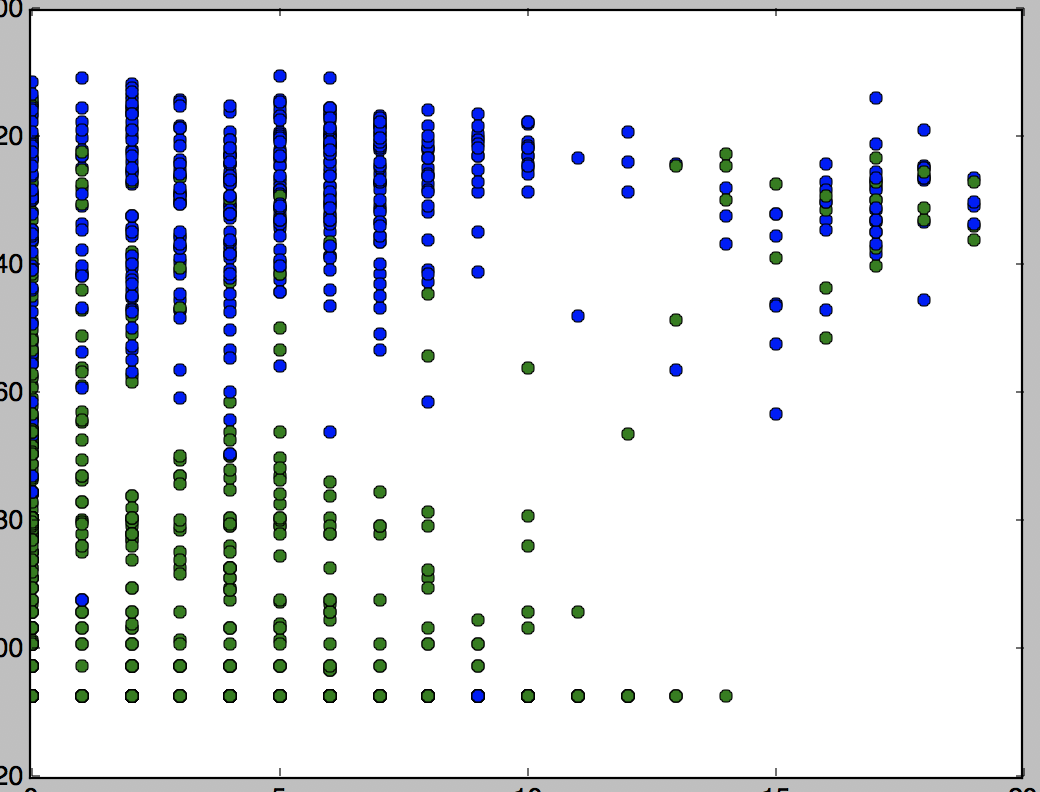
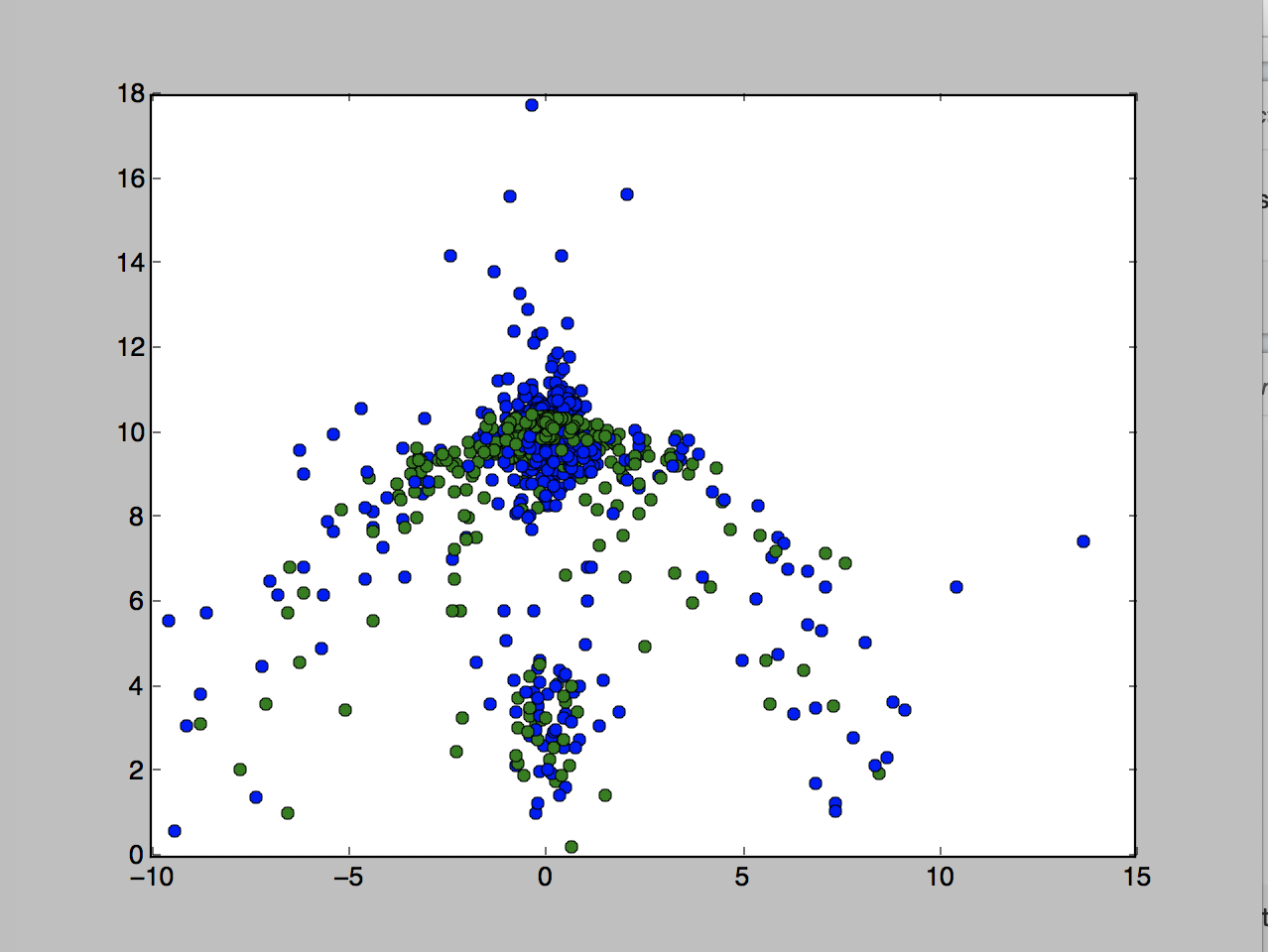
sTeam 6

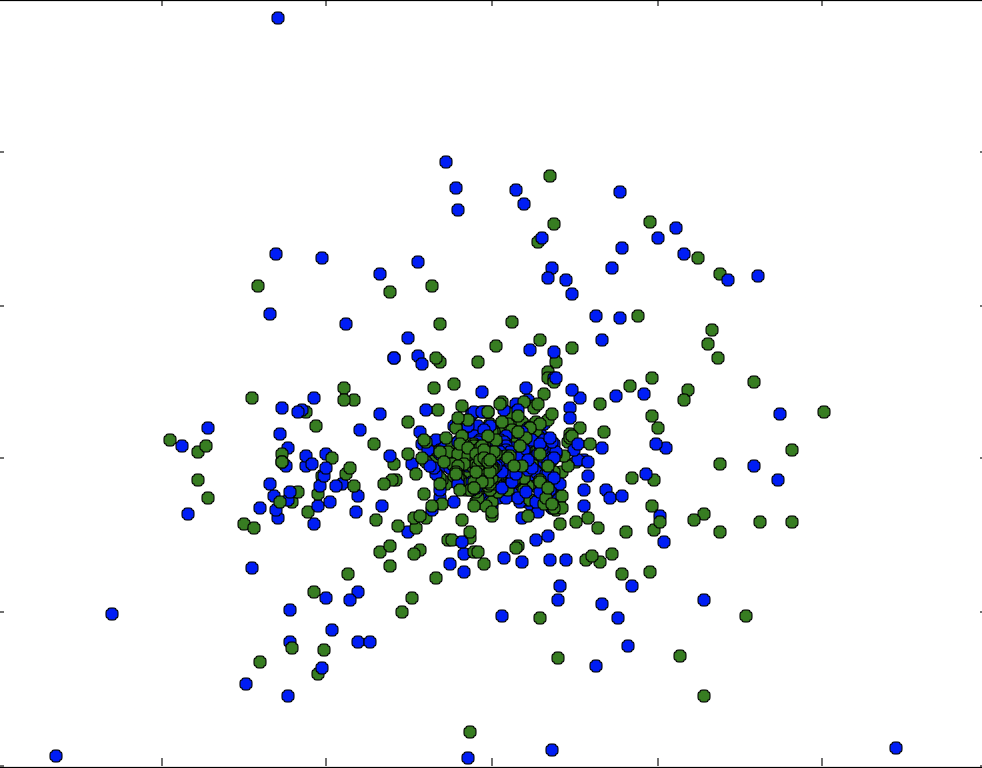
Part2 Write-up



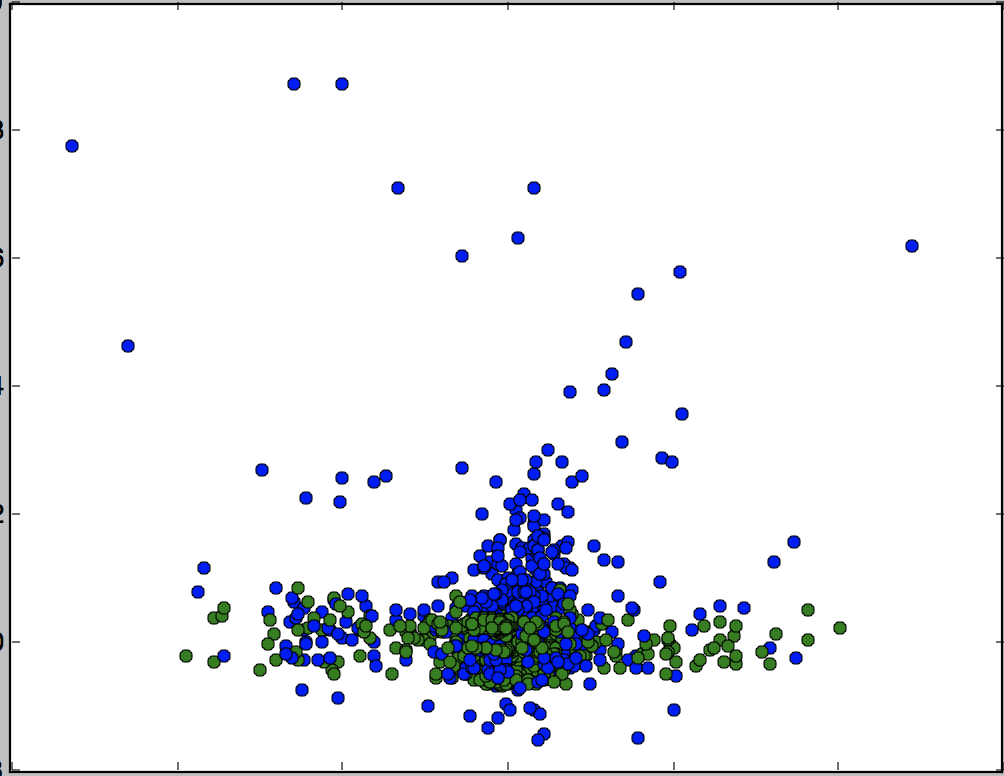
Zero Crossing X vs. entropy



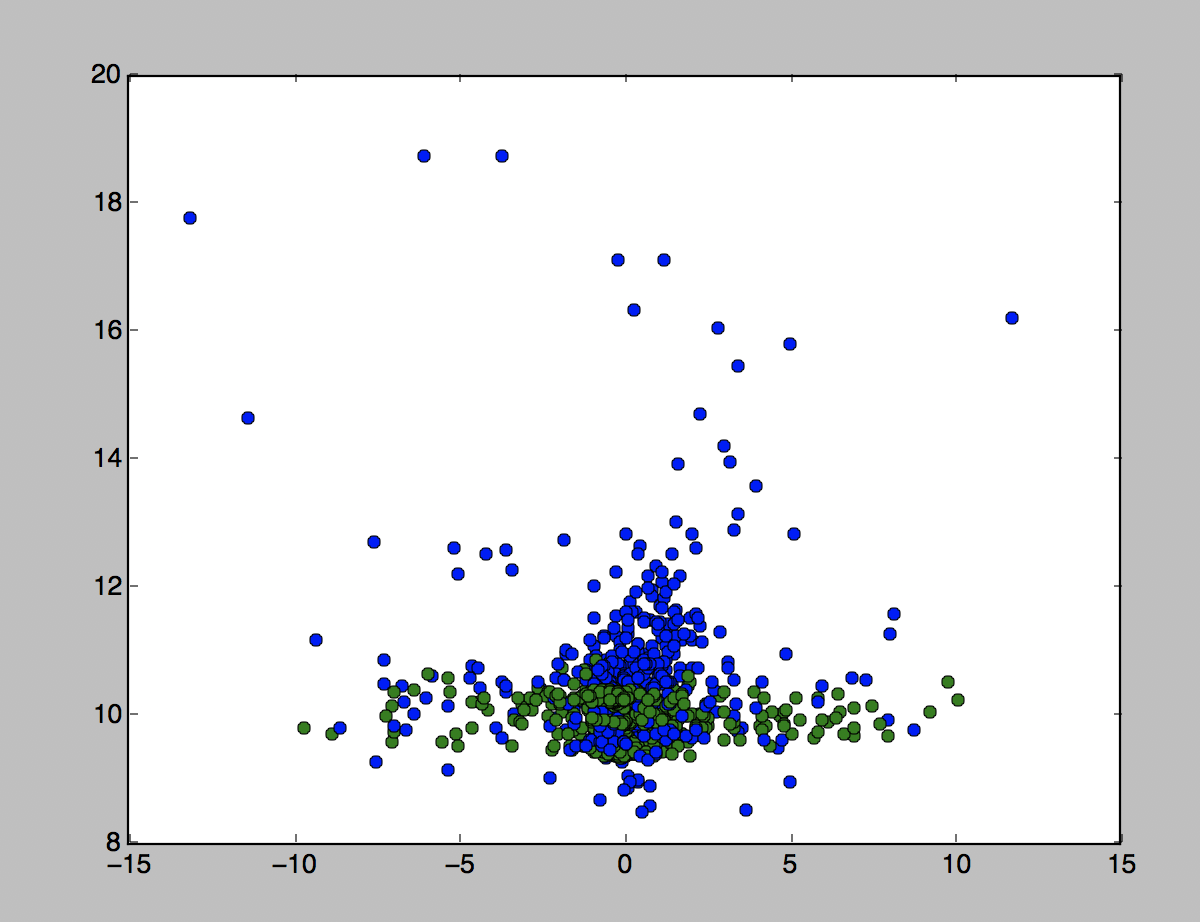
meanY vs. meanZ



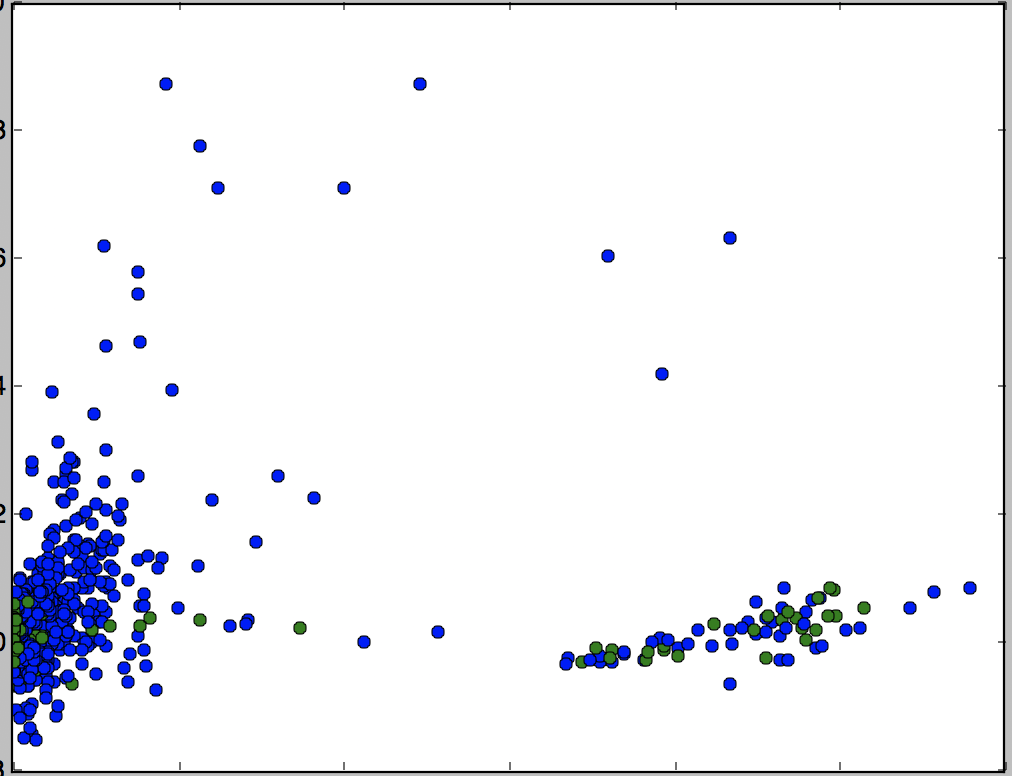
median X vs. median Y



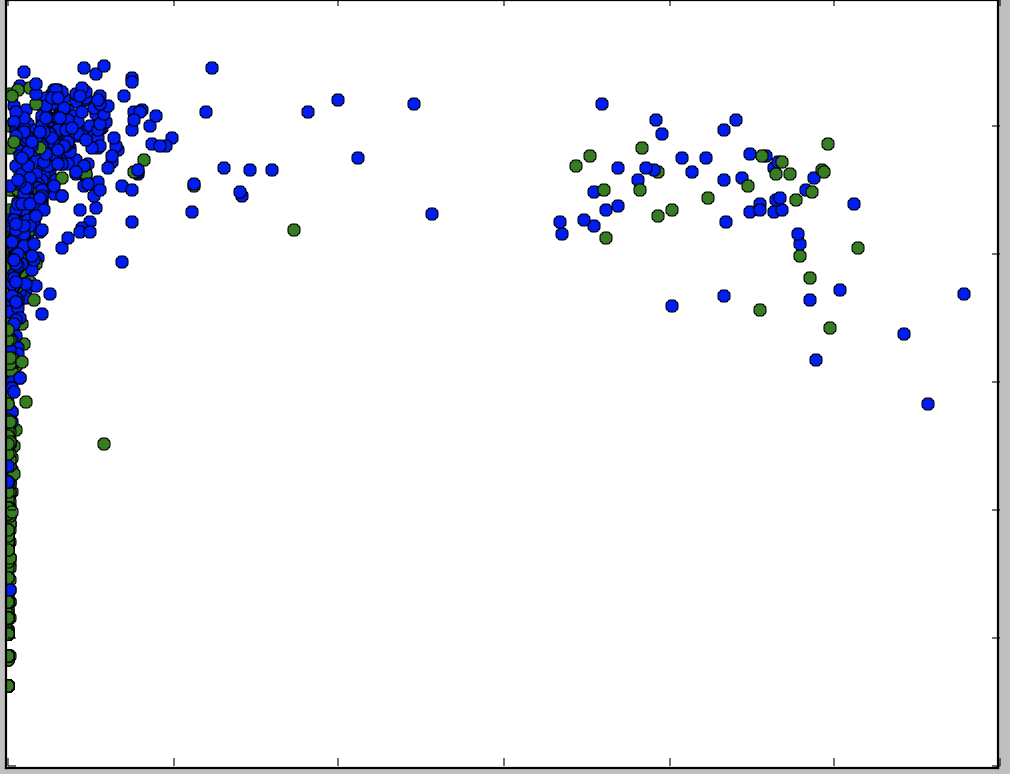
medianX vs. mean-magnitude



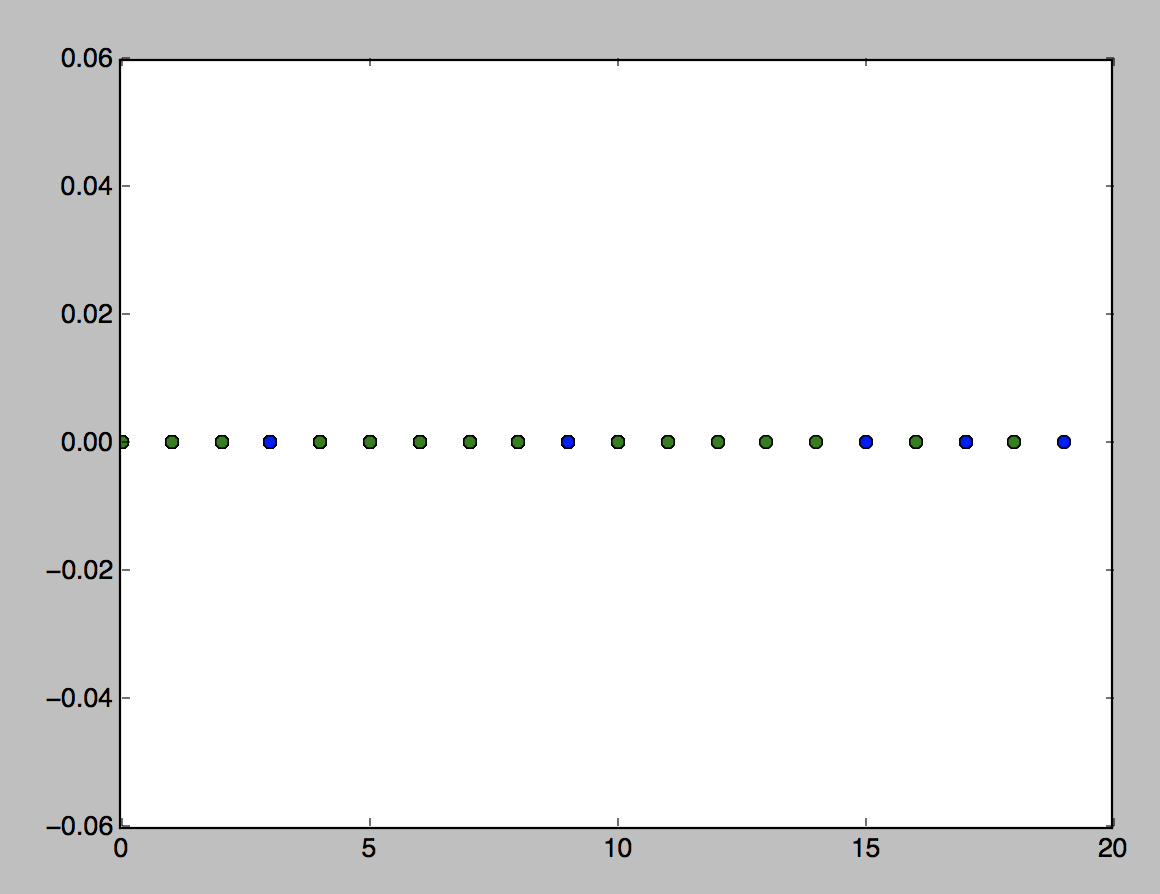
meanX vs. mean-magnitude



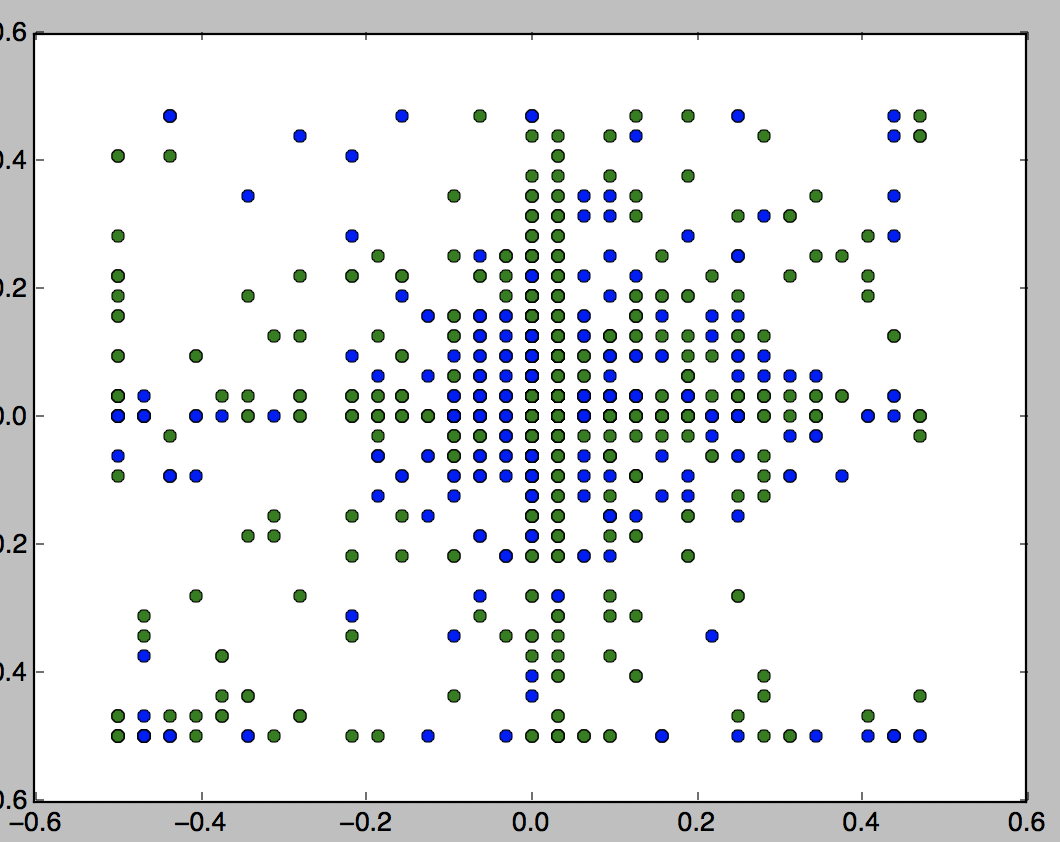
variancX vs. mean-magnitude



variance vs. entropy



zero Crossing X vs. peakx



FFTX vs. FFTY

Base on the visualization, the variance vs. entropy work best, since I can differentiate most of the point on plot by putting a line at the middle between 0.0 and 0.2. The one that works poorly is FFTX vs FFTY, because all the point on the plot just scatting around the graph, no way to separate those point by using one line.

**Part 3**

1. Train at least **four** decision tree classifiers varying the m ax\_depth and max\_features parameters. Evaluate each decision tree classifier and report the accuracy, precision and recall metrics averaged over 10-fold CV.

* MAX Depth = 3 & criterion = "entropy"

the average of accuracy is 0.946047678795

the average of precision is 0.932373237324

the average of recall is 0.942083980181

* Max Depth = 13 & criterion = "entropy"

the average of accuracy is 0.946047678795

the average of precision is 0.938712801121

the average of recall is 0.937819088152

* max\_feature = sqrt & criterion = "entropy"

the average of accuracy is 0.949184441656

the average of precision is 0.943264835306

the average of recall is 0.941992598312

* max\_feature = auto & criterion = "entropy"

the average of accuracy is 0.928481806775

the average of precision is 0.916724403214

the average of recall is 0.919305043774

1. Include the Graphviz image displaying at least two of these trees. Speak briefly to the visualization, comparing the trees.
2. Describe what effect each parameter has on your results. Why do you think that is?

Different parameter gives different results for accuracy, recall and precision, but they are basically pretty close. As I showed in question 1, I tried four features to calculate the average of precision, precision and recall. They are different, but the differences are really small based on decision tree. But when the max\_feature is equal “auto”, all averages become lower.

1. Is the decision tree classifier a linear or nonlinear classifier? Explain.

Decision tree isn’t a linear classifier, because it doesn’t have lines to classify the activities.

5. Train at least one other model and report the cross-validated accuracy, precision and recall metrics on the test data. Does it do better than the decision tree classifier?

I used SVM to train another data model. It is a little bit worse than the decision tree. However, I believe it depends which parameter we used for decision tree. When I try the max\_feature = sqrt. SVM performed much worse than decision tree.

Decision tree:

max\_feature = sqrt & criterion = "entropy"

the average of accuracy is 0.949184441656

the average of precision is 0.943264835306

the average of recall is 0.941992598312

SVM

the average of accuracy is 0.331869510665

the average of precision is 0.414649681529

the average of recall is 0.494291647807