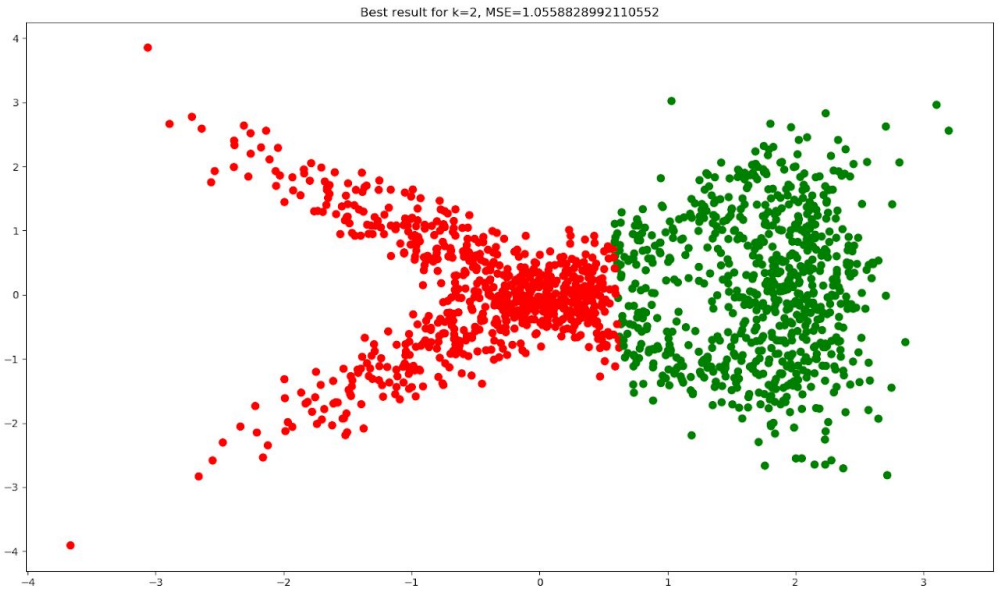
Charles Bolton
Machine Learning

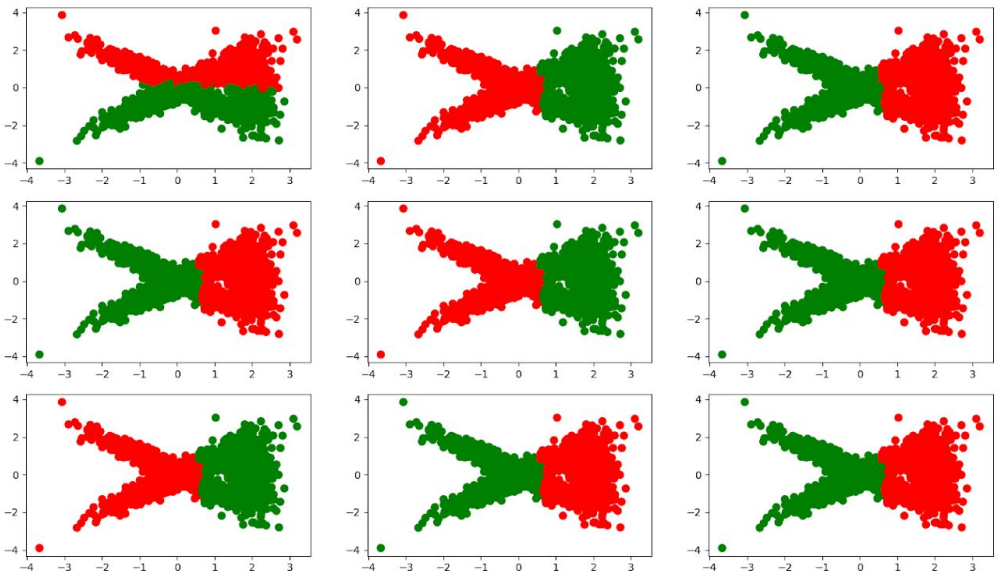
K-means and Fuzzy C-means Clustering

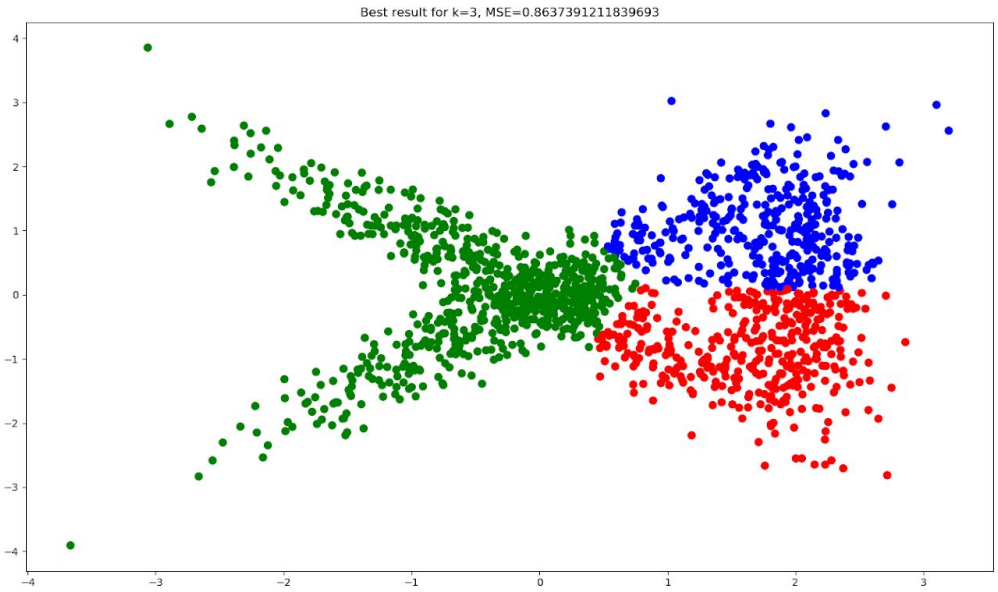
K-means:

For this program, I prompt the user for a number between 1 and 9 for *r*, which is the number of initializations for the algorithm. The algorithm then performs k-means clustering for values of k from 2 through 8. I have provided plots below showing 4 different things. For each value of *k*, there are nine plots for the final converged clustering. Then, from these nine, the plot with the lowest MSE is also highlighted. Finally, I took the average MSEs of each of the *r* initializations for each value of *k* and also plotted these results, which shows, rather unsurprisingly, that the average sum of squares error gets smaller with larger values of *k*.

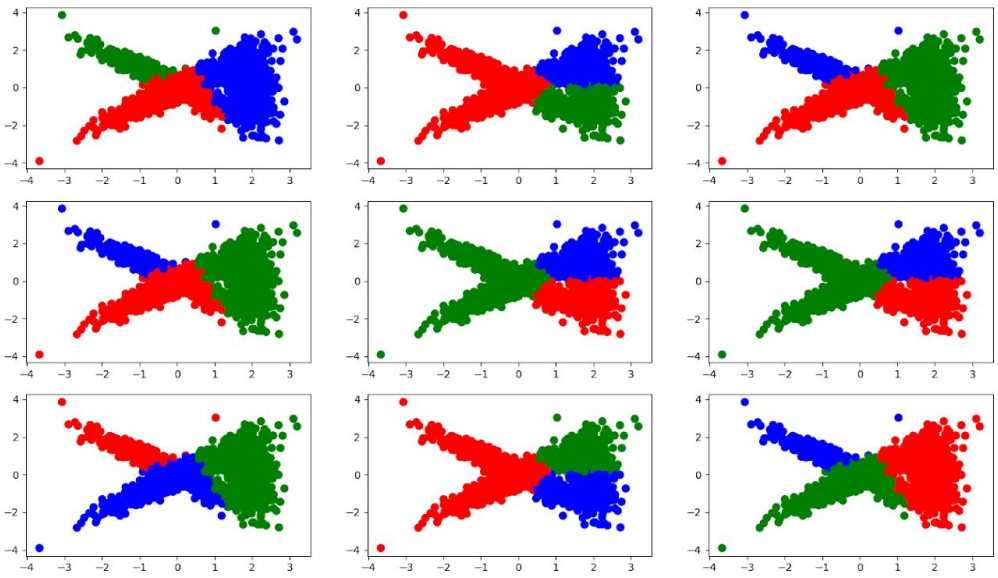


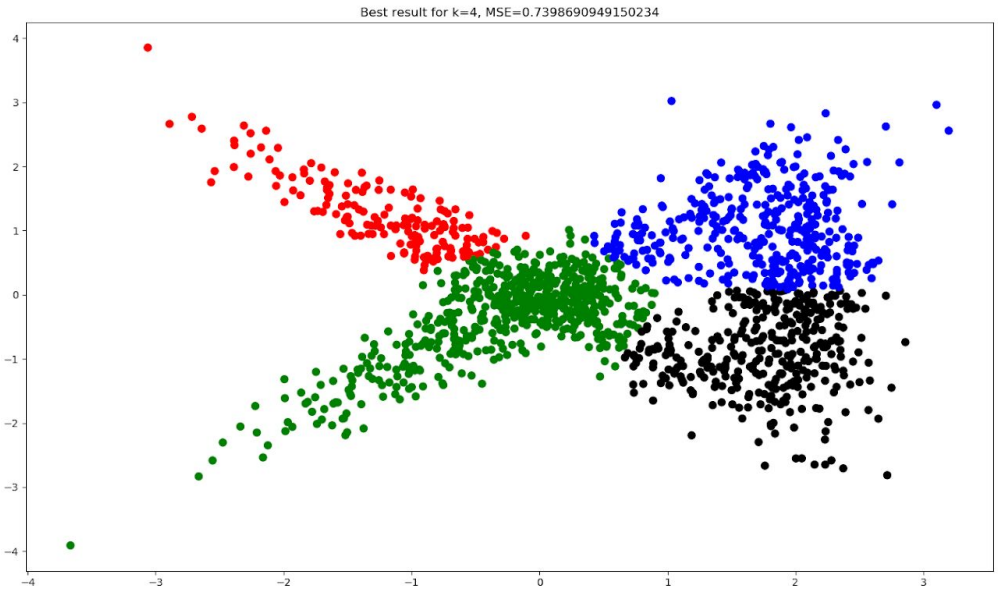
K-means clusters for k=2



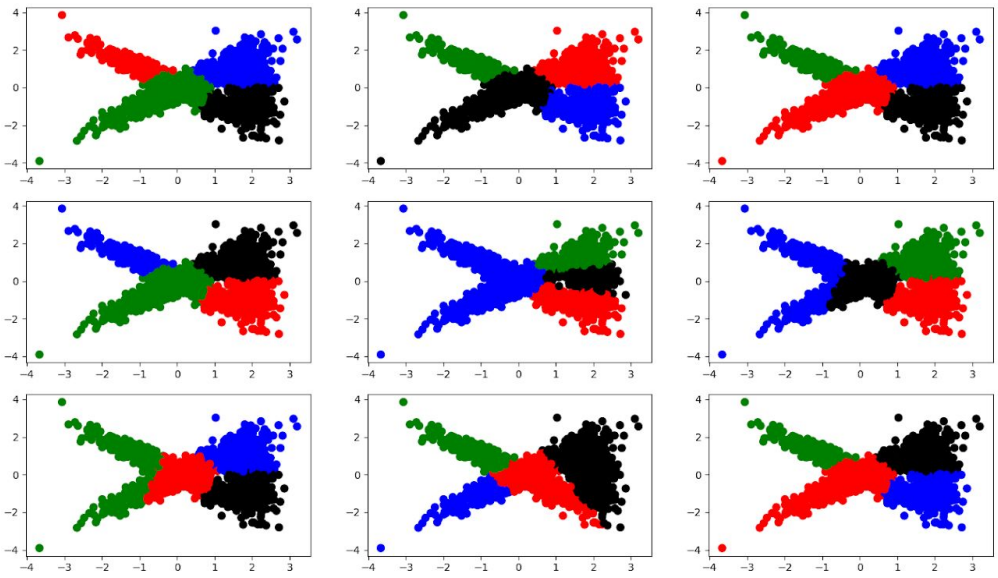


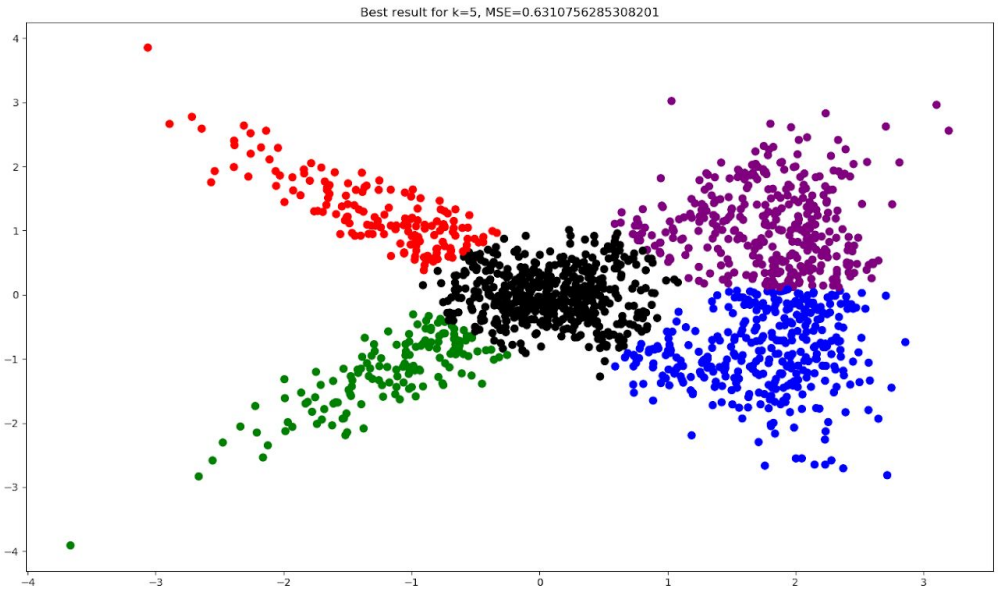
K-means clusters for k=3



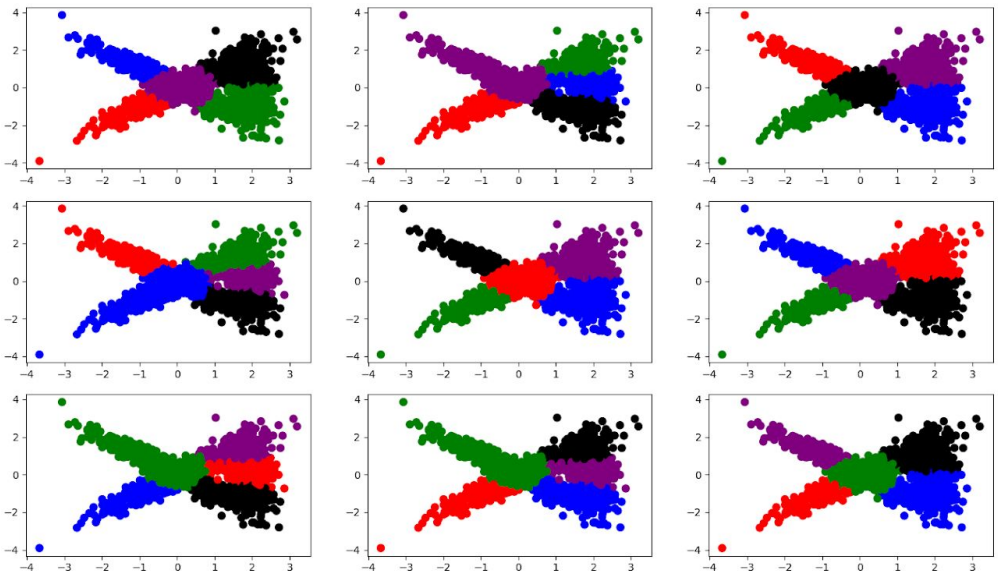


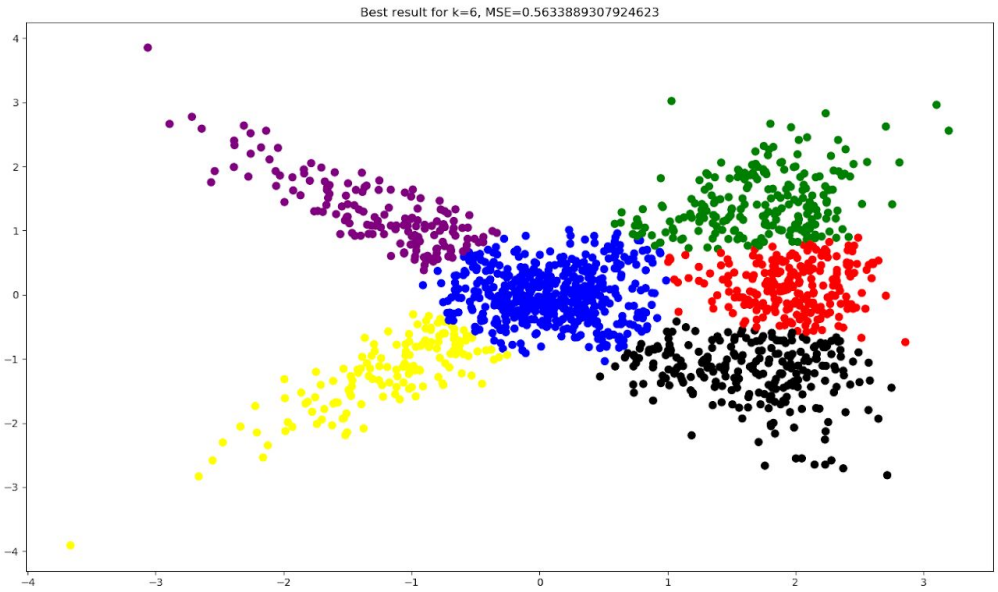
K-means clusters for k=4



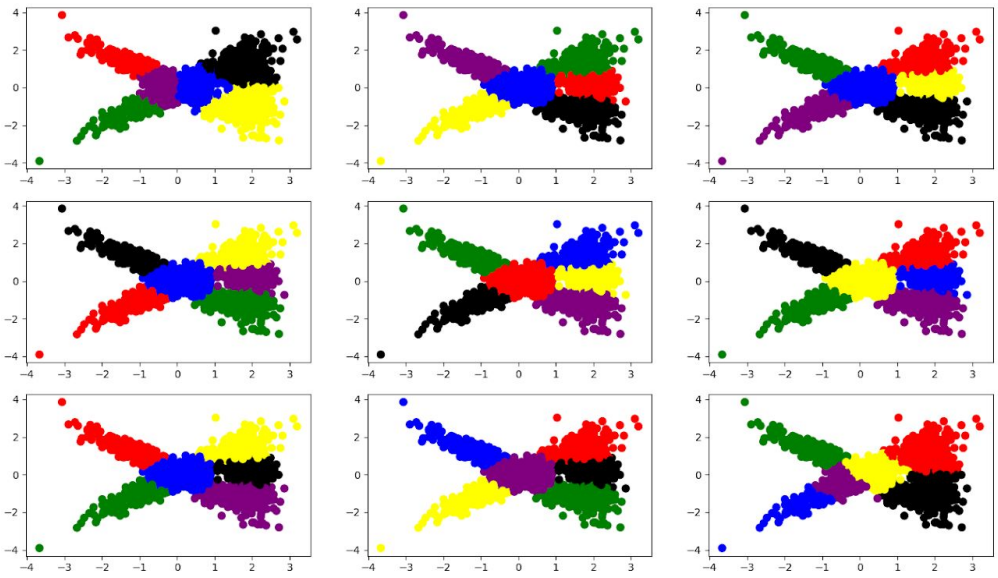


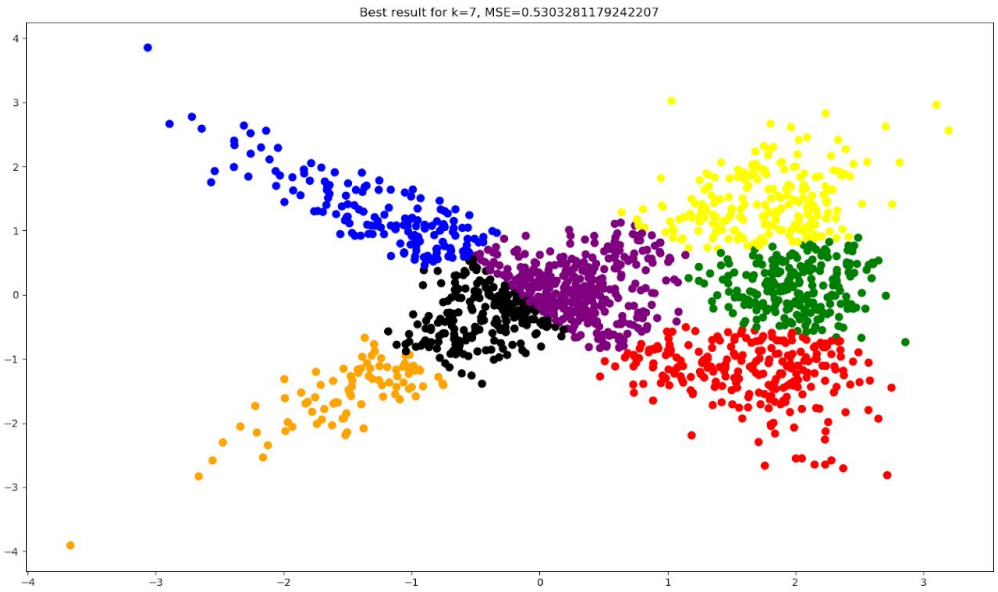
K-means clusters for k=5



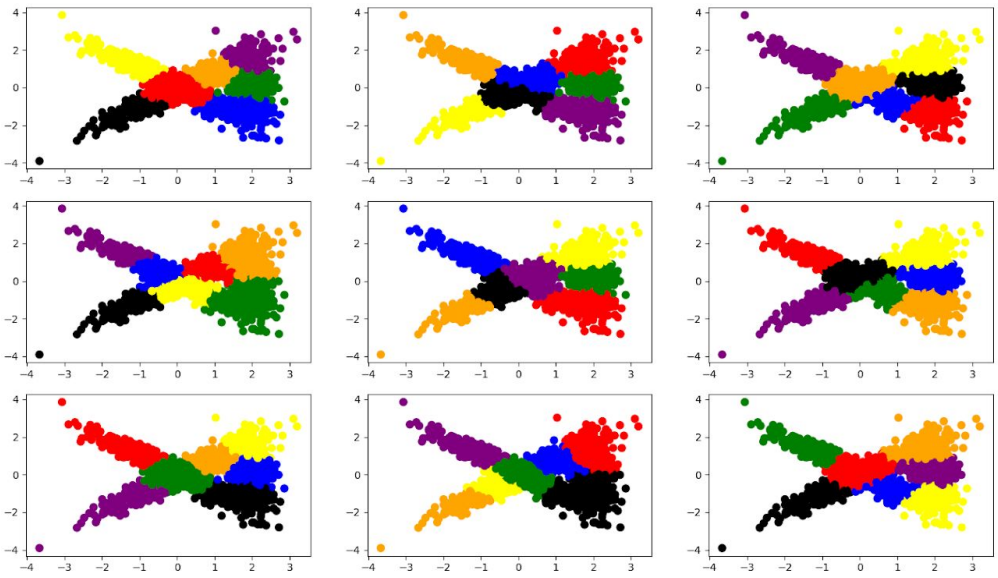


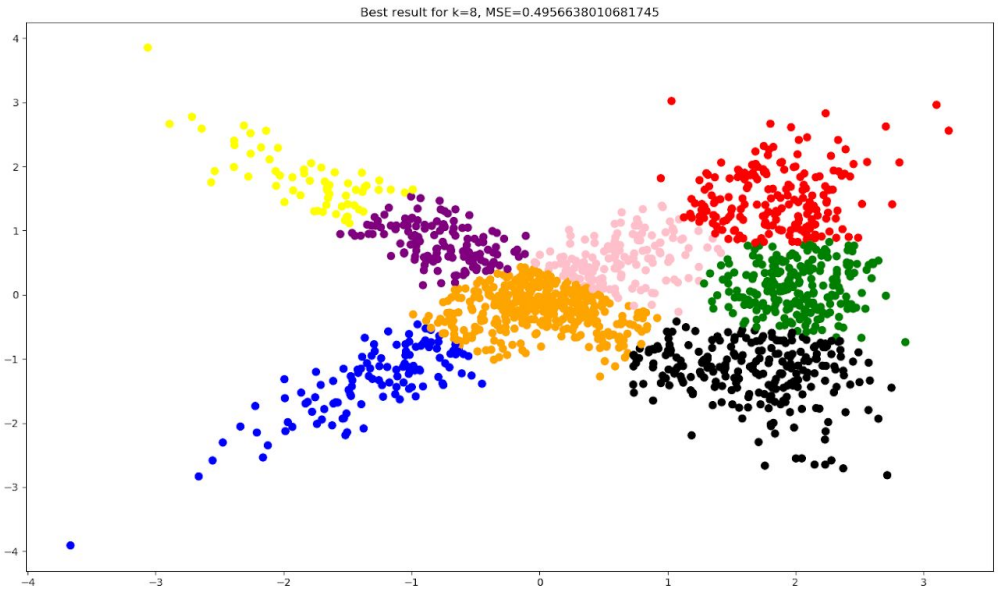
K-means clusters for k=6



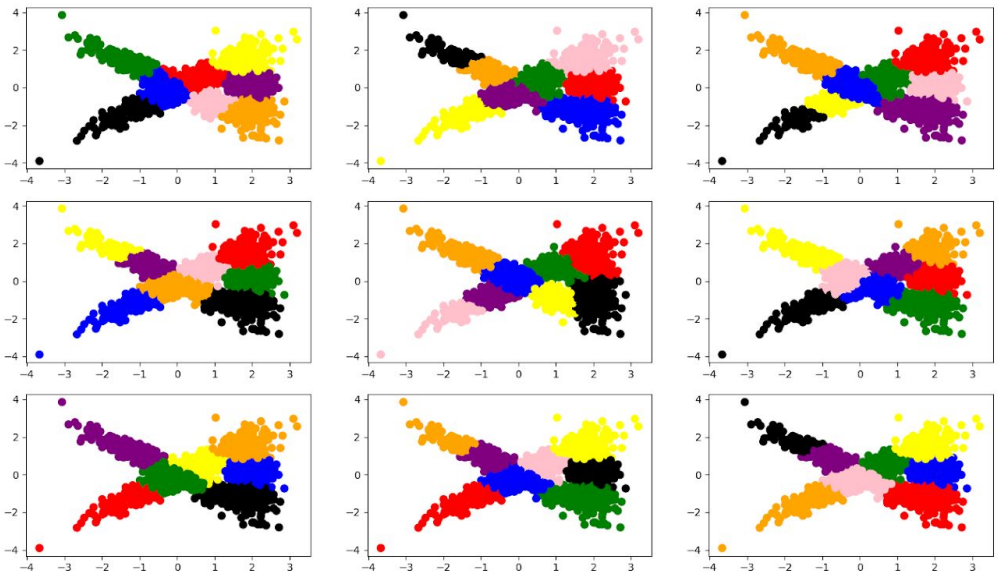


K-means clusters for k=7

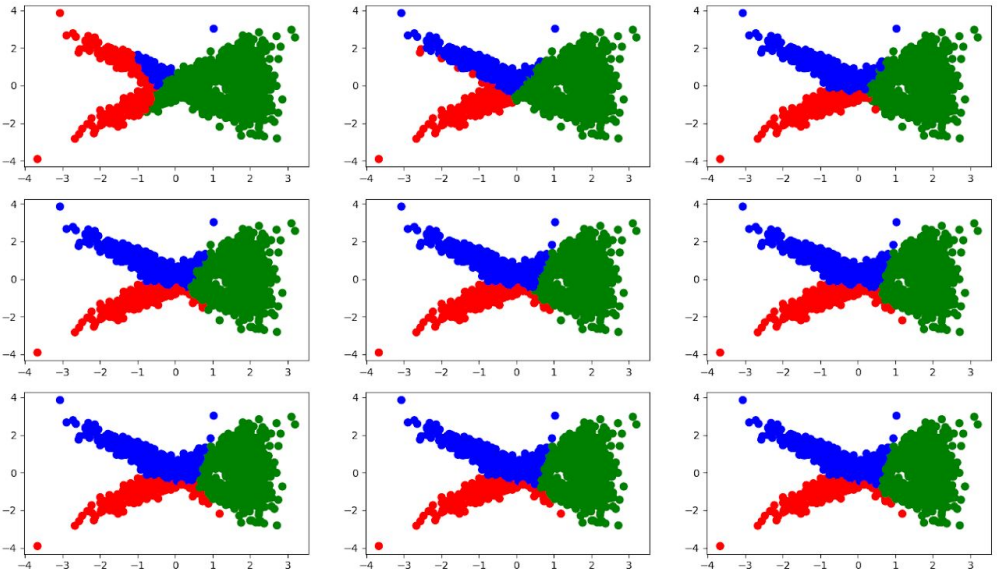




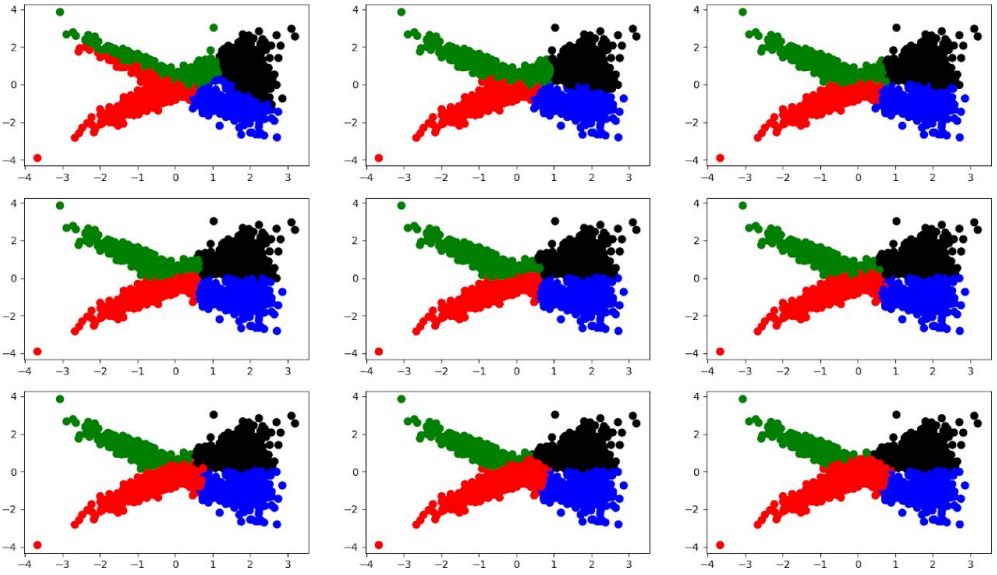
K-means clusters for k=8



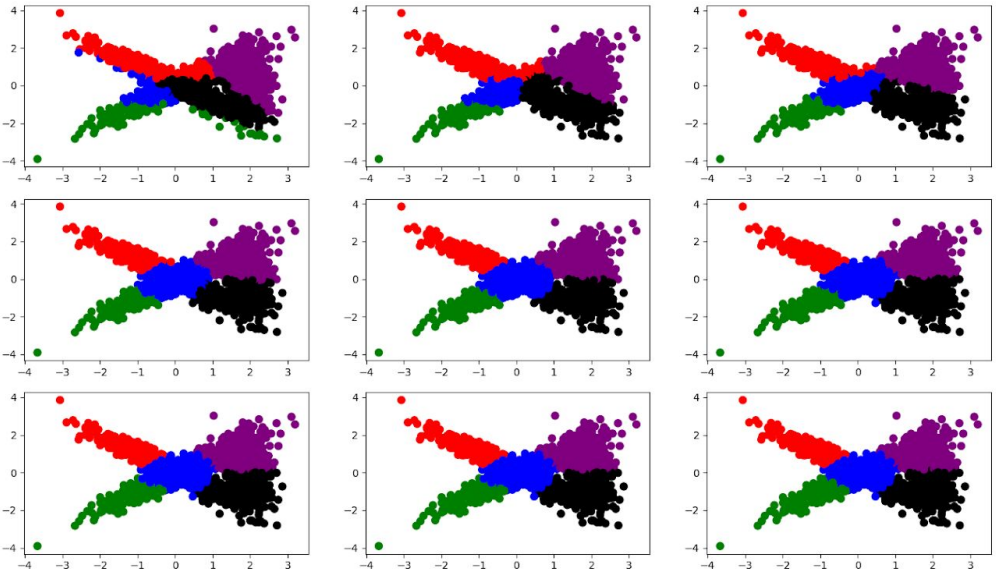
K-means clusters for successive iterations



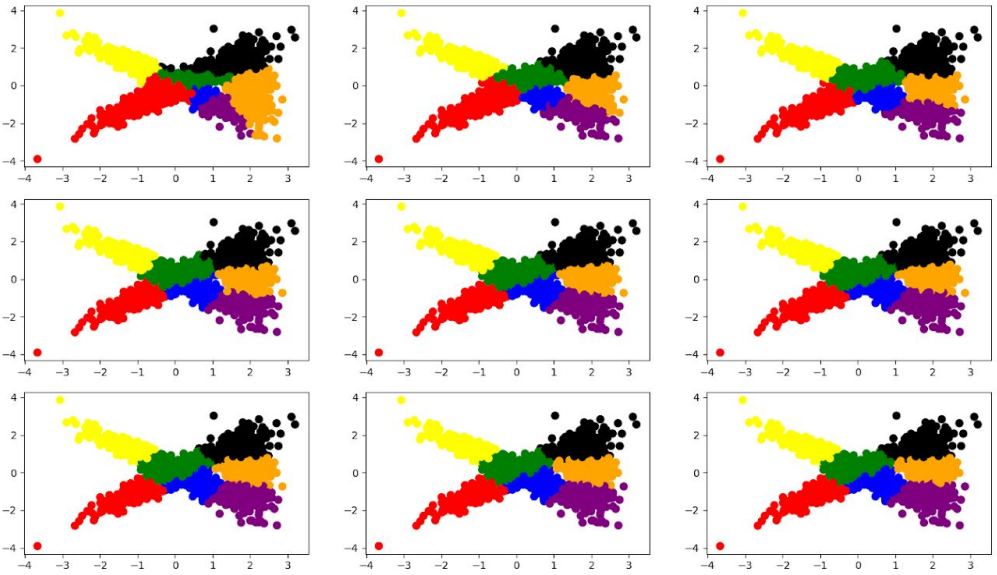
K-means clusters for successive iterations



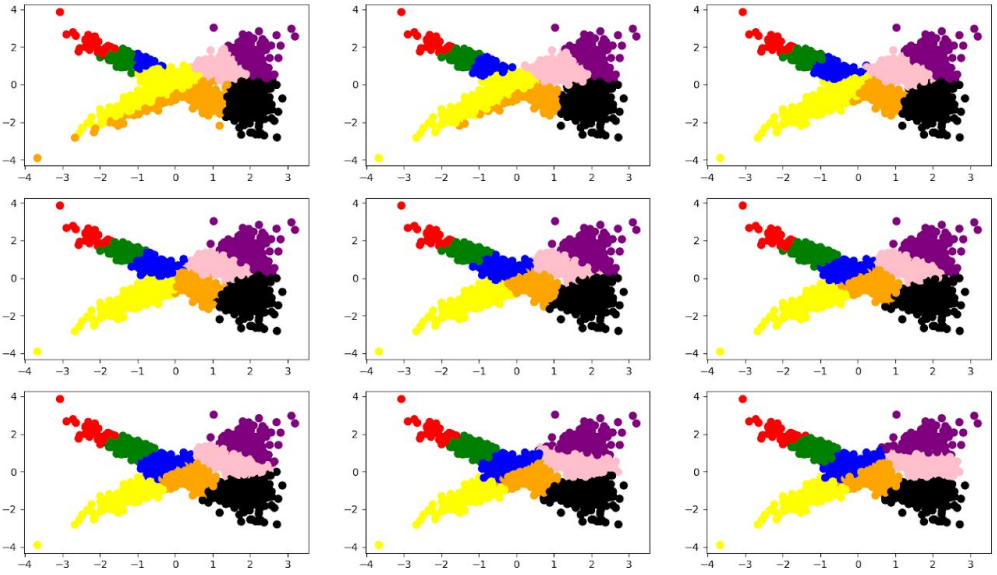
K-means clusters for successive iterations

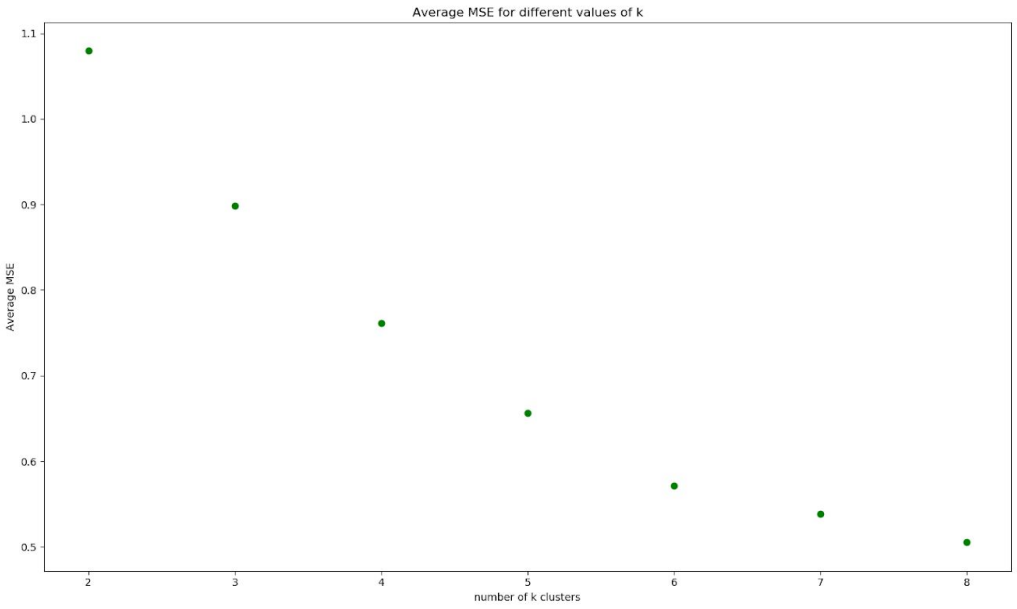


K-means clusters for successive iterations



K-means clusters for successive iterations

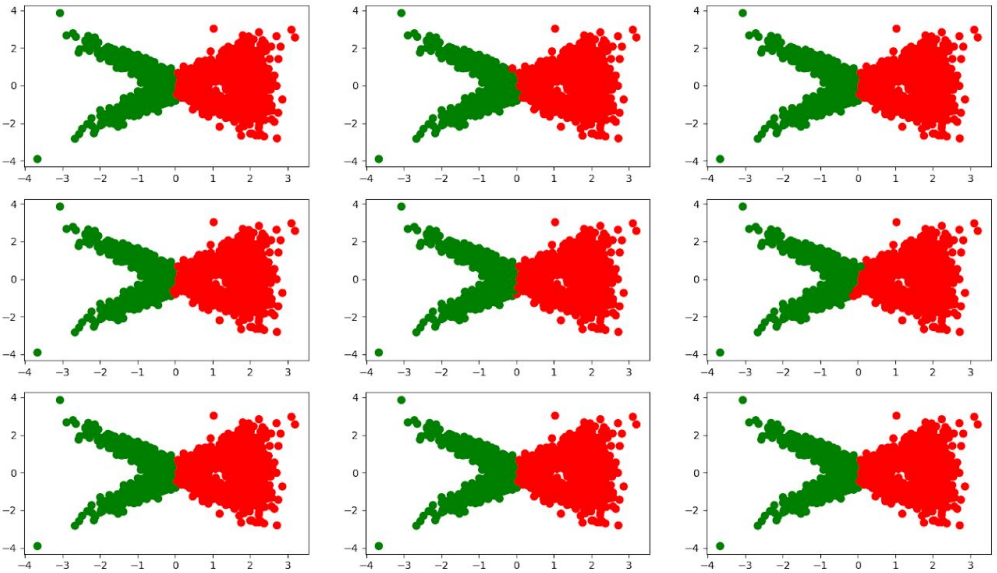




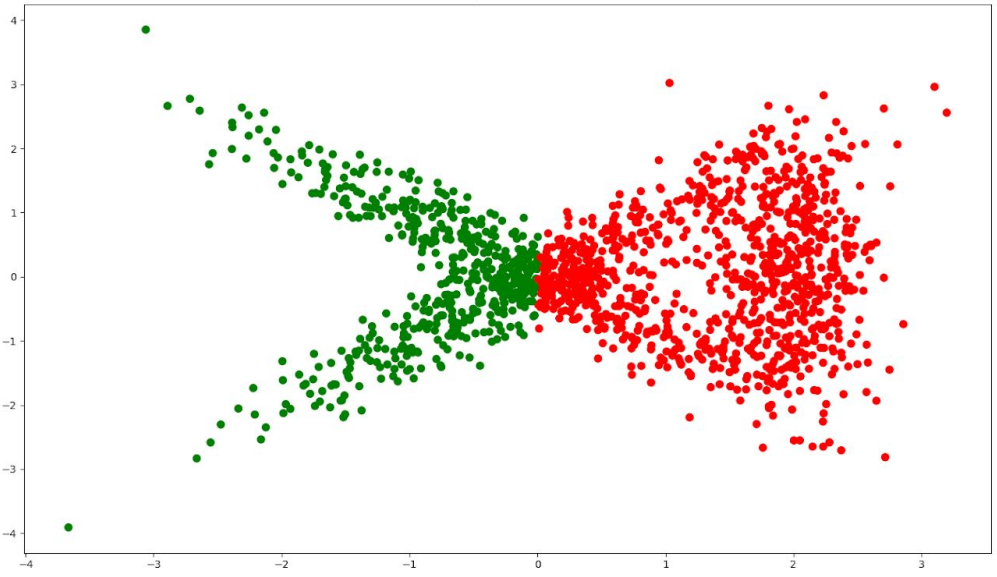
Fuzzy C-Means:

For Fuzzy C-Means I noticed that the average MSE did not go down as the number of clusters increased. In other words, the average error was about the same for any number of clusters. One thing I did notice was that sometimes one or more of the centroids tended to be “pushed” out, meaning that the weights on the data points were pulling none of the points in the direction of certain centroids. I’m not sure if this is a problem, but it tended to make some of the clustering scenarios have fewer than the specified number of clusters since certain centroids overpowered the other centroids.

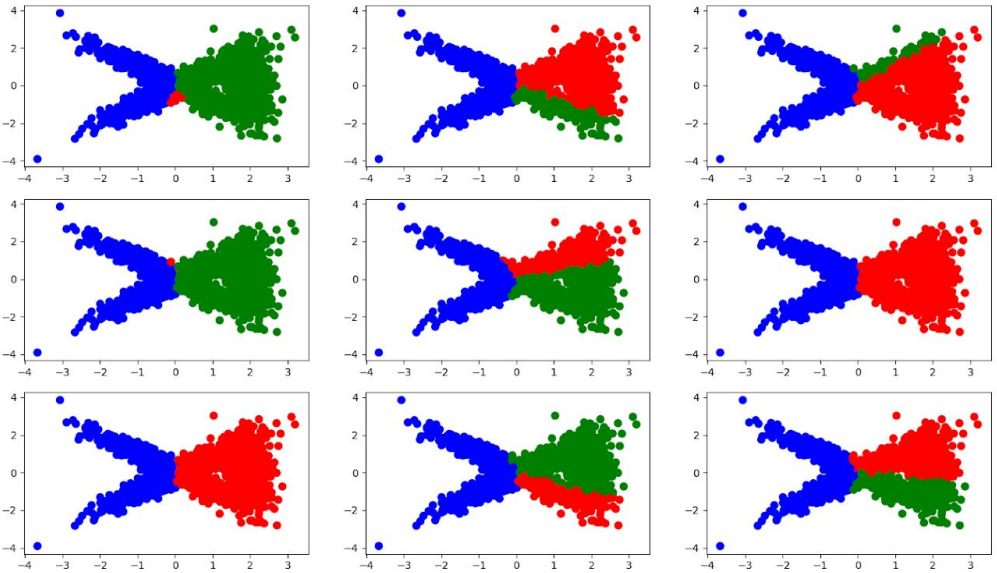
Fuzzy-C-means clusters for k=2



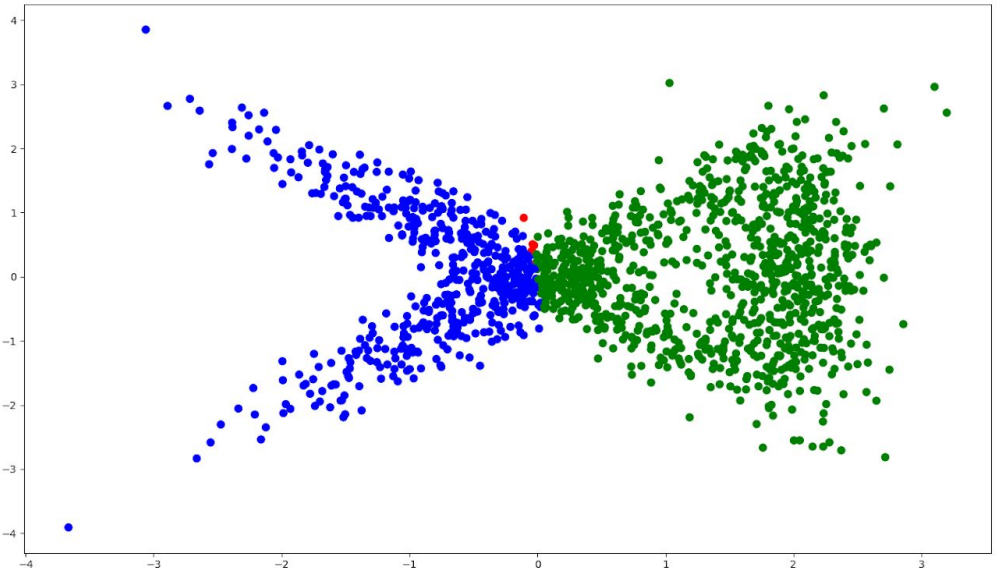
Best result for k=2, MSE=1.5238073950259323



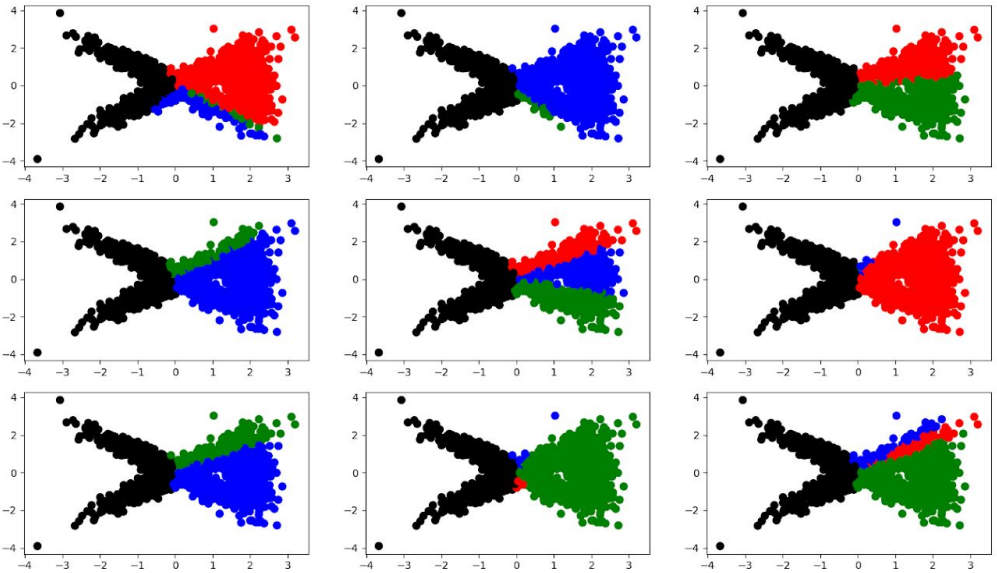
Fuzzy-C-means clusters for k=3



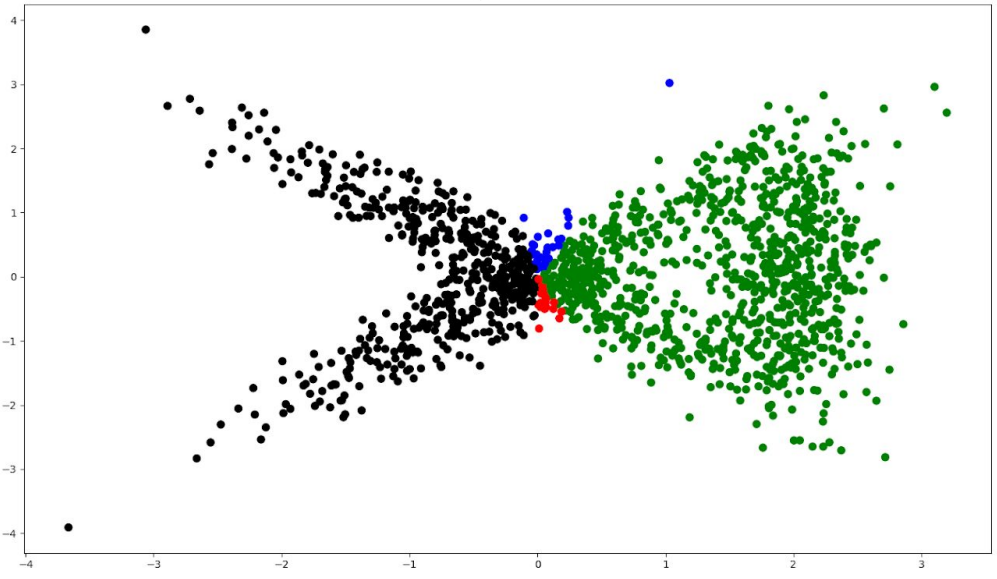
Best result for k=3, MSE=1.523784894009418



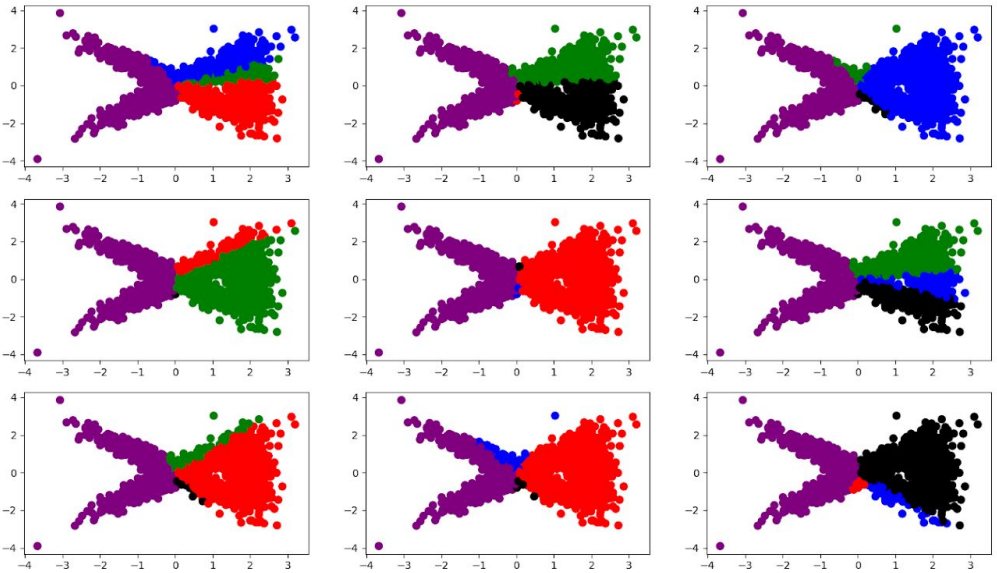
Fuzzy-C-means clusters for k=4



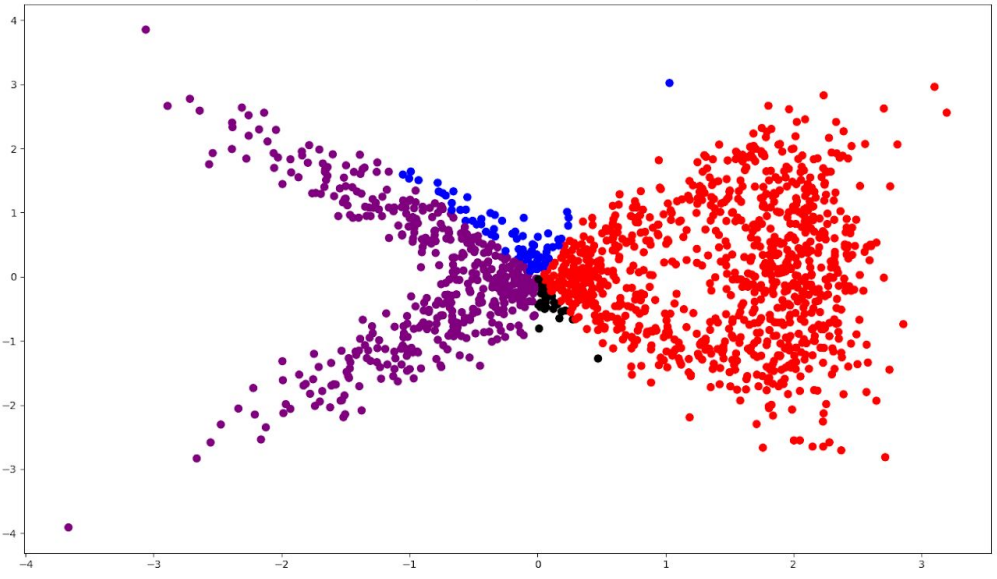
Best result for k=4, MSE=1.523792653529055



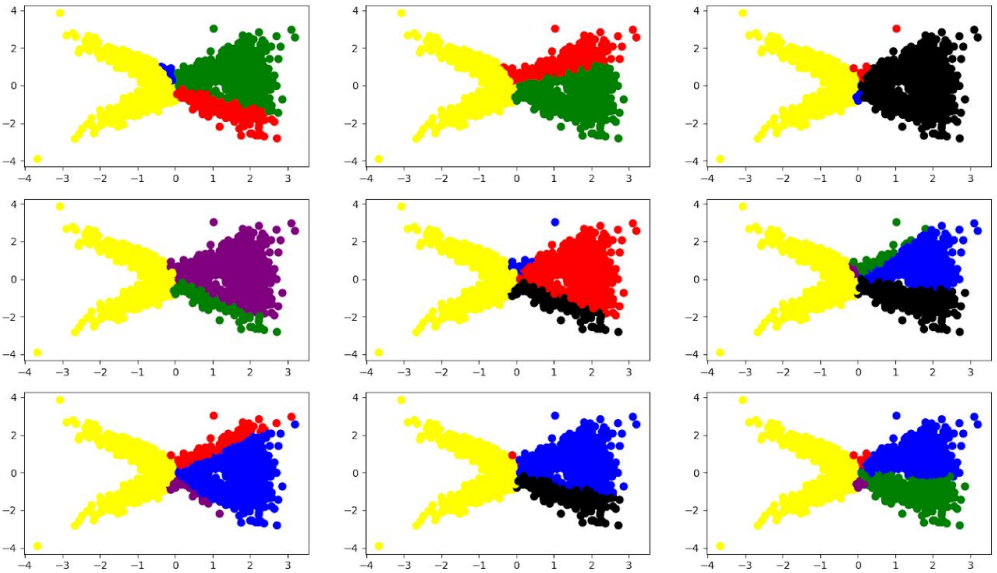
Fuzzy-C-means clusters for k=5



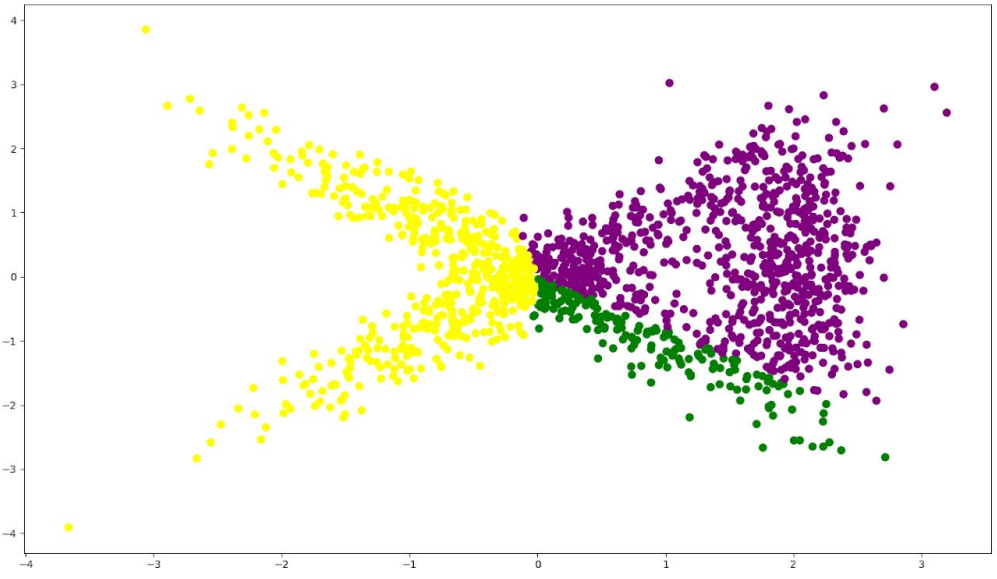
Best result for k=5, MSE=1.5237872388971685



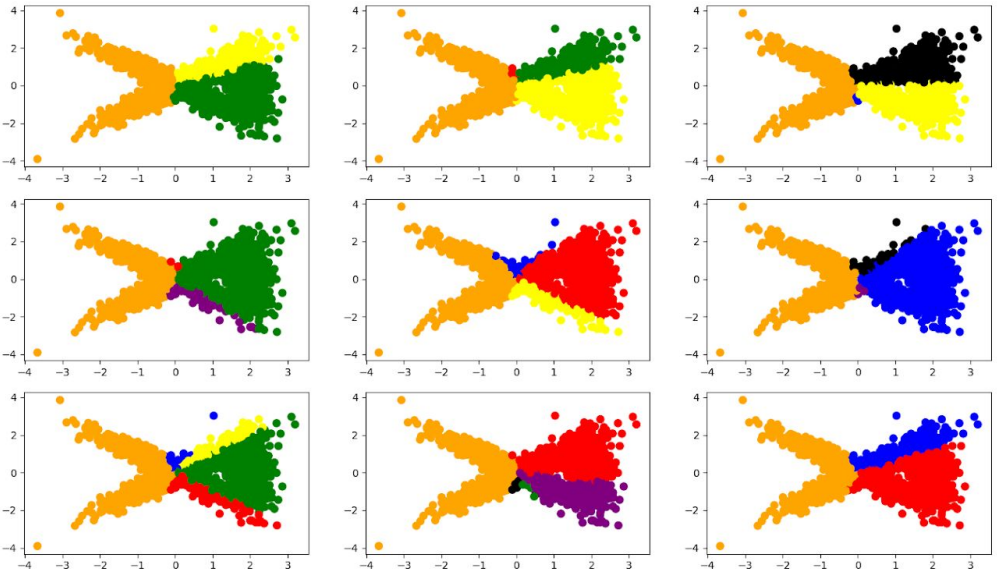
Fuzzy-C-means clusters for k=6



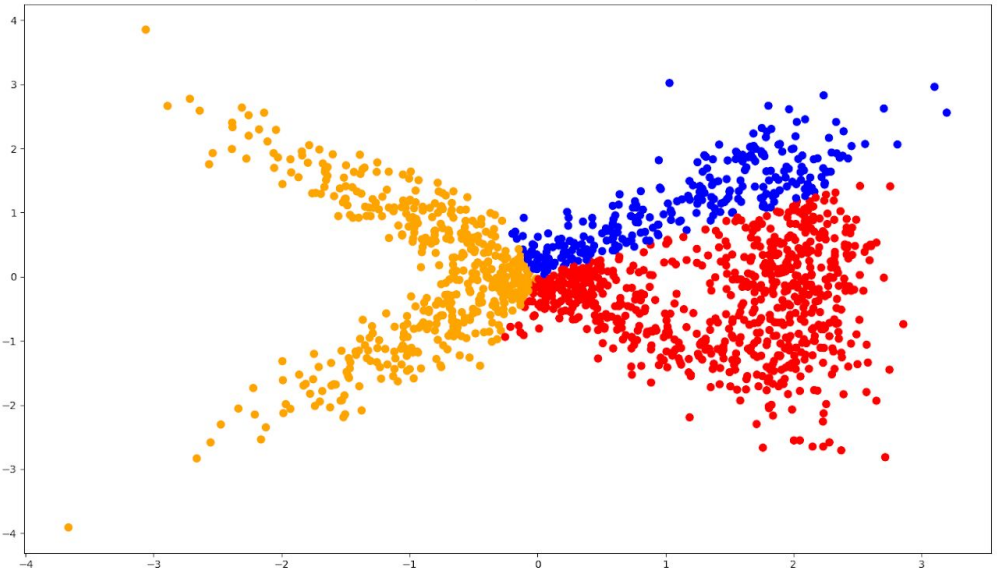
Best result for k=6, MSE=1.523781571123679



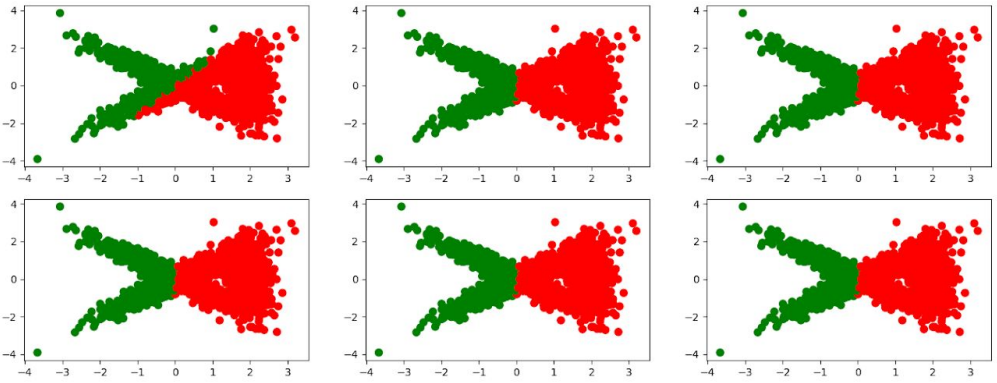
Fuzzy-C-means clusters for k=7



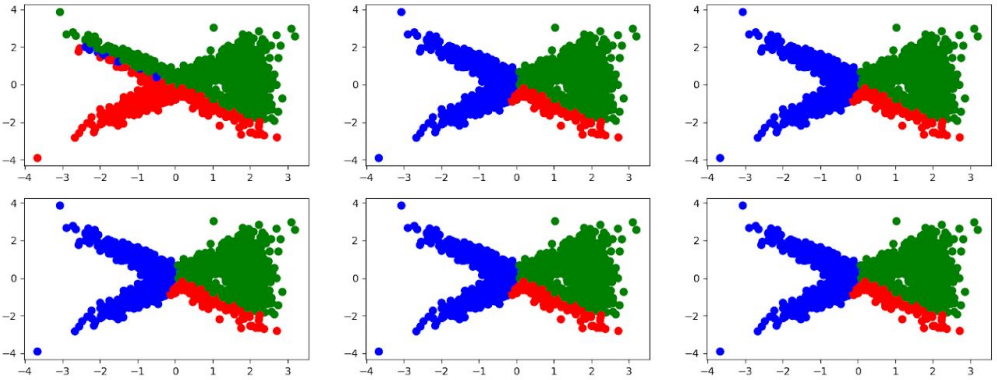
Best result for k=7, MSE=1.5237828755224214



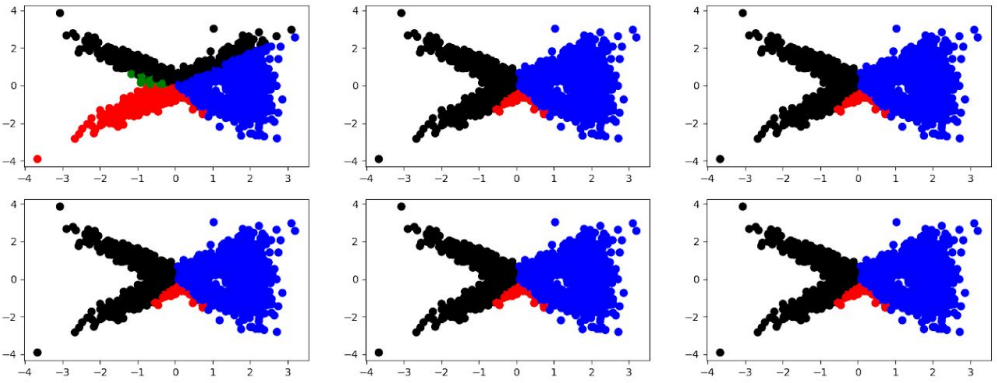
K-means clusters for successive iterations



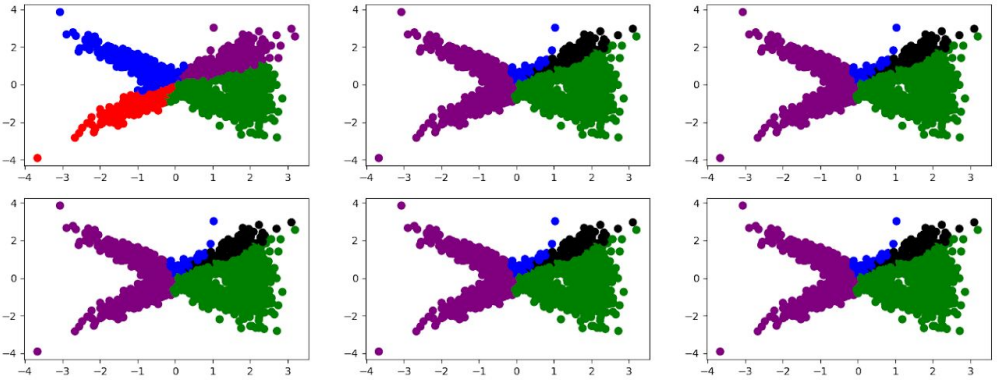
K-means clusters for successive iterations



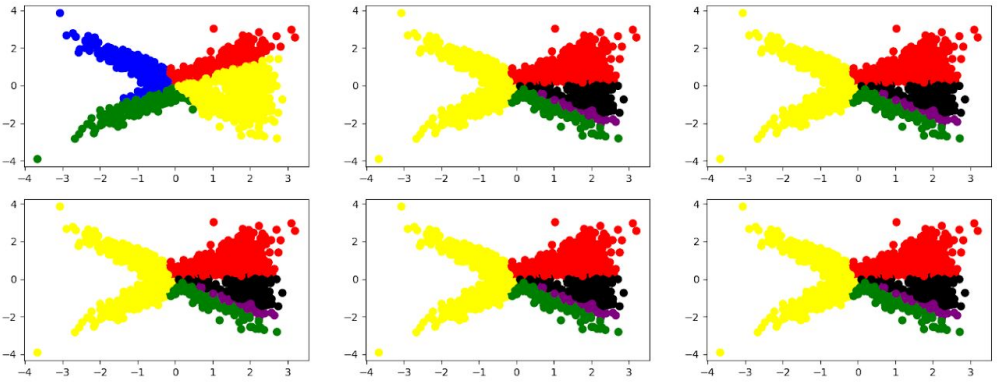
K-means clusters for successive iterations



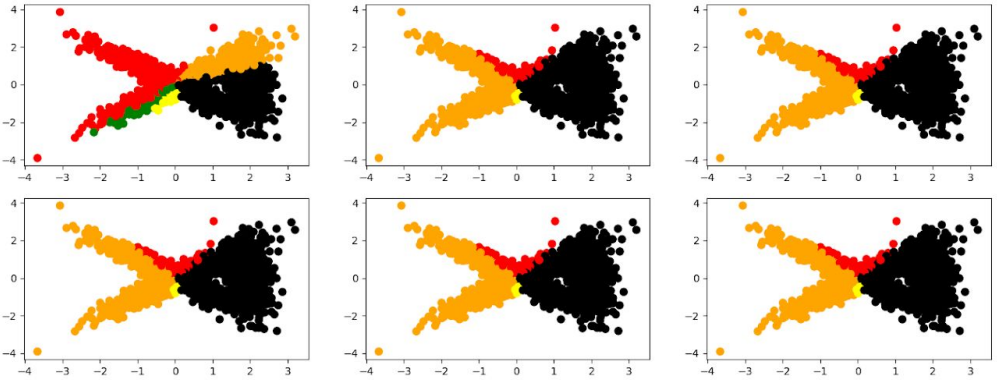
K-means clusters for successive iterations



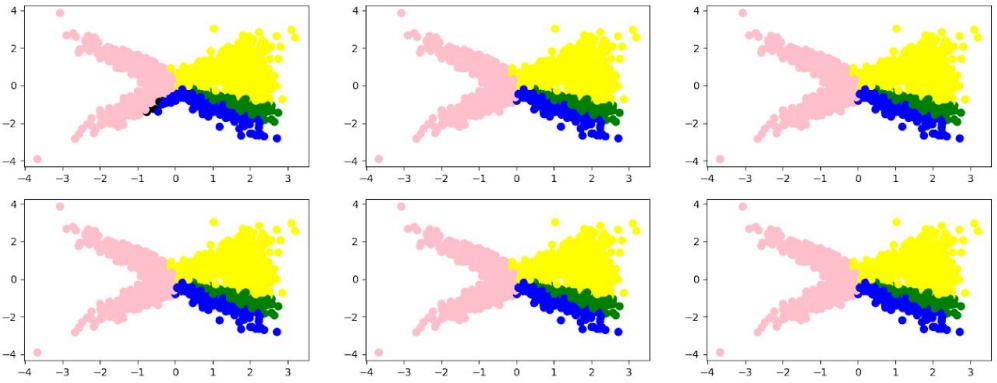
K-means clusters for successive iterations



K-means clusters for successive iterations



K-means clusters for successive iterations



Average MSE for different values of k