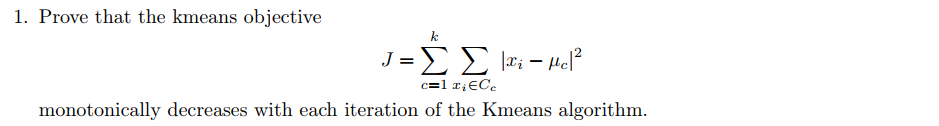
**Assignment 4**



**Answer:**

In the iterative phase of the K means algorithm we re-estimate our cluster center according to the following equation:

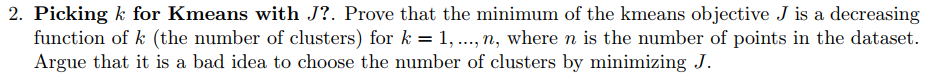
given that we have randomly initialized K cluster center. Now, if we can show that in the iterative phase we are taking the decision that converges to a minimum according to our objective function then we can say then the objective decreases with each iteration.

Now to show that our objective function is converging let us take 1st partial derivative of our objective function:

ivative of our objective function:n is converging let us take 1st decreases with each iteration.ase we are thaing the

Now we set the partial derivative to 0, to get our minimum:

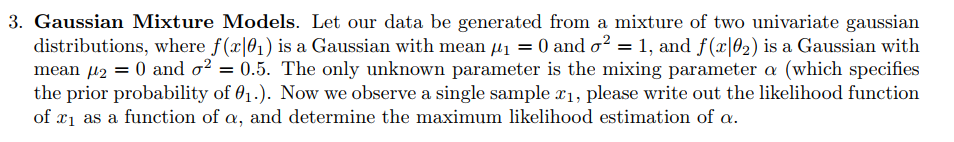
So, we can see that in each iteration we are taking a decision that guarantees to us to the local minima. We exit the algorithm when it’s not possible to minimize the objective function, if it’s possible to minimize the objective function then we re-estimate the center clusters. So, we can say that our objective function of Kmeans algorithm monotonically decreases with each iteration.



**Answer:**

The minimum of the Kmeans objective J is a decreasing function of k where k = 1...n data points. Suppose for some k and we have a non-decreasing J. Now if I add another cluster to our data point J will try to minimize again. And we will see that the new run of Kmeans is not yet converged so we are not at the minimal possible J for k+1 clusters. In the next step J will decrease as we added a new cluster. So, we can say that J is a decreasing function with respect to the number of clusters.

It’s a bad idea to choose the number of clusters by minimizing J. Cause until we reach k = n our objective function will keep decreasing and when k = n is reached we would get j = 0. It would always return K = n and it will completely overfit on our data.

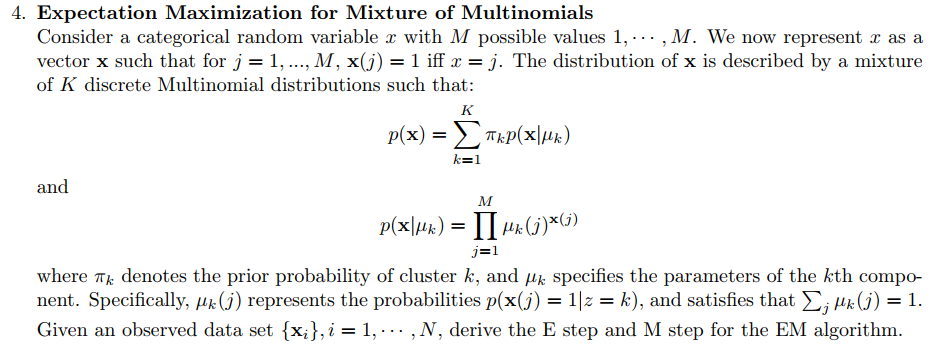


**Answer:**

Our sample is a mixture of 2 gaussian distributions N (0, 1) and N(0, 0.5). So, our probability density function can be generated using the information of gaussian distributions. So, the Likelihood function can be written like the following:

Now, to calculate the maximum likelihood we can rewrite the equation as the following:

Now the maximum likelihood of clearly depends on , where the sign of the scope depends on the gaussian distributions. If < 0 then we should choose else if we should chose



**Answer:**

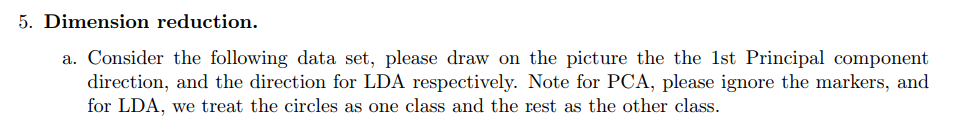
The E step and M steps for the EM algorithm are given below:

**E-Step**: Computing qi(zi) for all i = 1…k, which denotes the posterior of each cluster label.

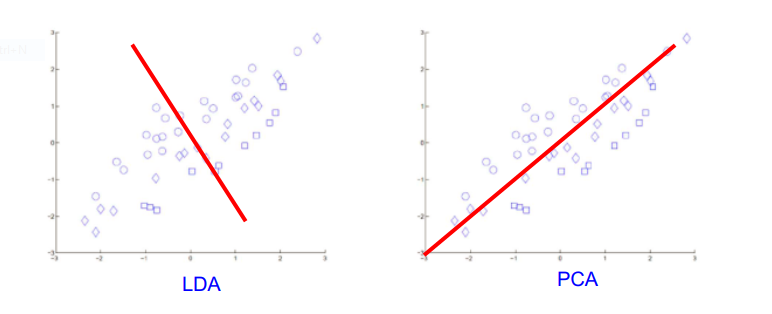
**M-step:**

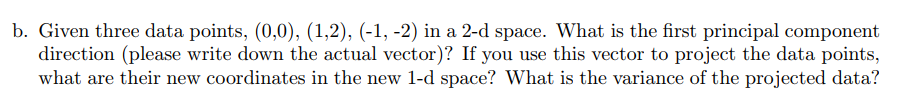
set

for updating theta,



**Answer:**



**Answer**:

The data points can be represented as the following matrix X,

Now, before calculating PCA 1 we need to calculate the covariance matrix of x, because in PCA our target is to maximize where Now using an online calculator covariance of x was calculated which is:

Now according to Lagrange, , where u is the unit vector onto whose directed we are trying to project our data. Now this equation can be rewritten as, which gives us the following:

Now to get the 1st eigenvector u1, we now do the following calculations:

Variance of this data is:

after projection = = (0,

Variance of projected data =

(ans.)