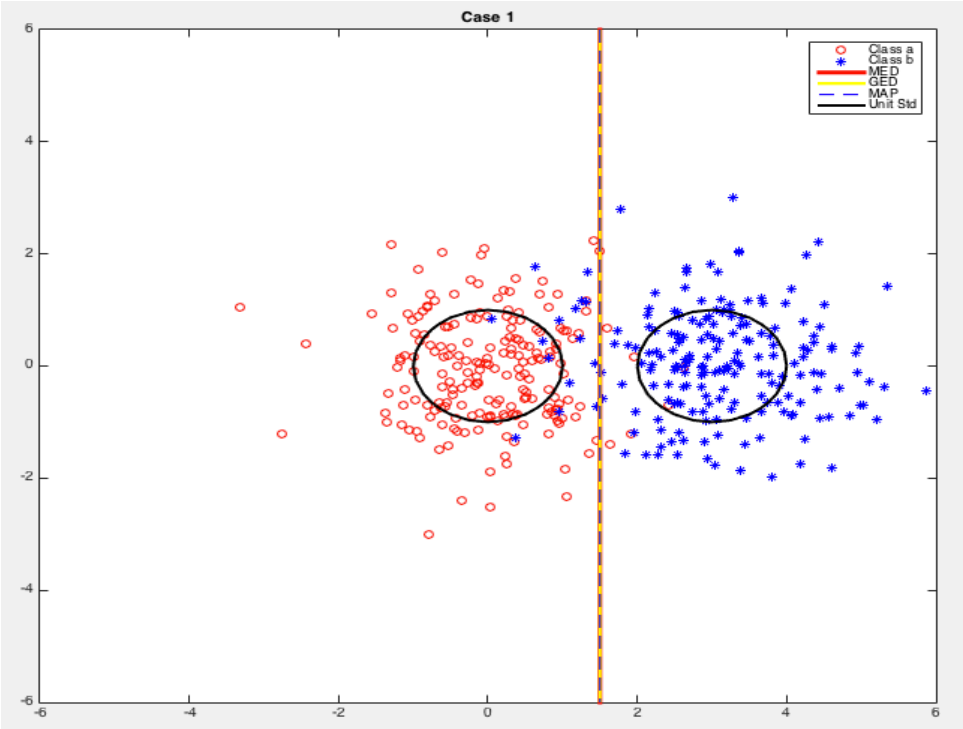
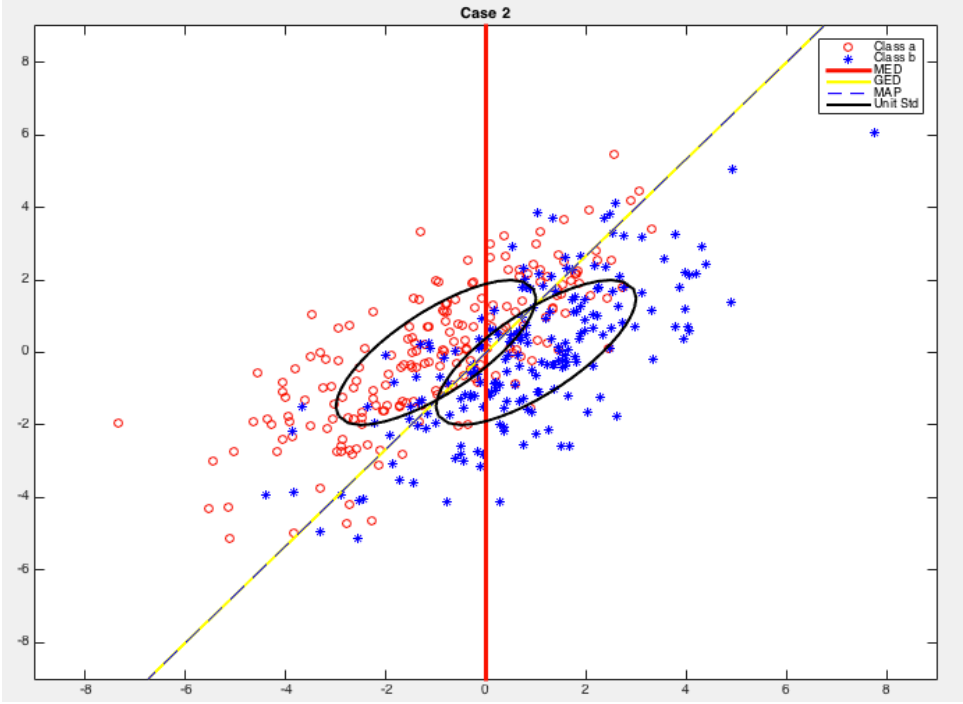
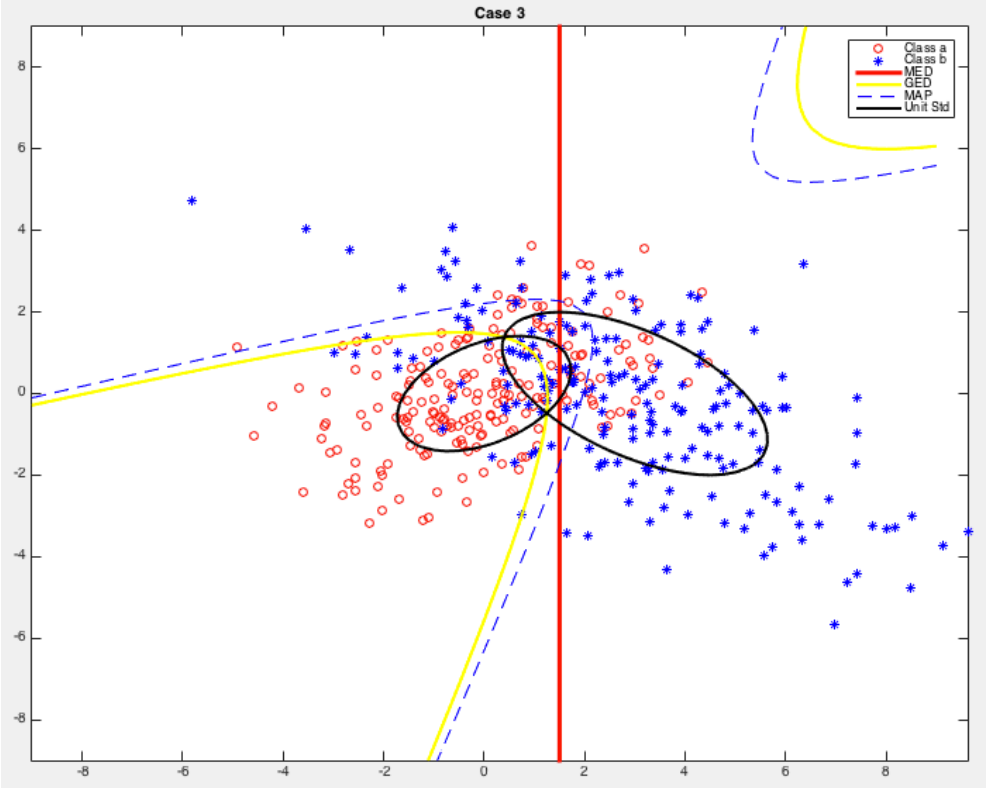
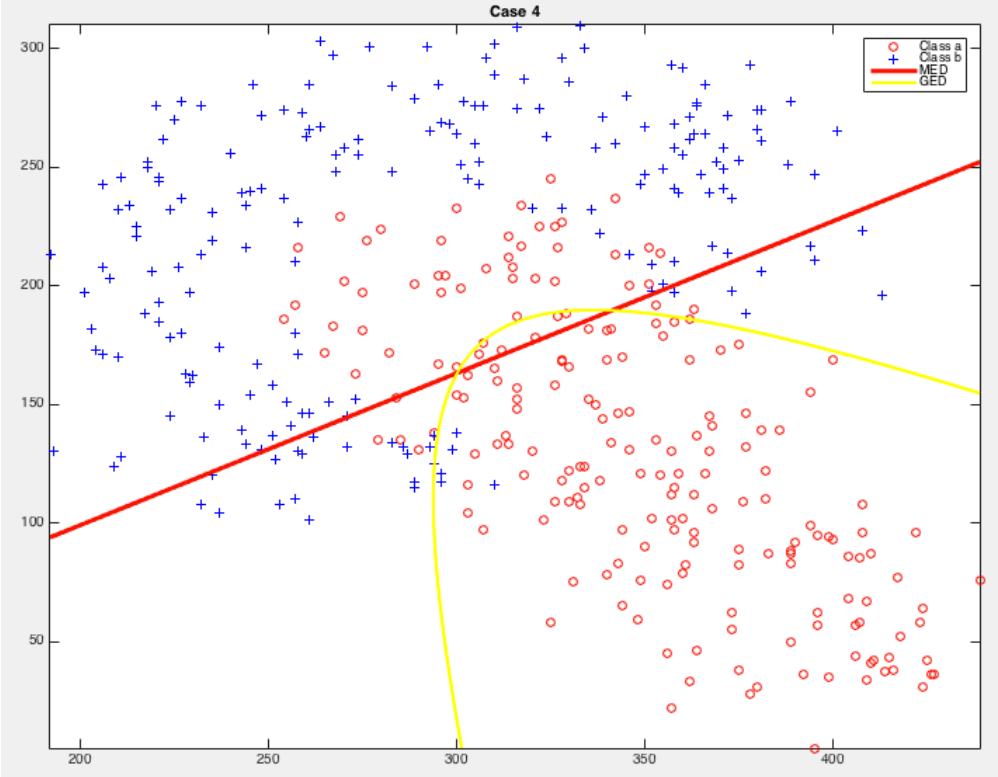
SYDE 675
Shala Chen
Assignment 1
Feb 1st 2017

Set One






Case 3
Class a
Class b
MED
GED
MAP
Unit Std
8
6
4
2
0
-2
-4
-6
-8
-8
-6
-4
-2
0
2
4
6
8

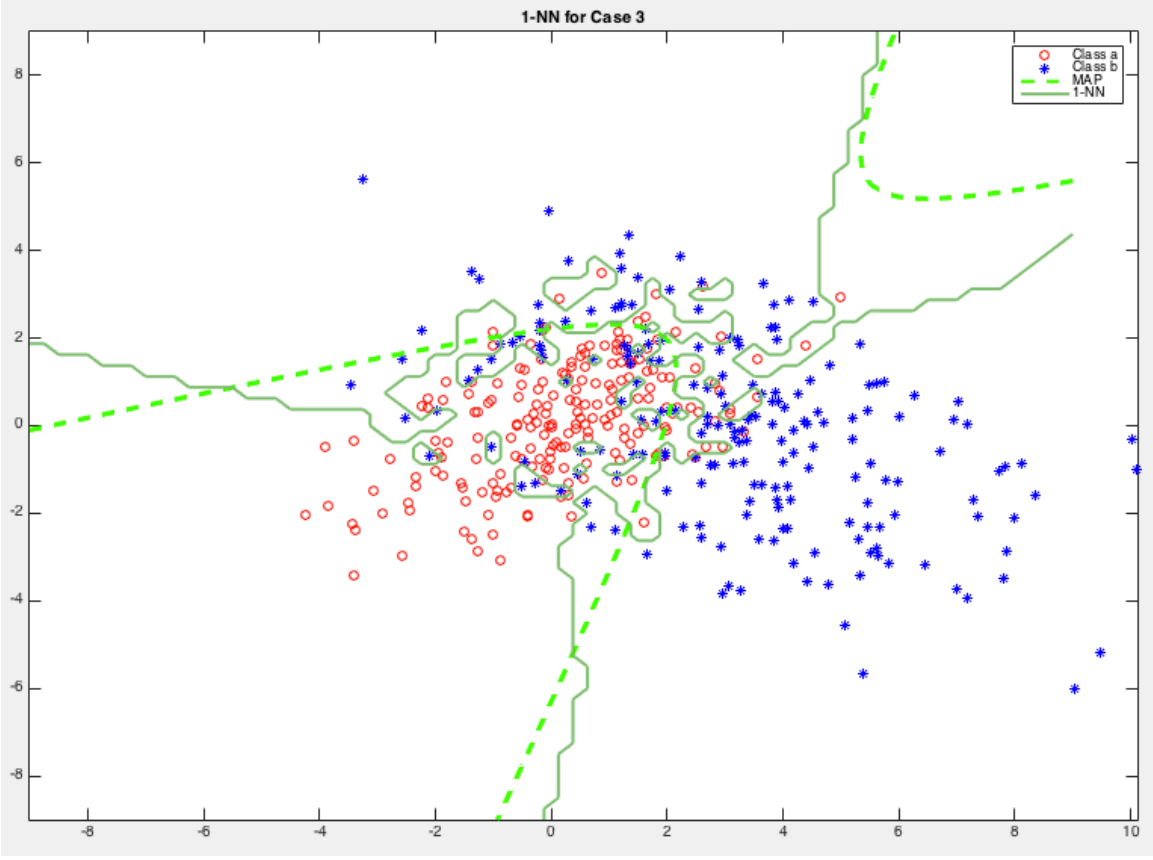

Case 4
Class a
Class b
MED
GED
300
250
200
150
100
50
200
250
300
350
400

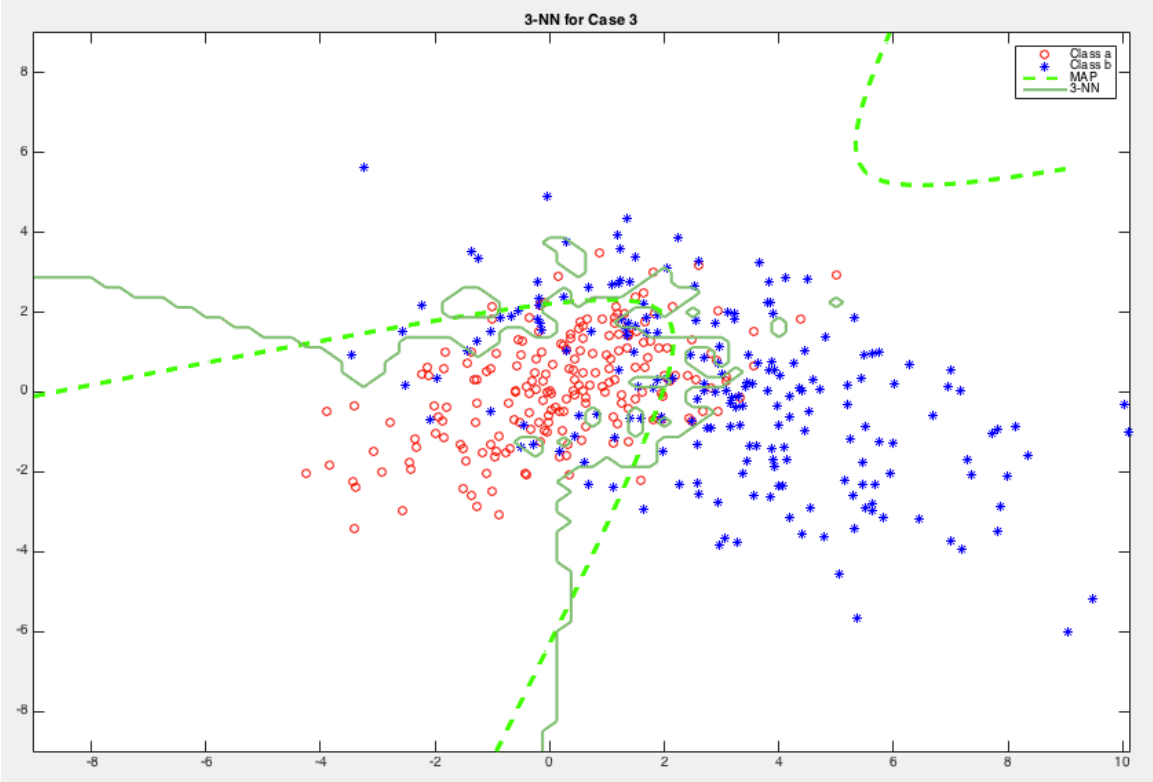
MED is dependent only with the mean of the data sets, so all the MED boundaries are straight lines in between the center of two data sets, no matter how the data points spread.

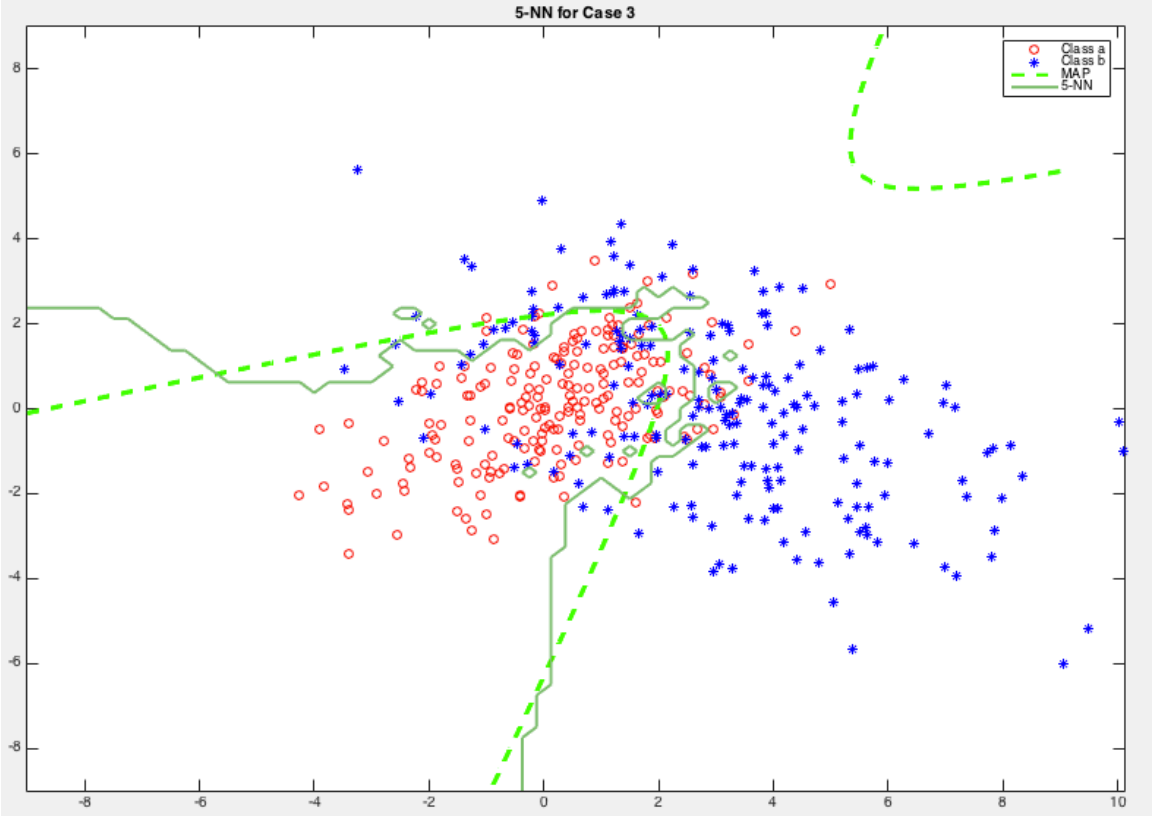
GED is related to covariance. With this classifier, the distances are determined by data's own covariance. The test points will be transformed by the class covariance, so the boundary could be a curve.

MAP classifier uses data pattern's posterior probability. When clusters are equally likely, 2 classes with the same covariance can have the same MAP and GED boundary. MAP classifiers also decide in favor of lowest and highest variances.

Set Two




3-NN for Case 3
Class a
Class b
MAP
3-NN


5-NN for Case 3
Class a
Class b
MAP
5-NN

1-NN boundary is determined by the distance between the test data point and the closest training data point with it. So that one outlier could make a huge difference on the classifier boundary. It is possible that one data point on the far upper right corner would label the whole area into its class.

3-NN classifier are more generous on the error. It uses the mean value of the 3 shortest distances, which avoids the impact of one or two outliers that are far from its own class.

5-NN classifier boundary is even more similar with the MAP boundary. It's smoother and more generous with the outliers. It's less sensitive to specific data points but focus on the whole pattern of the data sets.