Indian Institute of Technology Madras
Presents

NPTEL
NATIONAL PROGRAMME ON TECHNOLOGY ENHANCED LEARNING

Pattern Recognition

Module 02

Lecture 07

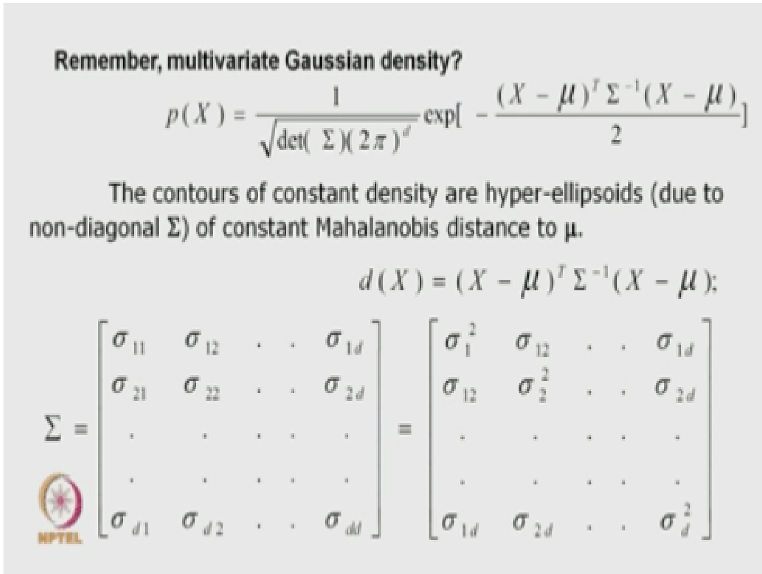
Normal Distribution and Decision Boundaries II

Prof. Sukhendu Das
Department of CSSE, IIT Madras

We are in the last class of the lecture series on pattern recognition, we have discussed about normal distributions which is a very important probability function used for probabilistic modeling in the filed of statistical signal processing, pattern recognitions image, video processing, any many other applications. Towards the end of that lecture we had also seen the scope of a distance based on the covariance matrix of the multivariate Gaussian distribution, we had seen that.

And now we will revisit that once more, look at that expression, and try to derive what we will call as decision boundaries to be used for the task of classification or pattern recognition specifically, let us look back into the slide.

(Refer Slide Time: 01:16)



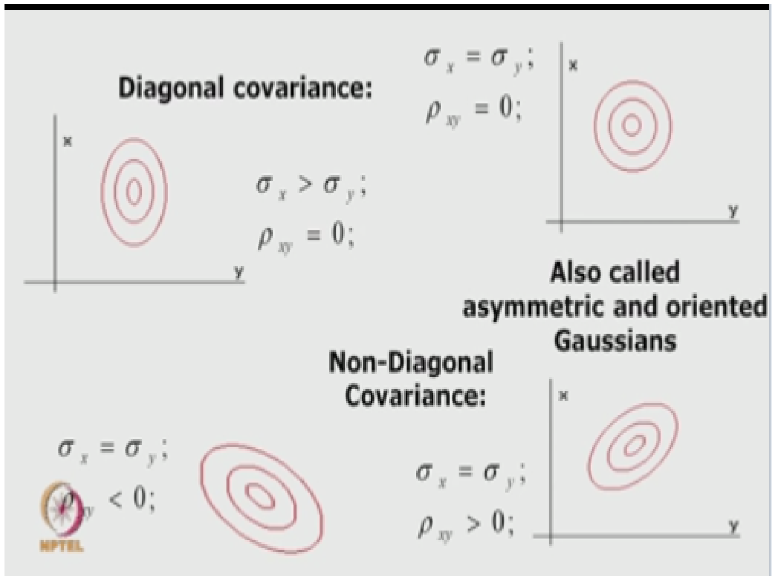
So this was the multivariate Gaussian density function. This is the normalizing term and the expression here within the exponent contains the mean of the data as well as the inverse of the covariance matrix here. These dictates the scatter of the data in several dimensions put together. Now we had seen at the end of the last class or what are called contours of constant density due to this non-diagonal elements of the covariance matrix which are of constant Mahalanobis distance we will define this very soon.

Distance to the mean of the data, and that is given by this function d(X). So this is where we almost stop, although we would have talked specifically about what are called hyper-ellipsoids, we will do that which is dictated by the non-diagonal terms of the covariance matrix. But let us know for the time being that the expression within the exponent term here is responsible for the distance.

And then we will see how this distance can be exploited for the task of pattern classification or pattern recognition. To wind up that slide, this is how the expression of the covariance matrix would look like. Here in the covariance matrix you have the diagonal terms which are the individual variances along the particular direction, and the half diagonal terms again, it is a symmetric matrix, they represent the corresponding covariance between two dimensions I and J.

So that means σij represents the covariance between dimensions I and J. So using this idea of the distance from the mean as given by the term called Mahalanobis distance to the μ or mean of the data. We will proceed and see now how this distance could be exploited to do certain task of pattern classification or pattern recognition.

(Refer Slide Time: 03:44)



Let us look at some examples of how the diagonal covariance term in two dimension, this is the slide which shows at a distribution in 2d(X) and Y very simply. And the covariance matrix has the diagonal terms which are equal, that is why you have σx=σy, and the half diagonal terms are 0. So the covariance is here strictly a diagonal, you have diagonal terms which reflect the corresponding variances and that too they are same, and the half diagonal terms is 0, it is strictly a diagonal matrix.

Go back and that particular case this is what you will have as the radar distribution, or you can see if the radar distribution is of nature like this, then you are bound to have equivalent to equal terms along the diagonal σx^2, σy^2 and so on, and the half diagonal terms are 0. This is also another example of a diagonal covariance term where the half diagonal term is 0 alright, but the variance or the standard deviation is given here along one particular direction is more than other one.

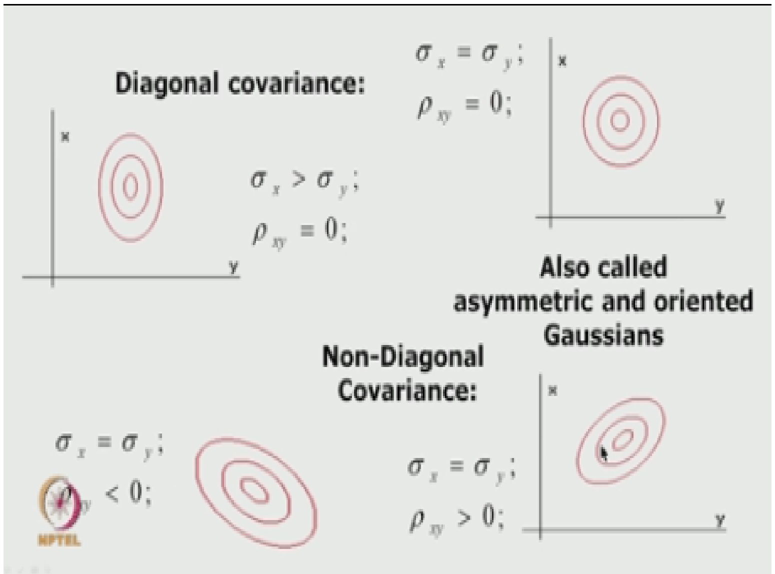
The variance along the X direction is more than that along Y, you can see that the scatter or spread along Y is much less than the scatter along X direction which is the vertical one and that is much, much more larger than the horizontal one. So these are two examples of diagonal covariance. What if non-diagonal terms are present? This is an example where $\sigma x=\sigma y$ that means the variances of standard deviations along X and Y are same.

However, you have a nonzero half diagonal term has given here, and that dictates that the scatter of the data will not now have a short of a inclination along a particular direction. It will have a trend which shows that as X increases the radius Y also increases, the reverse can happen. This is a case when the half diagonal term is negative that means when you can say when X is increasing, Y is decreasing or goes in the negative direction or vice versa.

So this is an effect of the half diagonal term, if the half diagonal term, the ρxy is given here is 0, then of course you will have situations has given, these are just examples to show. And the Gaussians which you will get, remember to what we are showing are basically iso-contours, iso-contour lines which are at a certain distance D given by that expression which we just saw in the last class or strictly the Mahalanobis distance from the mean.

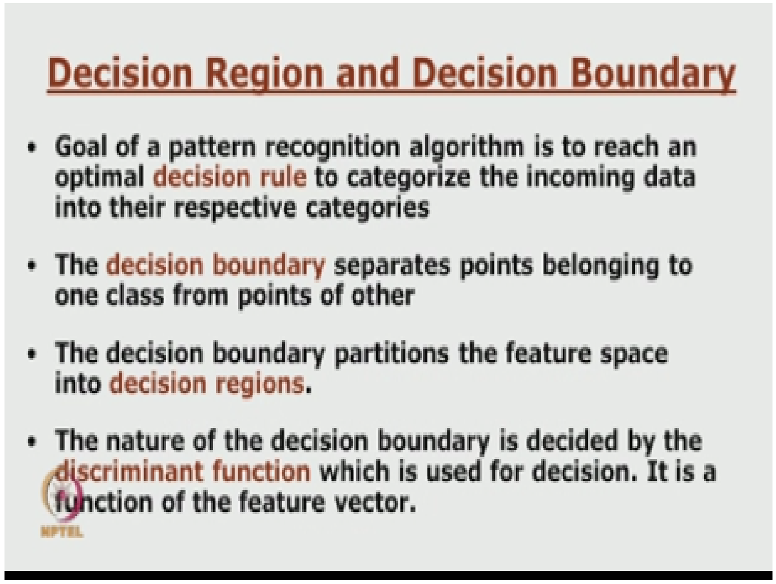
Those are the iso-contour lines which are shown there around the mean. However, the Gaussian function maybe symmetric, may not be symmetric. So we can also get asymmetric or oriented Gaussian functions and for those cases the iso-contour lines are as given in the figure here.

(Refer Slide Time: 07:07)



The bottom of your screen the 2 cases of diagonal covariance's also and the left hand side the diagonal covariance here is an example of an asymmetric or these 2 can be considered as oriented Gaussians these are just examples in 2D to illustrate the effect of the covariance matrix remember in general it is a we deal with large dimension data now before we go and use this corresponding expression of the distance and the probability of multivariate Gaussian density for the task of pattern classification or pattern recognition.

(Refer Slide Time: 07:54)



We will coin certain terms which are very important as shown in the slide now we are going to introduce 2 terms called the decision region and decision boundary we will actually call them and chart and DR and DV very soon remember the goal of a pattern recognition algorithm is to actually find an optimal decision rule to categories the incoming data into their respective categories.

This is a very generic statement I am making about pattern recognition remember one of the decision rules which you are expect is going to be Bayes rule based on Bayes theorem which we have already heard from Prof. Murthy so based on that Bayes decision rule when we categories samples into different classes or categories how does this distance based on the probability density function of a Gaussian or a multivariate Gaussian distribution play an important role.

And from that point of view what are regions and decision boundaries we will define those terms first and then we will see the mathematical expressions which will actually clarify you dough about these 2 techniques so let us look at back the decision boundary are the once which separate points belonging to one class from point belonging to the other so it basically separates 2 sets of points.

Let us say your discriminating 2 class of flowers white flower and a black flower very simple or certain fruits from vegetables and you have taken certain features to discriminate between these 2 categories of classes okay and we have seen remember this examples of what we have seen during the process of classification verses clustering that each instance or a sample is a point in a high dimension feature space.

So we can think of a boundary in-between which sits in-between 2 regions in the feature space discriminating between 2 class okay that region that boundary is called a decision boundary very simply so if that is decision boundary with separate points then of course decision regions or the once which are formed by portioning the feature space, so a decision boundary partitions the feature space into decision regions.

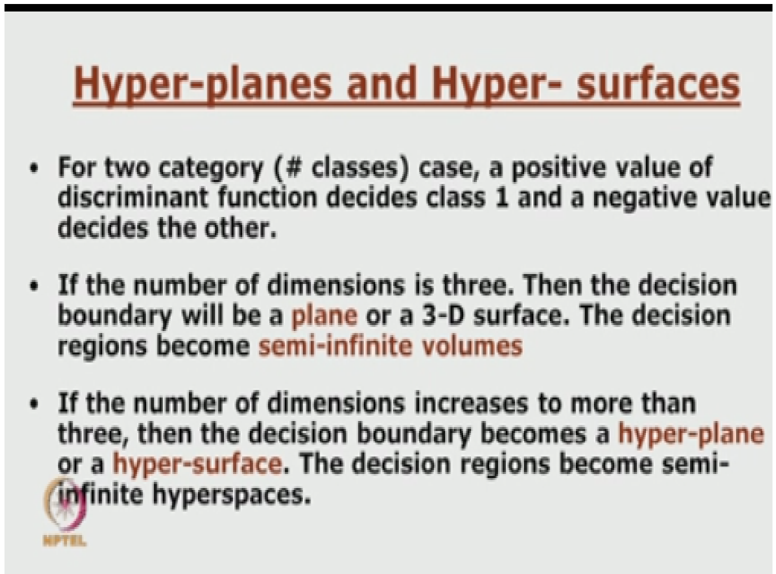
So think of two regions you can take a simple example of 2 different neighboring countries or 2 different sates from the field geography okay you have a or a line between 2 states 2 districts or 2 countries thinking that to be a decision boundary then what are the decisions regions the regions are those geographical maps belonging to 2 neighboring countries states or districts very simple example.

The nature of the decision boundary is decided by what is called a discriminant function this is where the multivariate probabilistic function will come we have to form a discriminant function which is to used for obtaining a decision what is a decision the decision of assigning a sample through a particular class based on a certain decision rule it could be the Bayes rule we can keep the Bayes rule in our mind of course that could be other methods of classification statistical UL methods.

Whatever it may be but for the time being let us say we have the simple Bayes rule we will revisit that for a movement okay and based on that Bayes decision criteria or decision rule will form a discriminant function is that discriminant function it decides the nature of the decision boundary okay so the discriminant function which is used for deciding what is the category of the incoming data that basically forms the decision boundary.

It is a function of the further vector which represent the sample of the instance so we have introduced 3 terms remember decision region is DR decision boundary is DV and discriminant function we may not call DF but remember we will hence for interchangeably use DR with decision region DB meaning decision boundary.

(Refer Slide Time: 12:45)



Decision boundary is create hyper planes or hyper surfaces that means if you take an example of 2 categories is a simple example we have already take let us say the number of class is 2 a positive value of the discriminant function decides that the sample belongs to class 1 and a negative value decides the other one in the last slide we say discriminant function is a function of the feature vector correct.

You think of any arbitrary function for the time being although we will form a lies and take some examples of discriminant function later on so what do you do you try to compute the value of the discriminant function based on a certain value of a feature vector which belongs to a particular test sample which you want to categories the discriminant function will give you value typically.

The decision of the discriminant function will tell you whether the sample belongs to class1 or class 2 class A or class B very simple we are talking about binary classification as it is called or the number of classes = 2 so your just 2 class it is like a white car versus a black car okay or apples versus oranges very simply okay that is want you want to discriminate class A class B class 1 class 2 okay so there is something like a decision boundary in between these two classes on one side you have decision region for class 1 you have another decision for class 2 with the db or decision within between we going to have a diagram of this very soon do not worry okay.

So what will be the discriminant of function do when it will evaluates samples for a particular regions a class 1 it will give positive values here for the other class it will give negative values here, so just look at the sign of the discriminant function you do not have to worry of the absolute value the sign of the discriminant function will tell you to which class it belongs, now you should be able to extrapolate and tell me what happens if the what happens to the value of discriminant function.

If this sample of the instants fall on the decision boundary if the sample or the instants falls on the decisions boundary the value of the discriminant function will be b0 fine very east good, so once w know this. If the number of dimension is 3 then the decision boundary will be a plane or 3D surface well in this case we are talking about linear decision boundary remember in one of the earlier slides we are taking about hyper ellipse voids and things like that some non linear write in the decision boundary typically if you take lines in a geographical amp which partitions two regions of states districts and countries where it really you will find the linear boundary where it really in the rare case you will get that.

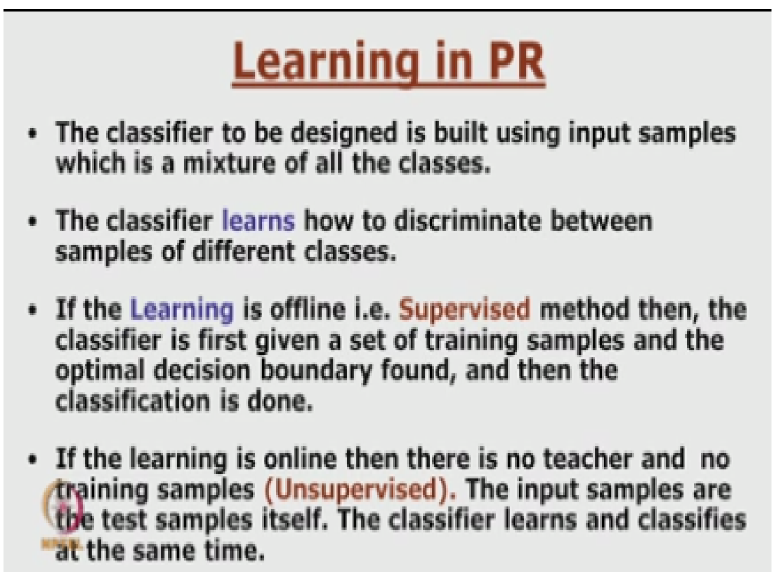
Mostly it is non linear irregular I should say okay so whether you will get a liner non linear dependents upon several factors in fact the correlates matrix will listen we will talk about that later on for the timing we will assume that this is the linear decision boundary and of the decagons boundary if the dimensionality of the problem is 2 then you are talking of a line in 3D discipline and of course in higher dimension we talk of this is hyper-planes and the decision regions then become what are called semi-infinite volumes this is a term which actually which is borrowed.

Sometimes in the field of geometry or computer graphics why it is a semi-infinite volume because one side you have the decision boundary the other side it is stretches to infinity goes to infinity okay so it is finite or non side infinite on the other, so it is a semi-infinite volume it just a terminology trying to give you an idea so if the number of dimensions increases to more than 3 then the decision boundary of the dv becomes a hyper-plane or a hyper-surface then the decision regions become semi-infinite hyper-surfaces.

Remember plane and surface plan is a special case of a surface is a plane or surface we call it as plane but a surface can be only non linear also so the linearity or non linearity is a separate issue linearity or non linearity is a separate issue we will talk about that later on because I already mention it dependents on the properties of the covariance matrix it is of the diagonal terms will be mainly responsible for this okay diagonal terms also have a role to play but of diagonal terms strictly will have a role here.

Okay and also the covariance matrix between classes so all these will dictate but if the decision boundary is linear then it is line in 2D will have a straight coming up a plane in 3D and hyper plane in higher dimension, if it is non linear it will be a curve a conic the quadratic curve let us say in 2D it will be an arbitrary conical surface in 3D and hyper surfaces in higher dimension, so just keep so these are something to do with decision boundaries.

(Refer Slide Time: 18:32)



So the terms which are given very closely associated with the process of pattern classification pattern recognition is the process of learning beaks when you are going to designer algorithm for pattern classification you have to learn something from the data samples are there are two methods of learning but what is this process of learning remember a classify this to be design is built using input samples which is a mixture of all the classes so the classifier is made to learn from different samples.

This is the process of learning this was talked about in the introduction part also in the lecture that means if you want to say discriminate between apples and oranges and let us you want to us color as a feature it is possible because apple will have a side reddish stings of the color the orange of course will have mix obvious color and if you use color as a feature in 3 dimension nd you want to learn this colors from samples it is possible to do with apples and oranges and what you want to do in this process of learning.

We will see here that the essential process of learning involves learning or designing the classifier and repeat again designing the classifier involves learning the discriminant function or parameters of the discriminant function which in turn dictates what are going to be your decision boundaries or decision regions decision regions and decision boundaries of the dual of each other once you found decision boundary you get regions or you want to establish the regions you get boundaries.

We will see that with the examples once you form to neighboring states you get a boundary or you draw boundary and create to regions okay and the decision boundary and or regions or dictated by the discriminant function, so the process of learning the discriminant function is the process of which you learn the parameters of the function from the input samples and it says that sometimes you need the samples from all the classes and categories to learn they could be given together or one after the other that is okay.

It does not matter but as long as you know that the sample belongs to a particular class you are fine so the classifier learns how to discriminate between samples of different classes if the learning is offline typically it is called supervised method well that is not the main criteria what I must say is if the classifier is first given a set of training samples and the optimal decision boundary or the DB is found and then the classification task is performed this method is called supervised learning.

Why it is called supervised learning well there is a teacher, the teacher tells you that these are the samples which belong to class 1 these are the other set of samples which belongs to class 2, please learn the set of features which discriminate one from the other find out what is discriminate function and the decision boundary so it is like a almost a teacher teaching you how to work out the process of classification, samples are given to you.

Like somebody the teacher in the kinder garden teaches you how to write the character A, how to write the character B that means samples are given many, many times in the book and then in board for students to write and practice that is the supervised method where there is a teacher explaining and telling you these are samples of particular class category. If you take ABCD then you are talking about words, characters to be very precise, okay.

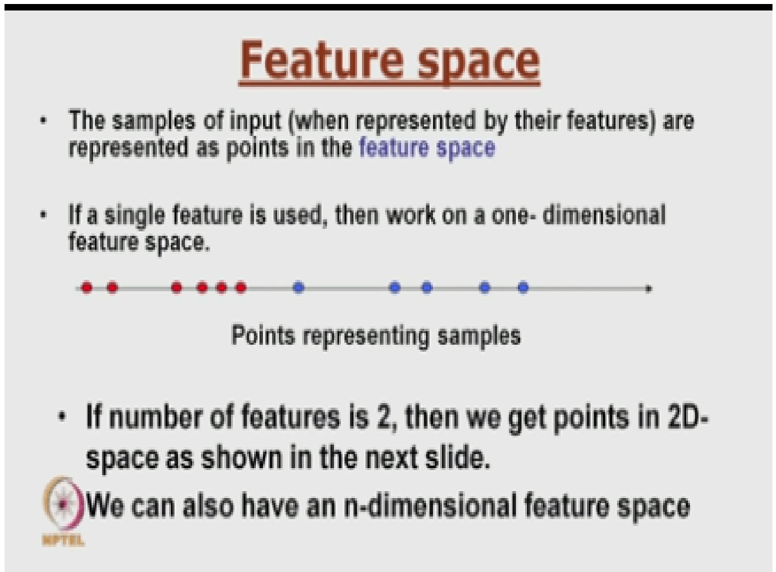
And that is the supervised method of learning so the classify learns from samples given a priory finds out the optimal decision boundary and then performs classification task based on a certain set of test samples using a set of test samples you perform the task of classification after you have learned you know supervised manner the optimal decision boundary.

What is the other method of learning, unsupervised there is no teacher, there are no training samples given the process is generally considered online and there is no teacher, there are no training samples. This unsupervised method of learning uses the input samples as the test samples themselves the classifier learns and classifiers at the same time.

So in this particular case of unsupervised method there is no teacher telling you this is how you should write character A, this is how you should write character B all the characters are given to you together to learn all the 26 A to Z, how could it, nobody tells you that this is A, this is B, this is C and so on you have a set of characters you should have to now discriminate yourself between A verse B verses C verses D and so on.

There are sufficient samples available for that but you are doing it learning and classifying simultaneously you do it together at the same time and that is the process of unsupervised classification no teacher.

(Refer Slide Time: 24:00)

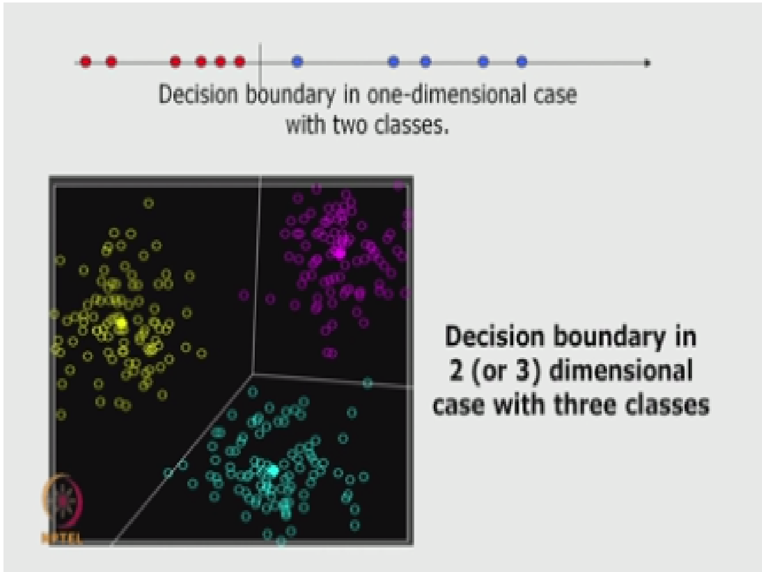


We talked about this feature space some time back so I am not going to introduce it in a big way again, but we all know that the samples of the input when represented by their features are represented as points in feature space. A single feature is one dimensional example, this is simple example but points represent sample and feature space. If I remember correctly we had used this slide earlier in a classification verses clustering example as well, okay.

Of course you could ask question right away where is the decision boundary here, where is the decision region will come to that in a moment. But of course the number of features could increase, and example 2 you get points in 2D space which we will see in the next slide you can have number of features as 3 then we go to 3 dimension.

And of course higher dimensions if the number of features are more than 3 or more in a typical n dimensional feature space is sometime called as the d dimensional feature space we can have decision boundaries, decision regions and discriminate functions which are form.

(Refer Slide Time: 25:03)



This is a typical example that is a small point which is actually shown by a line here, this is a point here which indicates a decision boundary in a one dimensional case of a two case problem you have a set of samples belonging to class 1 as marked as red points, class 2 as blue points anywhere in between you could have actually put the decision boundary the point in between okay, it depends up on the algorithm. Let us look at this example, each has been borrowed okay from a document, the decision boundary in 2 or 3 dimensional now this example shows three classes that means three classes, three set of instants marked in three different colors.

Magenta, yellow and blue I hope the color is clear I repeat again magenta is one color say class 1, yellow in class 2 and cyan if not blue cyan color is the class 3, okay. Now these points could be scattered in 2 dimension xy or you can visualize that you are sitting in a room this is one corner of the room and these are points in 3D space that is also possible, it is like a projection of a three dimensional world on to two dimensional space.

But the points may be actually in 3D okay, so in that case when you see this three lines I have attempted to draw three decision boundaries or dives separating one class to the other that means this decision boundary separates this sphere of classes this decision boundary separates these two this decision boundary separates these two, these could be lines in 2D, that could be lines in 2D or they could be planes in 3D.

Either way you visualize it is okay, that is why you look back into the slide I wrote that you can visualize this problem in either two or three dimension.

(Refer Slide Time: 26:56)

This is a hand drawn example, let us see another example of sample points in two dimensional feature space this is very easy to visualize I have just two sets of sample points belonging to class 1 as given here, could say I have used another mark for the different color to indicated as class 2 okay, so two dimensional space instant of x and y I have used f1, f2 and the two feature components of the feature which I have extracted from two sets of categories of sample.

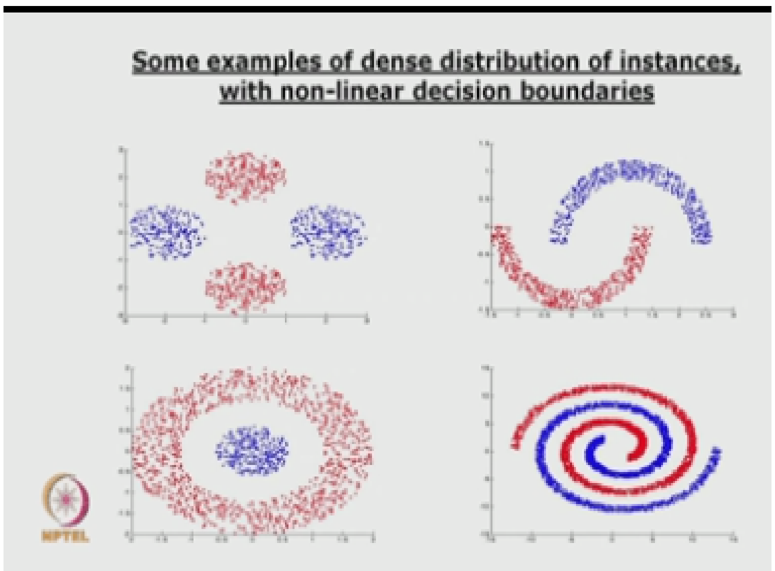
If you look back into the slide this is an easy problem to categories why, I can actually draw a line through the middle which I am not drawing here, but I am asking you to visualize that if I draw a line through the middle then that will be the decision boundary to discriminate between these two classes and I can actually use any short of simple classifier to find out this particular discriminate function to discriminate between these two classes and obtain this DB that is an easy problem.

But is it always the case that I can separate to set of points using what I call as a linear under session boundary which is a line 2d or a plane in 3d or a hipper plane a had dimension not always possible in this case yes but what if we look back in to slide I am introducing certain sample point belonging to class 3 third category of a class, and a samples have been marked by a different symbol with the different color as we are now look if you want to discriminate class 1 versus class 3 or class 2 versus class three you will not to be able to draw a lint anywhere in this space we discriminates and create 2d r and create 2dr or 2 decision regions discriminating one class versus the other.

Look back remember class one versus class2 is still it still possible to draw a dv between the two which is linear but for class one versus class 3 you need to probably draw non linear elliptical boundary here, the same to do with the class 2 versus class 3 it is possible that you may need a elliptical construed to discriminate between class 2 versus class 3, so that means sample belong to class 3 if you want to discriminate with respect to the other two classes you require a non linear decision boundary linearity linear decision boundaries will not solve the problems.

In such cases these examples from the class of problems which are call non linearly separable problems that means you cannot solve the problem of classification using a linear separable boundary or a linear decision boundary, not linearly separable or non linearly separable these terms are sometimes used in the tangibly by people okay. so class 2 versus class 3 look back again class 2 versus class three or class one versus class three are examples hand drawn of non linearly separable or to not linearly separable problems class one versus class two is a problem which is linearly separable. These are plans in two dimensional spaces I leave it for you to visualize the same in three d or higher dimensions.

(Refer Slide Time: 30:21)



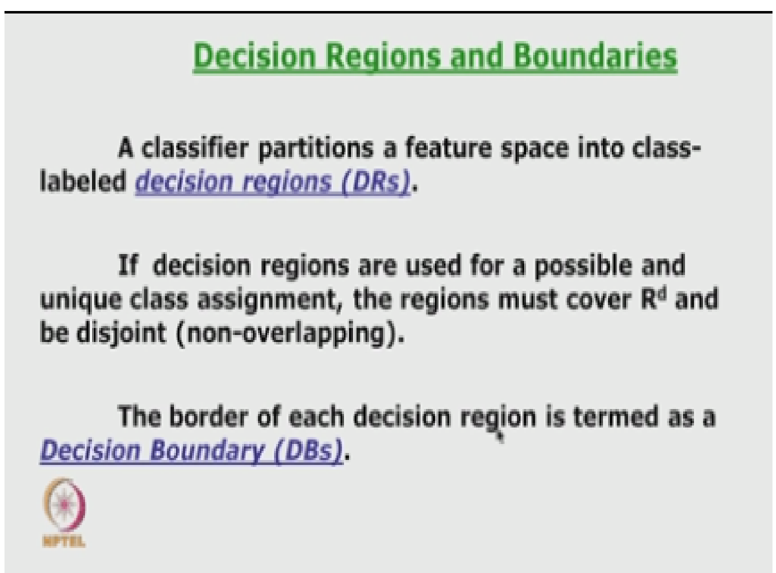
These are some nice synthetic examples which are typically available in many literature and books which shows examples of problems which you need a non linear decision boundary, this example shown dense distribution of instances belong to two different classes using two different colors red indicates one color the blue indicates one colors so this is one problem so what is the boundary you may need here, it is short of an elliptical boundary here okay you can see the red in a non linear boundary here between these two classes to separate one.

And ellipse will do here okay you need a very complicated boundary to discriminate between these two, situations can be even verse than this in all these examples which I have drawn and show so far the samples have actually not over lapped the samples have not over lapped means there has not been any amount of mixing between samples belong in to two different classes one and two assuming the one of this cases at the samples overlap then what amount of sophisticated classifier and feature space is used of course given a feature space you are using or designing a classifier your classifier is found to make some errors and mistakes.

However if the samples do not overlap it should be possible for you to design a sophisticated classifier which will legal an optimal decision boundary linear non linear whatever the case maybe I will repeat a good optimal classifier based on a good optimal design should be able to come up with the good linear or non linear decision boundary such that you can separate between two categories of classes.

As long as this samples do not overlapped like as given in the figure here although you do not have a linearly separable boundary to draw but it should be able to actually obtain good amount of classification.

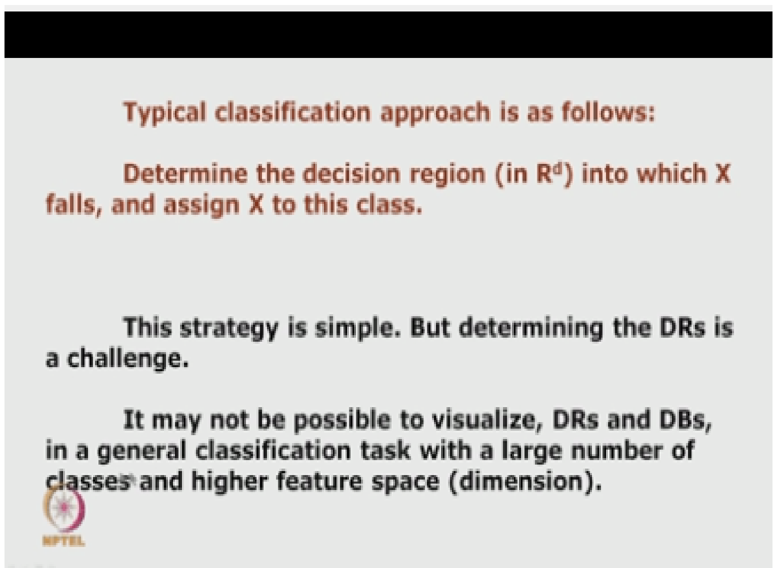
(Refer Slide Time: 32:35)



So continuing with discussion on DRs and DBs a classifier is expected to patrician the feature space in to class level DRs or decision regions and if decision regions are used for a possible unique class assignment which it will be then the region must cover the entire d dimensional space and must be disjoined or non overlapping these are certain desirable properties the border of each decision region is termed as a decision boundary.

In fact it is the border between a pair of adjutant decision regions to be very precise the border between two adjutant decision regions corresponding to two different classes are a pair of classes is called a decision boundary. Where really you will have boundary in caps rating one class in other side you do not have a class scope you have another class obviously which is we at least we does not belong to this particular class which is encapsulated by the decision boundary.

(Refer Slide Time: 33:32)



So typical classification approach is as follows, I am not talking of any particular algorithm right now for classification yet, I hope you understand that I am not entering in to a particular algorithm will pick up the base decision rule when it comes and try to link it with the multivariate Gaussian density but I am just getting in to the problem of classification, so what I am saying looking back the classification approach it is expected to yield good discriminate function obtain good decision boundary is between two decision regions.

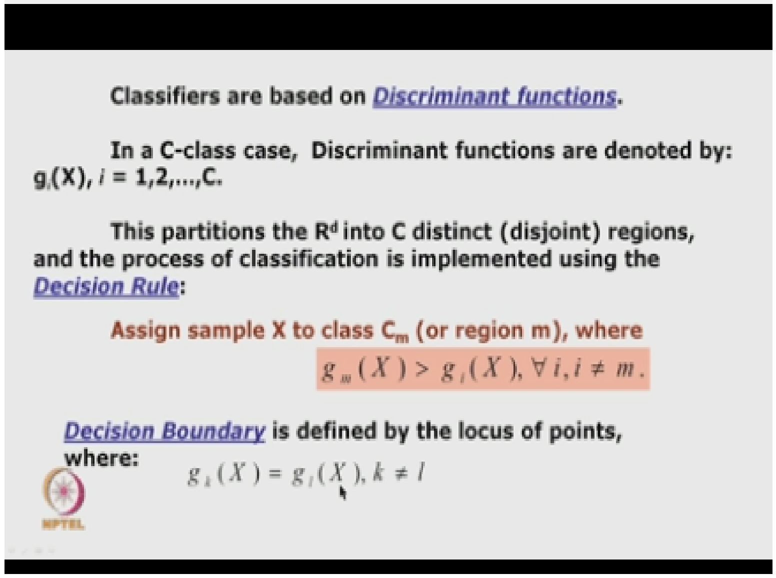
And it is to determinate a good dr in the dimension space on to which a particular sample acts or a instant x falls and assign X on to that particular class. This strategy is very simple, this short of approach for classification is simple the challenge is actually to determinate a decision regions and sometimes it cannot be possible to visualize all the dr and db in high dimensional space.

Often you will find that my diagrams will be a concentrating on samples in 2d or 3d for a 2 class or 3 class problem, so far you have seen and you will see that in this class today and in the next class following this the samples will be drawn in 2dimensional or 3 dimensional space at the most. Even when we had that animation trying to describe you the difference clustering and classification in 3d with 2 class problem okay.

But of course it is left to always the imagination of the community to visualize what is going to happen in the high dimension with large number of classes' okay there is lot of other issues which sometime it becomes difficult to visualize in lower dimension what can happen in higher dimension. So classifier or the design of the classifiers is basically was designed in the discernment function because the discernment function will give you the decision boundaries or db as well as the dr for the decision regions.

So design a classifier again I repeat is building discernment function or learning the parameter of a particular discernment function okay. Remembering that if you have a C class situation that means the number of categories are typically we are taking the number of class = 2 or 3, you will have as many discernment function as many the number of classes. Look back into this slide.

(Refer Slide Time: 36:08)



So classifiers are based on the discernment function and at the C class base the discernment function are donated by from here onwards till introduce g(ix) C indicating the number of classes and this partition the 3d space x c distinct disjoint regions and the process of classification is implemented in a simple decision rule it says it is nothing new to you, it has been done.

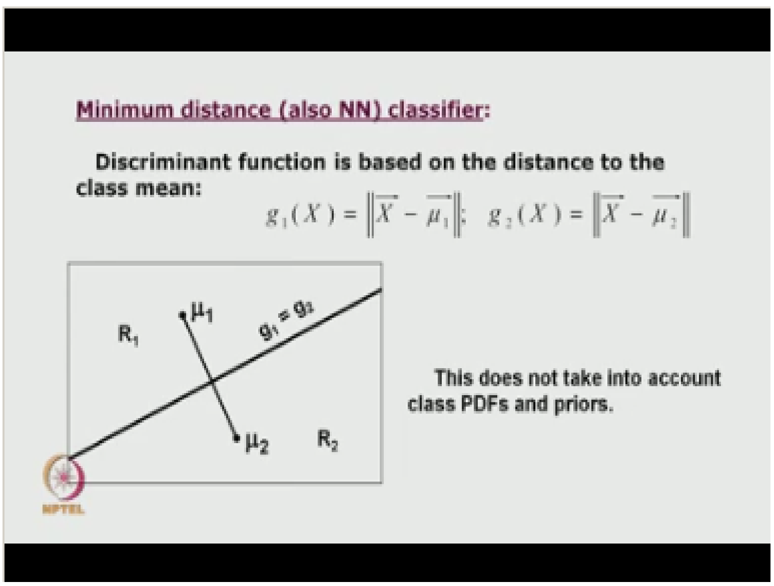
Along the lines of based decision rule also this can be put assign sample X to a class m or region m if the corresponding value of g is more than all the other values of g i≠ m that means what I need to do, after the classifier as learnt decision boundaries or the discernment functions I substitute into c different. What is c? c = number of classes okay, so I will have gi is the discernment function for the i^{th} class.

So if the c class there will be C gi so for all those C different gi I substitute the feature value x for that particular test sampling which I want to classify or discriminate or categorized or recognize I substitute, find out which value of the gi is the highest gi the peak or the maximum this is the simple policy but of course who will give you the Gi we will get into the mathematical details how to build the GI now.

What about decision boundary so it seems the decision rules are giving me the dr and the db is also given by this discernment function for any K and l ≠ to other, decision boundary is defined by the points where the correspondi8ng g are same, this we talked about already earlier. At the hyper plane the value of g neither will be positive nor negative you remember sometime back we talked about this the value of g = 0.

In this case of course in multi class classification problem while your c is more than 2 the of course you need to have c different regions, in some sense you can think of n c2 n is the number of classes those many decision boundaries and each of the decision boundaries a pair of gi will equal at each such in the decision boundary it will pair. So if you choose the boundary between 2 classes' k and l as given here the corresponding values of gi will be same.

(Refer Slide Time: 39:37)



A very simple classifier called a minimum distance classifier also called NN this is called the nearest neighbor it is not written in the slide but please make a note it is the nearest neighbor classifier okay minimum distance classifier. Discriminate function is based on the distance to the classifier very simply what I am computing here look at the expression of g the first expression of g which I am putting very simple distance of the sample from the mean for class one for one two distance of the sample from the mean of the second class okay.

Let us say in a two dimensional space these are the two corresponding class means μ1, μ2 and this is rational model which I get for g1 =g2 based on the expression which we had in the previous slide. So this is the region r1 decision region this is the r2 the decision region 2 around the mean for the corresponding class and the dv is where this is the most simplest example of a decisions boundary and a decision region for what is called as a minimum distance classifier or a nearest neighbor classifier the expression of g is very simple the distance between the sample x and the mean how far you are from the sample nearest sample nearest mean is what you will get as the class assignment.

Off course you must keep in mind one particular thing that the corresponding value of g in this particular is not going to maximize in that case you have to take 1/g if you want to really maximize it okay so you can think of minimum g for class assignment but it is okay we are interested in decision regions and decisions boundaries the only thing what we have done in this particular case is we have not taken into an account.

Any information about classes except the means it is like you are distinguishing apples from oranges you have a set of apples you have set of oranges that is all right you have the two class meets at these two points in feature space let us say but you do not know whether the samples of the apples are scattered around.

In a larger domain compared to this cater of the oranges how far is this distance between the two means are known but the individual scatter or water called in statically terms if we go back to the slide. This method of g do not take into account class profit density function of PDF which can be a model by a multi variant quotient density function and the example of it and the class pairs so without that if you do not want to use it many of distance the enabled classifier will work correctly but it is dangerous to sometime use this minimum distance classifier without using class priors into account you may get long wrong classification errors.

In classification if you do not take class priors or class distribution function which are not this has been discussed already earlier as introduction to classification class by professor Moorthy and myself earlier so to incorporate that we have to bring in the class distribution function bring in the probability and to do that let us do the classification problem using a short of a best classifier almost available at beginning of any book of pattern recognition or pattern classification.

Which is the best decision rule or base theorem and we will discuss this is in the next class how incorporation of the probability density functions using class pair conditional density functions etc within the base theorem give a much more better accurate process of classification and better estimate of the decision boundary and formation of the discriminal function and hence the decision regions for the purpose of classification we will end the class here today with this short note or the idea.

Which we have discussed today or to summarize the process of classification involves designing o0f discrete functions or learning the parameters of discrete functions the discreet functions will give the decision regions a boundary between two regions is called the decision boundary these decision boundary could be linear on linear depending upon the individual cater are spread of the class samples or data which will be reflected by the covariance matrix okay.

So we will see in the next class how the multi variant Gaussian distribution function in corporate using the base decision rule or the base theorem brings in the concept of decision boundary by exploiting in the distance of sample from the class that particular distance from we will come back again which you add in the beginning of the class today it will come back in the next class and the ideas whenever we talked about DBS, DR discriminate functions hyper plans hyper surfaces etc the concept will be much clear thank you very much.

Online Video Editing /Post Production
K.R.Mahendra Babu
Soju Francis
S.Pradeepa

Camera
Selvam
Robert Joseph
Karthikeyan
Ram Kumar
Ramganesh
Sathiaraj

Studio Assistants
Krishankumar
Linuselvan
Saranraj

Animations

Anushree Santhosh
Pradeep Valan .S.L

NPTEL Web &Faculty Assistance Team

Allen Jacob Dinesh
Bharathi Balaji
Deepa Venkatraman
Dianis Bertin
Gayathri
Gurumoorthi
Jason Prasad
Jayanthi
Kamala Ramakrishnan
Lakshmi Priya
Malarvizhi
Manikandasivam
Mohana Sundari
Muthu Kumaran
Naveen Kumar
Palani
Salomi
Senthil
Sridharan
Suriyakumari

Administrative Assistant

Janakiraman.K.S

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