Problem Set#2
CS7641 Machine Learning

Asif Rehan
arehan7@gatech.edu

Problem#2

K-means procedure can be viewed as a special case of the EM algorithm applied to an appropriate mixture of Gaussian densities model.

KMeans

$\pi_k=1/K$, and variances are shared spherical, $\Sigma_k=\sigma I$

- Assign cluster: In K-Means, the assigned cluster for x_i is the cluster k that minimizes the distance to it

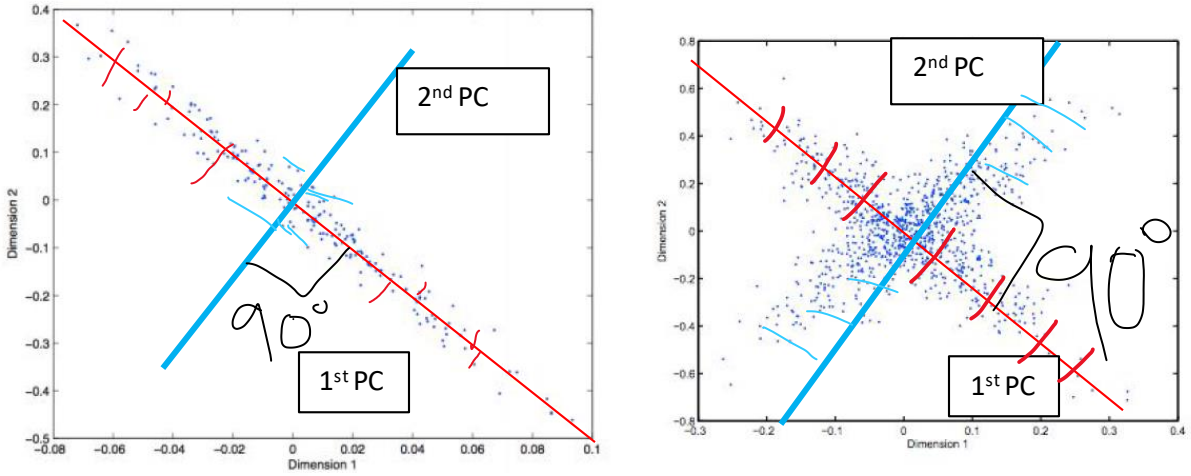
$$P(z_i=k \mid x_i) = \text{argmin}_{z,\mu} \sum_i ||\mu_{zi}-x_i||^2$$

- The update step requires checking all the samples clustered to one and taking their centroid as the new mean for cluster k

General EM for Gaussian mixtures

- EM with a variance and prior to $\Sigma k=\sigma I$ and $\pi k=1/K$ respectively can be similar to K-Means above.
- E: In general $P(z_i=k|x_i) \propto \pi_k N(x_i;\mu k,\Sigma k)$. By replacing the Normal distribution with $\sigma=1$, $P(z_i=k|x_i) \propto \text{argmin}\ (\exp\{-||\mu_k-x_i||^2\})$ which is similar to K-Means cluster assignment above
- M: Estimate new μ_k: $\mu_k=\sum_i P(zi=k \mid x_i)xi \,/\, \sum_i P(z_i=k|x_i)$ if we make the probability binary, 0 or 1, then this is again same as K-means update step above. Unlike K-means where only samples belonging to the cluster is used to recalculate the mean/centroid of the cluster, in EM all samples are weighted based on the probability that they belong to that class.

Problem#3



For left data, the first Principal Component is along the axis where the data has most variance. Then the second PC is orthogonal of it.

Problem#4

a. It might be produced by either Hierarchical Clustering with Single Link Clustering since it connects the clusters using their shortest distance or EM clustering since there could be two elongated Gaussian distributions that could perfectly capture the clusters.

Others will not work well because the shortest distance between two clusters will be shorter than the distance between two elements in the same true cluster.

b. EM, Kmeans will work as the shapes are spherical for this case. Others will not work well as the distance between items from different clusters will have shorter distance and they will be merged in the same cluster.

c. EM will work best in this case because it could handle two clusters that have such two elongated distributions. Others will cluster points around the overlap area into the same, so will not work well.