2

(a)

3 (a).

K-Means is a special case for EM algorithm where all covariance are 0, and mixture weights are equal.

K-Means is to hard assign data points to a particular cluster of convergence. It makes use of the L2 norm when optimizing (Min {θ} L2 norm point and its centroid coordinates).

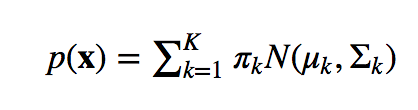
The objective of K- means is to minimize in class variance:

min J = 2

EM is to Soft assigns a point to clusters (so it give a probability of any point belonging to any centroid).  
It doesn't depend on the L2 norm, but is based on the Expectation, i.e., the probability of the point belonging to a particular cluster. This makes K-means biased towards spherical clusters.

The objective of GMM is: max P(x|π)

For GMM, for a set of samples with K class, The model has a Gaussian distribution [1]:

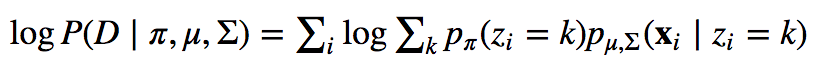


Where k = 1, We can define Z ∈ {1, ……., K}.

/Users/disheng/Desktop/Screen Shot 2017-11-03 at 12.49.39 PM.png

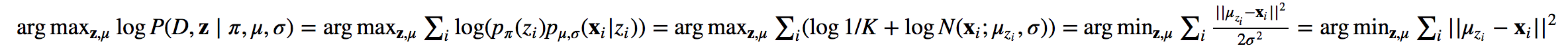
Here, p( z = k) = πk, and p(**x**∣z = k)=N(μk,Σk).

Now we define D = **x**1, …, **x**n. We estimate the parameter with Maximum Likelihood Function:



K – Means is just where all class has the same arbitrary mixture weights: πk =

The variances are spherical Σk=σI,



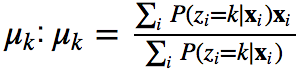
Here we see when πk = and Σk=σI, the objective of GMM and K-means align.

Based on above set ups:

The E step for K – means is

Screen%20Shot%202017-11-03%20at%201.10.48%20PM.png

The M step for K – means is



(b)

References:

[1] https://alliance.seas.upenn.edu/~cis520/wiki/index.php?n=Lectures.EM