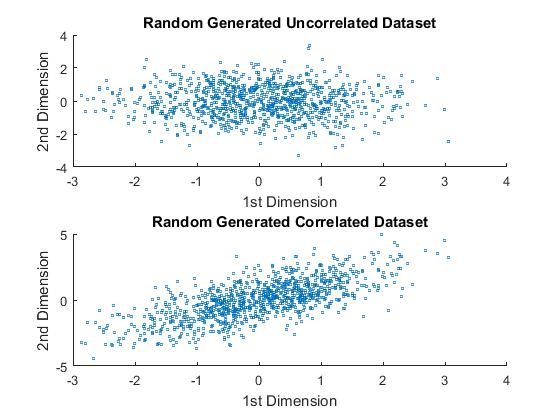
ECSE 517 Assignment 1

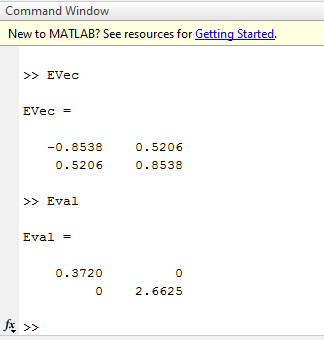
Sort Action Potential Waveforms using Principal Component Analysis(PCA)

# Introduction

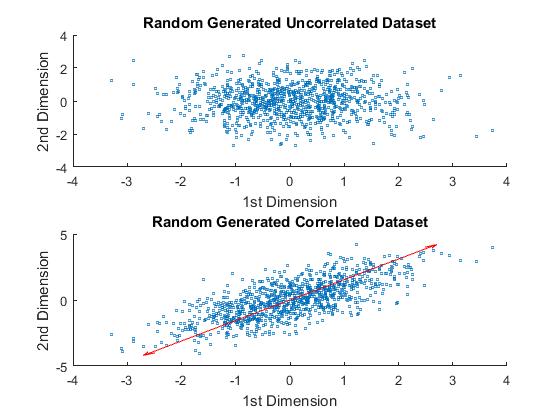
Two random sets of data were generated, one correlated and the other uncorrelated. The plots of each are provided below.



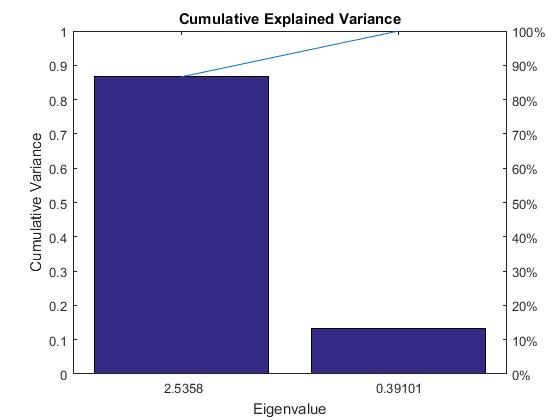
We find the eigenvector and eigenvalues of the covariance as follows –



The principal component is found to be 2.6625 as it’s the largest Eigenvalue. The plot with this (magnified) eigenvector is shown below.



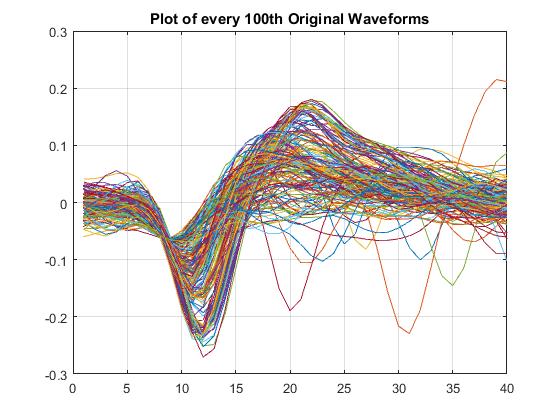
The cumulative explained variance from a pareto plot of the data is shown below –



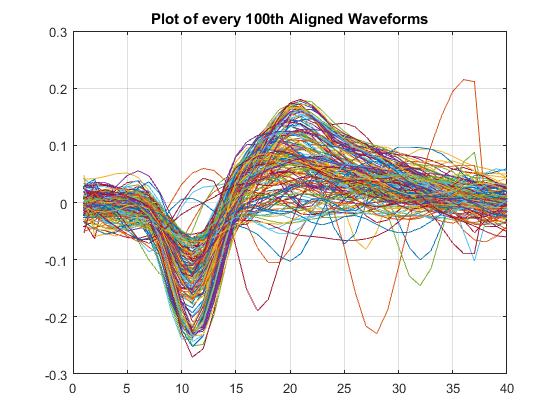
# Part A: Dimensionality Reduction

## Clean up data for Spike Sorting

The plot below shows every 100th Waveform produced from the raw data –

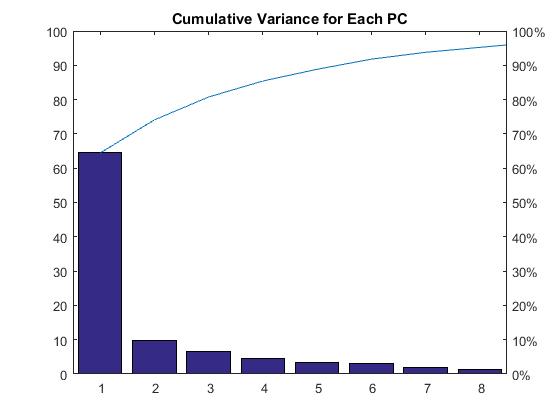


The mode for the minimum value was found to be 11. After carrying out alignment by finding the mode of the minimum of each wave and shifting accordingly the following output is observed -

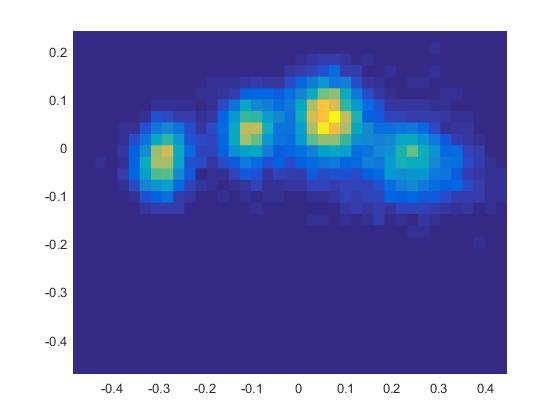


## Run pca on Awaves

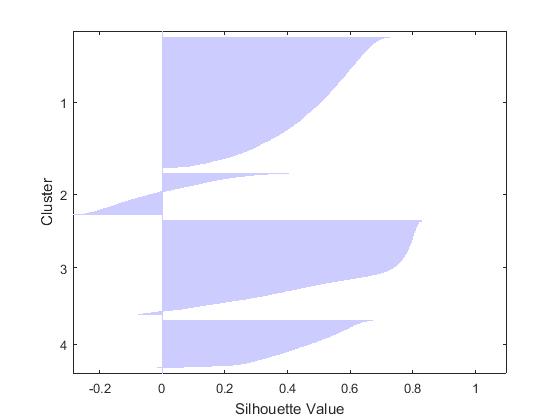
1. Below is a plot of the explained cumulative variance accounted per principal component -



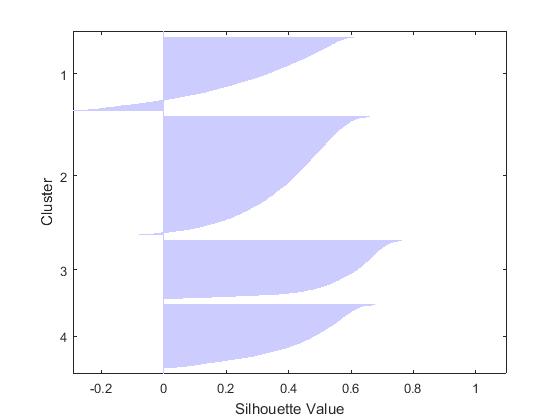
1. By 6 PCs, 90% of the variance is accounted for. By 8 PCs, 95% of the variance was accounted for. 23 PCs count for less than 0.1% of the variance.
2. The figure below shows us a plot of PC1 vs PC2. We can see here that there are 4 clusters.



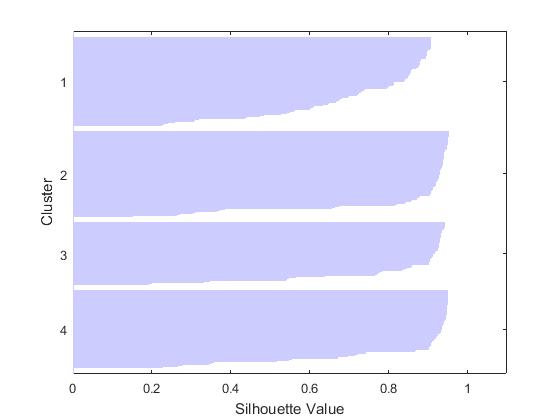
1. The plot of the waveform groups is as shown below.



Previously, we counted 4 clusters. Here we can see the four groups returned by kmeans. In this plot the values range from +1 to -1 were +1 indicates points distant from neighbouring clusters, 0 means points are not particularly in one cluster or another and -1 indicating that points are in the wrong cluster.



If we do the grouping of the Tw data using kmeans as well, we obtain the following plot –



From the above plot we can see that almost all the data points are in clusters and are very well separated from other clusters’ points as the Silhouette values are all close to 1. It also supports our second theory of having 4 clusters in our dataset.