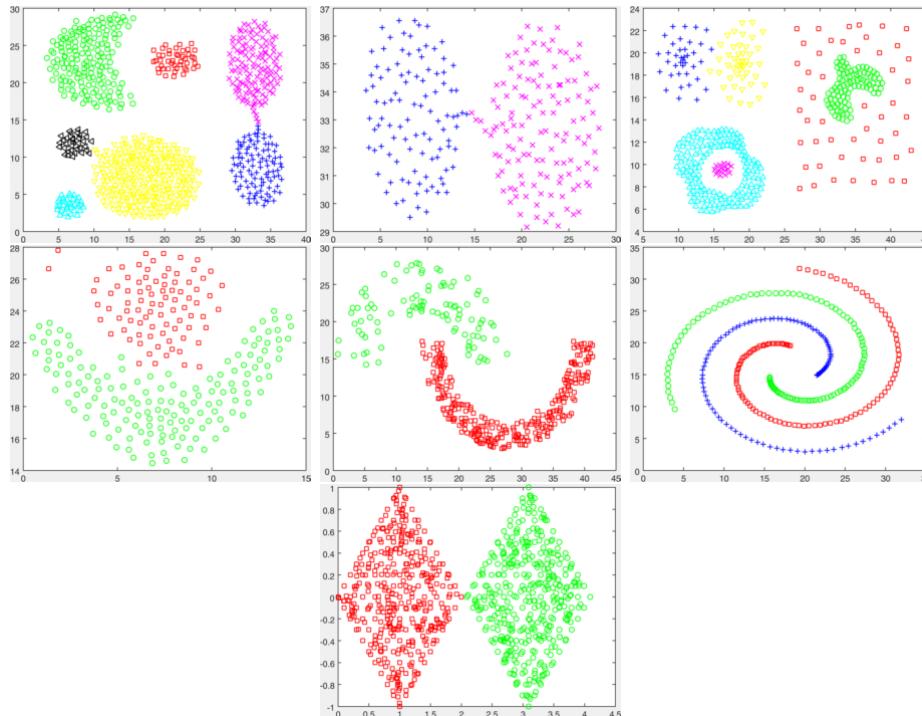# 11-755: MLSP
HW #3: Clustering and EM

Nikolas Wolfe

December 15, 2015

Problem 1: K-Means & Spectral Clustering

You are given a number of toy datasets. The visulized groud truth clustering configuration of these data: Aggregation (K=7), Bridge (K=2), Compound (K=6), Flame (K=2), Jain (K=2), Spiral (K=3) and TwoDiamond (K=2) are shown as follows:

Answer: I implemented this code in Java, using the Efficient Java Matrix Library (EJML) package to do matrix multiplication and Eigen/Singular Value decomposition. In general, spectral clustering works a lot better than K-means and is able to determine a lot of structure (such as the spiral pattern) that k-means is unable to find.

In terms of evaluation, I ran the clustering algorithms and generated the confusion matrix between the most common label that emerges in each cluster and their actual labels. A diagonal matrix in this case is of course perfect performance. The accuracy can be determined from looking at the counts of labelled points for each class.

Problem 2: Expectation Maximization