Indian Institute of Technology Madras
Presents

NPTEL
National Programme on Technology Enhanced Learning

Pattern Recognition

Module 06

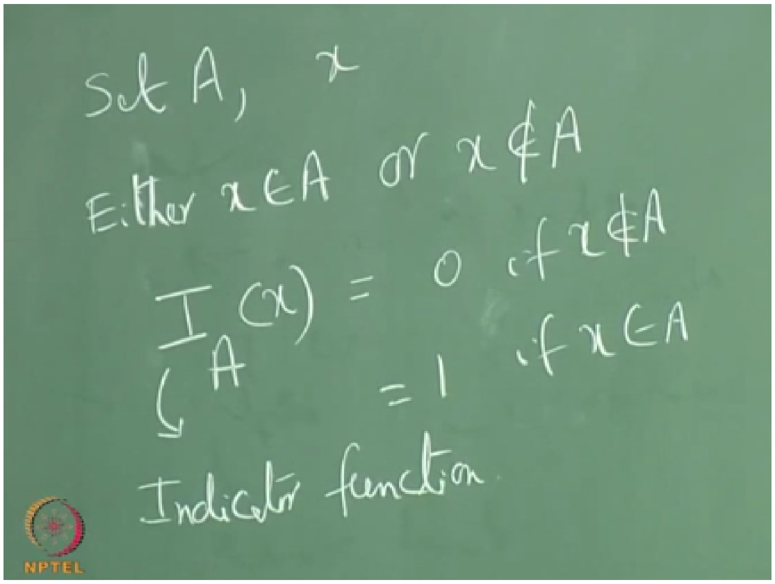
Lecture 07

FCM and Soft-Computing Techniques

Prof. C. A. Murthy
Machine Intelligent Unit,
India Statistical Institute, Kolkata

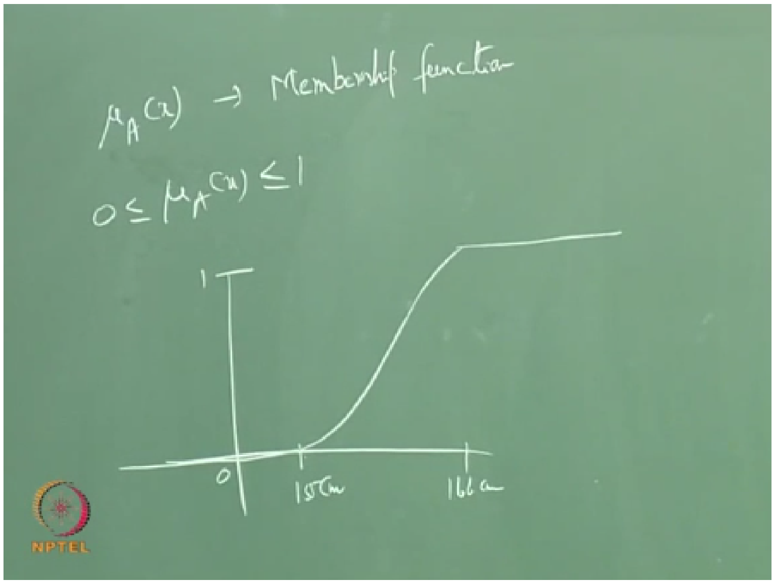
Today I will be talking about fuzzy C means algorithm I will be talking about fuzzy C means algorithm before that let me just say a little bit about soft computing techniques rare fuzziness is one of the components of soft computing techniques soft computing techniques were first introduced by someone named lottery today in 1965 actually he introduced fuzzy sets in his famous paper trust published information Sciences fuzzy sets is a variation from the ordinary sets.

(Refer Slide Time: 01:09)



Note that in ordinary sets if you have a set A and if you have a point x we know that either x belongs to A or x does not belong to only one of them is true another way of saying it is you have an indicator function of the set a where this IX = if x does not belong to A and 1 if X belongs to A I is said to be indicated function of the set indicator function so they defined fuzzy sets as.

(Refer Slide Time: 02:16)



You have a set A with the membership function value membership function which is μ for a fuzzy set A it takes values not only 0 and 1 but it can also take values in between 0 and 1 that is this is known as membership function membership function and the μAx takes values in the interval 0 to 1 not only 0 A1 it may take a value in between.

There are many examples one example is suppose you want to look at the set of all let us say tall persons and then if someone's height is let us just say five feet, five inches and would you just call him tall probably all the persons whose height is more than 5 feet 7 inches you may call them tall and probably all the persons whose height is less than let us just say 5 feet 2 inches you may call them not tall.

But from 5 feet 2 inches to 5 it is 7 inches you have you may not have a clear-cut idea of what exactly is tall probably. Let us say this is 155 centimeters and then say it is 150, 164 let us just say 166 centimeters yes I have taken some values so and here I am writing membership function say this is zero then says it is one if the height is 155cm probably you may not want to call the person as tall if it is 166cm surely you would like to call the person as tall and anything greater than 166 you would like to call the person as tall anything less than or equal to 155 it is surely not tall this is this one and between 155 to 166 probably you have like this and this is called membership function.

The membership first the attribute tall is will be increasing and it will go to one this many times in reality we use adjectives without clear-cut mathematical formulations without clear-cut mathematical formulation exact and precise mathematical formulation is out there and by using those adjectives when I speak to you or when you speak to me when use those adjectives are when I use those adjectives.

We understand each other when we are able to understand each other and we would like to have we would like to have our computer also to understand these words we would like to have the computer also to understand the verse that is the basis for soft computing I mean that is the basis for fuzzy set theoretic that is the basis for fuzzy sets.

It is also the basis for what is known as fuzzy logic what is known as fuzzy logic in the usual binary valued logic that is 0 or 1 a statement is either true or it is not true but in fuzzy logic that is a statement is true with some membership value rather just say 0.5, 0.6, 0.7 and it is not true with some membership value find again with some membership.

And like that you can have I mean remember ship functions in many problems which is actually the case in reality in reality we do not always make mathematically precise statements in fact many times most of the times we do not make mathematically precise statements even then we understand each other then we would like to have our computer also to understand the language which we are speaking then come then the analysis should be done in such a way that this understanding is possible.

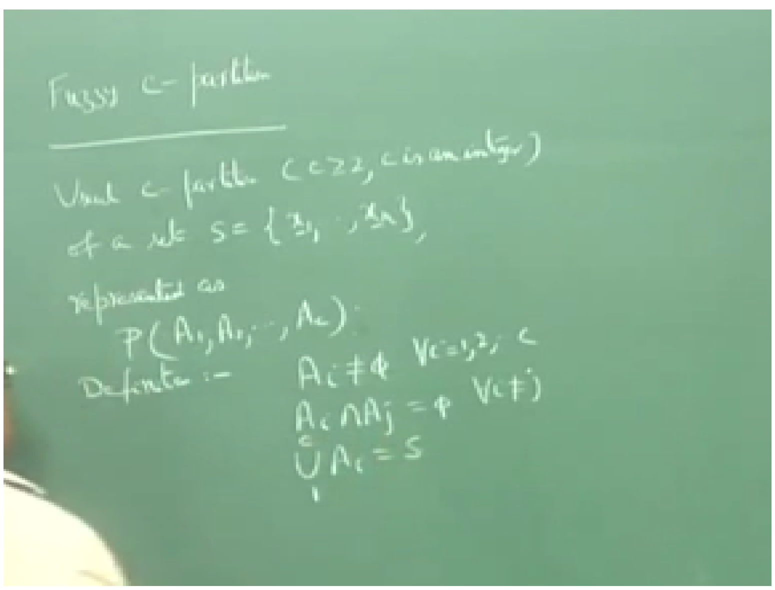
So one of the ways that the suggested eth 2 x 2 use these number shift values in the analysis so there are very many developments after fuzzy sets neural networks is considered to be a part of soft computing multi-layer perceptions and radial basis function networks you have principal component analysis networks these are considered to be part of soft computing genetic algorithms are considered to be part of soft computing rough sets are considered to be part of soft computing soft computing actually means.

That we can have input imprecise we can have ambiguous input and sometimes even in character C when we discuss many things sometimes we make incorrect statements but even then we understand each other always whenever I mean the basic significance of all these things is we would like to make our computer to understand and to behave in such a way as a human being since we sometimes make in correct statements still we understand each other then how do you make the computer also understand these sort of things.

Basically with this brief introduction to soft computing like fuzzy sets neural network genetic algorithms and rough search okay I am not going to discuss all these things in this class I will have only talked about fuzzy sets because I will be using them in the fuzzy seeming algorithm Fuzzy sets means you basically have some membership values you basically have some membership values in the membership values are used in several applications and that is the reason why I introduced fuzzy sets A set a is said to be a fuzzy set.

If the membership values they lie between 0 and 1 n and there is at least one point for which the membership value is strictly in between 0 and 1 if all the membership values are either 0's or 1's then it is going to be called as a crisp set if all the membership values are either 0's or 1's and nothing in between 0 and 1 then the set is called as a crisp set position is there exists at least one membership value which lies strictly in between 0 and 1 it is neither 0 or 1 at least 1x for which µix is strictly in between 0 and 1 so that is fuzzy sets.

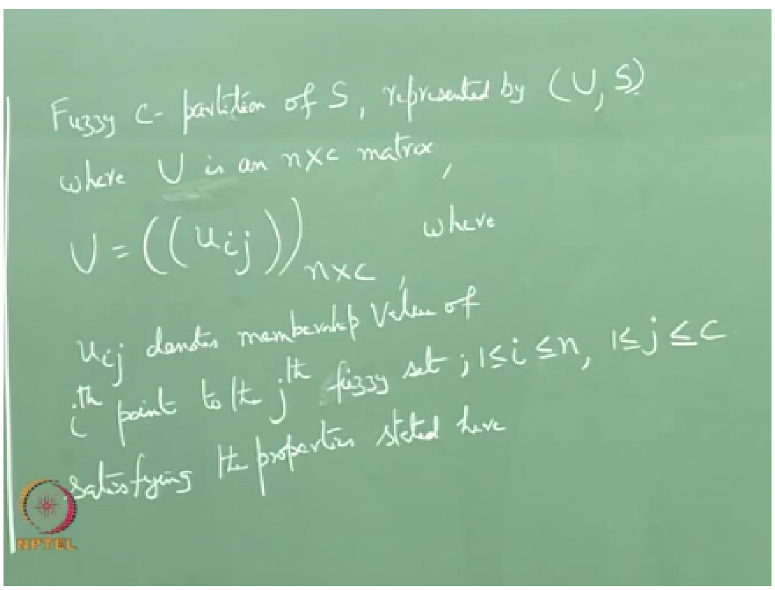
(Refer Slide Time: 11:45)



Now using these fuzzy sets we can also discuss what is known as fuzzy partition fuzzy C partition fuzzy C partition first let me tell you the meaning of ordinary partition that we all know but I will repeat it again then we will go for fuzzy C partition now the usual C partition C is ≥ 2 and C is an integer usual C partition of a set S is let us just say x1 to xm let us say is x1 to xn is what it is like this usual C partition of a set s represented as let me just call it P is for partition A1 A2 Ac you need to give partition of S x C subsets partition of S X C subset C is the number of subsets is what represented as this.

And the definition is what is the definition is first each Ai ≠ 5 right and Ai intersection Aj = right and Union right partition of S in to see subsets that is the urban air to AC if I represent the C subsets as a 180 AC then the definition is that each Ai ≠ ϕ and Ai intersection Aj is ϕ and Union Ai 1 to c = S so this is a usual C partition now what is fuzzy partition.

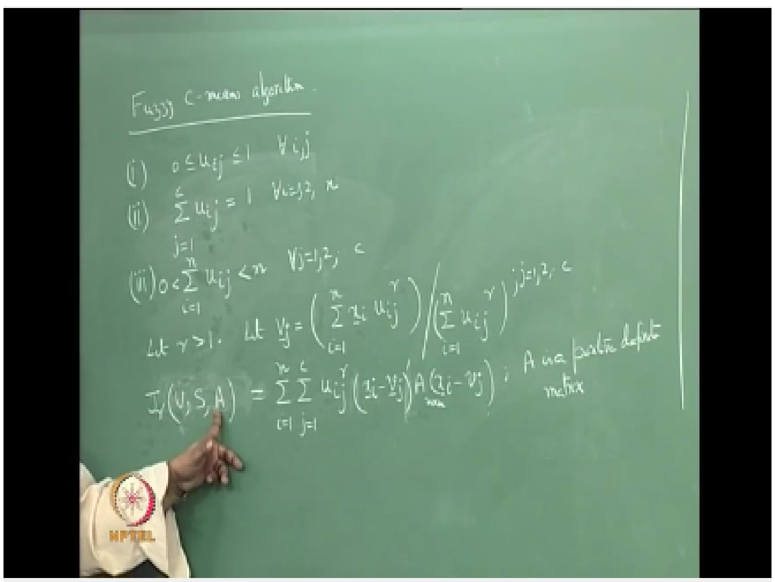
(Refer Slide Time: 14:43)



Now fuzzy C partition of s basically you need to give membership values basically you need to give membership values fuzzy C partition of S represented by represented by U is U S is the set okay where U is n / c matrix where U is and n/c matrix U is equal to so i^{th} row this column element is okay where uij denotes membership value membership value of i^{th} point to the j^{th} set j^{th} fuzzy set.

j^{th} fuzzy set naturally the number of points is n so I lies between 1 to n and C1 $\leq$ j $\leq$ c fuzzy C partition of S represented by US where U is an n / c matrix Uij denotes membership value of i^{th} point with it fuzzy sets okay satisfying the following properties satisfying the properties stated I do not I do not want to write below because I need to write there so satisfying the properties stated let me just try it state it here just one minute I will erase this portion.

(Refer Slide Time: 19:05)



The first property is naturally 0 $\leq$ uij $\leq$ 1 for all ij membership values, two $\sum$ uij, j = 1 to c this means you have taken diet point you fix the ith appoint its membership for that for the first set is ui1 a 1 no matter for the second set is ui2 number three for the jet-set is ui_c membership at the safe set is you I see, so this $\sum$it is =1 okay its $\sum$is =1 in the usual partition if your point X belongs to 1 of the sets it is not going to belong to any other set, so for that point if you sum up all the membership values 1 place you will get 1 and at all other places you are going to get 0 so $\sum$is 1.

Here the $\sum$ is 1 but it is not necessarily true that exactly at 1 place here you have 1 and at other places we have 0 you may get some fractional values here 3 is $\sum_I$ is =1 to n u_{IJ} that is you are fixing the set now and you are looking at the sum of all the membership values for a particular set, it should be strictly > 0 it should be strictly < n. The corresponding thing here in this properties is $Ai \neq \varphi$ for all I, that means every set has at least 1 point.

If a set has no point then all the membership values will be 0 then the $\sum$ also will be 0 right, if a set has no points all the membership values will be 0 $\sum$is also 0 so this is actually an extension of this property this is actually an extension of this property. So if you u_{ai} satisfies all these three properties then you say that is a fuzzy sea partition of the data set, that is a fuzzy sea partition of the data set.

Now when we have usual partition what did we do we have defined an objective function if you remember for the c means or k-means algorithm when we did it we defined an objective function what was the objective function? The objective function is you have this partition and we define V_I mean of a Ai is =1 to C we have got a partition for this partition for every I mean of a I okay is defined and then what did we do we looked at this double $\sum$of X -, suppose we have this set s and this sets is a subset of M dimensional space s is a subset of M dimensional space Euclidean space.

Then we can talk about mean of A_I sum of all the points in a divided by the number of points in a you are going to get mean of A_I then this is objective function what is this objective function? For each point X we just look at which cluster it belongs to it belongs to the cluster say A_I then for that point X you look at the mean of that cluster the mean is V_A find out the distance between X and Vi and then take the square of the distance, that you do it for every X and I and i is =1 to C.

So this is basically the within cluster distance, that partition that which provides the minimum within cluster distance is they can at the best partition, that partition that provides the minimum within cluster distance that is taken as the best partition. So basically this is to be minimized over all these partitions this is to be minimized over all the partitions, this is the minimum within cluster distance criterion and the usual C-means algorithm or K means algorithm I am using both these words C means and k-means.

The reason is that when this algorithm was proposed some time in 60 it was proposed ask-means algorithm and that is what people have been using it and fuzzy set theory people they devised this algorithm that was in late 70s and early 80s late 70s and by best debt and he used C for the number of cluster, he used C for the number of clusters. So number of clusters C K fine I mean you give some name for the number of clusters and then he devised the algorithm.

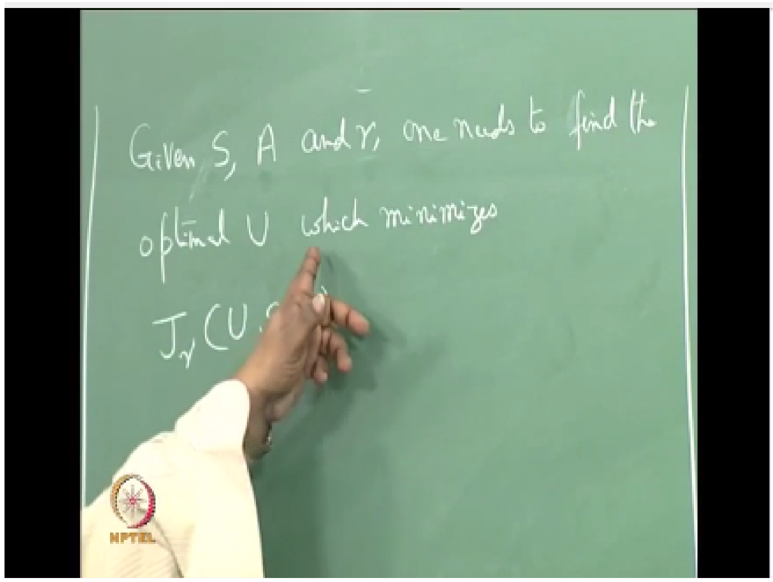
So c or k there is a number of clusters and here C is the number of clusters, so it is C means algorithm, now it is the minimum within cluster distance criterion is this and since we are not in a position to get the minimum within cluster distance, since we need to do the search over the number of such partitions is too many. So there the algorithm k-means or the C algorithm was devised, the C means algorithm was devised.

Here also in the fuzzy set theory setup also we have defined a partition and there is an objective function what is the objective function the objective function here is the following, we take a parameter let r ≥ strictly greater than 1 let r ≥ strictly greater than 1, let V i that is mean for the ith cluster is ∑ X_I into sorry let me write it as j is for the clusters, so Vj is ∑ X_I x i x ui whole' ^ r we are multiplying X_i buy you a j whole' ^ r, if ui_j are r either 0 are once the place when it has 1 the corresponding axis will be some divided by the number of such points.

The place where ui_j are 1 if you just sum up all those things that will give you the number of points in that cluster, that is you ui_j r^{Σ} that will give you the number of points and his ∑ of X_i divided by the number of point, so that is means ∑of X i divided by the number of points, so that is me this is the mean. Now objective function the objective function is dependent upon this value our objective function if you start with if you have A the c partition U, S and r is with respect to the parameter r what is this one?

This is actually going to be u_{IJ} ^{R} norm of actually xi - Vj' A and let me just write it like this (U, S, A) where is a positive definite matrix a is a positive definite matrix, note that it should have M rows and M columns, note that it should have M rows and M columns, if A identity matrix and you U, S, A are either 0 or 1 then actually what you are going to get is this, let me repeat it if A is identity matrix and you U, S, A are either 0 s or 1s and nothing in between then this ∑ is same as what I wrote here.

So this is we are looking at it in a much generalized setup, first this U it is not either 0 are 1 something in between and we introduced a positive definite matrix and there is an r so this is to be minimized for a given A and S given A sand R you need to get U which minimizes this that is the problem formulation.
(Refer Slide Time: 34:58)

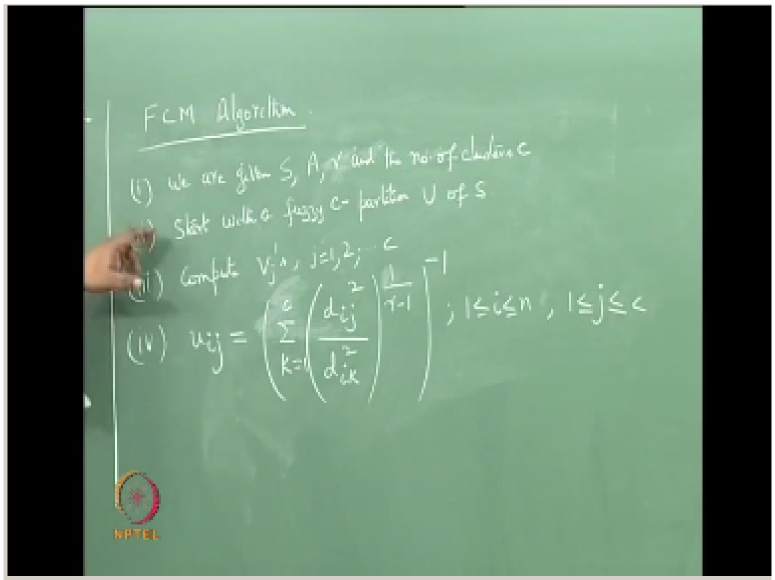


I will write it here given S the data set A the positive definite matrix and r 1 needs to find the optimal you which minimizes J are you given S is the data set at the positive definite matrix and r the exp term r is the exp term, 1 needs to find the optimal U which minimizes J are you as U is the membership value matrix, S is the dataset and A is the positive definite matrix. So till now we have been discussing the problem formulation in the physician's algorithm.

We have defined an objective function J we have defined an objective function J, this function depends for every fuzzy C partition you of the set S and for a given r > 1 and for a given positive definite matrix A the objective function value is given by this expression, we are supposed to find the best U given S, A and r one needs to find the optimal U are the best U best from the point of view of that you which minimizes Jr, U, S, A.

Now how does 1 do it before I go into the algorithm part I would like to mention a few things this algorithm was first proposed by James Bedeck, a very famous figure and a proof was also given to the algorithm that it converges and then the proof was modified and then a modified proof was given but then ultimately after a I should say a few iterations ultimately the correct proof was found the algorithm was same throughout in all these things the proof needs knowledge quite a bit of knowledge of mathematics including topology and other such fields I will not going to the proof of the algorithm I will basically tell you the steps of the algorithm.

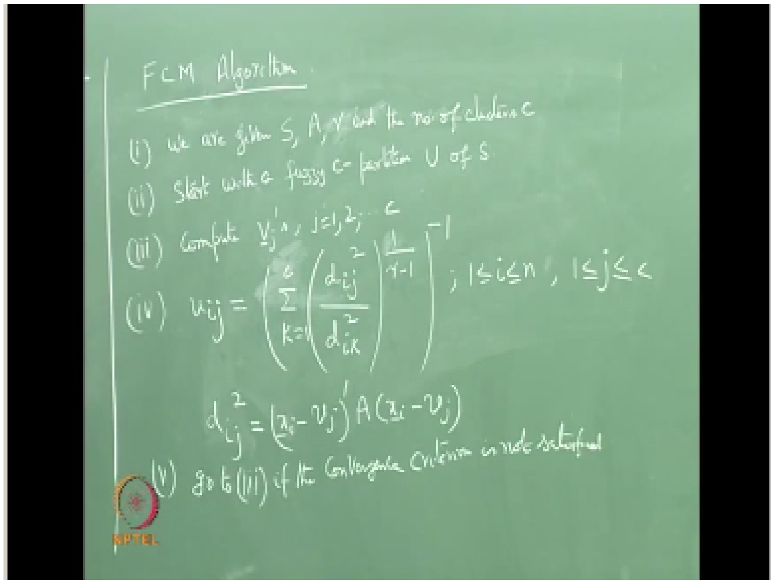
(Refer Slide Time: 38:31)



So this is f c m algorithm F for fuzzy c means fuzzy c means algorithm what are the steps first one is we are given the set S which has smaller number of points and it is a subset of RM m dimensional space we are given the positive definite matrix a and then exponent are I am saying that we are given.

Okay usually people choose the value of R some R greater than 1usually people choose the value of R something that is greater than 1 and people choose it as something which is very close to 1 from one point zero one some people make use to also but this value is generally chosen by users and of course the number of clusters see this C is same as the same we will start with a fuzzy c partition U of S.

(Refer Slide Time: 39:54)



So once you is there you can always calculate the means V J's what is the expression for me the expression for mean is this, this is the expression for me so you can always calculate this you can always calculate this expression so you can always calculate V and once you calculate VGS you can calculate you I guess how the expression for u IJ is this summation K =1 to C of D IJ square by D I K^2 K = 1 this K is 1 to C this thing to the power 1 by r – 1.

And hold to the power of minus 1 here R < 1 that is used then you can write R - 1 what are this dij ^{2} and DA K^2 actual K = 1to C and D IJ square is this portion dIJ ^{2} is this portion that is you understand why it is done as D D is the distance, distance of the earth point to the jet mean that is di J^2 this is di a square in the numerator and the denominator di K^2 K = 1 to see this thing whole to the power 1 /r - 1 $^{-1}$ okay.

So you compute given you are computing V's from there you and from you again we then again you and again V the fifth step is go to three if the convergence criterion is not satisfied go to three if the convergence criterion is not satisfied what is the convergence criterion you have now a previous fuzzy C partition and present for the C partition two fuzzy C partitions you take the difference that is again a matrix okay.

And then take the difference take the modulus values if each of them is say less than say some epsilon then you stop which otherwise you do this you can so you can just go on and on and on doing till that is satisfied you can have a difference stopping criterion also here I use the stopping criterion on use you can do the same thing on visa also if two consecutive set of vectors means the previous mean.

And the present mean of the same cluster if the distance between them is less than Epsilon and if that happens for all the C clusters they also even stop it you do that or you do this it doesn't matter we just go on and on doing this so convergence for, for the convergence criterion you need to have a small value epsilon and you need to calculate the difference between either that to fuzzy C partitions are that to see mean vectors right either you have to calculate.

The difference between the two fuzzy C partitions consecutive two for the C partitions are consecutive two fuzzy C means if the each difference you say less than epsilon then you stop it otherwise you go to that calculate the next one just go on doing and the convergence theorem says that this process really converges that is if you go on and on doing this thing then you will surely.

I mean ultimately the you I guess the previous usual presently whether they will become same or the means also the previous mean the present means they will become saying ultimately the differences they will go towards zero that is basically the convergence theorem now the next point is what sort of clusters are we going to get right so we have got an algorithm and we can use it and it is converges.

So you get a you write what you may do is that since it is fuzzy clustering and you are ultimately interested in getting the ordinary clustering you can proceed in a few ways one way is that all the membership values which are greater than0.5 okay are for each point what you can do is that find the maximal membership value to whichever cluster it goes you there run okay give that thing one and the rest are 0 this is a wave which is followed by many.

Okay in that way you will get a crisp clustering that is one thing that you can do it you can also do one thing you can keep the fuzziness as it is many times you need overlapping clusters what this physicians algorithm does is ultimately it provides you overlapping clusters okay because the same point has membership to more than one class so what you are all timidly going to get is overlapping clusters.

Now when you get overlapping clusters in many situations were lapping clusters are important why they are important you do not want to say that it is always this cluster are always that cluster may be you do really have I mean that point maybe actually put in more than one cluster why do you want to force it to go to only one question maybe keeping the information that it can go to both the clusters may help you these sort of situations.

They arise many times in I will tell you one example with this sort of situation arises many times these things they occur in the satellite image processing when you have taken when you take the image of a place from the satellite say the resolution is say let us just say even if it is one meter resolution suppose you take the bridge on the Knavery River okay the photograph is taken from above so you see the bridge.

And under that there is water now that those pixels they can go to either water or they can go to bridge they can go to concrete structures or they can go to water you may want to keep these two different two clusters information intact you may not force it to go to always to either water or to I mean either the water are to the other one other cluster the reason is that if someone is interested in crossing the river by boat.

Then he would want the information that this is water if you are interested in crossing the river by railway line and by using railway line then you need the bridge so why do you force it to go to either always this are two that you may want to keep this overlapping information intact so that wherever you need.

The that part then you may want to keep it there are you understanding me you need not force it always to go to one of the clusters you need not force it to go to always one of the clusters keep it as it is so that only when ultimately you are forced to be a classification are forced to do clustering depending on the surround depending on some information then okay in this dependent under these circumstances you would like to use this water.

So we use the water these sort of things are extremely necessary in very many real-life problems in very many real-life problem for the same thing you may have more than one alternative you may want to keep both the alternatives intact you may not force yourself to go to only one of them so that exactly at the right at the right time you will decide which alternative you need to take.

Okay in fact this is the reason for I mean fuzzy set theory based classification or fuzzy set theory based clustering is has become popular because of this particular reason you may you may not always want to force it to go to one of the clusters are one of the classes you may do it at the right point of time depending on the situation there okay.

So you can keep the overlapping information intact so that you can you will use it for future purposes depending on the problem and thirdly there is one more thing what sort of clusters are we going to get they are essentially again convex sort of clusters if, if this a is the usual identity matrix okay then you will get basically convex sort of clusters again the problems regarding non convexity.

And other things they need to be properly addressed in this context with this I stopped okay we have question show to choose the value of our exponent I do not want to always I do not want to make a strong statement that our should be chosen always in this way or in that way I do not want to make some such statement what I can say is that people choose our sometimes many persons choose our in some, some particular race.

For example someone may choose R as something very close to one someone point zero one because you may want to get something very close to the usual k-means algorithm as you can see far is actually equal to one then this is like k-means algorithm right the usual k-means algorithm so that is why one of the reasons that people choose R to be very close to one is this they want to make this thing to be as close to k-means as possible unusual k-means algorithm.

So they may choose R to be a value that is very close to one if your question is what is the performance of this algorithm for different values of R how do you compare the performance of this algorithm for different values of R, R has some impact naturally or has some impact on the final output R has some impact on the final output, output changes if R value changes that is also true I think except.

That I do not want to make any statement on the I mean if R is very, very large than what is going to happen so these things I do not want to talk about it now because it depends on the data set there are very many other things involved here so I don't want to make a statement about what will happen if R is very, very large and very, very small okay other questions.

This is one of the most popular algorithms in I mean by algorithms are physicians algorithm many people apply this one in fact probably the word fuzzy is misnomer the$\sum$UI J =1 for each point some other memberships for different clusters is 1 then basically you are looking at a probabilistic setup right probability of the highest point going to the cluster one is some value cluster two is some value.

So some of this probabilities is equal to one so probably the word fuzzy is a misnomer because you are making it the summation has one so maybe that is also one of the reasons why many problems like this method many problems like this method okay because of that and it has nice mathematical proof thought he proof is complicated I went through each and every step of the proof, proof is good but it is a complicated one other questions right thank you.

End of Module 06 – Lecture 07

Online Video Editing / Post Production
M. Karthikeyan
M. V. Ramachandran
P. Baskar

Camera
G. Ramesh
K. Athaullah
K. R. Mahendrababu
K. Vidhya
S. Pradeepa
D. Sabapathi
Soju Francis
S. Subash
Selvam
Sridharan

Studio Assistants
Linuselvan
Krishnakumar
A. Saravanan

Additional Post – Production
Kannan Krishnamurty & Team

Animations
Dvijavanthi

NPTEL Web & Faculty Assistance Team
Allen Jacob Dinesh
Ashok Kumar
Banu. P
Deepa Venkatraman

Dinesh Babu. K.M
Karthick. B
Karthikeyan. A
Lavanya. K
Manikandan. A
Manikandasivam. G
Nandakumar. L
Prasanna Kumar. G
Pradeep Valan. G
Rekha. C
Salomi. J
Santosh Kumar Singh. P
Saravanakumar. P
Saravanakumar. R
Satishkumar. G
Senthilmurugan. K
Shobana. S
Sivakumar. S
Soundhar Raja Pandian. R
Suman Dominic. J
Udayakumar. C
Vijaya. K.R
Vijayalakshmi
Vinolin Antony Joans

Administrative Assistant
K.S. Janakiraman

Principal Project Officer
Usha Nagarajan

Video Producers
K.R. Ravindranath
Kannan Krishnamurty

IIT Madras Production

Funded By
Department of Higher Education
Ministry of Human Resource Development
Government of India

www.nptel.ac.in