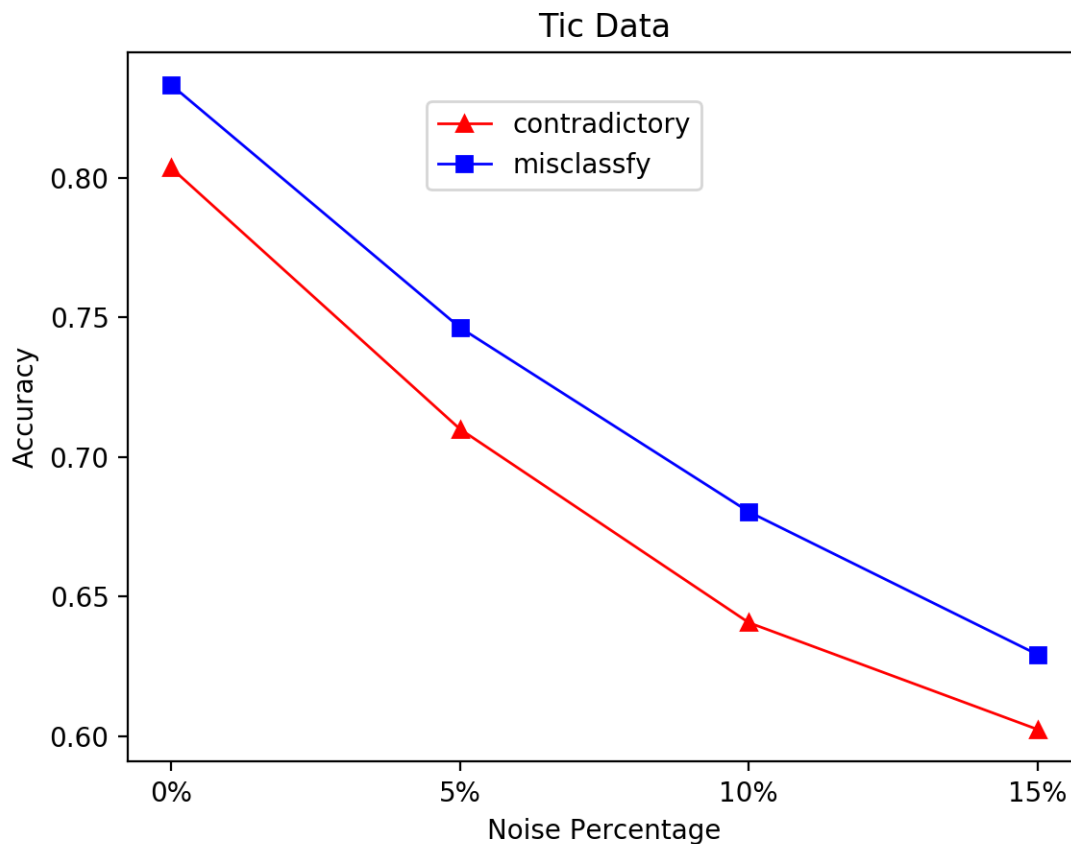
	Mean Accuracy	Variance Of Accuracy
Glass	0.6983522727272724	0.015390040676652892
Tic-Tac-Toe	0.8379676418439714	0.002231092475817049

From the result, we can see that Tic-Tac-Toe has a better result than glass data. It is because the difference between continuous data and split data. It is hard to split continues data and may cause over fitting due to the continuous data.

Question2:

a.





B. How do you explain the effect of noise on the C4.5 methods?

By comparing two figure, we can see that noise has more effect on tic-tac-toe data than glass data. By computing the decision tree for both data, we can know that glass data tree has less braches that the other one. The noise data may effect less on glass data branches. As a result, the accuracy of glass data will be less than tic-tac-toe data when we have noise. In addition, we can see that the accuracy is stable even though we add more noise on it. It is because c4.5 algorithm go over all branches in the tree. It uses all information that provided by the tree, so it is not that sensitive for the noise

By comparing the data between pruning and un-pruning, we can see pruning decision tree has a slower decrease when we adding noise. It is because we can delete useless information when we do the pruning.

Question3:

a).

	10	50	150	392
Accuracy	0.5801	0.8077	0.9315	0.9477

b).



From the first picture, we can see that it is near the central part of picture. In addition, the features are near the edge of numbers in picture. It may be because the distance between edge of different numbers are large. With the dimension increasing, the feature will cover more and more central part. The side part of picture will not be considered as best feature because the data of them are all zero and there is no difference between different numbers on that part. Those features will be deleted during the backward search.

c).

	10	50	150	392
Feature Select	0.5801	0.8077	0.9315	0.9477
LDA	0.84	0.84	0.84	0.86

From the result, we can see that LDA has a better performance when the selected feature number is less or equal than 50. After that, feature selection has a better performance when the selected feature number is larger than 150.

This is because LDA is a feature subtraction method. It will calculate the matrix for inter-class and intra-class and get the new reduced data from original data according to those matrices. It performs more information than feature selection even though they have the same length of reduced data. As a result, it has a better performance when we have a small d .

However, reduced data from LDA also perform some useless or negative effect data. It considers data like background data from matrix and uses them to classify test data. For feature selection, when the dimension is large enough, it will cover features that can separate each class as much as possible. There is no noise feature in this reduced data. It will have a better performance than LDA when the dimension is large enough.