Cem Ergin

COMP 421 HW#7 Report:

Before starting, I’ve read the chapter concerning Clustering and re-watched the classes concerning the subject. To get a better grasp I also searched through some online material and found out that Gaussian Mixture Model algorithm are a specific form of Expectation Maximization. Since we are working with data created suing bivariate Gaussian densities in this assignment, I thought it would be appropriate to use Gaussian Mixture Model algorithm and used that approach to find the appropriate clusters.

I’ve constructed two separate functions for the required the K-nearest Neighbors algorithm and Gaussian Mixture Model algorithm. I did this to be able to test and control any parameter that they require. For example, alpha of my KNN and smoothing variable of my GMM functions and maximum number of iterations they will execute can be varied. Furthermore, both of them can be configured to plot graphs after calculations are done. I used the KNN function to initialize the Responsibility and Mean value matrixes as requested but if they are not provided the GMM function initializes randomly.

I tested some different alpha and smoothing variables to see their effects on the results. I specifically spent some time playing with GMM function’s the threshold value which breaks the calculation loop when the difference of calculated error of an iteration and the previous iteration are smaller than this threshold. I saw that when I decreased this threshold value the algorithm starts to output bizarre solutions. For example, when I choose values smaller than 0.1, the cluster in the middle grows unexpectedly. For values larger than 0.1, the algorithm behaved expectedly but the accuracy of its behavior wasn’t really satisfying. I think that the current setting seems to work well.

I drew the confidence ellipse for the actual means and covariance with red & solid lines and for the values GMM function outputs with black & dashed lines. Since I used random colors to color the point clusters, in some executions they may have similar colors but most of the time all clusters are clearly distinguishable.