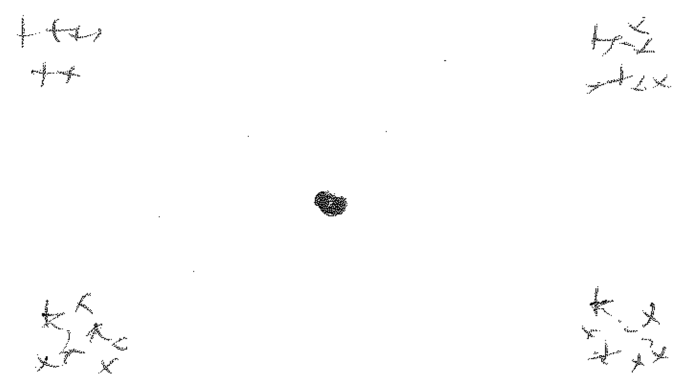
HW 4

(1) K-Means is a better options since with GMMs, the main advantage is to handle soft-clusters (points belonging to multiple clusters) & the trade-off is a more complex algorithm (EM). Therefore, if the clusters are well-defined K-Means would be a more efficient option.

Of course, GMMs & K-Means are related in the sence that as the iterations go to $\infty$, GMMs (with hard assignments) is equivalent to K-Means.

(2) No, K-Means cannot guarantee convergence to the globally optimal solution. Even if we know the number of true clusters, the initialization plays an important role in K-Means. With bad initialization we could end up with poor clusters.



In the above figure, if we initialized 4 cluster centers (since we know there are 4 clusters) right in the middle as shown, we may converge to a case where, one cluster center is assigned all points & the other 3 cluster-centers are empty (no points assigned to them)

(3) Two ideas could be effective

a). Detect ~~[illegible]~~ sparse clusters i.e, Clusters with very few datapoints ~~[illegible]~~ ~~[illegible]~~ mean that these points are dissimilar to all other points in the dataset

b). Datapoints ~~[illegible]~~ far away from its ~~[illegible]~~ cluster center are also anomalies since they are dis-similar to points that are "closest" to them in the Clustering.

(4)

Since we are using 3 components in the GMM, we are assuming that there are 3 clusters in the data.

The first data-point has a higher probability of being generated by component 1, which means it is more likely to be in the first cluster.

Similarly the second ~~component~~ data point, is more likely to belong to cluster 2,

The third data point has more uncertainty, i.e, it can belong to any of the 3 clusters, with almost similar probability.

The component coefficients give us an idea about the "strength" of each cluster. That is how likely is it that a random point belongs to a cluster.

In this case,

$$\alpha_1 = \frac{0.8 + 0.7 + 0.3}{3} = \frac{0.18}{3} = 0.6$$

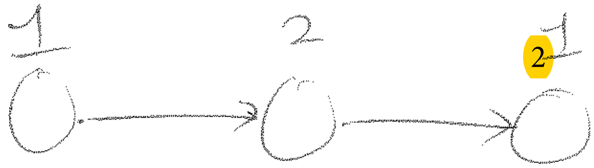
$$\alpha_2 = \frac{0.4 + 0.1 + 0.1}{3} = 0.2$$

$$\alpha_3 = \frac{0.1 + 0.2 + 0.3}{3} = 0.2$$

$\therefore$ Component-1 (or cluster 1) is in some way the most dominant cluster/component.

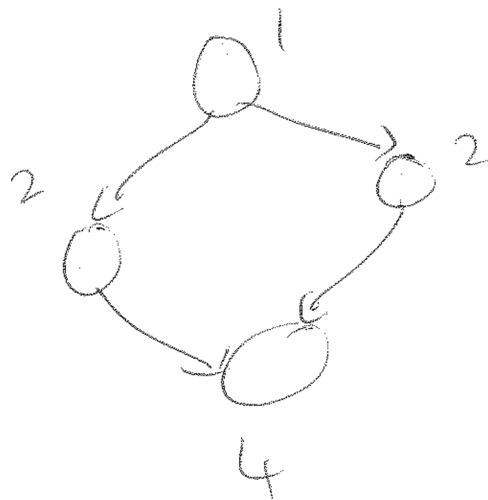
5)

a)




= 5 4 parameters.

b)




= 9 parameters.

6) a). C is conditionally independent of A given B

$$P(C \mid A, B) = P(C \mid B)$$

b) D is conditionally independent of A given B, C

$$P(D \mid A, B, C) = P(D \mid B, C)$$

7)

a) $(1-0.3) * 0.1 + (1-0.8) * 0.8 = (1-0.5)$

b). $0.8$

c) $P(E=1 \mid A=0, B=0, C=1, D=1)$

$= P(E=1 \mid D=1) = 0.5$