***HomeWork 4***

***Question 1.***

This problem is all about applying chain rule and finding the terms

Differentiating with respect to the above terms individually and substituting back, we get,

Next, we need to find,

***Question 2.***

***a)***

Log-likelihood is given as,

***b)***

E step: We need to find:

***c)*** M step:

We need to maximize the below equation in this step,

Take the derivative, set that to zero and compute the optimal

***Question 3.1***

Representing each point in the co-ordinate system as we have that,

Therefore,

Hence,

***b)***

K-mediods is less influenced by outliers in the data when compared to k means. K means is sensitive to outliers since an object with an extremely large value may distort the distribution of data.

K-mediods is computationally more expensive than k-means and the average execution time taken by K-means algorithm is greater than K-mediods.

In K means, it aims in minimizing the objective function,

The basic strategy of KMediods clustering algorithms is to find k clusters in n objects by first arbitrarily finding a representative object (the Medoids) for each cluster. Each remaining object is clustered with the Medoid to which it is the most similar. K-Medoids method uses representative objects as reference points instead of taking the mean value of the objects in each cluster. The algorithm takes the input parameter k, the number of clusters to be partitioned among a set of n objects.

K means algorithms are known to be fast clustering algorithms, however they are sensitive to the choice of starting points and are inefficient for solving clustering problems in large datasets.

K-Medoids algorithm is computationally harder than KMeans due to computing the medoids using the frequency of occurrences. K-Medoids has the potentially important characteristic which centers are located among the data point themselves.

***4.1***

***b)***

We see that K-Means work fine for Blob but it fails for Circles data set. It is because, for Circles data set, we can’t linearly separate the data on a 2-d plane. So, we need to project the data to a higher degree and use a hyperplane to separate the clusters. If we use kernel approach, then we can eliminate this problem.

***4.2)***

***D.***

For K = 3

#of iterations taken is = 217

maximum marginal log likelihood obtained for training is = -1920.92

maximum marginal log likelihood obtained for validation is = -5269.56

For K = 5

#of iterations taken is = 259

maximum marginal log likelihood obtained for training is = -1879.35

maximum marginal log likelihood obtained for validation is = -4752.54

For K = 7

#of iterations taken is = 272

maximum marginal log likelihood obtained for training is = -1947.50

maximum marginal log likelihood obtained for validation is = -4979.54

For K = 9

#of iterations taken is = 591

maximum marginal log likelihood obtained for training is = -1959.11

maximum marginal log likelihood obtained for validation is = -5289.47

For K = 11

#of iterations taken is = 455

maximum marginal log likelihood obtained for training is = -1933.32

maximum marginal log likelihood obtained for validation is = -5636.40

***E.***

I would select K=5 because the maximum marginal log likelihood for both the training set and validation set is comparatively the best among others. Hence the same.

**Collaboration:**

<http://mattmazur.com/2015/03/17/a-step-by-step-backpropagation-example/>

<https://chrisjmccormick.wordpress.com/2014/08/04/gaussian-mixture-models-tutorial>

<https://www.youtube.com/watch?v=Ih5Mr93E-2c>

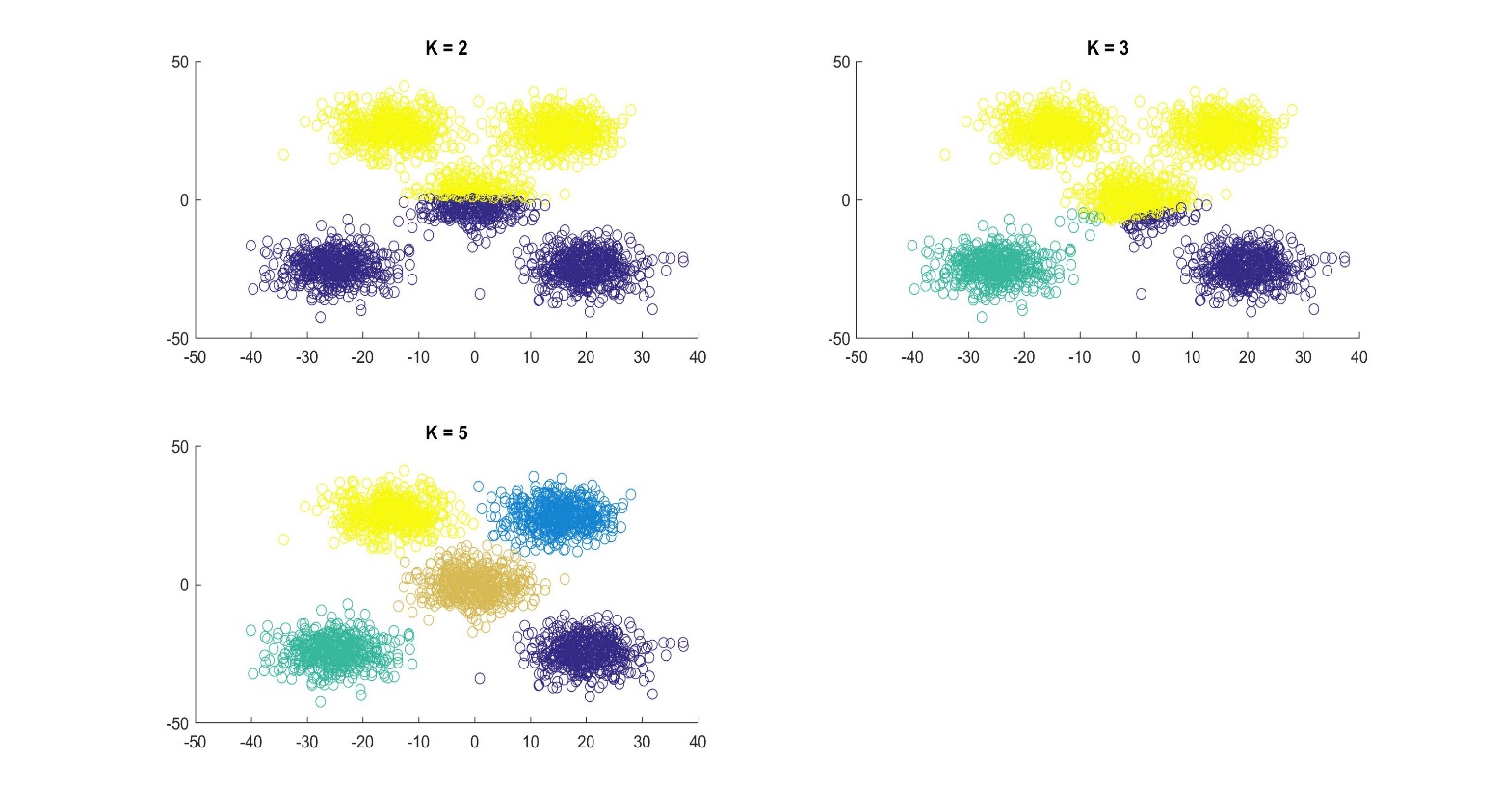
<https://www.youtube.com/watch?v=REypj2sy_5U>

<http://www.cs.cmu.edu/~tom/10701_sp11/>

<http://www.cs.cmu.edu/~ninamf/papers/cluster-chapter.pdf>

<http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.259.1427&rep=rep1&type=pdf>

***4.1*** For blob,



For Circle

