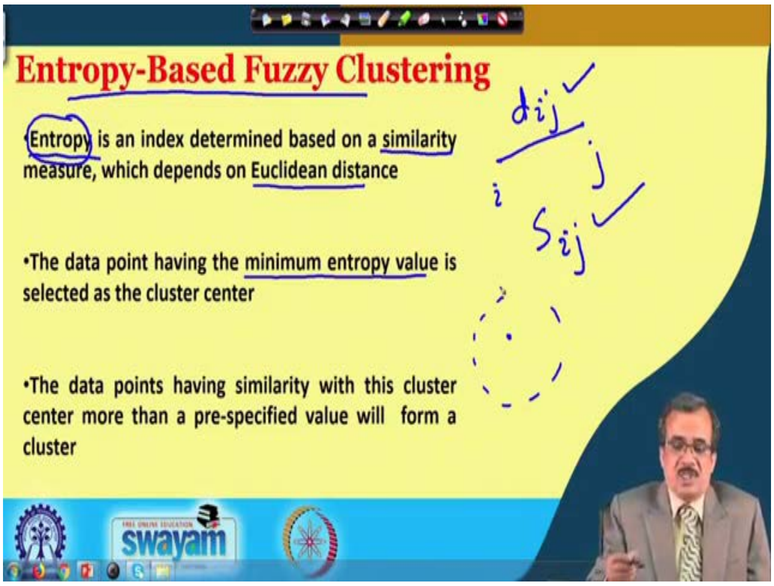
Fuzzy Logic and Neural Networks
Prof. Dilip Kumar Pratihar
Department of Mechanical Engineering
Indian Institute of Technology, Kharagpur

Lecture – 15
Applications to Fuzzy Sets (Contd.)

(Refer Slide Time: 00:14)



We are discussing fuzzy clustering algorithms. Now, we have already explained the working principle of fuzzy C-means clustering with a suitable numerical example. Now, we are going to start with another very popular algorithm, which is known as entropy based fuzzy clustering algorithm. Now, here, we are going to use a term, that is called the entropy and which is nothing, but an index. Now, this particular index is used just to identify which one should be the cluster centre.

Now, supposing that we have got a large number of data points in multidimensional form and our aim is to identify, which should be the cluster centre. Now, what you do is, we use the concept of this particular entropy, and this entropy value, if you want to calculate, we will have to take the help of one value that is called the similarity value. And, similarity is based on the numerical value of Euclidean distance, now supposing that I have got two points say point i and another point say point j. So, very easily, I can find out, what is the Euclidean distance and that is denoted by d_ij. And, if I know the

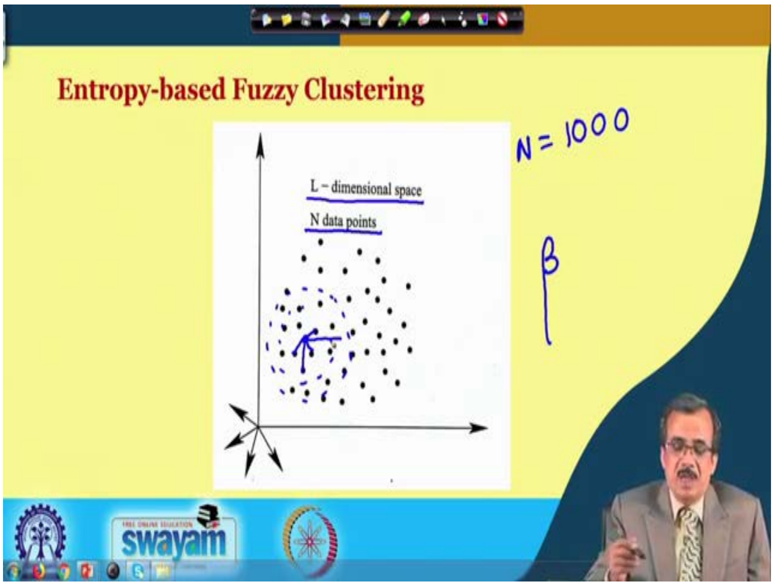
Euclidean distance, I can find out the similarity between these two points, that is denoted by S_ij.

Now, if the distance between the two points is more, the similarity will be less and vice-versa. And, as I told, once I got this particular Euclidean distance, we are in a position to calculate the similarity. And, if I know the similarity information, we can find out this particular entropy, and I have already mentioned that this particular entropy is an index, which helps us to decide, which one should be the cluster centre.

Now, the point which is having the minimum entropy value is selected as the cluster centre. Now, once I have got this particular cluster center, supposing that I have got this is the cluster center. Now, surrounding this particular cluster center, we have got a number of data points, now we will have to take one decision. So, out of all the data points, which are going to enter that particular cluster.

Now, what I do is, we try to take the help of similarity, once again and there will be some threshold value of similarity. Now, the data points surrounding this particular cluster centre, which are found to have similarity greater than equals to some pre-specified value will be encouraged to enter this particular cluster. Now, this is the way actually, we do the clustering and this particular thing, we are going to discuss in much more details.

(Refer Slide Time: 03:36)



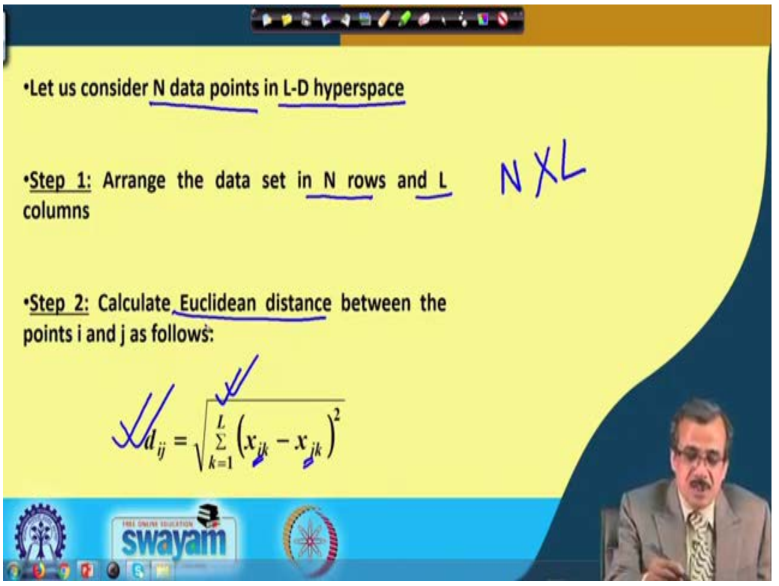
Now, here, I am just going to take one typical example, supposing that we have got a large number of data points. So, capital N number of data points and supposing that this particular capital N is equal to say 1000. So, we have got 1000 data points and these data points are in higher dimension say L dimensional space; that means, if I want to represent a particular point, I need to have capital L number of numerical values.

Now, what we do is, by using this entropy-based fuzzy clustering, as I discuss, the first thing we do is, we try to identify, which one should be the cluster centre. Now, supposing that we have calculated entropy for each of the data points, and now how to calculate I will be discussing in much more details. But, supposing that we have got the entropy values, the data point which is having the minimum value of entropy will be declared as a cluster centre. Now, supposing that let me assume that, this particular data point is having the minimum entropy value. So, this is nothing, but the cluster center.

Now, the moment, we declare that this is the cluster center, and now surrounding this, we will have to find out actually a few members which will also belong to this particular cluster. Now, what we do is? So, we try to compare the similarity of the data points with this cluster centre. So, how to find out the similarity? So, what we do is, we calculate the Euclidean distance and using the Euclidean distance information, we try to find out the similarity.

Now, if this particular similarity with respect to the centre of the cluster is found to be greater than or equal to some threshold value denoted by β, then we allow those points to lie within this particular the cluster. So, this is the way, actually we do the clustering using the entropy-based fuzzy clustering. Now, I am just going to discuss in much more details.

(Refer Slide Time: 05:52)

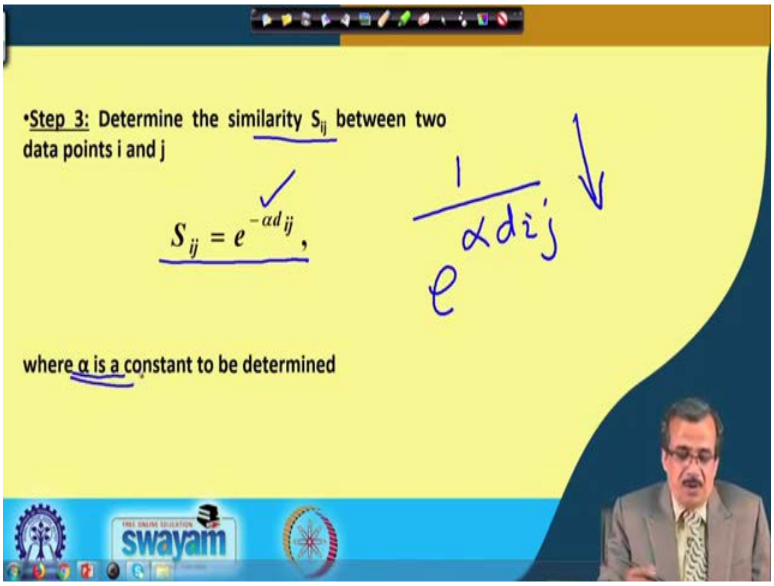


Now, let us consider that we have got capital N number of data points and these data points are in L dimensional hyperspace and our aim is to actually carry out this particular clustering; that means will have to divide these data points into a few clusters and these clusters will be fuzzy clusters. Now, we take the help of a few steps now, step 1: we arrange the data point in N rows and L columns. So, N rows means I have got capital N number of data points and each data point is having L dimensions; that means, L numerical values.

And, we take the help of one matrix and its dimensions are actually $N \times L$. So, there are N rows and L columns. Now, step 2: we calculate the Euclidean distance between the two data points i and j, say using these particular well-known formula. So, $d_{ij} = \sqrt{\sum_{k=1}^{L} (x_{ik} - x_{jk})^2}$, it is very simple.

Now, here k varies from 1 to L, L is nothing, but the total number of dimensions. Now, here, we consider k equals to 1 to L and I have got two points, one is your i and another is your j, now dimension-wise we try to find out the difference square of them and after that we add them up and we take the squared root of that. So, this is the way we calculate the Euclidean distance between the two points i and j, now once, I have got the Euclidean distance between the two data points.

(Refer Slide Time: 07:47)

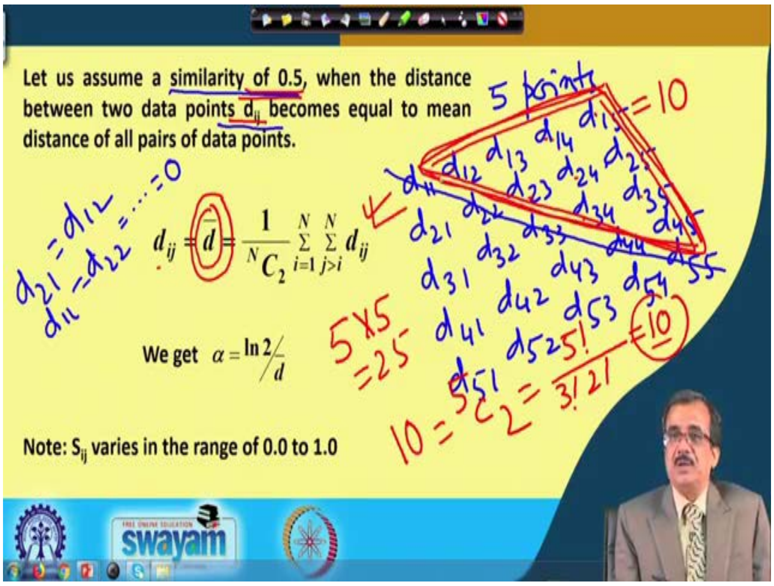


Now, actually, we will have to find out what is the similarity among them. Now, as I told that if the Euclidean distance between the two data points is more, their similarity is less and vice-versa. Now, let us see how to represent this particular relationship in the mathematical form. Now, in step 3; we try to find out the similarity S_ij between the two data points, that is your i and j. Now, the way the similarity and this particular distance relationship has been written is as follows, like $S_{ij} = e^{-\alpha d_{ij}}$.

Now, if d_ij is more then this becomes 1 divided by e raise to the power alpha dij. So, if d_ij is more so, this particular expression is going to be reduced; that means, your similarity is less. So, if the two data points are too far. So, that similarity will be less and vice-versa. Now, here, I just want to mention that this particular relationship is actually not the unique.

Now, I can write down this particular relationship between the Euclidean distance and similarity in a slightly different way also, but this is actually the method, the proposer used, so I am just going to use the same expression, that is, $S_{ij} = e^{-\alpha d_{ij}}$, where α is a constant and the value for this particular constant is to be determined. Now, how to determine the value of this particular α, that I am going to discuss.

(Refer Slide Time: 09:49)



Now, here actually to determine the value of α, we assume a similarity of 0.5, when the distance between the two data points that is d_ij becomes equal to the mean distance of all pairs of the data points. Now, let me take a very simple example, supposing that I have got only 5 points here. So, if I have got only 5 points, what could be the possibilities of the distance values, the distance could be your d_11 then comes d_12, d_13, d_14, d_15.

Then, the distance between 2 and 1, 2 and 2, then comes 2 and 3, then d_24 then comes your d_25 then comes d_31, d_32. So, d_33, d_34 then comes d_35 then d_41, d_42, d_43, d_44 then comes d_45 then d_51. So, d_52 then comes d_53 then comes d_54 and then comes your d_55.

Now, here so, d_12 means, what is the distance between 1 and 2 and d_21 is the distance between 2 and 1. So, we assume that your d_21 is equal to your d_12, similarly d_41 is equal to your d_14 and so on. Now if you concentrate on the diagonal elements that is d_11, d_22, d_33, d_44 and d_55 so, these diagonal elements if we concentrate. So, the distance between 1 and 1 or 2 and 2 so, these are all equal to 0.

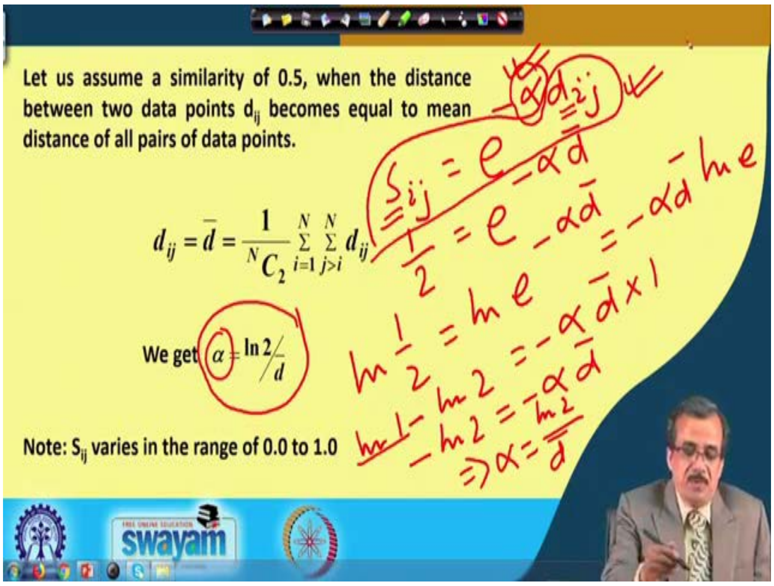
So, distance between 1 1, 2 2 and so on and that is equals to 0. And, moreover we have already consider the d_21 is nothing, but d_12; that means, your if I just concentrate on only one side of this principle diagonal, I will be able to find out the distance values for example, if I just concentrate on these the distance values; that means, your d_12, d_13, d_14, d_15, d_23, d_24, d_25, d_34, d_35 and d_45. So, I will be able to find out the distance values.

Now, here, we consider 1 2 3 4 5 6 7 8 9 10. So, we consider the 10 distance values. Now, here, if I just know these particular 10 distance values, my purpose is served and I did not calculate all 5 multiplied by 5, 25 distance values. Now, these particular 10 is actually nothing, but is your ${}^5C_2$ and ${}^5C_2$ if we calculate, this is 5 factorial then comes your 3 factorial, 2 factorial. So, this is nothing, but 5 multiplied by 4 that is 20 divided by your 2 and this is nothing, but 10.

So, if I know these 10 information, my purpose will be served. And, what we do is, we determine your the $\bar{d}$, that is the mean distance. So, we consider 1 divided by ${}^NC_2$. So, this ${}^NC_2$ is nothing, but 10, because here N is equal to 5, according to this particular example. And, then, we try to find out summation i equals to 1 to N, summation j is greater than i to N, d_ij; that means, we consider only one side of this particular the triangle and this is the way, actually we can find out, what should be the average distance value, that is nothing, but is your $\bar{d}$.

Now, let me come back. So, we assume a similarity of 0.5, when the distance between the two points that is d_ij becomes equals to $\bar{d}$. So, d_ij equals to $\bar{d}$ and this is nothing, but this particular expression. Now, here actually, if I just derive, I can find out what should be the suitable expression.

(Refer Slide Time: 14:43)



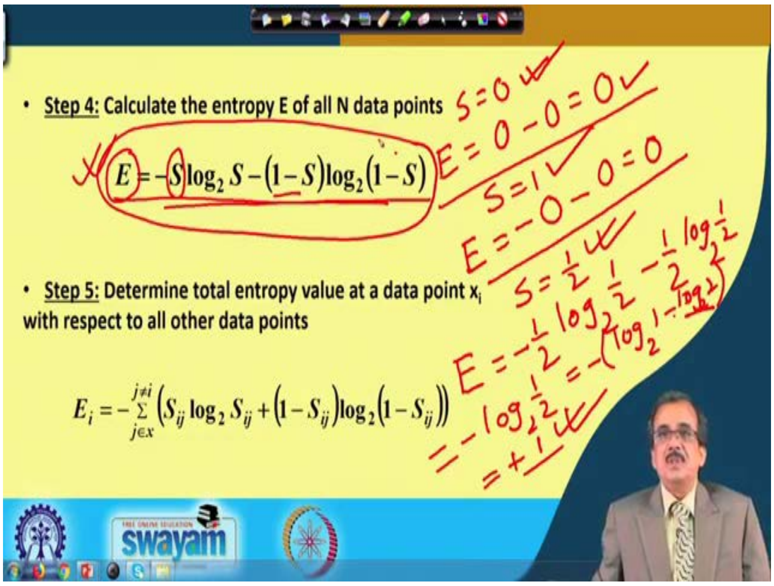
For this particular α, now let me try to derive here the similarity. So, if I just see the expression for the similarity that was nothing, but S_ij. So, e raise to the power minus alpha d_ij.

So, this is the relationship between the similarity and the Euclidean distance and our aim is to derive this particular the expression for α. Now, what we do , we consider a similarity of 0.5. So, 0.5 is nothing, but say 1 divided by 2 and that is nothing, but is your e raise to the power minus α and these d_ij is nothing, but the mean distance, that is, $\bar{d}$.

Now, if I take log like the log base e on the both sides. So, I will be getting $\ln\frac{1}{2} = \ln e^{-\alpha\bar{d}}$ now this can be return as your $-\alpha\bar{d}$. So, $-\alpha\bar{d}\ln e$ and this can be written as ln 1 minus ln 2 and that is nothing, but - α . So, $\bar{d}$ now ln e, that is, log e base e that is equals to 1 and here, log 1 is equals to 0. So, I can find out $-\ln 2 = -\alpha\bar{d}$ and from here, I can find out that $\alpha = \ln 2 / \bar{d}$ and this is the expression which I have written.

So, I can find out the numerical value for this particular α and once I have got the numerical value for this particle α, very easily, I can find out the relationship between your similarity and the Euclidean distance values using this particular expression. So, this is the way actually we will have to calculate the similarity. Now, once I have got this particular similarity.

(Refer Slide Time: 17:03)



Now, here, we will have to find out what should be the entropy. Now, to determine the entropy, that is the index, what I do is, we use this particular expression in step 4 and remember one thing, this particular expression has been actually designed based on one philosophy. Now, I am just going to discuss the philosophy first and which is the reason behind defining this relationship between entropy and similarity. So, this indicates actually the relationship between this entropy and the similarity S. Now let me try to concentrate here, let us see, if I take some suitable value for the similarity, then what happens to the entropy.

Now, let me assume that S is equal to 0, now if I take the S is equal to 0; that means, your similarity equals to 0; that means, the distance between the two point is actually very high and the two points are too far and let us see what happens to entropy. So, by using this particular the expression, now if S is equal to 0, what will happen to the entropy? So, E is nothing, but minus so, I am just going to put S is equals to 0 and here if I put S is equal to 0, log 0 is not defined. So, it is undefined, but it is multiplied by 0 so, its contribution will be actually 0.

Then, comes your minus 1 minus S so, S equals to 0. So, I will be getting 1 here, then log base 2 S equals to 0. So, log base 2, 1 and log base 2, 1 is once again equal to 0. So, I will be getting 0. So, this is nothing, but 0. So, if I put S equals to 0. So, I am getting entropy is equal to 0, now let me put another extreme value for this similarity, let me put

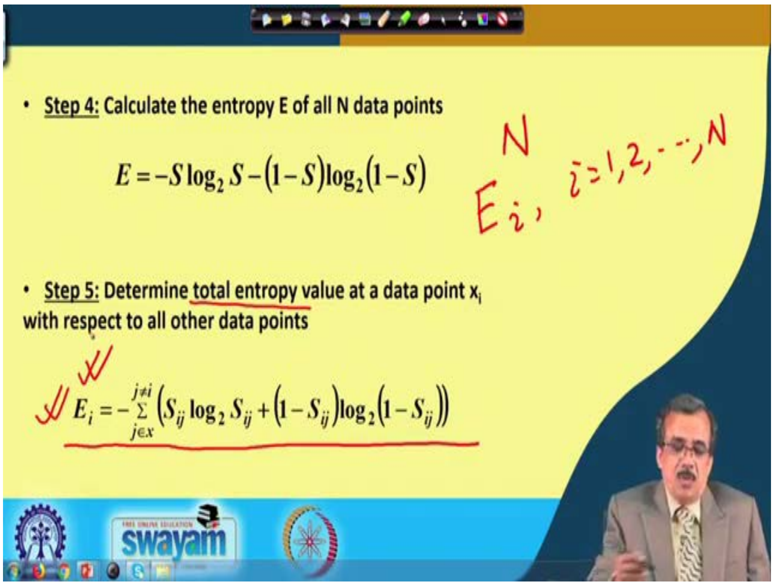
S is equal to 1; that means, your the similarity between the two data points is equal to 1; that means, the two data points are exactly similar and the distance between them is equals to 0.

Now, if I put S equals to 1 in this particular expression, I will be getting that entropy is nothing, but so, $-1\log_2 1$. So, that is equals to 0 then comes your S equals to 1 so, this will becomes 0. So, once again, I will be getting 0. So, far S equals to 1; that means, when the similarity between the two data point is equals to 1; that means, your the Euclidean distance is equal to 0, then also the entropy becomes equal to 0.

So, for the two extreme conditions, when the similarity equals to 0 and similarity equals to 1, the entropy becomes equal to 0. Now, let us try to find out, when S is put equal to say 0.5 or half let us see what happens, because S is in between now. So, we calculate actually the entropy. So, this is nothing, but $-1/2$ then comes your $\log_2 \frac{1}{2}$. So, S is equal to half so, this will becomes half then comes your log base 2. So, this is nothing, but half. So, this can be written as your $-\log_2 \frac{1}{2}$. So, this is nothing, but is your minus. So, $\log_2 1 - \log_2 2$.

Now, log 2 base 2 is actually your 1 and your $\log_2 1 = 0$. So, I will be getting minus 1 here and I have got here minus. So, this minus and minus so, this will become equal to plus 1. So, entropy becomes equal to 1. So, when similarity is 0.5. So, entropy becomes equal to your 1.0. So, based on this particular philosophy so, this relationship has been derived and once I have got this particular thing, now, you are in a position to find out the total entropy for each of the data points.

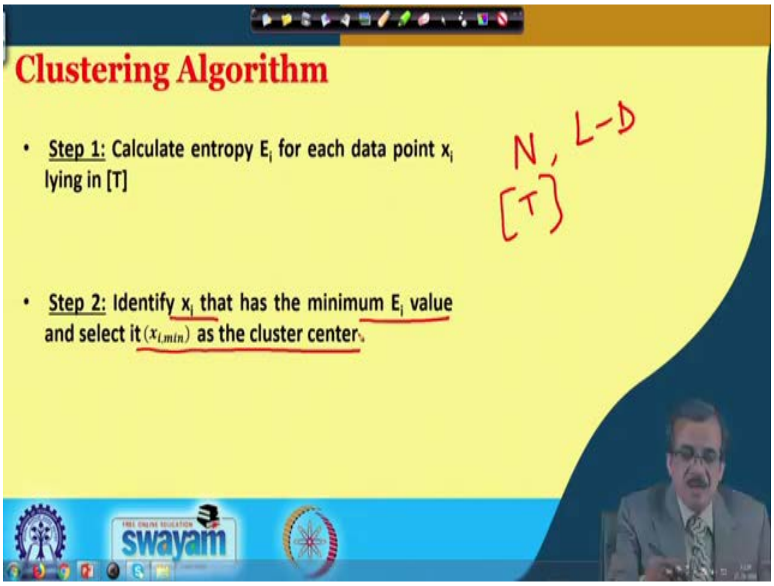
(Refer Slide Time: 22:05)



Now, you have got capital N number of data points, for each of the data points, I will be able to find out what is E_i, where i varies from say 1, 2 up to your N. So, for all the data points, I will be getting the entropy values, that are the index values and to calculate the total entropy actually, this is the expression which I will have to use.

So, $E_i = -\sum_{j \in x}^{j \neq i} (S_{ij} \log_2 S_{ij} + (1-S_{ij}) \log_2 (1-S_{ij}))$. So, this is the way actually, we can find out what should be the total entropy for each of the data points and once I have got this information of the total entropy for each of the data points.

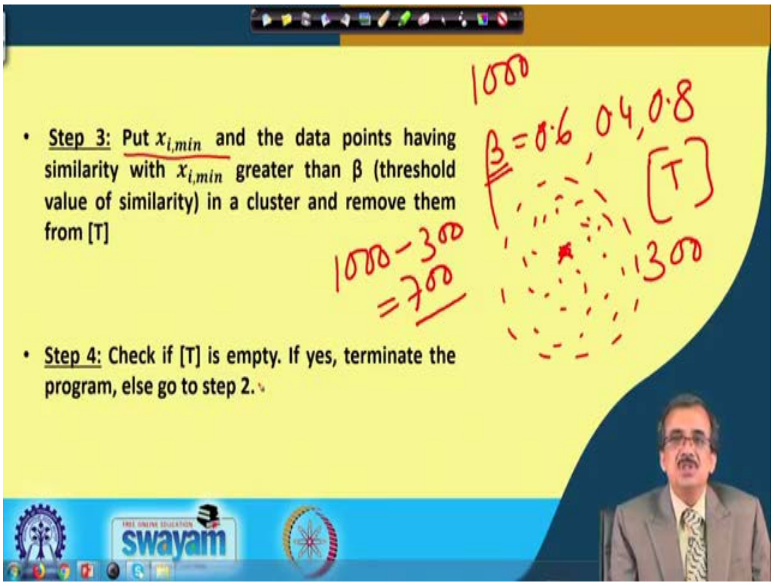
(Refer Slide Time: 23:11)



Now, I am in a position to give the statement of the different steps of this algorithm. So, we are going to discuss the steps to be followed in this clustering algorithm. Now, step 1; we calculate entropy E_i for each of the data point x_i lying in T hyperspace. So, as I told, we have got capital N number of data points in L dimensions So, all such data points, I am just going to represent by T hyperspace data points, ok.

Now, what you do is? So, we try to calculate the entropy the total entropy for each of the data points using the method which we have already discussed. And, once I have got the whole information of this entropy for all the data points, now you are in a position to identify a particular data point, which is having the minimum value of entropy. So, step 2; we identify x_i that is got the minimum entropy value and that is declared as your the cluster centre. So, we declare that particular point as the cluster centre, which is having the minimum value of entropy.

(Refer Slide Time: 24:39)



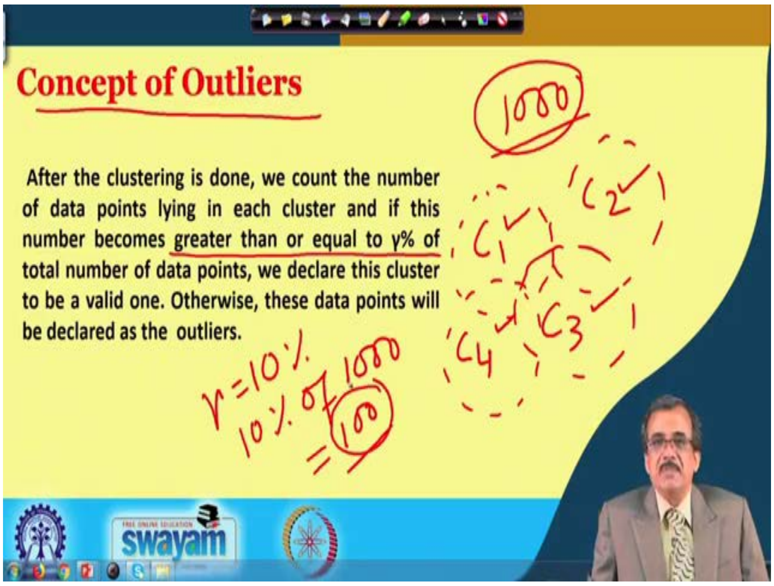
And, once I have got that particular cluster centre in step 3, actually what I do is, we put xi minimum and the data points having similarity with x_i minimum greater than your β and β is nothing, but the threshold value of similarity.

So, the user is going to define. what should be the threshold value of similarity for example, say this could be 0.6, it could be 0.4, it could be 0.8, and so on and what you do is, we have got this particular cluster centre and surrounding this, we are going to define one cluster. So, what you do is, the data points which are having similarity with this cluster centre greater than or equals to the threshold value, are allowed to enter to this particular cluster.

So, finally, I will be getting this particular cluster. So, let me repeat in step 3; we put x_i minimum and the data points having similarity with x_i minimum greater than β in a cluster and we remove them from your the T hyperspace supposing that initially I had 1000 data points and in the first cluster supposing that, 300 data points are entering. So, I have got 1000 minus this 300 that is nothing, but 700 data points remaining and with the help of these remaining 700 data points, I will try to form the second cluster, third cluster, and so on.

Now, step 4 checks, if T is empty if it is yes, you terminate the program; that means, all the data points have been put into some clusters, else we go to step 2; that means, we repeat from step 2 to step 4. So, this is the way actually we do the clustering using the entropy-based fuzzy clustering.

(Refer Slide Time: 26:48)



Now, here, I just want to tell you that this clustering algorithm is very flexible and if I just change the value of this β a little bit there will be actually too much change in the obtained cluster and that is why, this clustering algorithm is very flexible.

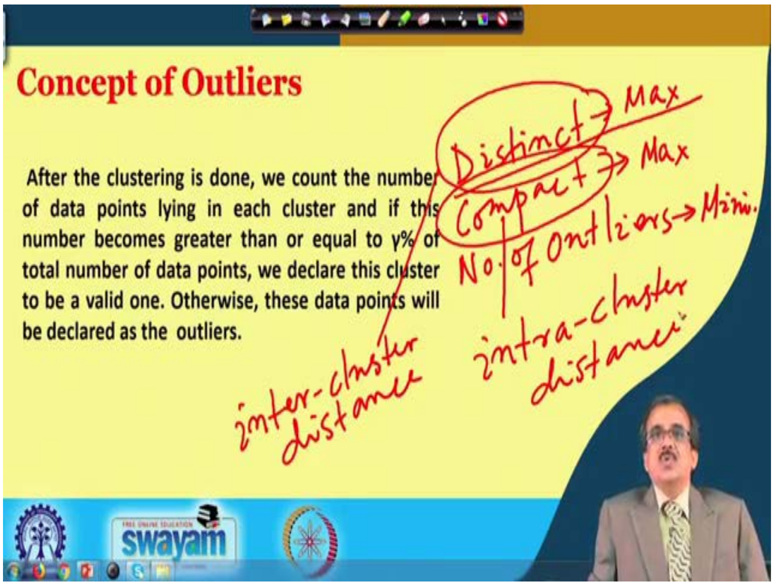
Now, here, I am just going to discuss one concept, that is your the concept of outlier. Now, actually if you see so, supposing that say I have a got 1000 data points and this 1000 data points have been divided into a few clusters, for example, say this is one cluster, this is another cluster, this is another cluster, this is a fuzzy cluster. So, there could be overlapping also.

So, there could be another cluster say C_4, now what we do is, we try to count the number of data points present in each of the clusters. So, we try to find out, how many data points are there in the first cluster, that is C_1, the second cluster whose centre is C_2, that number of data points in third cluster, number of data points in fourth cluster. And, supposing that these 1000 data points have been actually clustered into four clusters and we know how many data points are present in each of the clusters. So, after that actually, what we do is, we try to define whether all the clusters are valid clusters or there could be a few outliers; outliers means actually those data points, which do not belong to any of the clusters.

So, what I do is, we try to count the number of data points present in each of the clusters and if that particular data point is found to be greater than equals to γ percent of the total data points, then we declare that this particular cluster is a valid cluster. Now, if I take say γ equals to 10 percent that is one-tenth; that means, 10 percent, that is a 10% of your 1000, that is nothing, but your 100 points. That means, to declare a particular cluster a valid one, there must be at least 100 points in that particular cluster, supposing that there are only 30 points in a particular cluster, so-called cluster. So, we define those 30 points are nothing, but the outliers and we, in fact, do not consider that particular cluster having only 30 points as a valid cluster.

Now, this is the way actually, we do the clustering and as I have already mentioned that while doing this particular clustering. So, we will have to be very careful, so that this particular cluster becomes a distinct one.

(Refer Slide Time: 30:00)



So, we try to define the distinct cluster. So, the cluster has to be distinct, this has been already discussed that the cluster has to be compact and at the same time there should not be any outliers. So, the number of outliers number of outlier should be as minimum as possible. So, our aim is to minimize the number of the outliers and in ideal condition the number of outlier should be equal to 0 and our aim is to maximize the distinctness and our maximize the compactness also. Now, these I have already discussed, to measure the distinctness of the clusters, we consider inter-cluster distances, this I have already discussed in the last lecture.

So, to measure the distinctness actually what I do is, we consider actually inter-cluster distance. To measure the compactness, what I do is, we consider the intra-cluster Euclidean distance values. So, this is the way actually, we try to find out what should be the distinctness, what should be the compactness and we can also measure the number of outliers. And, let me repeat, our aim is to reach the clustering, which ensure the maximum distinctness, maximum compactness and a minimum number of outliers. Now, this is the way actually we use this particular entropy-based fuzzy clustering.

Thank you.