

# Pattern and Speech Recognition WS2015-16

## Exercise 6

Atanas Poibrenski(2554135), Marimuthu Kalimuthu(2557695), Furkat Kochkarov(2557017)

December 11, 2015

### EM algorithm for Gaussian Mixture Models

#### Exercise 1

- 1 Done.
- 2 See “data\_preparation.m”

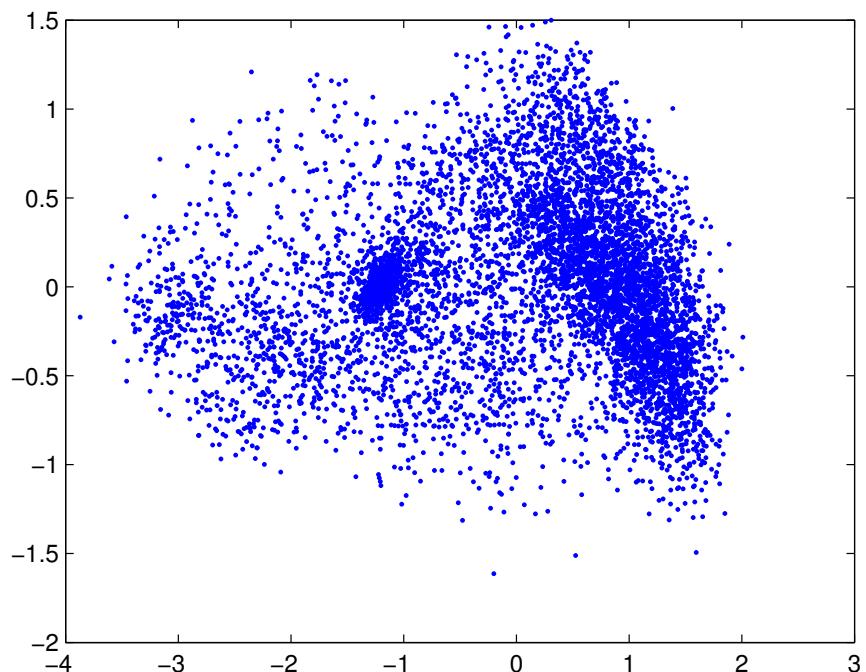


Figure 1: Transformed data distribution

- 3 See “GMM.m”
- 4 See “GMM.m”
- 5 See “GMM.m”

- 6 See “GMM.m”. We calculated the covariance matrix of the transformed data in (2) and from that we can observe that it is indeed a diagonal matrix.
- 7 In each iteration, after updating the *mean*, we compare it with the *mean* of previous iteration. If the difference between them is less than the threshold of 0.0001 ( $10^{-4}$ ), we stop further iterations.
- 8 Plots for K=2,4,10 and multiple runs.

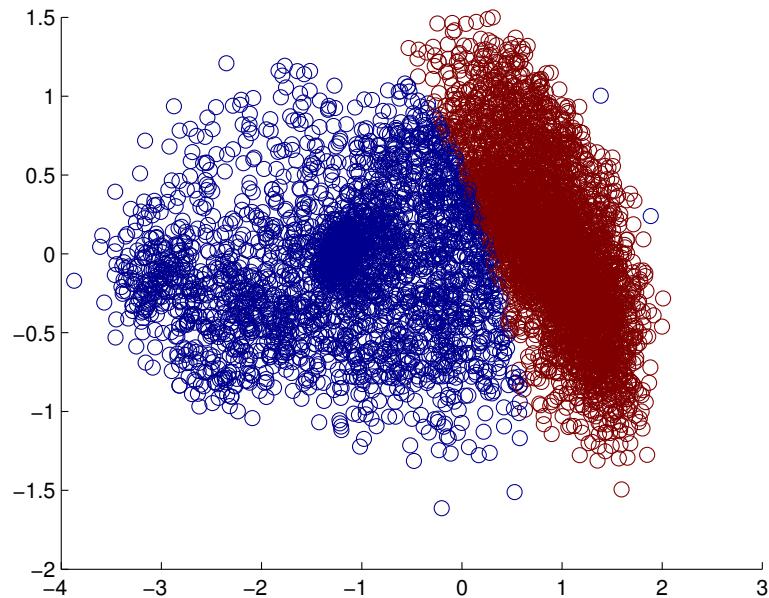


Figure 2: GMM clusters for K=2

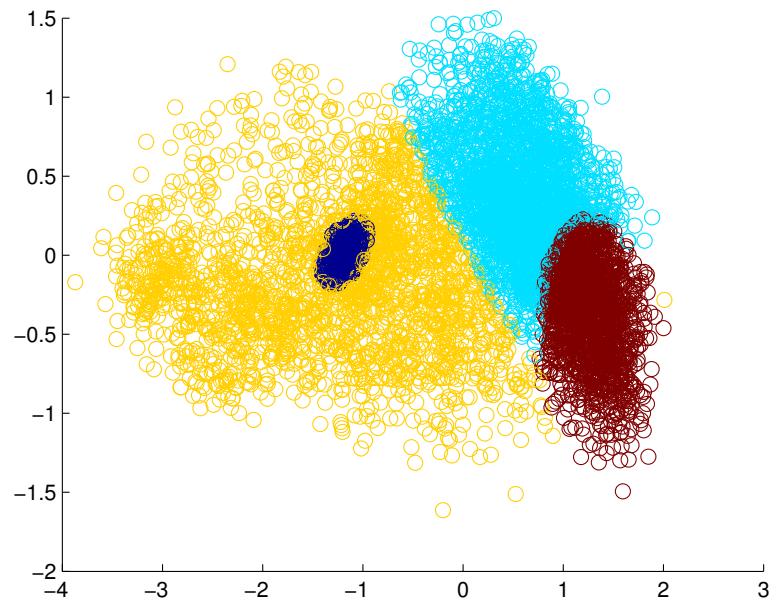


Figure 3: GMM clusters(run-1) for K=4

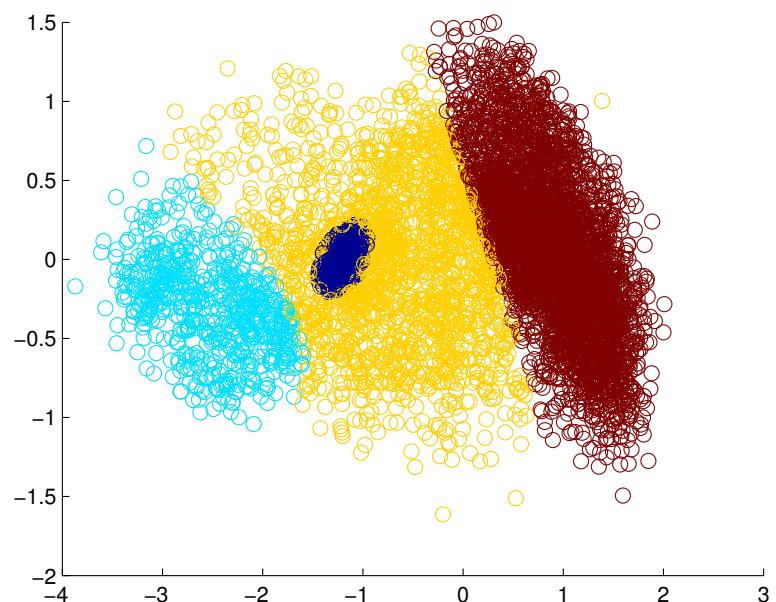


Figure 4: GMM clusters(run-2) for K=4

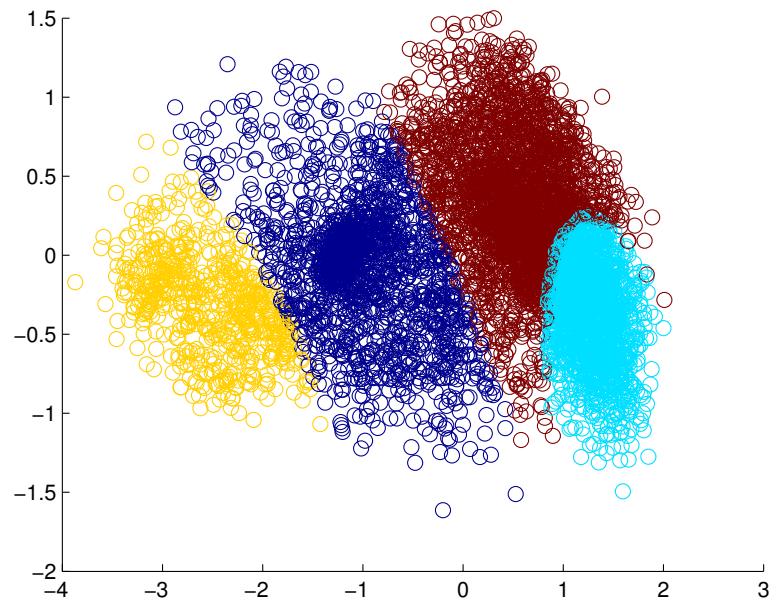


Figure 5: GMM clusters(run-3) for K=4

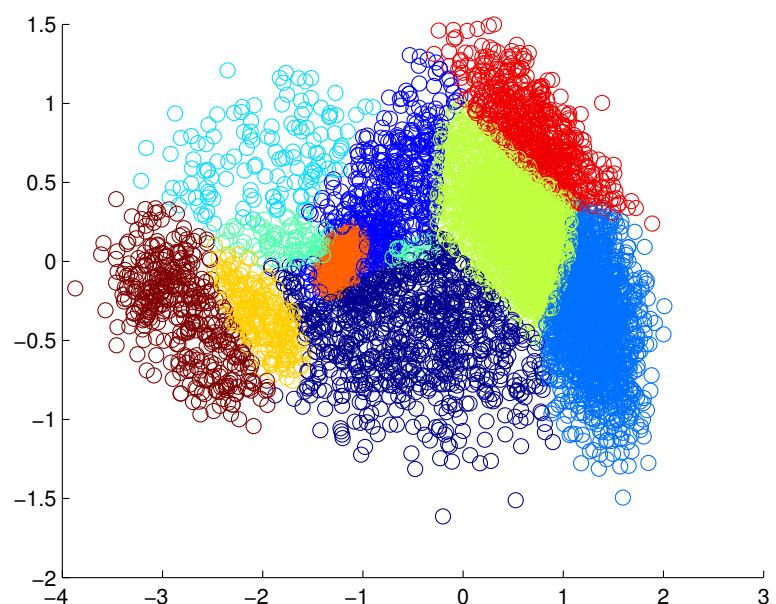


Figure 6: GMM clusters(run-1) for K=10

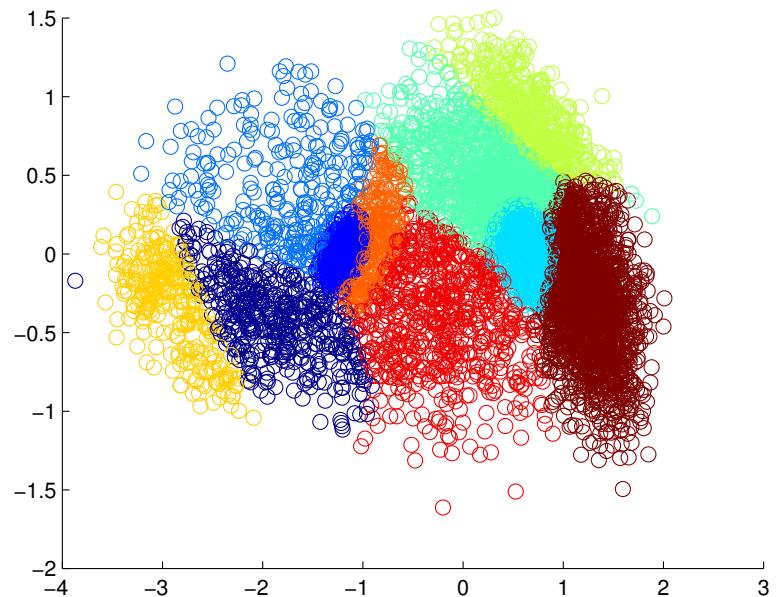


Figure 7: GMM clusters(run-2) for  $K=10$

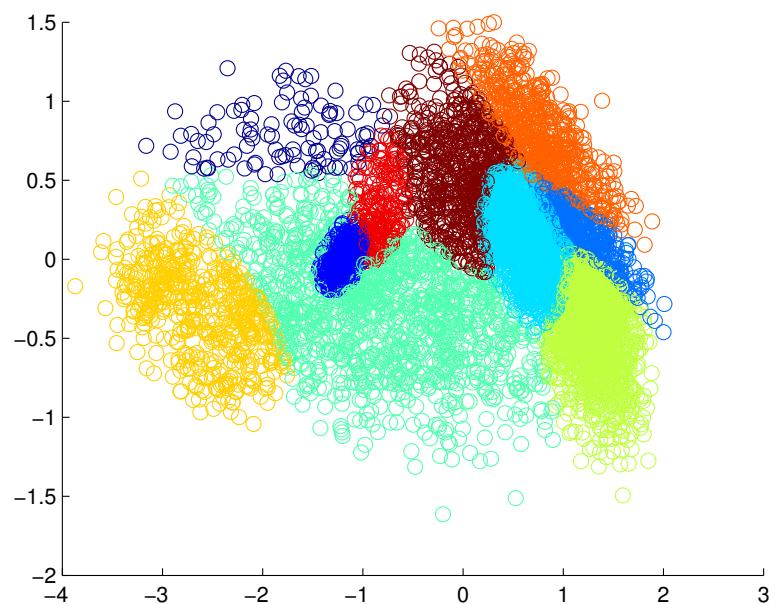


Figure 8: GMM clusters(run-3) for  $K=10$

- 9 GMM is a soft-clustering which assigns a probability for each data point to which potential cluster it may belong to. The boundaries between the clusters is not linear.

## 10 Plots for K-Means

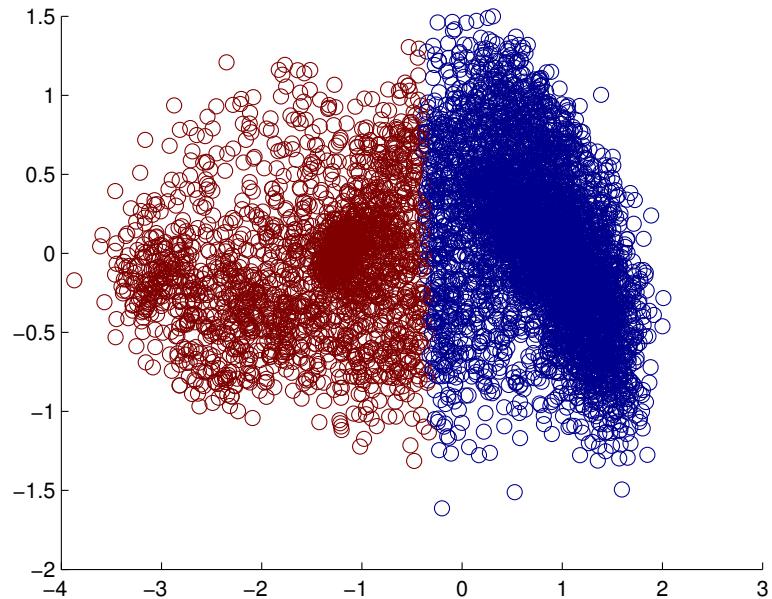


Figure 9: K-Means clusters for K=2

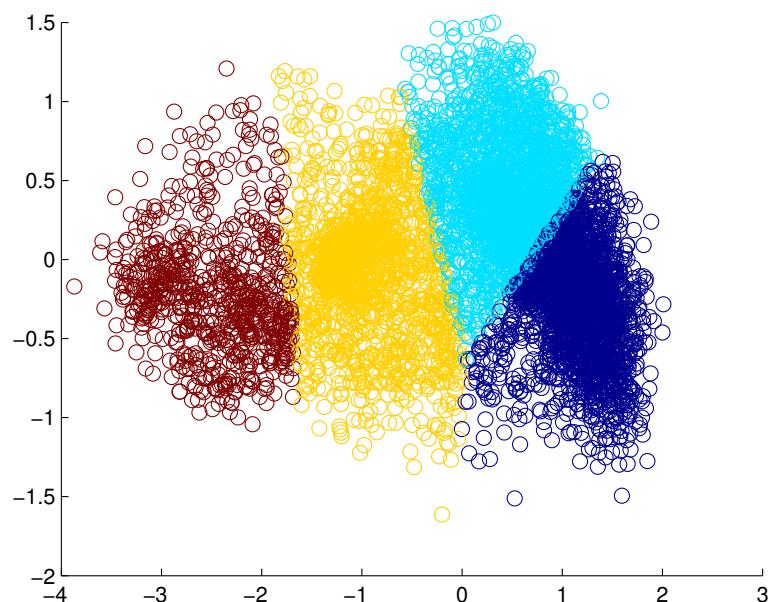


Figure 10: K-Means clusters for K=4

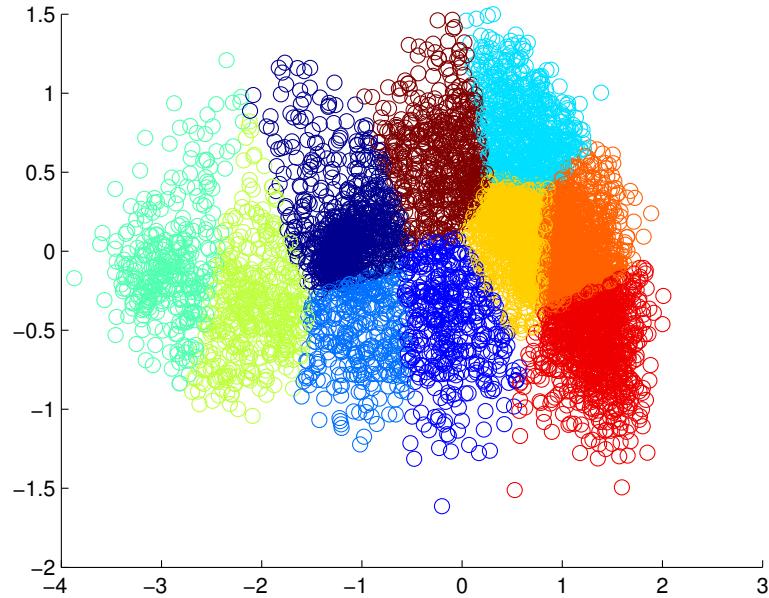


Figure 11: K-Means clusters for K=10

#### Differences:

- K-means clusters the data points with linear decision boundary like voronoi diagram whereas GMM provides soft clustering with non-linear decision boundary.
- In K-means, which is a hard clustering, the data points either belong to one cluster or the other. But GMM gives a probability for the data points being in a particular cluster.(In our case, we finally assign the data points to the cluster with highest posterior probability.)
- From the above figures, we can see that GMM allows clustering inside clustering whereas this is not possible with K-means.

#### Exercise-2; BONUS

The center of mass in each cluster of GMM is relatively the same as the center of mass of each cluster in Kmeans.

The plots are given below.

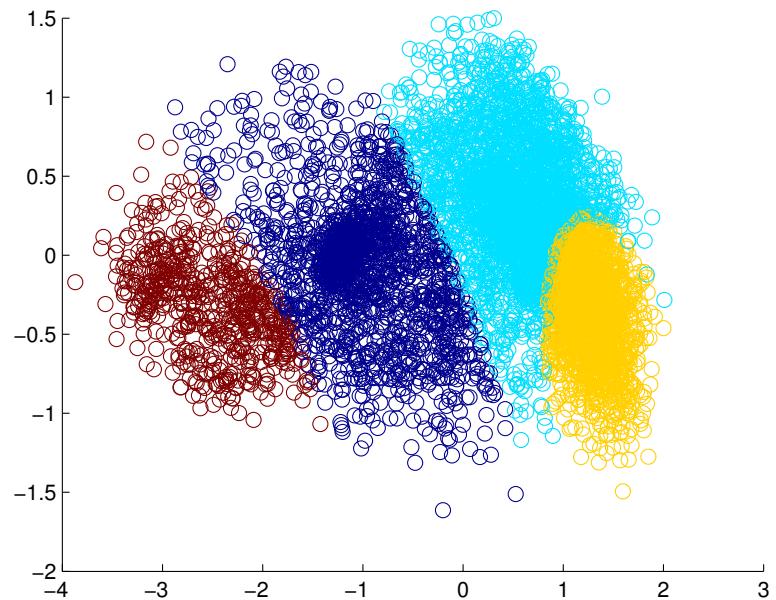


Figure 12: GMM2 clusters for K=4

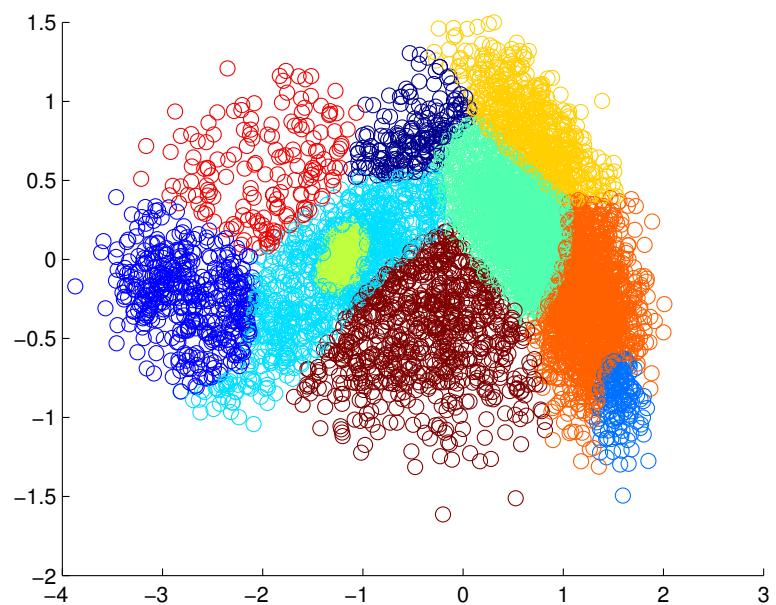


Figure 13: GMM2 clusters for K=10