**Signals, Patterns & Symbols: Coursework One**

Themes to perhaps be mentioned in the introduction:

* Classifying versus clustering.
* 5 different features in an unspecified no of object classes
* 150 points in training vs 15 in test (10%).
* Perhaps mention significance of this figure, too small/big?
* Cluster points to get class labels to train classifier.
* Explain differences between data sets.

Week 17:

Part a)

* How did we identify features? Include plot matrix.
* Explain which features **best separate classes** and why we chose them. What did the two columns and three point clusters of data mean in reference to features and classes.

Part b)

* KMeans algorithm - an explanation of how it works?
* Why did we parameterise k = 3.
* Justify whether visual inspection is good enough to set k = 3.
* We used KMeans to generate the class labels and generate centroids. Explain our plot in reference

Part c)

* We used KNN to classify test data, how effective was it. Did it generalize well, overfit or underfit? What would have been better and why? Do we theorize maximum likelihood to be better.
* Explain why the Voronoi tessellation boundaries are the way there are

Part d)

* Which techniques did we employ to disrupt normal operation of k means:
* Did we identify any other methods to break k means. Dima mentioned there were 3
* 1. Difference of starting points of centroids
* 2. Changing the number of iterations
* Is it a local optima for parameter 1)
* Explain why our technique worked. What about the Voronoi plot.
* Show the graph.

Week 18

OBJECTIVE: Compare nearest centroid classifier (simple) with maximum likelihood

Part a)

* Mean/covariance describe size of our covariance matrices why are they this big. Visualise and show ellipse graph. Explain mahalanobis distance value of 6, why 6?

b) Show the decision boundaries graph. Why do they have the shape they have. Systematically go through each line and explain why the decision boundary is there. Do they agree with the classification that

We should demonstrate a clear understanding of the relevance of TEST versus TRAINING DATA.

In order to demonstrate the quality of the different classifiers, which of the classifiers generalized well i.e. did not under or overfit on the TEST data, ignoring the training data.

Let’s examine the ML classifier versus the KNN/Kmeans