Course Name: **Machine Learning and Deep learning - Fundamentals and Applications**

Professor Name: Prof. M. K. Bhuyan

Department Name: Electronics and Electrical Engineering

Institute Name: Indian Institute of Technology, Guwahati

Week-9

Lecture-32

Welcome to NPTEL online course on machine learning and deep learning fundamentals and applications. In my last class, I explained the concept of k-means clustering, which is an unsupervised clustering technique. In the k-means clustering algorithm, first I have to randomly select k number of centroids. After this, I have to assign data points to the centroids based on the nearest neighbor distance and after this I have to recompute the centroids and this process I have to do iteratively until the convergence condition is not satisfied. And that is the fundamental concept of the k-means clustering. In the k-means clustering, I will be getting the hard decision boundaries.

That means a particular data sample is assigned to a particular cluster. There is no possibility that a particular data sample may belong to another cluster. Today in this class, I will be explaining the concept of the fuzzy k-means clustering. In this case, the concept is very similar to the k-means clustering.

Only one difference is in this case, I will be getting the soft decisions. That means a particular data sample may belong to another class also. That means a particular sample may belong to another cluster also. So that possibility is defined by the fuzzy membership grid. The fuzzy membership grid lies between 0 and 1.

Suppose the fuzzy membership grid is 0.9, that means there is a high possibility that a particular sample may belong to another class. So that consideration I will be considering in case of the fuzzy k-means clustering. And fuzzy k-means clustering and the k-means clustering has many applications, particularly the clustering of data points. And also in case of the suppose image processing applications, this can be applied for image segmentation.

The fuzzy k-means clustering or simple k-means clustering can be applied for image segmentation. This is one important application. et us discuss about fuzzy k-means clustering, which is a soft decision based clustering technique. So in my last class, I explained the concept of the k-means clustering. You can see I have considered some data points and randomly I am selecting some centroid.

So in this example, I am selecting three centroids. One is this one, the red one, blue and the green one. So these three centroids I am selecting randomly for clustering. After this in the step number 1, assign points to the clusters based on the minimum distance. So I have to find the distance between the data points and the centroids and based on the minimum distance, I can assign a particular data point to a particular cluster center.

That is the centroid. After this, I have to recompute the means. And finally what we have to do, repeat the steps 1 and 2 until the convergence condition is not satisfied. So that is the concept of the k-means clustering. In my last class, if you remember, cluster centers I am defining by μ and by C.

So in one algorithm, I have shown the cluster center by μ_i and also in another representation, I have shown as C_i. But today I am considering this cluster center as θ. So I am considering θ as a cluster center or the centroid. So θ we are considering. So in the k-means clustering, first we have to select the cluster centers θ_1, $\theta_2 \ldots . \theta_K$

So these are the centroids. So I am considering suppose this is a vector. So these are cluster centers. And the input vector is $\bar{x}$. That is assigned to a particular class.

The class is suppose ω_j. And corresponding to this class, the cluster center is θ_j. So that means $\bar{x}$ is assigned to the cluster center θ_j based on the condition.

The condition is if $d^2(\bar{x}, \bar{\theta}_j) < d^2(\bar{x}, \bar{\theta}_i) \forall i \neq j$. So based on this distance measure, I can assign the data point $\bar{x}$ to the cluster center theta j.

And that is actually corresponding to the class ω_j suppose. So class information is not available, but we are selecting the centroids randomly. So in case of the k-means clustering, what we have to consider, we are having the clusters.

Suppose this is one cluster, this is another cluster, and maybe this is another cluster. So after doing the k-means clustering, I am getting these clusters.

So this is suppose corresponding to the cluster center θ_1, this is corresponding to the cluster center θ_2, this is corresponding to the cluster center θ_3 like this. So we are getting the hard decision boundaries, the hard boundaries we are getting. So these boundaries, the hard boundaries or that means, this is also called the crisp. That means we are getting the discrete boundaries. The meaning is that this suppose if I consider this particular sample, this particular sample will belong to the cluster corresponding to the cluster center θ_2.

There is no possibility that this particular sample may belong to another classes. Suppose it may belong to this class also, there may be some possibility or this point may belong to this cluster also there is some possibility, but that possibilities we are not considering. So one particular data point is assigned to a particular cluster. That is the hard boundary or hard decision or the discrete boundary we are getting in case of the k-means clustering. In case of the fuzzy k-means clustering, we will be considering the possibility that means a particular sample may belong to another class, classes also.

So this possibility is defined by the membership grade. So I will be explaining this concept. So as I told you the k-means clustering has many applications. Like here I am showing one example corresponding to the image segmentation. You can see this is the input image and we are doing the clustering based on the intensity information.

So in the second figure, you can see the result of the clustering and in this case we are considering the fissure is the intensity value of the pixel. So based on this fissure that is the grayscale intensity value, we can do the clustering. And in the third case also what we are doing based on the color information we can do the clusters. So clustering based on the color. So this k-means clustering algorithm can be employed for image segmentation.

And this is also another example for image segmentation. So you can see I am showing the original image and corresponding to this original image you can see I am considering K=2. K=2 means I am considering 2 centroids and corresponding to this you can see the results of the segmentation.

This result. Similarly, corresponding to K=3 you can see the result of the segmentation K=3.

That means we are considering 3 centroids. So K=10 you can see the results. So if you see the quality of the image in the last row. So you can see the quality of the image corresponding to K=10 is better than the K=2.

So that means the compression depends on the value of K.

If k is equal to high that means I am getting the good quality image segmented image and if k is equal to 2 the quality is not significant and you can see all the segmented outputs corresponding to the original image the original image already I have shown here. So now directly I will explain the concept of the fuzzy k-means clustering. So what is the difference between the crisp case and the fuzzy case. So suppose if I consider a set A okay then suppose we are considering a variable suppose the domain is X and suppose x_i whether it is an element of X x_i is an element of X.

So we are defining the membership grade that is defined by $U_A(x) = 0 iff x_i \notin A$.

$U_A(x) = 1 iff x_i \in A$. A is a set. This is the example of the Crisp set. So what do you mean by Crisp set? I can show you pictorially. So suppose in this side I am plotting x and in this side I am plotting the membership grade.

This is a membership grade. So this is I can write this is a membership grade. So if I consider a Crisp case it is the representation of the Crisp case. So the high value is 1 because it has two value either 1 or 0. So if x_i is an element of A then the output is 1 that means the membership grade is 1 and if x_i is not an element of A then the membership grade is 0. So this is actually corresponding to the Crisp case.

And in case of the fuzzy set what I can show you in case of the fuzzy I can consider a curve like this. This is the representation of a fuzzy set. This is one membership function I am showing like this. So it corresponds to the fuzzy set. So here you see the membership grade it may be 0.5 also or it may be 0.9 also. So the membership grade lies between 0 and 1 corresponding to the fuzzy set. So the membership grade lies between 0 and 1 in case of the fuzzy set it may be 0.9 it may be 0.8 and in case of the Crisp set only two values we are considering either 0 or 1 that is the Crisp set.

So this briefly I am showing to show the distinction between the Crisp set and the fuzzy set.

Now directly I will come to the fuzzy KMS clustering. So let us move to the next slide. So fuzzy KMS clustering it is actually the soft decision based partitioning. That means the meaning is one element may belong to more than one set.

I can write one element may belong to more than one set.

So it depends on the membership grade. Suppose the membership grade is 0.9 that means there is a high possibility that particular data point may belong to another cluster also. In this case we are considering these classes ω_1 ω_2 ω_3 these classes we are considering. So X- is the feature vector X- is assigned to the class ω_1 and corresponding to this the membership grade is $u_1(x\text{-})$ that is the membership grade that is actually the degree of belongingness.

Similarly $\vec{X}$ may be assigned to the class ω_2 and in this case the membership grade is $u_2(\vec{x})$.

So these are actually the membership grade. Membership grade means the degree of belongingness. So for fuzzy K-means clustering we are considering the centroid. The centroid is $\vec{\theta}_j$ that is parameterized representative of the jth cluster. So in this case we are considering $\vec{\theta}$ as a cluster center. So $\vec{\theta}_j$ means the parameterized representative of the jth cluster.

So this vector $\vec{\theta}$ I can write like this $[\vec{\theta}_1^T \vec{\theta}_2^T \ldots \vec{\theta}_K^T]^T$. So all these centroids that means K number of centroids we are considering. Since, we are considering the fuzzy k-means clustering similar to the k-means clustering. In this case also we are considering K number of centroids. So $[\vec{\theta}_1^T \vec{\theta}_2^T \ldots \vec{\theta}_K^T]^T$ these are the centroids the K number of centroids.

And we are considering the matrix the matrix is U that is the N x K matrix. So what is N? N means number of patterns that means number of data points. What is the meaning of K? K is nothing but the number of classes that means number of classes means the centroids. So randomly I have to select K number of centroids.

So move to the next slide. So in my previous slide I defined a matrix U and that is the N x K. N is nothing but the number of patterns and K is nothing but the number of classes or number of centroids. So (i,j) element of U, U is the matrix is $u_j(\vec{x}_i)$ that is the (i,j) element. So now we are considering the distance we may consider the Euclidean distance. So distance between the input vector $\vec{x}_i$ and the $\vec{\theta}_j$ $\vec{\theta}_j$ is nothing but the centroid that is we can find the similarity or dissimilarity between the vector $\vec{x}_i$ that is the input vector and the $\vec{\theta}_j$ that is the centroid.

We find the similarity between $\vec{x}_i$ and $\vec{\theta}_j$. Now we are defining the fuzzy distortion or the cross function the fuzzy distortion or the cross function $J(\vec{\theta}, U) = \sum_{i=1}^{N} \sum_{j=1}^{K} u_{ij}^q d(\vec{x}_i, \theta_j)$. So we are defining the fuzzy distortion or the cross function like this. This u_{ij} already I told you this u_{ij} is nothing but the membership grade that is the degree of belongingness that means the meaning is what is the meaning of membership grade Grade of membership of $\vec{x}_i$ in the jth cluster. So that means what we are considering the grade of membership of $\vec{x}_i$ in the jth cluster.

So for all the clusters we have to determine the grade of membership. In this case you can see I am considering one parameter here the q is the parameter if I consider q=1 that is actually the hard decision or maybe I can write the Crisp that is the Crisp case q=1 and if q > 1 that corresponds to the fuzzy decision. So we may consider q=2 q=3 like this and typically we can consider q=2 that is the fuzzy decision for fuzzy decision q should be greater than 1 and if I consider q is very very high q is suppose very very high suppose q is equal to infinity then actually it is the total fuzzy total fuzzy that means total ambiguity that we cannot distinguish these patterns. So that is not the ideal case the case is q should be greater than 1 that is the fuzzy decision and if q=1 that is nothing but the simple k-means clustering. So if I consider q=1 that is nothing but the simple k-means clustering that is the hard decision or the Crisp decision I am getting.

So these are the parameters the q is a parameter after this I am considering some constraints in case of the fuzzy k-means clustering. So move to the next slide. So some of the constraints we are considering so what are the constraints $\sum_{j=1}^{K} u_{ij} = 1$. So for all the clusters we are considering the membership grade $\sum_{j=1}^{K} u_{ij} = 1$. So where this the membership grade should lies between 0 and 1.

So $i = 1,2,\ldots,N$ that means N number of data points and $j = 1,2,\ldots,K$ that means K number of centroids and also we are considering the condition the $0 < \sum_{i=1}^{N} u_{ij} < N$. So $j = 1,2,\ldots,K$. Actually here you can see what actually we are determining what is the meaning of this is the grade of membership of $\bar{x}_i$ in the jth cluster is related to to the grade of membership of xi to the rest of K-1 clusters. That means we are considering the grade of membership of $\bar{x}_i$ in the jth cluster is related to the grade of membership of $\bar{x}_i$ to the rest of remaining clusters. So these constraints we are considering for the fuzzy k-means clustering.

Now let us write the algorithm of the fuzzy k-means clustering. So move to the next slide the algorithm fuzzy k-means clustering. So first like that simple k-means clustering we have to randomly select the centroids. So choose $\bar{\theta}_j(0)$ is the initial estimate for $\bar{\theta}_j$.

So $j = 1,2,\ldots,K$. So randomly I have to select the centroids that is the initial estimate for $\bar{\theta}_j$. So K number of centroids I have to consider. After this I have to initialize the iteration number. So t=0 that is the iteration number.

So I am considering these two loops. So first loop is for i = 1 to N because we have N number of data points for j = 1 to K, K number of cluster centers and we are considering this is the membership grid in the iteration t $u_{ij}(t) = \frac{1}{\sum_{p=1}^{K}\left(\frac{d(\bar{x}_i,\bar{\theta}_j(t))}{d(\bar{x}_i,\bar{\theta}_p(t))}\right)^{\frac{1}{q-1}}}$. So we have to determine the membership grid. So this expression I can determine from the cost function the fuzzy cost function I can determine this membership grade.

So equation number 1 I will later explain how to determine this one. So we have to determine the membership grade like this and after this end for i end for j this is in the first iteration what actually we are doing we are finding the distance between $\bar{x}_i$ and $\bar{\theta}_j$. So these are the clusters all these clusters from $\bar{x}_i$ we are finding the distances to all the clusters. So suppose this is $\bar{\theta}_j$ so we are finding the distances between $\bar{x}_i$ and $\bar{\theta}_j$.

So based on this actually we can determine the membership grade. So if you see the expression for the membership grade in the numerator you can see what we are doing we are finding the distance between $\bar{x}_i$ and the $\bar{\theta}_j$. In the denominator what we are determining we are considering the distance between $\bar{x}_i$ and all the clusters because the p is from 1 to K so for p=1 we have to determine the distance between $\bar{x}_i$ and $\bar{\theta}_1$ for p=2 I have to find the distance between $\bar{x}_i$ and the $\bar{\theta}_2$ like this for K number of clusters I have to determine this that is in the denominator. In the numerator of the expression of the membership grade I have to find the distance between $\bar{x}_i$ and the $\bar{\theta}_j$ so that I have to determine. After this move to the next slide we have to consider the next iteration because this is about the membership grade how to determine the membership grade based on this we can determine the degree of belongingness the membership grade means the degree of belongingness and after this I have to recompute the centroids. So move to the next slide so we are considering the next iteration that is t=t+1.

Now for j = 1 to K K number of clusters parameter update or I can say the centroid update. So centroid is given by $\bar{\theta}_j = \frac{\sum_{i=1}^{N} u_{ij}^{q}(t-1)\bar{x}_i}{\sum_{i=1}^{N} u_{ij}^{q}(t-1)}$ and end for j. So we have to update the centroids that means I have to recompute the centroids that is very similar to the simple k means clustering. After this I have to consider the convergence condition so what is the convergence condition I have to consider until a termination criteria termination criteria is met. So termination criteria is $||\bar{\theta}(t) - \bar{\theta}(t-1)||$ and we are considering one parameter that is the threshold is epsilon we are considering.

So in this expression actually this actually represents any vector norm any vector norm and this epsilon is a small user defined parameter. In the k means clustering what we considered for the termination criteria if there is no significant changes of the position of the centroids in two successive iterations then I can stop the iteration I can stop the algorithm. Similarly in this case also we are observing if any significant changes of the values of the θ, θ means the centroids in two successive iterations that means $||\bar{\theta}(t) - \bar{\theta}(t-1)||$. So we are observing that whether in two successive iterations any significant changes in the position of the centroids if it is less than epsilon then we can stop the algorithm that is the termination criteria and that is very similar to the k means clustering.

So this is the k means the fuzzy k means clustering algorithm. Now we have defined the membership grade in the equation number 1 in the previous slide. So if you see the equation number here in the previous slide we define this membership grade this is a membership grade. So we determine the membership grade like this in equation number 1 that is the degree of belongingness. So how to determine this u_{ij} and that I can show you so you have to do some mathematics. So let us see how we can determine that expression for the membership grade.

So we have the expression for the fuzzy cost function. This expression already you know so fuzzy distortion function or the cost function we have shown. So this is the fuzzy cost function that already I have explained. Now we are considering the Lagrangian function. So $J(\theta, U) = \sum_{i=1}{}^{N}\sum_{j=1}{}^{K} u_{ij}^{q} d(x_i, \theta_j) - \sum_{i=1}{}^{N}\lambda_i(\sum_{j=1}{}^{K} u_{ij} - 1)$.

So we are considering this after this we have to differentiate J with respect to the membership grade suppose we are considering delta u_{rs}. So if I differentiate this one then I will be getting $q u_{rs}^{q-1} d(x_r, \theta_S) - \lambda_r = 0$. After this we have to do some mathematics I am not showing all the steps. So finally you will be getting the membership grade $U_{rs} = \frac{1}{\sum_{j=1}^{K}(\frac{d(x_r,\theta_S)}{d(x_r,\theta_S)})^{\frac{1}{q-1}}}$ this variable r is considered to consider the number of samples and s is considered to show number of clusters.

So finally we are getting the expression for the membership grade like this. So this is how you can determine the membership grade in the previous slide. So in between you have to do some mathematics I am not showing all the mathematics here you can get in the book but this is the procedure to determine the membership grade. So you have to consider the Lagrangian function. In this class I explained the concept of the fuzzy k-means clustering which is a soft decision based clustering technique. In fuzzy k-means clustering we have to determine the membership grade that is the degree of belongingness.

That means a particular data point may belong to another cluster that depends on the membership grade that is the degree of belongingness. If the membership grade is very high suppose 0.9 or 0.8 that means there is a high possibility that particular data sample may belong to another cluster.

So that is considered in case of the fuzzy k-means clustering. In the simple k-means clustering we are not considering that aspect but in the fuzzy k-means clustering we are considering that belongingness. And you can understand the fundamental difference between the fuzzy k-means clustering and the k-means clustering. So let me stop here today. Thank you.