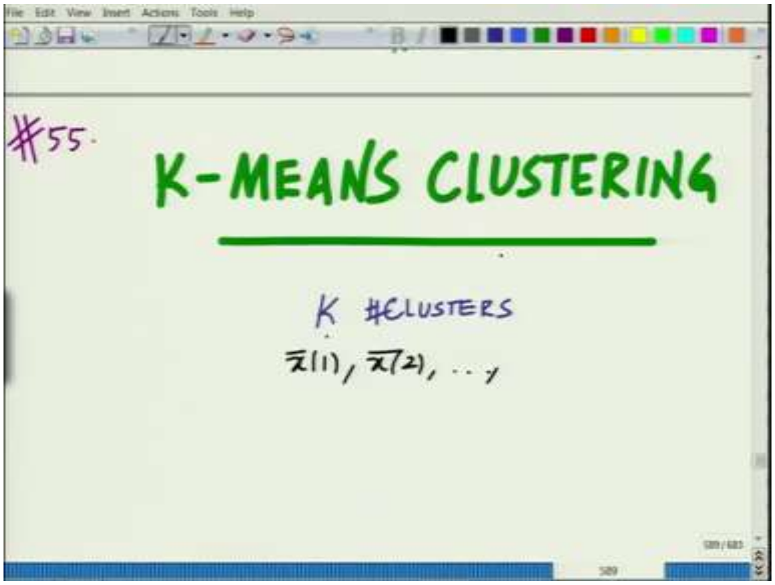
**Applied Linear Algebra for Signal Processing, Data Analytics and Machine Learning**
**Professor. Aaditya K Jagannatham**
**Department of Electrical Engineering**
**Indian Institute of Technology, Kanpur**
**Lecture No. 55**
**K-Means Clustering Algorithm**

Hello! Welcome to another module in this massive open online course. So, we are looking at K-Means Clustering which is essentially as we said one of the many algorithms to perform clustering that is grouping, a given unlabeled data set. That is large data set without labels into meaningful, logistically similar clusters or sets of points. This is known as clustering and this is the part of unsupervised learning because we are working with unlabeled data.

(Refer Slide Time: 00:42)



#55.

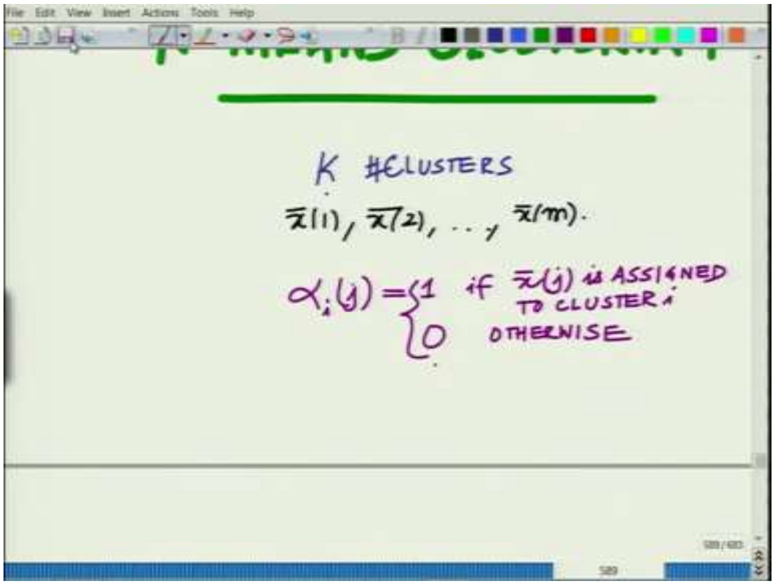
K-MEANS CLUSTERING

K #CLUSTERS

$\bar{x}(1), \bar{x}(2), \ldots,$

So, we are looking at the K-Means algorithm and let us continue, K-Means for, we are looking at the K-Means algorithm for clustering and remember the idea here is you have the, K is the number of clusters and remember we have the points x 1 bar or x 1 bar, x 2 bar up to, we have I believe x bar m.

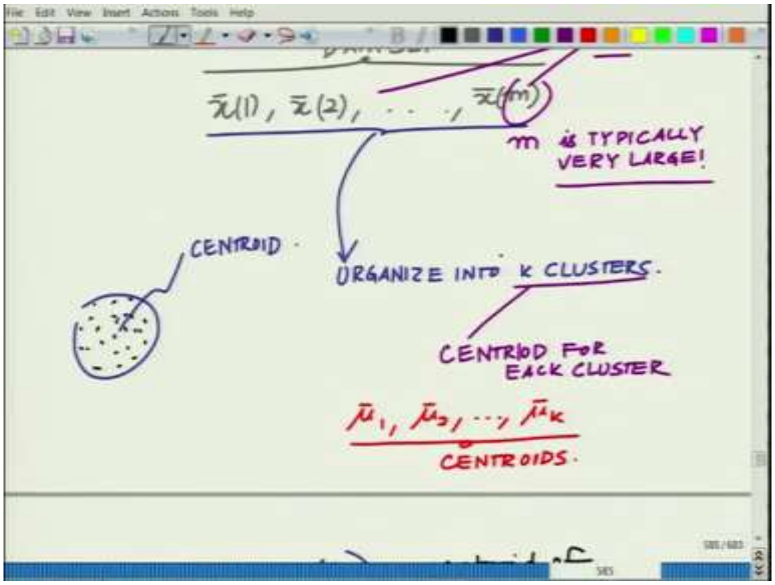
(Refer Slide Time: 01:51)



K #CLUSTERS

$\bar{x}(1), \bar{x}(2), \ldots, \bar{x}(m)$.

$$\alpha_i(j) = \begin{cases} 1 & \text{if } \bar{x}(j) \text{ is ASSIGNED TO CLUSTER } i \\ 0 & \text{OTHERWISE} \end{cases}$$



$\bar{x}(1), \bar{x}(2), \ldots, \bar{x}(m)$

m is TYPICALLY VERY LARGE!

ORGANIZE INTO K CLUSTERS.

CENTROID

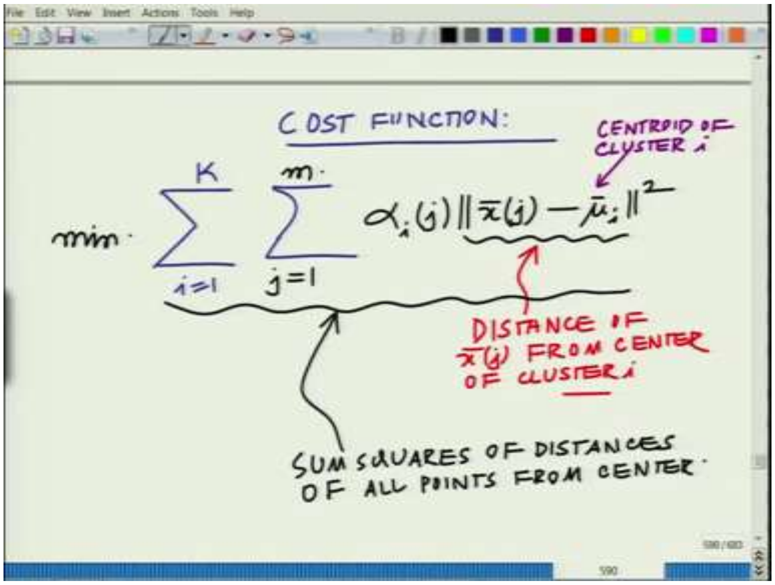
CENTROID FOR EACH CLUSTER

$\bar{\mu}_1, \bar{\mu}_2, \ldots, \bar{\mu}_K$

CENTROIDS.

This is that number of points and the idea remember is to determine these parameters alpha i j, where, which is equal to, that is basically this is equal to 1. If x bar j is assigned to cluster i and 0 otherwise. So, that is essentially, these are the parameters. These are, we can say the assignment parameters that have to be determined.

(Refer Slide Time: 02:44)



COST FUNCTION:

$$\min \underbrace{\sum_{i=1}^{K} \sum_{j=1}^{m} \alpha_i(j) \underbrace{\|\bar{x}(j) - \bar{\mu}_i\|}_{\text{DISTANCE OF } \bar{x}(j) \text{ FROM CENTER OF CLUSTER } i}^{2}}_{\text{SUM SQUARES OF DISTANCES OF ALL POINTS FROM CENTER}}$$

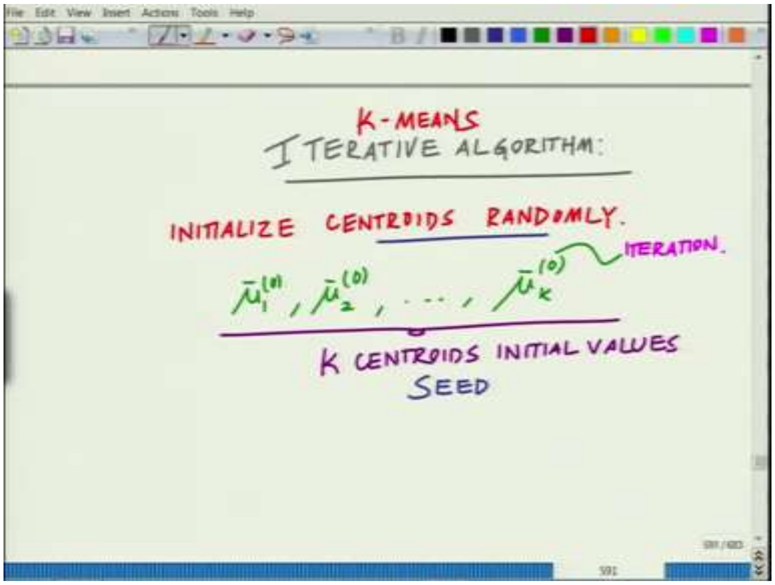
Now we have therefore summation. Let us look at the cost function that we would like to minimize for this cluster assignment. We would like to find the cluster assignment such that this cost function is minimized. What is the cost function? You want to sum over all the clusters, i equal to 1 to K, sum over all the points i equal to 1 to m, j equal to 1 to m, alpha i j x bar j minus mu bar i. So, we want to minimize, this essentially distance, if you look at this, this norm of x bar j, now you remember mu bar i this is the centroid of cluster i.

So, what we want to do is, now this is essentially the distance or you can say the norm or square of the distance or basically the norm is the distance of x bar j from center of cluster i and therefore if you look at this cost function, this, entire this thing is nothing but total square of the distance, sum square of distances of all points from centers of the clusters.

Or centroids of their clusters, from centroids of their clusters, so that is, so we want to find essentially it is very intuitive. So, we want to find the optimal assignment parameter alpha i j such that these distances, I mean that is the idea, the point is that you want to find this cluster such that these clusters are comprised of closed points.

So, that if you look at the distances of these points from the centroids of these clusters and you look at the sum of the squares of the distances that should be as low as possible which essentially implies that each of these points is close to the centroid of its own cluster and that is how the clustering is performed. And this is performed in an iterative fashion.

(Refer Slide Time: 6:35)



K-MEANS

ITERATIVE ALGORITHM:

INITIALIZE CENTROIDS RANDOMLY.

$$\underbrace{\bar{\mu}_1^{(0)}, \bar{\mu}_2^{(0)}, \ldots, \bar{\mu}_K^{(0)}}_{\text{K CENTROIDS INITIAL VALUES SEED}}$$

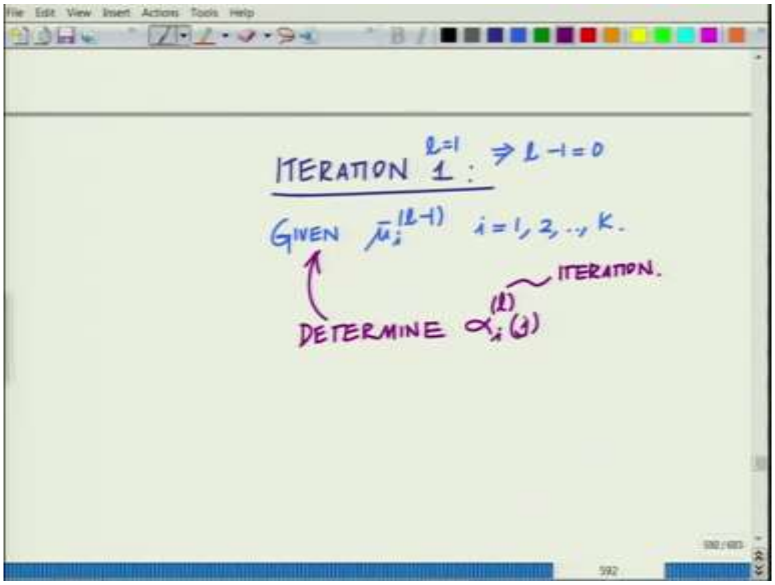
ITERATION.

So, K-Means is an iterative algorithm to basically minimize this cost function. Remember we want to minimize, the aim is to minimize this cost function and we would like to perform an iterative algorithm. What is the iterative algorithm? Now pick the clusters, start with or we say initialize these cluster centroids randomly or uniformly, choose some feasible sort of assignment. So, we will start with, remember these are the centroids mu bar 1 mu bar 2 0. These are the centroids of the K clusters in the 0 th iteration.

So, this as usual, this is our iteration index, which is, this is iteration 0 so this is even before the iteration begins. So, this is the initialization, these are the K centroids, so we will say initial values. These are the initialized, and these are then, and these are also termed as seeds basically. So, you have seeded this. So, these are the seed or this is the seed of your algorithm and so on. So these are the initial values of the centroid.

So, we have started with, so this is the, these are initialized randomly and this how we are initializing our K-Means algorithm which is essentially an iterative algorithm. K-Means clustering. So, this is your iterative algorithm. This is the K-Means algorithm, which is an iterative algorithm for clustering. Now let us look at what are the iterations.

(Refer Slide Time: 09:24)



ITERATION $\overset{l=1}{1}$ : $\Rightarrow l-1=0$

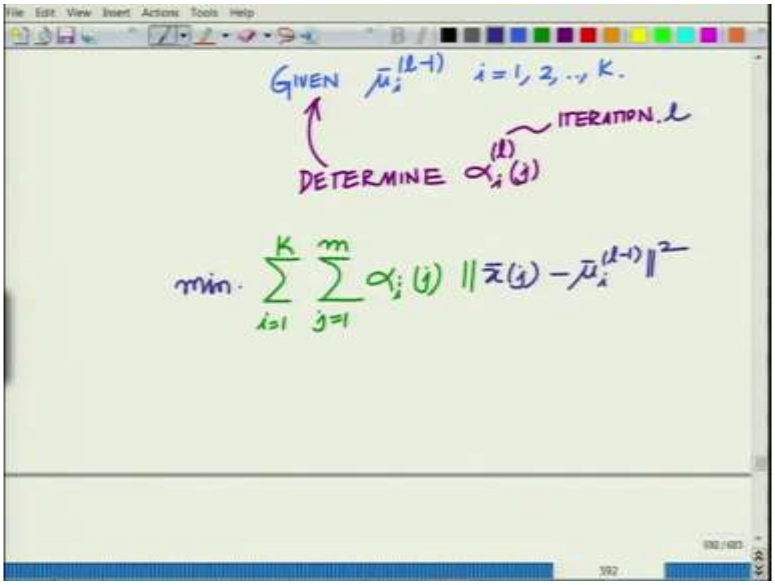
GIVEN $\bar{\mu}_i^{(l-1)}$ $i = 1, 2, .., K.$

DETERMINE $\alpha_i^{(l)}(j)$ ~ ITERATION.

Now iteration 1, what happens in iteration 1, now what we would like to do is given mu bar i l minus 1. So, we have let us say iteration index is l, l equal to 1 to begin with, which implies l minus 1 equal to 0. So, we have all these mu bar i l minus 1 or i equal to 1, 2 up to K. Now, determine alpha i j the assignment in the l th iteration. This is remember, and to begin with in this iteration l equal to 1. So, we have mu bar j, mu bar i l minus 1 that is mu bar i 0.

And that is technically what has been determined in the previous iteration l minus 1 th iteration and to begin with we have all the values mu bar i 0 and now we determine alpha i, j, l. That is basically the assignment parameters of all the points in the l th iteration and how do we determine that again we have to minimal cost function.

(Refer Slide Time: 11:02)

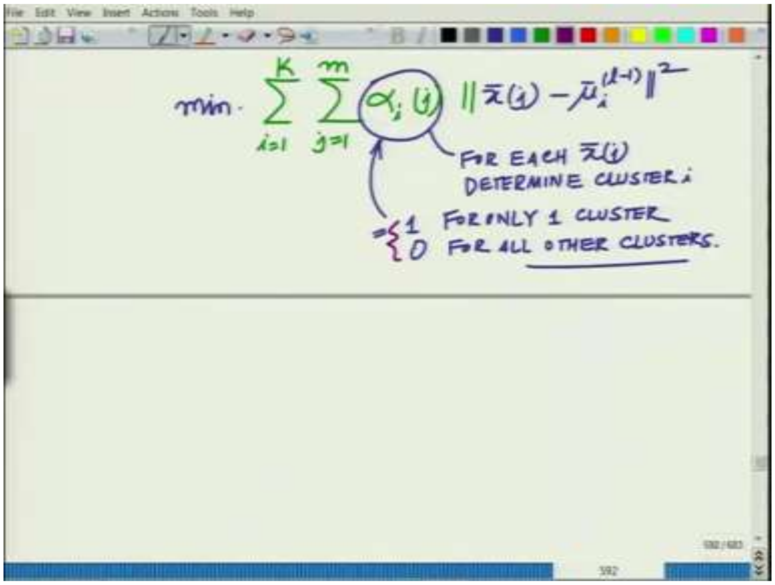


GIVEN $\bar{\mu}_i^{(l-1)}$ $i = 1, 2, \ldots, K.$

ITERATION $l$

DETERMINE $\alpha_i^{(l)}(j)$

$$\min \sum_{i=1}^{K} \sum_{j=1}^{m} \alpha_i(j) \, \|\bar{x}(j) - \bar{\mu}_i^{(l-1)}\|^2$$



$$\min \sum_{i=1}^{K} \sum_{j=1}^{m} \alpha_i(j) \, \|\bar{x}(j) - \bar{\mu}_i^{(l-1)}\|^2$$

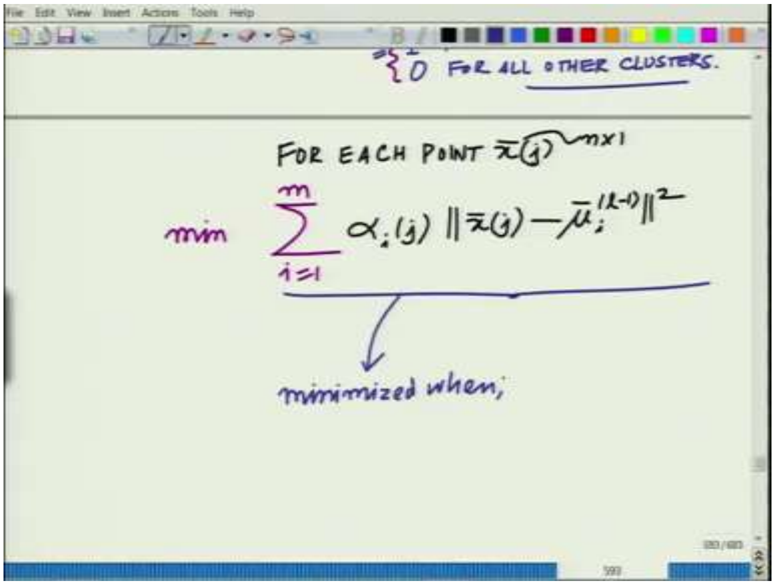
FOR EACH $\bar{x}(j)$ DETERMINE CLUSTER $i$

$$= \begin{cases} 1 & \text{FOR ONLY 1 CLUSTER} \\ 0 & \text{FOR ALL OTHER CLUSTERS.} \end{cases}$$

This is iteration l and we have to minimize the cost function. Remember we have the cost function. Our cost function this is essentially i equal to 1 to K, j equal to 1 to M, alpha i, j norm x bar j minus mu bar i l minus 1 square. Now determine these are essentially what we have to determine. So, now look at for each point for each point, now for each point j look at what is the alpha, let us compute what is this alpha i j for each j. For each j, each point j that is basically each j or you can say for each x bar j determine cluster i and for that cluster alpha i j will be equal to 1 and for all the other clusters alpha i j will be equal to 0.

And how do we do this? That is very simple, observed and now note that alpha i j, if you assign it to alpha i j remember it is equal to 1 for only 1 cluster, for each j only one cluster 0. So, in that sense this is a hard partitioning algorithm what I mean by that is in each iteration it assigns a point to only one particular cluster. It cannot assign it to multiple clusters. So, alpha i j equal to 1 only for cluster and it has to be 0 for all the other clusters for any point, and therefore, now if you look at the cost function.

(Refer Slide Time: 13:34)



$0$ FOR ALL OTHER CLUSTERS.

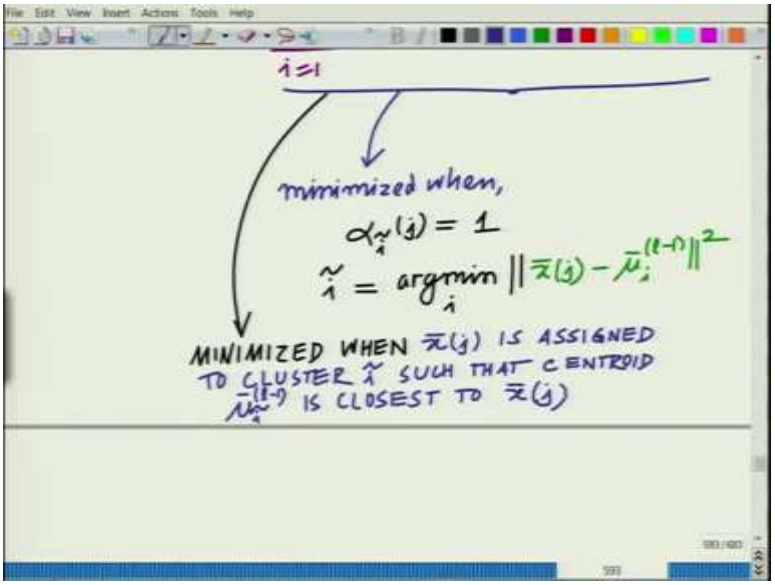
FOR EACH POINT $\underbrace{\bar{x}(j)}_{n \times 1}$

$$\min \sum_{i=1}^{m} \alpha_i(j) \, \|\bar{x}(j) - \bar{\mu}_i^{(l-1)}\|^2$$

minimized when;

For each point that is you look at the summation over all i equal to 1 m for each point. Now we are looking at it from the perspective of each point j or you can say x bar of j. This is basically we said this is I think n cross 1 dimensional vector. So, we take the sum of the distances from all the clusters heads and you can clearly see this is equal to, this is minimized, this is minimum if, observe that this is minimum if alpha i, this is minimum when, or this is minimized when, again let me write it technically and, minimized when?

(Refer Slide Time: 15:06)



$i=1$

minimized when,

$$\alpha_{\tilde{i}}(j) = 1$$

$$\tilde{i} = \underset{i}{\arg\min} \|\bar{x}(j) - \bar{\mu}_i^{(l-1)}\|^2$$

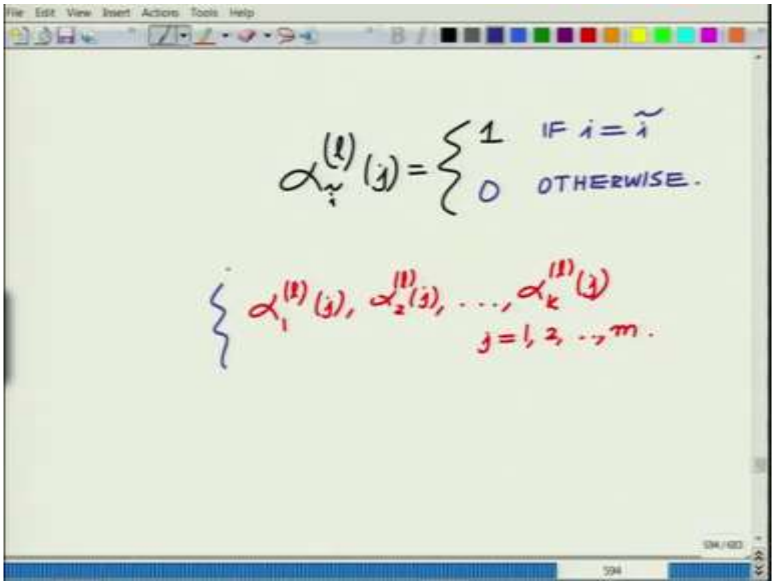
MINIMIZED WHEN $\bar{x}(j)$ IS ASSIGNED TO CLUSTER $\tilde{i}$ SUCH THAT CENTROID $\bar{\mu}_{\tilde{i}}^{(l-1)}$ IS CLOSEST TO $\bar{x}(j)$

This is minimized when we assign alpha i tilde j equal to 1 such that i tilde equals the minimum of x bar j minus mu bar i l minus 1 square. That is you are choosing from all these distances x bar j minus mu bar i l minus 1 that is given all the centroids. You are essentially choosing that centroid which is the closest to x bar j. So, you are choosing, you are assigning x bar j to that cluster whose centroid is closest to, and naturally if that distance is minimum this whole cost function you can see is minimized because all the alpha i j's are 0, you will only have alpha i tilde j equal to 1 corresponding to the cluster with the closest centroid.

And therefore, that is essentially, this is minimized by assigning, so the technical way to see this is, although it is very, a little complicated when you look at the mathematics. This is minimized the simpler way of saying this is, this is minimized when x bar j is assigned to cluster i tilde such that the centroid mu bar i tilde as computed in iteration l minus 1 is closets to x bar j. That is the philosophy.

So, you are assigning x bar j to, you are assigning the point x bar j each point x bar j for that matter to the cluster i tilde such that corresponding centroid mu bar i tilde as computed in the iteration l minus 1 is the closest to x bar j, that gives us the assignment.

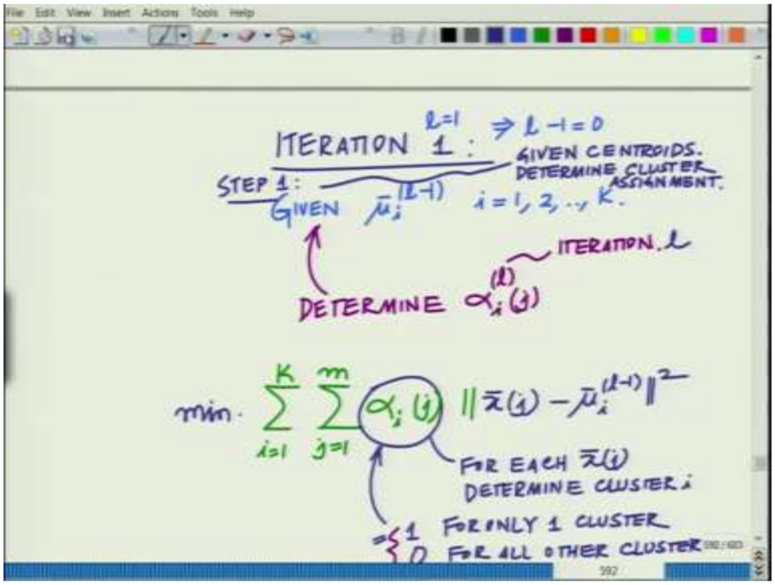
(Refer Slide Time: 17:45)



$$\alpha_{\tilde{i}}^{(l)}(j) = \begin{cases} 1 & \text{IF } i = \tilde{i} \\ 0 & \text{OTHERWISE.} \end{cases}$$

$$\left\{ \alpha_1^{(l)}(j), \alpha_2^{(l)}(j), \ldots, \alpha_k^{(l)}(j) \right.$$

$$j = 1, 2, \ldots, m.$$



ITERATION 1: $l=1 \Rightarrow l-1=0$ GIVEN CENTROIDS. DETERMINE CLUSTER ASSIGNMENT.

STEP 1: GIVEN $\bar{\mu}_i^{(l-1)}$ $i = 1, 2, \ldots, K.$

DETERMINE $\alpha_i^{(l)}(j)$ ITERATION $l$

$$\min \sum_{i=1}^{K} \sum_{j=1}^{m} \alpha_i(j) \left\| \bar{x}(j) - \bar{\mu}_i^{(l-1)} \right\|^2$$

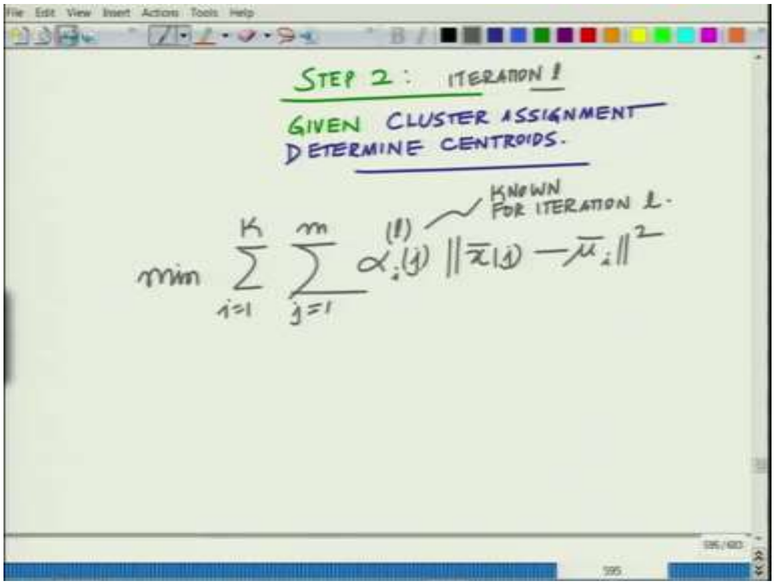
FOR EACH $\bar{x}(j)$ DETERMINE CLUSTER $i$

$$= \begin{cases} 1 & \text{FOR ONLY 1 CLUSTER} \\ 0 & \text{FOR ALL OTHER CLUSTER} \end{cases}$$

That is alpha i tilde in iteration l 1 if i equals i tilde 0 otherwise. So, we are assigning each, and at the end of it, therefore now we have obtained alpha 1 l j, alpha 2 j of l and alpha k of l, j for j equal to 1 2 remember we have total of m points. So, this is essentially what this gives us. This is, you can call this as the first step.

This is essentially what is obtained in, you can call this as the first step. So, in the iteration you have the step one that is given centroids or given centroids determine cluster assignment. Now step two will be given the cluster assignment, now we have obtained the cluster assignment. Now we will come to step 2.

(Refer Slide Time: 19:51)

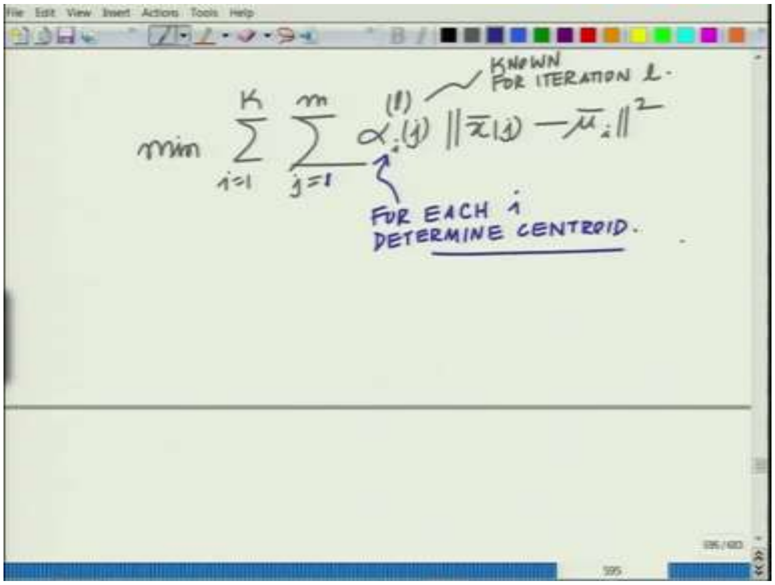


STEP 2 : ITERATION $l$

GIVEN CLUSTER ASSIGNMENT DETERMINE CENTROIDS.

$$\min \sum_{i=1}^{K} \sum_{j=1}^{m} \alpha_i^{(l)}(j) \left\| \bar{x}(j) - \bar{\mu}_i \right\|^2$$

KNOWN FOR ITERATION $l$.



$$\min \sum_{i=1}^{K} \sum_{j=1}^{m} \alpha_i^{(l)}(j) \left\| \bar{x}(j) - \bar{\mu}_i \right\|^2$$

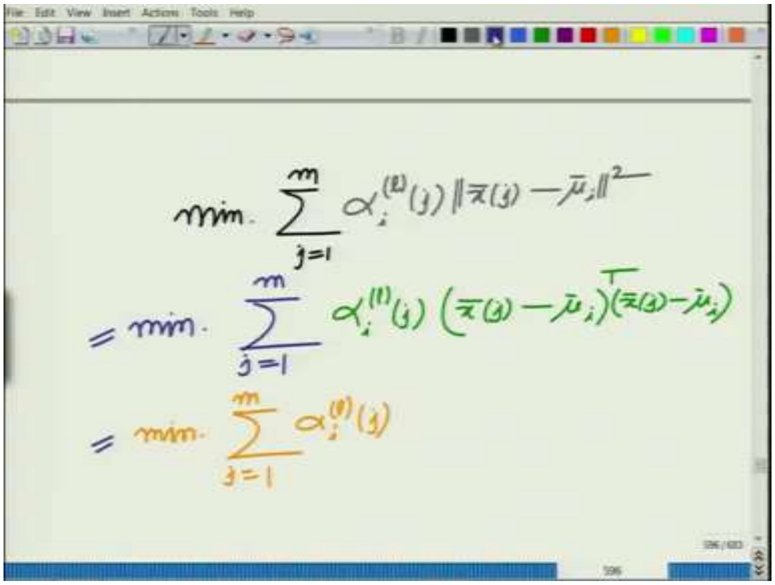
KNOWN FOR ITERATION $l$.

FOR EACH $i$ DETERMINE CENTROID.

Now what is step two in this iterative algorithm? Let us say step two is given the cluster assignment, the reverse given cluster assignment we have to determine the centroids, that is the step two. So, we are given the cluster assignment, now determine the centroids and that can be done as follows. Now we go back to our cost function which is essentially minimize summation i equal to 1 to K, j equal to 1 to m alpha i j l norm x bar j minus mu bar i whole square.

Now we know these, these are known, now these are known for this, remember we are still in iteration l, so this is step two iteration is the same iteration l th iteration. So, iteration index is still the same l and now what we have to do is, now determine this. Now we will look at this for each, now look at this for each, previously we fixed the point and assigned it to a cluster. Now we will fix a cluster and determine the centroid. So, we will have for each i determine the centroid, and how do you determine the centroid?

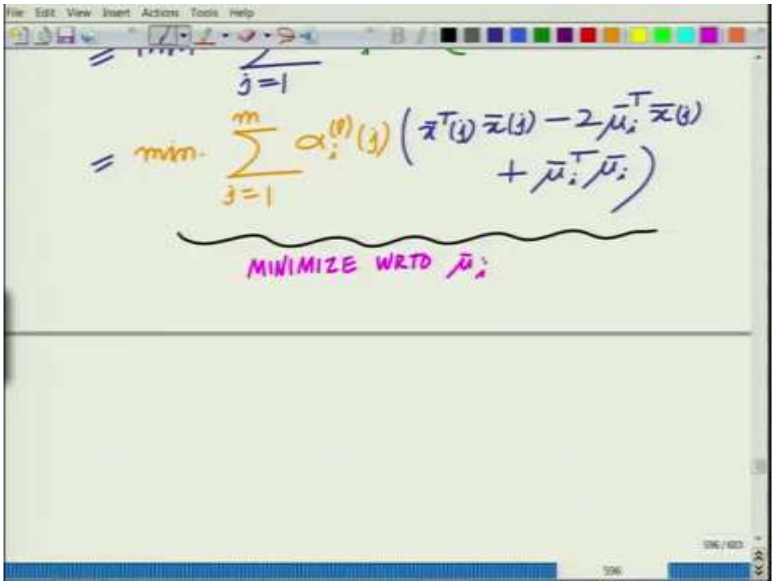
(Refer Slide Time: 22:25)



$$\min. \sum_{j=1}^{m} \alpha_i^{(l)}(j) \|\bar{x}(j) - \bar{\mu}_i\|^2$$

$$= \min. \sum_{j=1}^{m} \alpha_i^{(l)}(j) \left(\bar{x}(j) - \bar{\mu}_i\right)^T \left(\bar{x}(j) - \bar{\mu}_i\right)$$

$$= \min. \sum_{j=1}^{m} \alpha_i^{(l)}(j)$$



$$= \min. \sum_{j=1}$$

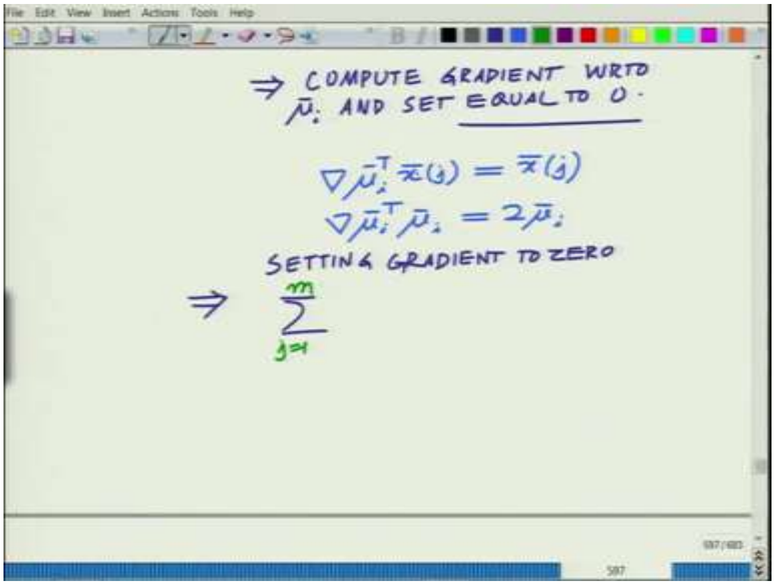
$$\underbrace{= \min. \sum_{j=1}^{m} \alpha_i^{(l)}(j) \left(\bar{x}^T(j)\bar{x}(j) - 2\bar{\mu}_i^T \bar{x}(j) + \bar{\mu}_i^T \bar{\mu}_i\right)}_{\text{MINIMIZE WRTO } \bar{\mu}_i}$$

Again now consider this as summation over j for each i, so we have j equal to 1 to m, we have alpha i l j norm x bar j minus mu bar i whole square and we have to minimize this, which I can write as minimize sum j equal to 1 to m alpha i l j, x bar j minus mu bar i transpose x bar j minus mu bar i, which now you can also write as minimize j equal to 1 to alpha i l j x bar j minus twice.

We have minus twice, we can write this as mu bar i transpose x bar j plus mu bar i transpose mu bar i and now if you look at this one.

Now once again we have to minimize this with respect to mu bar i. Now remember we have to minimize this whole thing, remember you have to minimize with respect to mu bar i and therefore the way to do it is.

(Refer Slide Time: 24:48)



⇒ COMPUTE GRADIENT WRTO $\bar{\mu}_i$ AND SET EQUAL TO 0.

$$\nabla \bar{\mu}_i^T \bar{x}(j) = \bar{x}(j)$$

$$\nabla \bar{\mu}_i^T \bar{\mu}_i = 2\bar{\mu}_i$$

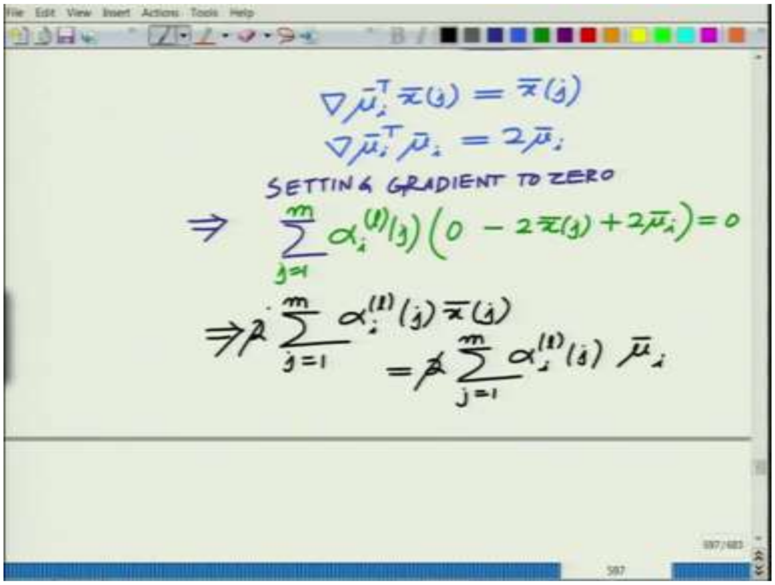
SETTING GRADIENT TO ZERO

$$\Rightarrow \sum_{j=1}^{m}$$

Compute gradient with respect to mu bar I. The reason is because mu bar, remember the only thing that we are saying over here is basically mu bar i is a vector and how do you minimize with respect to a vector, compute the derivative with respect to each component of the vector which is nothing but the gradient and set equal to 0. And we use the following principles that is if you have gradient of mu bar i transpose x bar j this is equal to x bar j and gradient of mu bar i transpose mu bar i this is equal to twice mu bar i.

Therefore, when you look at this now the gradient of this will be, if you look at this term the gradient of this will be basically, simply x bar j and the gradient of this term gradient. So, we are lighting the gradient so, and the gradient of this term, this will be twice mu bar i. So, essentially this implies setting gradient to 0, so we will have summation of j equal to 1 to m. Now the first term does not depend on mu bar i, so this gradient will be equal to 0. So, the first term we will have gradient 0.

(Refer Slide Time: 27:23)



$$\nabla \bar{\mu}_i^T \bar{x}(j) = \bar{x}(j)$$

$$\nabla \bar{\mu}_i^T \bar{\mu}_i = 2\bar{\mu}_i$$

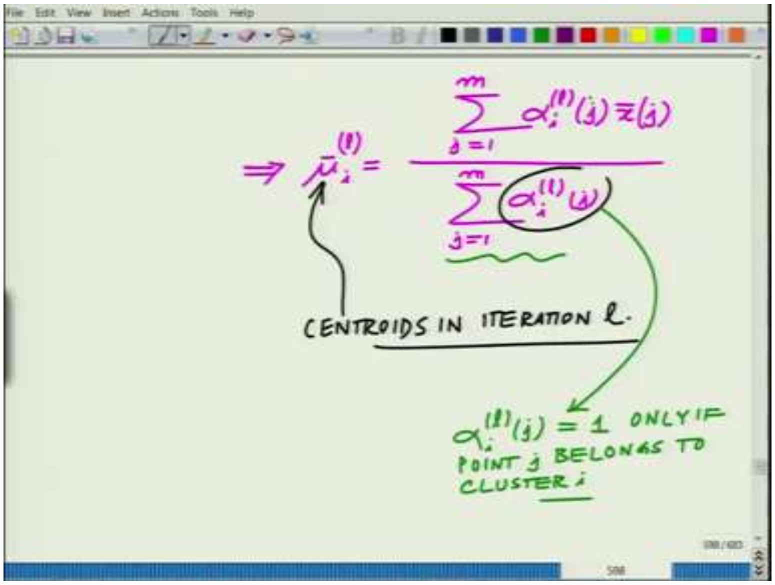
SETTING GRADIENT TO ZERO

$$\Rightarrow \sum_{j=1}^{m} \alpha_i^{(l)}(j)\left(0 - 2\bar{x}(j) + 2\bar{\mu}_i\right) = 0$$

$$\Rightarrow \cancel{2} \sum_{j=1}^{m} \alpha_i^{(l)}(j)\,\bar{x}(j) = \cancel{2} \sum_{j=1}^{m} \alpha_i^{(l)}(j)\,\bar{\mu}_i$$

So, this will be summation j equal to 1 to m alpha 1 alpha i l of j of each point j into 0 minus 2 x bar j plus twice mu bar i this is equal to 0. And this implies, now solving this, this implies summation j equal to 1 to m alpha i l of j twice x bar j is equal to twice summation j equal to 1 to m alpha i l j, mu bar i and this 2's cancel and therefore what you get is a very interesting solution.

(Refer Slide Time: 28:19)



$$\Rightarrow \bar{\mu}_i^{(l)} = \frac{\sum_{j=1}^{m} \alpha_i^{(l)}(j)\,\bar{x}(j)}{\sum_{j=1}^{m} \alpha_i^{(l)}(j)}$$

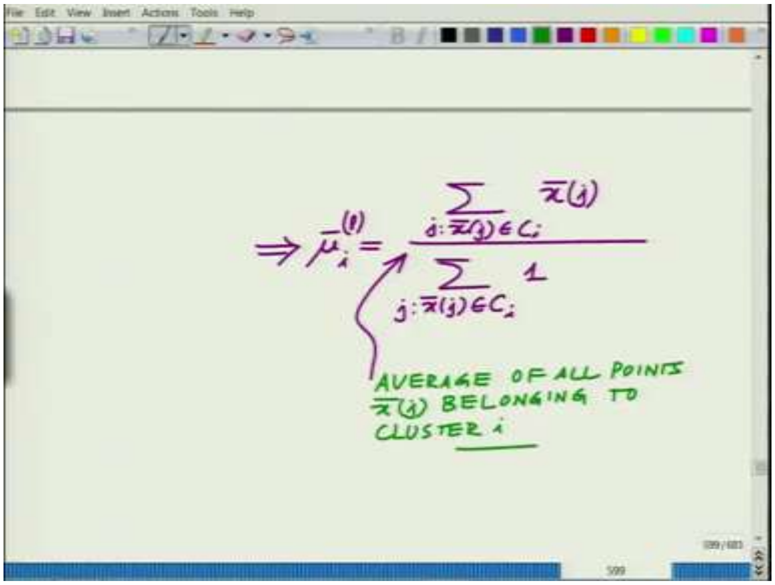
CENTROIDS IN ITERATION $l$.

$\alpha_i^{(l)}(j) = 1$ ONLY IF POINT $j$ BELONGS TO CLUSTER $i$

This is essentially, if you look at this mu bar i equals summation j equal to 1 to m alpha i l of j over summation j equal to 1 to m alpha i l of j into x bar of j and this is the centroids in the l th iteration. So, this these are centroids. This completes centroids in iteration l and if you look at this you will observe something interesting. Remember alpha i l j equal to 1 or alpha i l j equal to 1, let me write this a little bit more elaborately or little bit more clearly, remember alpha i l j equal to 1.

This quantity alpha i l of j equal to 1, only if point j belongs to cluster i, this is equal to 1 only if point j belongs to cluster i or class i and therefore this implies, that this is, simply what it means is we are going to weigh only that, that is for, so for each i you are only going to add 1. If j belongs to cluster I.

(Refer Slide Time: 30:35)



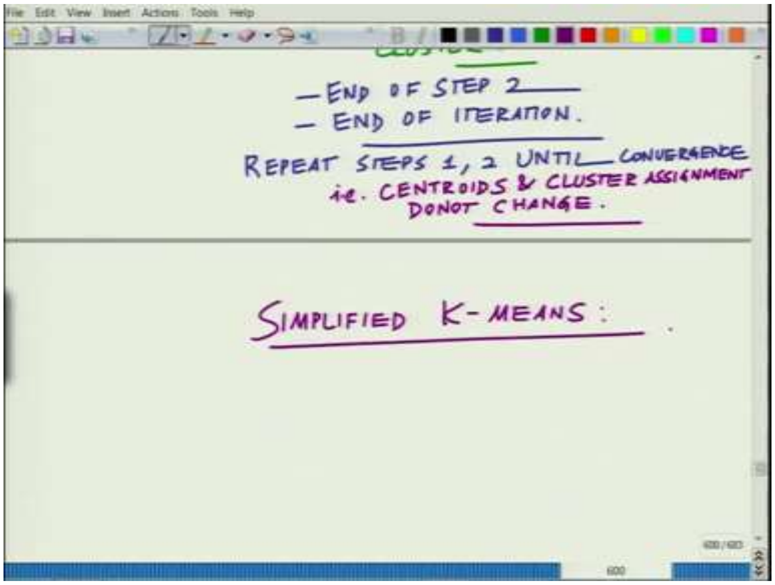
$$\Rightarrow \bar{\mu}_i^{(l)} = \frac{\sum_{j:\bar{x}(j)\in C_i} \bar{x}(j)}{\sum_{j:\bar{x}(j)\in C_i} 1}$$

AVERAGE OF ALL POINTS $\bar{x}(j)$ BELONGING TO CLUSTER $i$

So, this is going to be simply mu bar of i this is equal to summation over all j. Such that if x bar j belongs to this cluster i, you will add 1 that is you will add 1 corresponding to all points that belong to cluster I, and simply in the denominator, this is simply summation of j, such that x bar j belongs to class i, you will have, that is it and you will have x bar of j.

Which if you look at this, it is simply and this is the l th iteration and this is simply, if you look at it the average or the sample mean of all points x bar j belonging to average of all points x bar j. that is you take all the points x bar j, which belong to cluster i take their mean or take their average that is essentially the centroid corresponding to that. So, step 1 you compute cluster assignments, step 2 we compute the centroids based on the cluster assignment and we iteratively repeat these things, and these iterations continue and now the iterations continue.

(Refer Slide Time: 32:23)



— END OF STEP 2 —

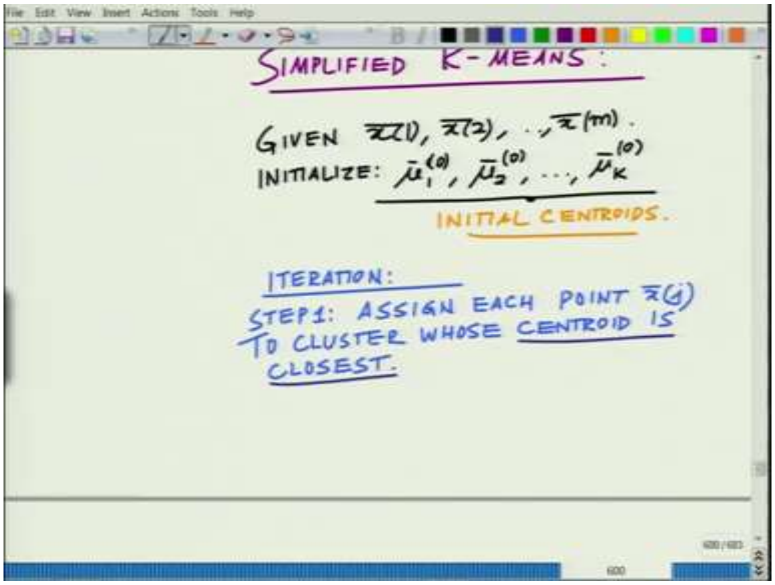
— END OF ITERATION.

REPEAT STEPS 1, 2 UNTIL CONVERGENCE
i.e. CENTROIDS & CLUSTER ASSIGNMENT
DONOT CHANGE.

SIMPLIFIED K-MEANS :

So, this is the end of step two, in fact I would also say end of the iteration and repeat, now what we do is repeat the iteration, repeat steps 1 comma 2 until convergence, and what do we mean by repeat steps 1 comma 2 that that is centroids and cluster assignment do not change. That is what happens, if you keep doing this iteratively after certain point the centroids in the clusters assignment remain fixed at that point you stop the algorithm and if you look at this, this is a very the intuitive way of saying, this is very simple instead of now that we have done all mathematics so intuitively it is basic, so intuitively or simplified us put it simplified K-Means.

(Refer Slide Time: 34:22)



SIMPLIFIED K-MEANS :

GIVEN $\bar{x}(1), \bar{x}(2), \ldots, \bar{x}(m)$.

INITIALIZE: $\bar{\mu}_1^{(0)}, \bar{\mu}_2^{(0)}, \ldots, \bar{\mu}_K^{(0)}$

INITIAL CENTROIDS.

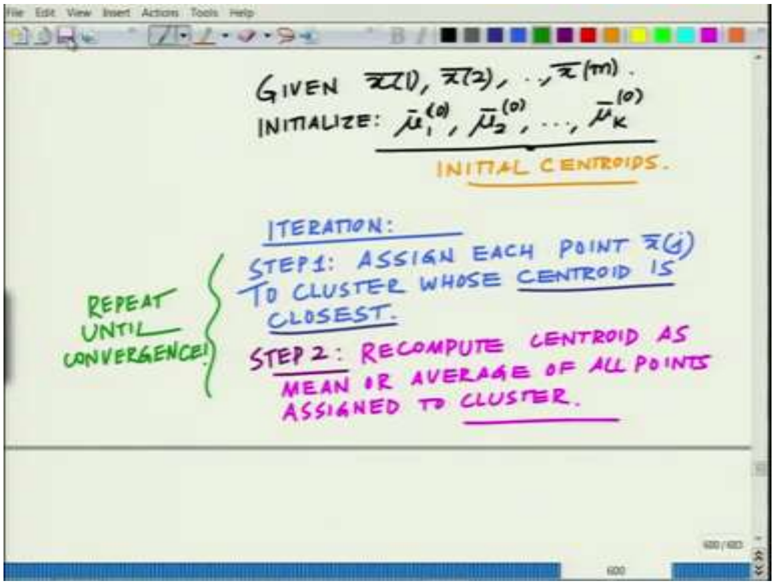
ITERATION:

STEP1: ASSIGN EACH POINT $\bar{x}(j)$
TO CLUSTER WHOSE CENTROID IS
CLOSEST.

Let us say this is simplified K-Means which is essentially given x bar 1, x bar 2, x bar 1 determine start or initialize mu bar 0 that is the initial centroids mu bar this is would be K of 0. These are the initial centroids, now iteratively, now what are the iterations iteratively, step one would be assign each point x j bar, each point x bar j to cluster whose centroid is closest, this is the idea.

So, you look at, so for each point you look at the distances to all the centroids, of each point look at the distance to all the centroids, assign it to the cluster with whose centroid is closest. Step 2 is simpler.

(Refer Slide Time: 36:33)



GIVEN $\bar{x}(1), \bar{x}(2), \ldots, \bar{x}(m)$.

INITIALIZE: $\bar{\mu}_1^{(0)}, \bar{\mu}_2^{(0)}, \ldots, \bar{\mu}_K^{(0)}$

INITIAL CENTROIDS.

ITERATION:

REPEAT UNTIL CONVERGENCE!

STEP1: ASSIGN EACH POINT $\bar{x}(j)$ TO CLUSTER WHOSE CENTROID IS CLOSEST.

STEP 2: RECOMPUTE CENTROID AS MEAN OR AVERAGE OF ALL POINTS ASSIGNED TO CLUSTER.

Take all points in every cluster average it. Step 2 for each cluster recompute centroid or each, so step 2 is recompute centroid as mean or average of all points assigned to the cluster, recompute the centroid and iteratively repeat this, and we repeat these steps until convergence. So, we compute, basically assign start with initialization of the centroids, assign each point to the cluster with the closest centroid. Take the average of all these points in each cluster to obtain the new centroids.

Again repeat, in the next iteration once again assign the points, each point to the cluster with the closest centroid, take the average to compute the centroid keep repeating it, at some point the clusters will not change that is, the closest centroid will be the centroid in that particular cluster itself and therefore, the average that is the centroid will again be the average of all the same points and therefore, things will sort of freeze and not change in fact the algorithm converges fairly quickly as you will observe and many have observed.

It is a very simple algorithm, easy to probably, I mean the math might be seem a little complicated but the idea is very intuitive, very easy to understand, probably one of the easiest clustering algorithms to implement and therefore very popular in machine learning. So, let us stop this here, let us continue our discussion on other applications of linear algebra in the subsequent modules. Thank you very much.