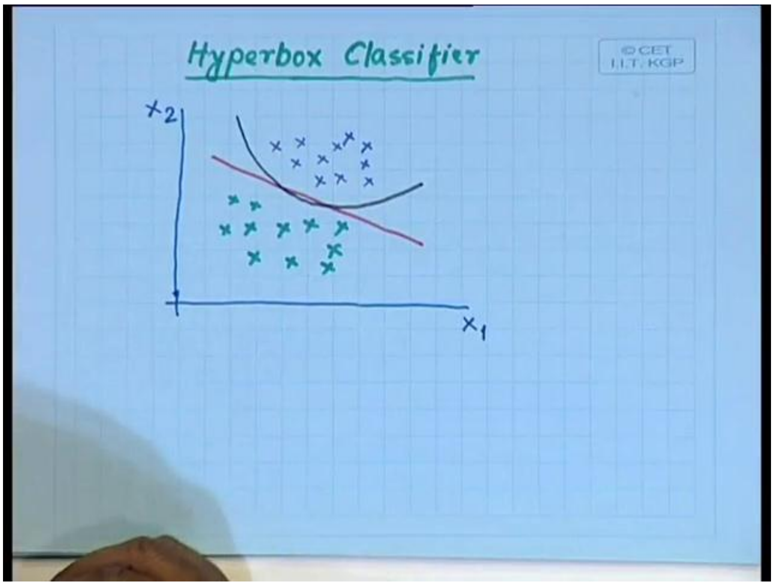
Pattern Recognition and Application
Prof. P. K. Biswas
Department of Electronics and Electrical Communication Engineering
Indian institute of Technology, Kharagpur

Lecture - 30
Hyperbox Classifier

Good morning. So, today what we are going to discuss about is the hyper box classifier. Earlier, we had just introduced the problem in the sense that so far the kind of classifiers we have talked about it separates between 2 classes of the regions belonging to 2 classes, which are simply connected.

(Refer Slide Time: 00:41)



In the sense that I have a set of feature vectors belonging to 2 classes, which are given in this form. Again I am taking this example in 2 dimension having 2 feature components x 1 and x 2. If the feature vectors for the training vectors belonging to 2 classes say omega 1 and omega 2, they form clusters like this, so this is 1 cluster. This is another cluster. 1 of the clusters is taken from say plus omega 1. The other cluster is formed by the points taken from another cluster say omega 2.

Then the regions belonging to these 2 different classes; omega 1 and omega 2, they are to be simply connected. I should be able to have a simple boundary, either a linear boundary or a non-linear boundary separating these 2 classes.

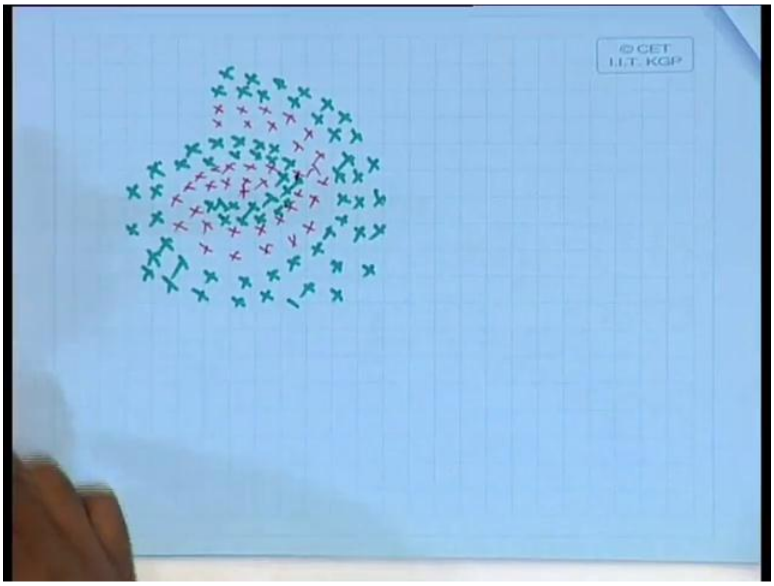
So, either I can have a boundary or linear boundary, something like this in which case the classes will be called linearly separable. I should be able to have a non-linear boundary something like this in which case the classes are linearly non separable. So, I should be able to have either a linear boundary or a non-linear boundary. So, accordingly we have talked about different types of classifiers starting from our base classifier, whether it is base minimum base classifier or base minimum error classifier and so on. We have seen that in a particular case, when all the classes have the same coherence matrix or the points taken from all the classes, they have the same coherence matrix. At the same time, the feature components are statistically independent or different feature components are statistically independent.

Then, we can have a linear boundary separating the 2 classes, omega 1 and omega 2. In general, if different classes have arbitrary coherence matrices or the points belonging to different classes of arbitrary coherence matrices in that case, the classes are not linearly separable. So, you can have a non linear boundary. Mostly, what we have is a quantity founded. Then we have discussed about other approaches for designing of linear classifier like approach or relaxation criteria, list approach, mini minimum square error based criteria based design approach.

In all these cases, we have talked about how we can design a linear classifier that separates 2 different regions by a linear boundary. Lastly, what we have talked about is the neural network approach. If I have a single layer perceptron or a single layer neural network having multiple number of outputs, then what we have is we separate the classes or the classes are separated with linear boundaries.

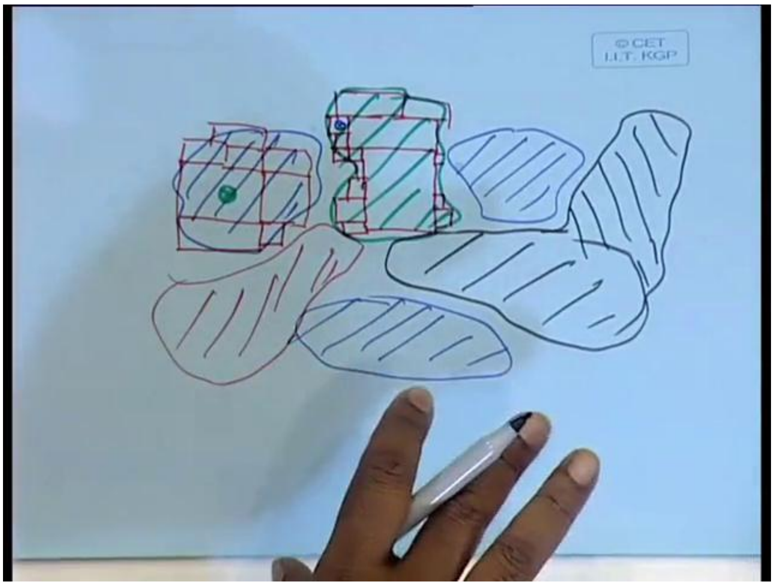
If the classes are not separable by linear boundaries in that case, we have to go for multilayer neural network or multilayer perceptron. The non-linear boundaries are actually replaced by piecewise linear boundaries. So, this multilayer neural network what it does? It approximates a linear boundary by piece wise linear boundaries. So, in all these cases, the regions belonging to different classes, they have to be simply connected. As we said earlier, we have a point of distribution belonging to different classes something like this.

(Refer Slide Time: 04:48)



So, these are the distribution points belonging to 2 different classes where you can assume that all the red points belong to 1 class, whereas the green points belong to another class. Not only that, I can have some arbitrary distribution something like this as well.

(Refer Slide Time: 06:34)



I have a set of points somewhere over here, which belong to 1 class. I can have set of points somewhere over here that belongs to the same class. I can have another set of points somewhere over here, which also belong to the same class. In between, I can have, so this is the set of points which belong to 1 class. This may be the set of points that belong to some other class. This may be the set of points, which belong to some other class. This set of points may even be extended like this.

So, if the pointer distribution is a complicated point distribution something like this, you find that neither linear boundary, obviously a linear boundary cannot separate among these classes. Even a non linear boundary is also unable to separate between different classes. So, how we can design a classifier that can even classify the point distributions or the class distributions, which are given like this? So, I just pointed out during our introductory lectures that we can move for that. That is what I do is you divide this space into a number of hyper boxes something like this. Different hyper boxes may be having different dimensions.

So, when you are here, so in this 2 dimensional example, I can assume that those hyper boxes are represented by rectangles. So, I have a set of rectangles, which represent 1 particular class. I have another set of rectangles, which belong to another particular class. If somehow we can assign the class level to this different rectangles, so I can say that this are the rectangles, which represent class say omega 1. These are the rectangles, which represent class omega 2. Similarly, I can have a set of rectangles over here that will represent class omega 3. I will have a set of rectangles covering this region, which will belong to class omega 4 and so on.

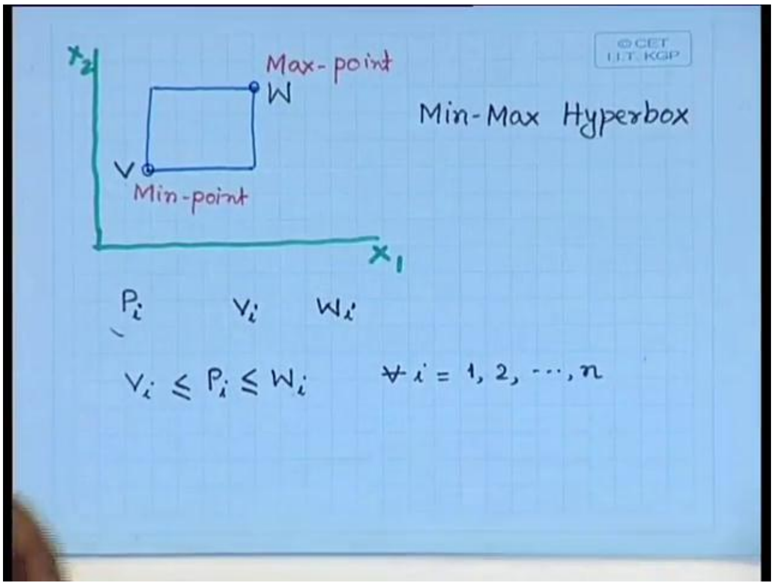
So, during the training process during the training of the classifier, we have to identify these set of rectangles or the set of hyper boxes. So, next time during the testing or when you actually go for classification of unknown feature vectors, I simply see the position of the unknown feature vector with respect to these hyper boxes, which are already generated. So, if the unknown feature vector falls somewhere over here, so this is the location of the unknown feature vector. Then I can immediately say that because this unknown feature vector is lying within a rectangle, which is leveled as class omega 1.

So, this unknown feature vector also belongs to class omega 1 or else the unknown feature vector falls somewhere over here. Then immediately I can because of this unknown feature vector falls within a rectangle, which is leveled as class omega 2, so this unknown feature vector also belongs to class omega 2. So, that way, I can define or I can generate a set of hyper boxes based on the training samples, which I have given.

Those hyper boxes are given the class levels based on from which training classes belonging to which class that particular hyper box has been generated.

So, that becomes the class level at a particular hyper box. Once I get the class of the hyper boxes than while testing or while in actual operation, I just find out that in which of these hyper boxes that unknown feature vector is situated. So, this unknown feature vector gets the same class level as that of the hyper box. So, these hyper boxes are to be generated during the training operation of the classifier. Now, suppose I have a situation in something like this.

(Refer Slide Time: 11:47)



I have 1 hyper box somewhere over here. Now, what is the advantage of these hyper boxes or the advantage of the rectangles? Once I divide the space, the feature space into a number of hyper boxes or into a number of rectangles, you find that such rectangles or such hyper boxes can be defined using just 2 points. If I know what is the left bottom point and if I know what is the right top point that is sufficient to define the hyper box. So, 1 of the points, this point left, because if I plot it against our conventional 2 dimensional co ordinate system, so this is my X 1 feature. This is the X 2 feature.

You find that corresponding to this left bottom point, both of the features X 1 and X 2 are minimum. Similarly, corresponding to this top right point, both X 1 and X 2 feature values are maximum. So, accordingly this point is called a min point and this point is called a max point. So, once I define this rectangle or the hyper box using this min point and max point, such hyper boxes are called min max hyper boxes. This is done to see whether an unknown point is within this min max hyper box or not suppose this min point is represented by vector V, which will have X 1 co ordinate and X 2 co ordinate, which are minimum among all the points.

It is also represented by max point, which is W, which also has X 1 co ordinate and X 2 co ordinate. These are maximum among all the points within this hyper box. Now, so singularly, if I have n dimensional feature space or the feature vectors are n dimensional feature vectors, then all the n co ordinates for this min point will be minimum among all the points. Similarly, all the co ordinates for the max point, all the n co ordinates for the max point will be maximum among all the feature vectors, which are given. So, if I say that a point P, I want to check point P belongs to this hyper box or not.

So, let us consider the ith dimension say, I have this unknown point P. I want to check whether this point P belongs to this hyper box or it does not belong to this hyper box. Suppose let me take that Pi to represent that ith dimension of this feature vector. It is ith dimension, 1 out of the 10. Similarly, for the min point say, Vi represents the ith min point that means the minimum value of the ith co ordinate. Similarly, Wi represents the maximum value of the ith co ordinate. So, this is ith dimension ith component of the maximum.

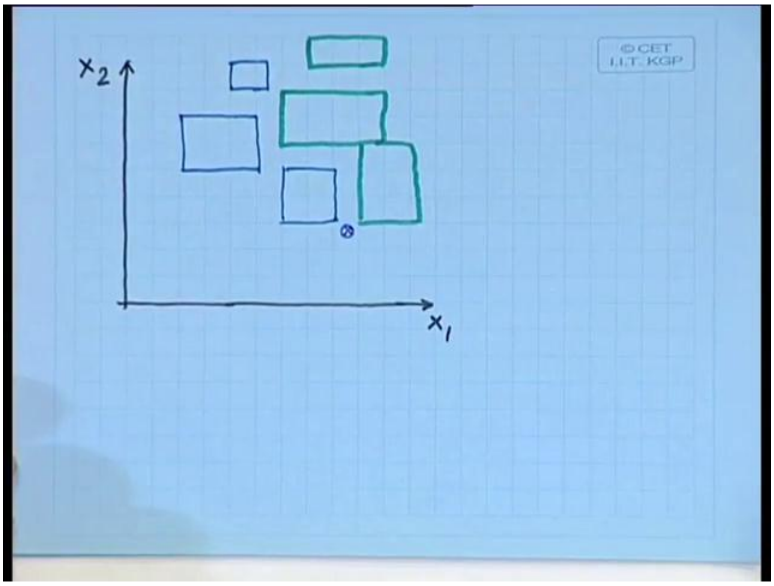
So, if this unknown feature vector P has to lie within this min max hyper box, then obviously I have to have the condition that Vi has to be less than or equal to Pi, which has to be less than or equal to Wi . This has to be true for all i, wherein from 1, 2 up to n. So, for all the components, all values of i Pi has to lie in between Vi and Wi. So, if it is so, then we can say that this unknown feature vector P lies within this hyper box, min max hyper box given by the min point V and the max point W.

So, this definition itself tells you that how to generate a min max hyper box that is given a set of points. I want to find out a min max hyper box, which encloses all the points within that set. So, for doing that, what I have to do is I have to find out every component the minimum value of every component among all the feature vectors, which are there in the set. So, if I have 50 points within the set for each of these feature vectors, I have to find out the first component. So, I will have 50 first components. Minimum of that will give me the first co ordinate of the min point.

Similarly, coming to the second dimension, I will get 50 components coming from 50 different feature vectors. I have to find out minimum of that. That gives me the second co ordinate of the maximum min point. Similarly, the maximum values will give me the corresponding co ordinates of the max point. So, given a set of points, for every component, the minimum values will give me the min point. The maximum values will give me the max point. You find that that follows from this particular condition that P lays within the min max hyper box.

Then, every component of P has to be within the minimum and maximum values of the corresponding component. So, that is within min point and max point, min component and max component of the corresponding hyper box. So, by checking this condition, I can easily find out with a point or an unknown point belongs to a given hyper box or not.

(Refer Slide Time: 18:54)



I come to a situation that I suppose to have a situation something like this. This is my X 1 dimension and this is my X 2 dimension. The hyper boxes are generated from the 10 examples. As the hyper boxes are generated from the 10 examples, it is not necessary that set of hyper boxes that you generate that will cover the entire space. It is quite possible that a part of the future space will not be covered by any of the hyper boxes because it depends on the distribution of the 10 examples.

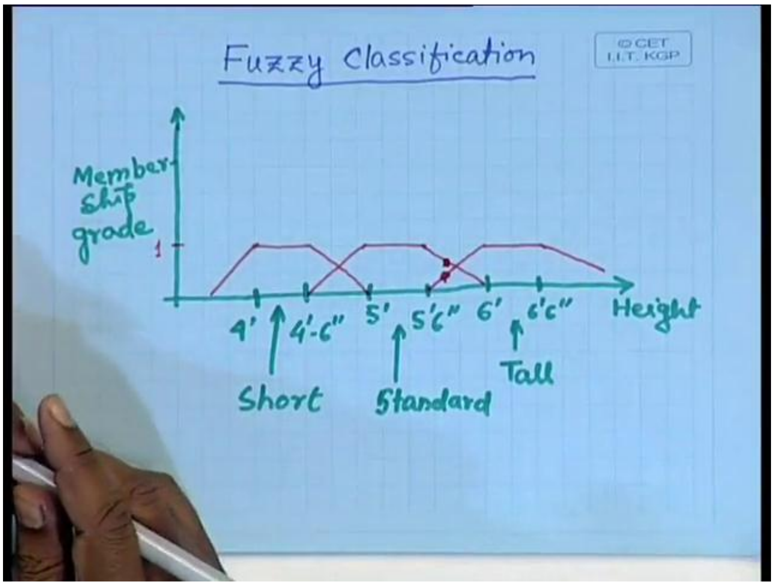
So, suppose I have a situation something like this that I have a hyper box somewhere over here, which belongs to class omega 1. I have another hyper box somewhere over here, which also belongs to class omega 1. I can have another hyper box somewhere over here, which also belongs to class omega 1. I can have a hyper box over here, which may belong to class omega 2. I may have a hyper box over here, which also belongs to class omega 2. I may have a hyper box over here, which also belongs to class omega 2.

So, you find that given these hyper boxes, there is a lot of space in between the hyper boxes, which is not covered by any of the hyper box. So, if there is a feature vector unknown feature vector because these hyper boxes are created using the 10 examples. So, it is obvious that all the 10 examples, which have given, they are covered at least by 1 of the hyper box. Given is an unknown sample, if an unknown sample falls somewhere over here, so this is the position of the hyper unknown sample. This unknown sample does not belong to any of the hyper box. So, should we say that this unknown sample does not belong to any of the classes?

So, it is quite logical that if I can have some measure by which I can classify this unknown sample as well, so 1 of the approach is what we discussed about the minimum distance classifier. That is you try to find out what is the distance of this unknown sample from defined hyper boxes, so by calculating the distance, you find out that which hyper box is nearest to this unknown sample.

I can classify this unknown sample to the nearest hyper box or the class of the nearest hyper box. That is 1 of the way. We go for minimum distance classification sort of thing combined with this hyper box classifier, the other approach is which appears to be more logical. This is because that is what in is in line with how we think or how we classify without actual thing. So, that is what is called fuzzy membership or fuzzy clarification.

(Refer Slide Time: 22:34)



I will not go into the details of fuzzy theory, but I will just introduce what is the concept of fuzziness. How this fuzziness can be combined along with our hyper box concept? So, we can have a definition of fuzzy hyper box or fuzzy hyper box set. So, the concept of fuzziness is something like this. Let us take a simple example that given a set of persons, live example persons, we want to say that who are the persons who are tall, who are the persons who are short and who are the persons who are say medium height.

Now, while doing this, we never take a tape. Find out what is the height of a person. Accordingly, we classify whether the person is tall or the person is short or we also say that the person is very tall. We say that the person is tall, but not so tall. Similarly, he is short, but not so short. So, when we decide something or when we classify something, we do not really classify in the hard manner. It is not that. If the height of the person is above 6 foot, 5, 6 inches, he is tall. We never do that. What we do is some sort of approximation or given set of classes. So, we want to find out what is the likelihood that a person belongs to one of the classes.

So, that is how we say that a person is very tall, the person is tall person is not so tall, he is very short, not so short or short things like that. That means when we are doing some sort of classifications in our classification, we are incorporating some sort of fuzziness. So, if I out something like this that from the horizontal axis, I put the height of the person. On the vertical axis, I will put something like membership grade.

So, our classification will be or the classification rule will be something like this. See if the person is his height is between say 6 feet and say 6 feet 6 inches, I will say that the person is tall. So, this is the region. When I say that the person is tall, roughly if it is between 4 feet to say 4 feet 6 inches, then the person is short and somewhere in between 5 feet to say 5 feet 6 inches, we say that he is medium or say standard.

This is because in India, most of the persons are within this category 5 feet to 5 foot 6 inches. So, I can say this as a standard or medium. So, the membership grade can be put something like this that if the person is tall, I can say that these are 3 different sets. 1 set is tall, tall, standard. The other set is short. So, I just want to classify like this that whether the person belongs to this set tall or the person belongs to this set short or the person belongs to this set standard.

So, if a person has a height, which is within 6 feet to 6 feet 6 inches, then obviously the person is tall. So, the membership to this particular set tall has to be equal to 1. If the height is in between 4 feet and 4 feet 6 inches, the membership to this state has to be 1 that is maximum. If it is within this membership to this set has to be 1 that is maximum. You find I have a lot of space between 4 feet 6 inches to 5 feet. Similarly, it is 5 feet 6 inches to 6 feet. Similarly, it is beyond 6 feet 6 inches below 4 feet.

So, these are the different height regions, which are not marked as belonging to any of these sets. So, I will have a maximum value equal to 1, membership equals to 1. When the height is between 4 feet to 4 feet 6 inches, the membership given to this particular set is 1. Between 5 feet to 5 feet 6 inches, the membership given to this set is 1. Between 6 feet and 6 feet 6 inches, the membership given to this set is 1 that is maximum. The remaining region can be done like this. If the height is less than 6 feet, depending upon the difference from 6 feet, I can reduce the membership grade. So, I can have a membership grade something like this, so over here in this region.

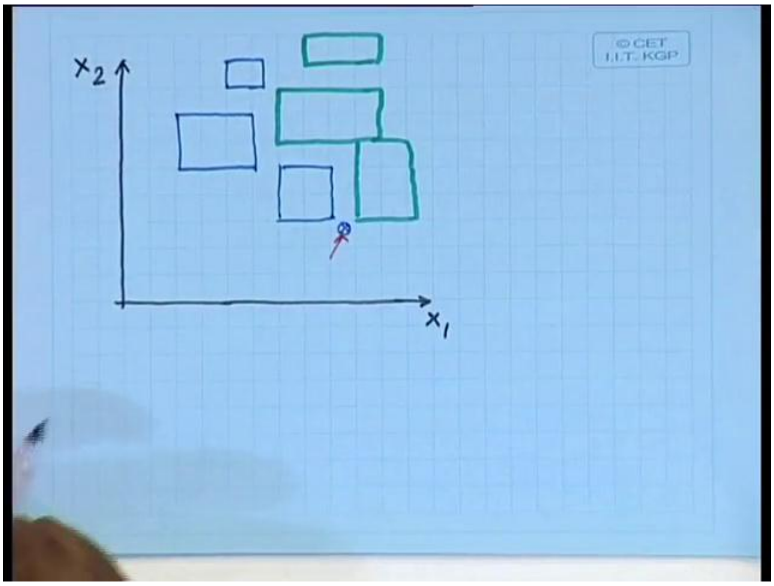
Let us consider 5 feet 6 inches to 6 feet. If the height is around 5 feet 9 inches, so I will be somewhere say somewhere over here at this position. So, in this location, you find that the person gets 2 membership values in 2 of the classes. So, the membership value to class tall is this to say tall is this much membership value, say short is this much. So, this itself consider as further classification because it tells me that what is the likelihood that the person is tall and what is the likelihood that the person is short.

This measure itself or that way I generate a vector actually. So, membership to class short falls to 0 at this point. At this point, membership to class short will be 0, membership to class standard will be this much, the membership to class tall will be this much. So, actually I get a 3 dimensional vector, 3 dimensional vectors of membership values to different classes. So, that is what fuzzy membership is. This fuzzy membership itself can be taken as the classification, which is called fuzzy classification.

If I want to make a hard decision that is decide about whether the person is stall or the person is short or the person is standard, then what I have to do is that I have to put a max operator. I have to find out what is the maximum of membership values. The membership value to which ever class is maximum, I classify the person. So, in fuzzy logic, taking this maximum operator is nothing but fuzzy union.

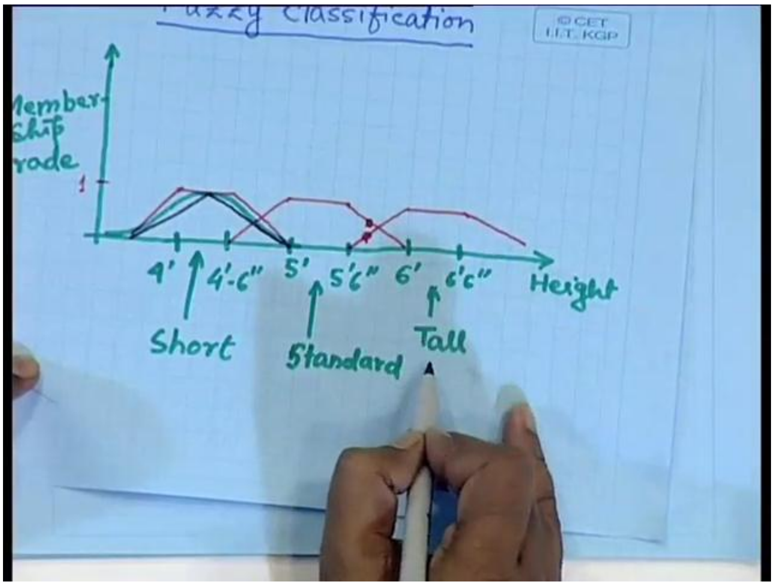
So, fuzzy union gives you the maximum of the fuzzy membership values. On the other hand, if I take the fuzzy intersection, it gives the minimum of the fuzzy membership. So, I said I will not go into the details of this fuzzy logic or fuzzy algebra, but just for our application what is required, I am discussing that. Now, after giving introduction to this fuzzy membership values, let us see at how we can make use of this concept in our hyper box classification.

(Refer Slide Time: 31:57)



So, what I will do is give him the unknown feature vector. I will try to find out that what its membership value to different hyper boxes is. That membership value or the state of membership values themselves can be used as the classification output. I can find out the maximum of these membership values. I can classify this unknown object to the class of the hyper box to which its membership value is maximum. Now, when I talked about this sort of membership profile, this membership profile is not unique. I can have other types of membership profiles as well. This is what is simplest the membership profile can be something like this.

(Refer Slide Time: 32:52)

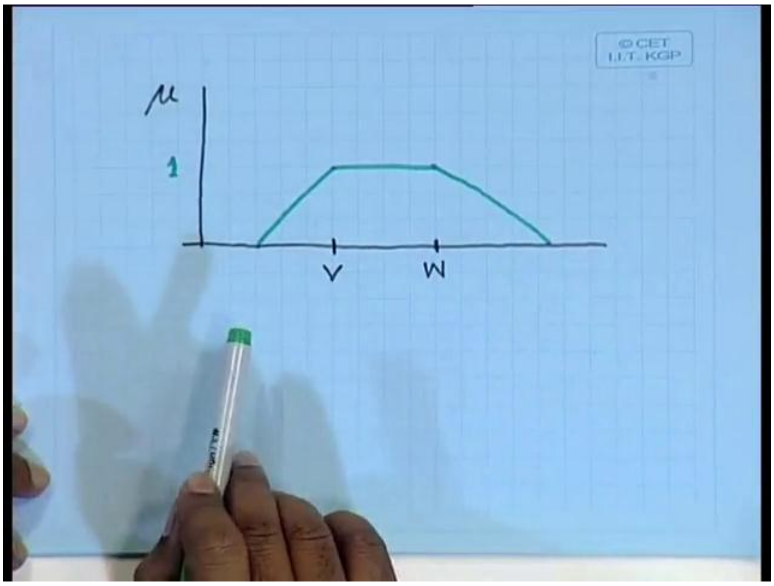


It can be a triangular membership something like this. There are different types of membership profiles that can be used. This is something like your Gaussian curve. What is mostly used is this sort of membership that is trapezoidal membership. You will find that this is nothing but a trapezium. What is most commonly used is the trapezoidal membership.

Now, the question is how to compute such membership functions. So, you can consider in our case, this point may represent our main point corresponding to this particular class tall. This point will represent max point. So, as we said, if the unknown feature vector falls within the hyper box that is within the min point and max point, then the membership value to that particular class will be highest. That is equal to 1.

As it falls beyond min point or beyond max point, not within the hyper box, membership value has to be reduced depending upon its distance from the hyper box and the distance. I can compute as distance from the min point. If it falls to the left of min point or I can compute the distance to be distance from the max point. If it falls to the right of the max point, so the kind of situation that I have is like this.

(Refer Slide Time: 34:43)

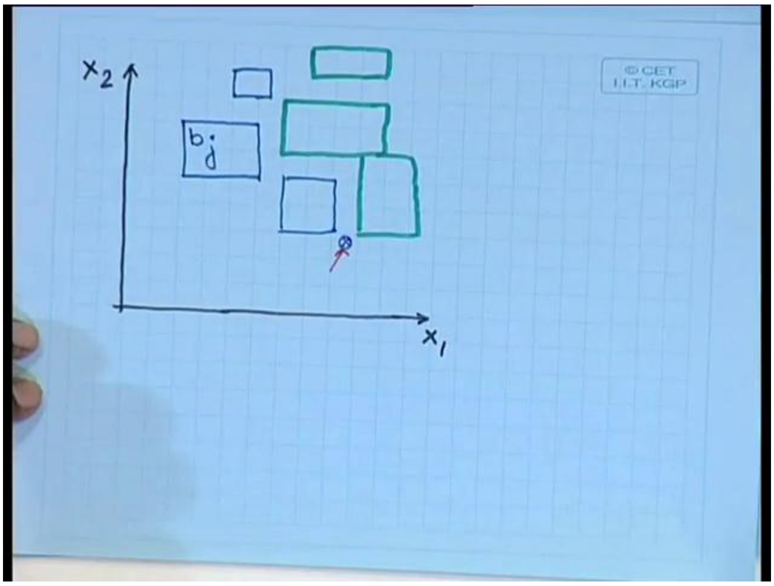


So, let us say that this is where I have the min point V. This is where I have the max point W. so, within the min point and max point, the membership value that will be computed will be maximum. That is equal to 1 and beyond min point and max point, the membership value has to be reduced something like this.

So, this is my on this axis. I have the membership value and the fuzzy membership value is usually represented by mu. That is what is conventionally used. I can use some other letter as well. That is not a problem, but if I use something else, I have to define what I am using. If I use this symbol, it is known to everybody that it represents fuzzy membership function. So, how do I compute this sort of membership function given the min point and the max point that is V and W?

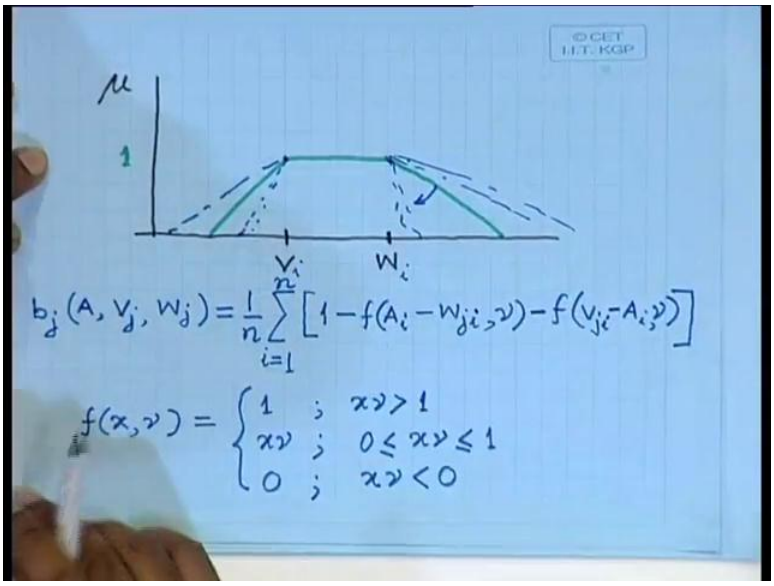
I am just writing in 2 dimensions. Actually, I have multi dimensional feature vector. So, both of these min point and max point, V and W will be multi dimensional vectors. So, if I define a function like this, so as I have multiple number of hyper boxes, so I have to identify which hyper boxes.

(Refer Slide Time: 36:36)



Suppose I consider a j hyper box. The membership function that will be computed by this j hyper box is represented by b j. So, I have to find get an expression of this b j and b j will depend upon 3 terms or 3 of the variable, 3 arguments; 1 is the min point of b j. This is because b j is a hyper box that is min point of b j, which we will represent as V j. The max point of the hyper box b j will be represent as W j. Our input feature vector, which is because depending upon the position of the input feature vector, I have to find out what is it is membership to the hyper box b j.

(Refer Slide Time: 37:26)



So, I can have a functional representation something like these say b j arguments are suppose my input feature vector A. Then V j, which is min point and W j, which is the max point are the 3 arguments on which this b j will compute the membership function. This membership function can be computed like this. If it is a n dimensional feature vector or the feature space that I have is n dimensional feature space, then I have to compute the membership function considering all the dimensions.

So, I define a function something like this 1 minus f A i that is ith dimension minus W j i that that is the ith component of max point W j. This max point W j is the max point of my j th hyper box. Then a term gamma, I will explain what this gamma is minus f V j i minus A i. The term gamma is going out of the screen light. So, like this, take the summation of this. I have n defined dimensions. This I am representing in ith dimension. So, take the summation of this for i is equal to 1 to n to consider the entire dimension and take the average. This is not complete because I have not defined what these functions f are.

So, I can define these functions f like this that f x gamma this will be equal to 1, if x into gamma is greater than 1. This will be x into gamma, if x into gamma lies between 0 and 1. This will be equal to 0, if x into gamma is negative or less than 0. This is how I define the function f. Now, see what it implies with respect to this particular diagram. If the point A, I am considering only the ith dimension that is true for all other dimensions. If I consider the vector A, which is the unknown feature vector here falls within the hyper box, I am considering the ith dimension. So, let me put this as V i, this as W i and A i will be somewhere in between.

So, if A i lies somewhere in between that means A i will be greater than V i. A i will be less than W i. Now, come to this term A i is less than W i. so, A i minus W j i will be less than 0. Similarly, A i is greater than V i. So, V j i minus A i that will also be less than 0. In this functional definition, x is nothing but A i minus W j i or V j i minus A i. So, both these terms being negative, my x into gamma will also be negative. Gamma is a positive quantity. So, this x into gamma that term will also be negative.

So, I am in this region in this definition because x into gamma is less than negative is less than 0. It is negative. So, f x gamma becomes equal to 0. So, both these functional values f A i minus W j i gamma and f V j i – A i gamma, both of them are 0. As both of them are 0, this summation or upper single term in this summation determines A value 1, which is found over i equal to 1 to n. So, I get value equal to n because this 1 is being added n number of times t, the average divide by n i get a value equal to 1.

So, that clearly shows if the point lies between min point and max point, then the membership value, which is computed will be equal to 1. Now, let us come to the other case. Suppose that the point lies to the other side to the left of here. If the point lies to the left of V i, then A i will be less than V i. At the same time, A i will be less than W i. So, if A i is less than V I, then this term is positive because this is V j i minus A i and A i is less than V i. So, this term is positive coming to this term. A i is also less than W i. So, A i minus W j i will be negative. Since, that this term is negative, so the value of this function is equal to 0 as per this definition.

The value of this function will be this or this depending upon what is the value of V j i minus A i times gamma. So, if V j i minus A i times gamma is greater than 1, then the value of this f x gamma is equal to 1. If this value is equal to 1, then what does this term give me? 1 minus 1 is equal to 0. If V j i minus A i times gamma is in this region, it is between 0 and 1. Then the output given by this term is 1 minus x times gamma. So, 1 minus x times gamma what is x? x is nothing but the distance from the mean point. So, that is my x.

So, it is 1 minus x times gamma that means I get a straight line with a slope equal to gamma. That is this straight line. Similarly, if the point falls to the right of W i, in that case that means A i is greater than W i. A i is also greater than W i. So, A i minus W j i term will be positive, whereas V j i minus A i, this term will be negative. As this term is negative, so this function functional value will be equal to 0. This functional value will be either this or this.

So, if the distance from W j i of the point times gamma is very large, it is greater than 1, then the output will be 0. This functional value the membership value will be equal to 0, whereas if the distance times gamma is in the range of 0 to 1, so the output is 1 minus this. Again, I get a straight line with a slope, which is equal to minus gamma. So, these 2 are different straight lines.

Now, see what this gamma does over here. Over here, gamma is the slope of the straight line. Over here, minus gamma is the slope of the straight line. That means the gamma;

this particular parameter gamma is indicating how steep this line is. If the gamma is high, then the straight line will be something like this. If the gamma is high, then the straight line will be something like this. The slope will be very high.

If the gamma is small, then the straight line will be something like this. That is what indicates the degree of fuzziness. When the straight line is bending in this direction, the slope is becoming more and more. That means we are moving more towards hard classification or crisp classification as it is called. If the value of gamma is less, then the crispness or hardness is reduced. At the same time, the fuzziness is increased. So, gamma is a parameter, which controls the degree of fuzziness. If the value of gamma is more, then the fuzziness is more. If the value of gamma is less, then the fuzziness is also less.

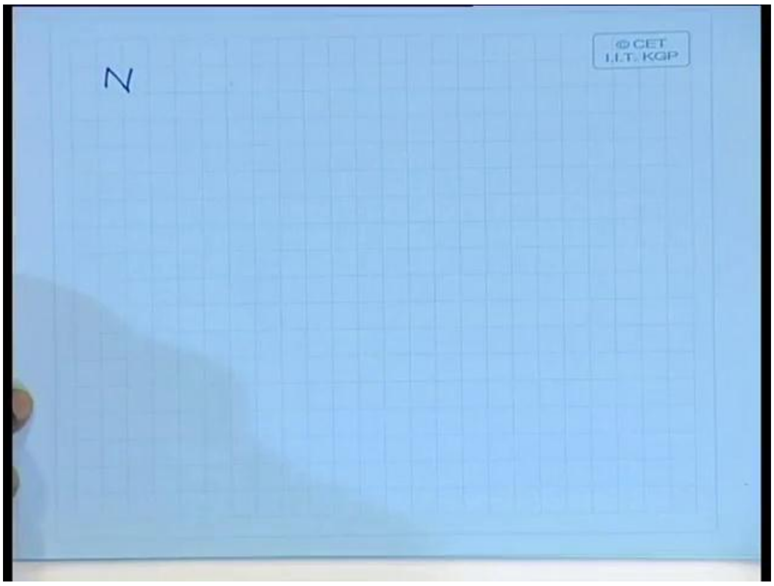
So, using this particular expression for different hyper boxes, I can compute the fuzzy membership function of this unknown feature vector to all the different hyper boxes. Now, to find out the fuzzy membership function of this unknown feature vector to a particular class, I will consider the membership functions computed by different hyper boxes belonging to that class.

So, here if I consider that this blue boxes represent plus omega 1 and the green boxes represent plus omega 2, I will compute the fuzzy membership function of this unknown point from each of these 2 boxes and take the maximum of these membership functions. That gives me what is the membership value of this unknown feature point to the plus omega 1.

Similarly, I compute the fuzzy membership function, fuzzy membership value of this unknown feature point from to each of these green boxes. Take the maximum of them that gives you what is the membership value of this point to this particular class. Once I get that, I get 2 membership values; 1 belonging to 1 for plus omega 1 other 1 plus omega 2.

So, these 2 membership values themselves as I said can be used as fuzzy classification. If I want hard or crisp classification, then I have to find out what is the maximum of these 2 and whichever is maximum, I have to classify this unknown feature vector to that corresponding class. Now, how does this algorithm will actually work? I have a set of training samples.

(Refer Slide Time: 49:31)



I have capital N number of training samples. Let us assume these training samples are coming from different samples. So, if I have 2 classes, some of these samples are taken from class omega 1 and some of the examples are taken from class omega 2. So, what I will do is I will consider this training samples one after another and while doing so, I will go on creating the hyper boxes.

So, when I consider the first training sample, I do not have any hyper box created till that time. So, what I will do is I will create the first hyper box. I have a single point. So, for the first, the first hyper box, I will put the min point and the max point to be the same. That means it is a point hyper box, which is same as the coordinates of that particular feature vector.

Then, I consider the second point in the second point because I am considering the points arbitrarily. It is not necessary that second point will come from the same class. So, if it comes from the same class, then I have to go for expansion of the hyper box. So, I have an initial point. Then I have a second point. So, I have to go for expansion of the second box to include this second point.

If it belongs to another class to the second class, then instead of expansion of the previous hyper box because previous hyper box belongs to some other class, instead of expanding the previous hyper box, I have to create a second hyper box. Again, it should be a point hyper box because this is the first point coming from the second class. So, I have to create a second hyper box, which is again a point hyper box for which the min point and max point is same. I have to go on doing that wherever possible. I have to expand the existing hyper box to include the new point which is considered; otherwise I have to create a new hyper box.

Now, while I expand the hyper box, there is a danger the danger in the sense if I come back to this figure see over here, we should not. If it is already within that hyper box that means it is already covered. If it is not covered, suppose in this particular case, I go on expanding this particular hyper box. If I expand this hyper particular hyper box, you find that it covers points from another class. So, when I expand an existing hyper box, then there are ways that the amount of error we will incorporate that will go increasing.

So, 1 extreme case is every point is considered as a point hyper box. The other extreme case is all the points irrespective of its class belongingness is put into the same hyper box. So, if every individual point is taken as a point hyper box, I do not have any error. In the other case, I have maximum error in between maximum and minimum. So, if I allow the hyper boxes to expand in un- constant fashion, then that risk of incorporating error is more.

So, I have to put some constant in the way the hyper boxes can be expanded. So, any time, I get a new point belonging to a class of an existing hyper box, I should not expand the hyper box, which is already existing in an un constant fashion in an uncontrolled fashion to include this new point. So, there should be some constant up to which I can expand a hyper box; beyond which I cannot expand a hyper box. So, I will stop here today. We will continue with this discussion in the next class.

Thank you.