Indian Institute of Technology Madras
Presents

NPTEL
NATIONAL PROGRAMME ON TECHNOLOGY ENHANCED LEARNING

Pattern Recognition

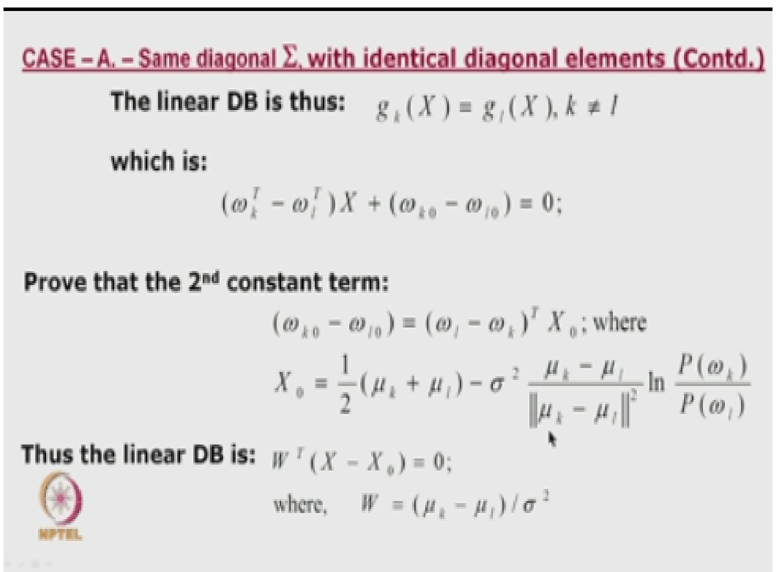
Module 02

Lecture 11

Linear and Non-Linear Decision Boundaries

Prof. Sukhendu Das
Department of CSSE, IIT Madras

(Refer Slide Time: 00:20)



Yeah, so to summarize the expressions of the linear decision boundary under the one of the simplest case of same diagonal covariance matrix as well as identical diagonal elements that means along every dimension you have the same variance, if you a D dimensional feature vector, then all the D different features have the same variance along their respective dimensions and they basically means a scatter and spread of color as a feature as same as the height, same as weight as another feature, same area for a texture or something else.

So in such cases let us look back to the expression of the linear decision boundary which is of course given by this expression where the two discriminant functions are identical with k0=L is given by this particular expression which is a linear form and this we have done, just seen that the expression of the linear decision boundary is also conveyed in a Christ form in this pattern, where the X0 actually is given by this expression.

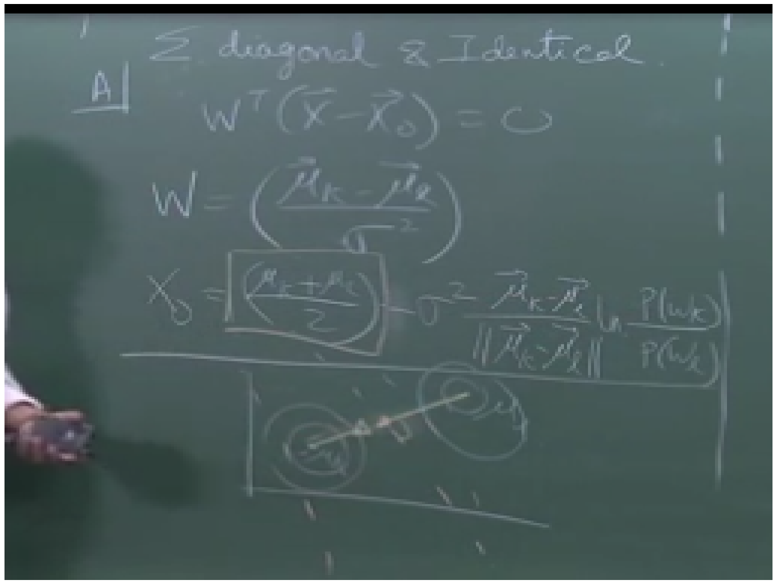
I left it this for your home assignment taken assignment, and you should be able to derive this based on the expression given here on the top. So that is finally here, so you have this as the expression of the linear decision boundary whether W is given by this difference between the mean divided by σ^2. In the more general case when we had taken the covariance matrix as an identity matrix, then of course it was just the difference of the two means, because the $\sigma=1$.

So now this is also a little bit of a more generalization where the matrix is diagonal like an identity matrix alright. But you have variances which is not equal to unity. And they are same, they are identical diagonal elements as given here, so that is the same as the σ. So that is the one which will be sitting here and the X0 has given here.

So let us write these expressions in the board, because we will be comparing figures, we had seen some decision boundary or examples of decision boundaries. In the last class we will see a few of them more, and we will compare two or three different cases of variance of the covariance matrix, we will take an arbitrary covariance matrix where it will not be diagonal, you will have half diagonal terms.

But what is the effect of class pair in such cases, what is the effect of the covariance matrix had to be identical across different classes or if it is varying across, it is not identical across different classes then what are the effects. So those will be the different cases we will study, we will write the equations on the board for your understanding for the case A, and then we will move on to the other cases.

(Refer Slide Time: 03:05)



What did we get as W, it is μk difference between two means divided by σ^2, is not it. So this is what we have, and what is the expression of X0 we got, see originally this should have been the perpendicular bisector, so it is the point between the two means, so it should be (μk+μL/2)+ can you give that term. Well, there should be a variance which will come here, this is the factor which depends on the class pairs.

So what that expression and I just formed, it should be proportional to σ in some sense or the variance. Then it is a vector, it moves along the line joining the two means, so you should have μk-μL there us a reason behind this, and it is norm, so this is a unit vector, well I am not putting, if you want you have all of these as vector signs. Then of course, this should be easy to guess, others are log of corresponding ratio pairs, was it PK/PL or PL/PK.

Let us have a look at the expression it is K/L. So this is what we have just remember, and this is what we are considering as case A, and in some sense if you want to have a two dimensional diagram of this, if there is a corresponding, if this is the μ class mean two dimension for class K, if this is the corresponding μL for class L. And since the σ is diagonal, so the isotropic lines which indicate the multivariate Gaussian function will be is conclude circles or hyper spheres.

So I will just throw one or two of them, we had this diagram in the last class. This is for one particular class, the same thing holds good for the other class as well. What is the vector μk-μl, so this is that vector, we will go back to slides and check if this is correct or not. So this weight is in the previous case, in the most simplest case when we had taken this to be an identity matrix, this term was not there, so W was basically this.

This and we also talked of this, that this W is actually normal to be claim, which is a linear decision boundary in this case it is a line plane in three dimension and hyperplanes in this. So the decision boundary will be, decision boundary if you take this particular term here, I probably just change the position of this arrow, so that I will just put it here. So this is the term, say this if you want to look at this term it should be a covariance here.

So this particular term is said as this here, so this is the point here, this is the linear decision boundary which is basically orthogonal to this. This is the point if the class means are same, class pairs, I am sorry, is the class pairs are same, then this term will vanish of course you can say that if this also 0 it can vanish but that is no meaning because you know that these are the diagonal terms in the covariance matrix.

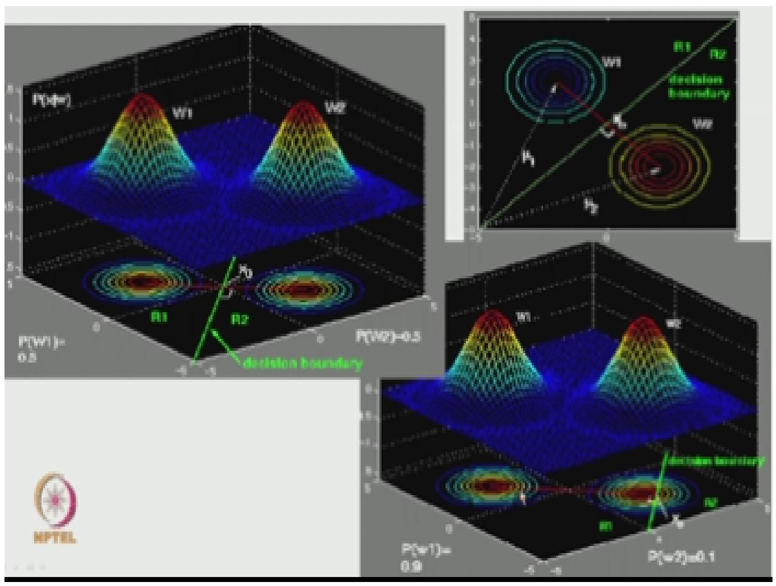
And we are excepting some non 0 variance of course you in the color we cannot have all the features same from different samples this is just union vector okay so basically is that I am moving up and down depending up on the ratio log priors ratio here and a larger value of the class prior on the numerator will actually push towards the L okay remember this is a vector so it is indicating this is point so + and – in some order towards kl and I think is there a negative sign here.

Yes that is all this is a negative sign so negative sign means if this is more it will move towards that side okay positive side we will move towards k okay and that is what we had seen with some of the examples so we will leave this with this diagram here and if you want to draw a linear decision boundaries for cases when K and L are there you can have something like or even for a large value you can have it like this for that DPA might come here if the Wl is very large compared to WK.

Say 0.9 or more then the linear team boundary can even come here at this point a okay it can pass through this point as well if the trial for the class K is more than okay so this is the these are some of the diagrams which we have seen in slides in the last class in 2 dimensions we will going to see a few more of such cases today before we move on to the case where we have obituary nom diagonal.

Obituary covariance matrix σ even with of diagonal term so this a special case where the covariance terms in the of diagonal terms for the covariance matrix are equal to 0 and that is when we could write the weight vector $\mu k - \mu L / \sigma^2$ which cannot be possible because we need to give the covariance matrix overall in the expression which we will see.

(Refer Slide Time: 09:19)



But before that we will just have a look at some other diagrams which are available in some public domain so this is one of the cases which we had just discussed remember this is a the corresponding class conditional probability distributions for w1 and w2 two different class which we considered as K and L and you can look at the 2 class mean this is similar to the diagram which I have drawn on the board here the corresponding iso contra lines of constant distance from the corresponding class means.

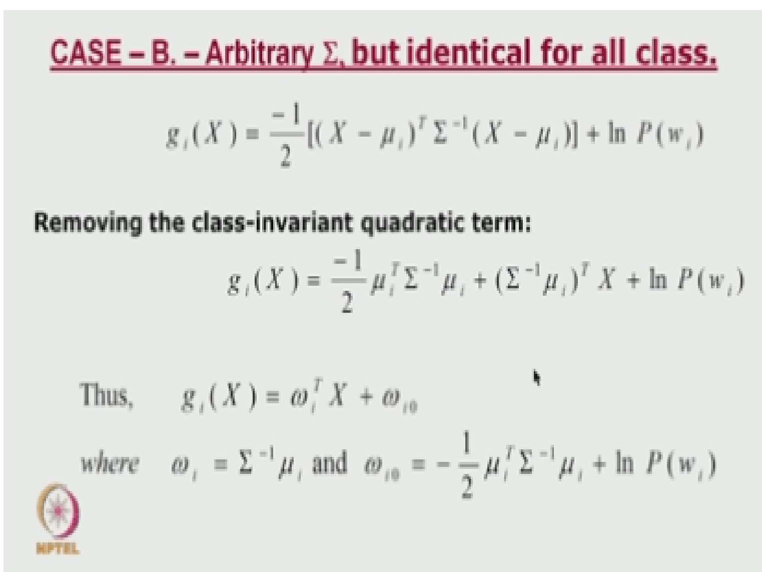
And this is the case of decision boundary with the 2 class pairs of same 1 is 0.5 the other is also 0.5 so this is the very simplest k so when the decision boundary is normal is aligned normal to the lines having the class means okay this is a the same diagram when we are looking at this plane projection of these 2 distributions here the same thing is what we are saying at this movement 2 distributions here 2 class mean vector $\mu 1$, $\mu 2$ and the decision boundary orthogonal to this particular line okay.

This the case which I was talking about at the end sometime back that if one of the class priors is made equal to 0.9 and you can see that and the other class prior is value to the point 1 so we had decided this that the decision boundary will not now stay at the centre point or mind joining to the class means it will of course remain perpendicular to the line joining through the class mean but it would not pass through this point closer moved to in fact it is crossed beyond the mean for the class 2 and this is where the decision boundary and this is the decision boundary and this is the case because the class prior for the class 1 is more than class 2 of course the other factor which you had seen in the expression is that the variance storm okay.

The larger the variance storm you have more is the shift or drag of this linear decision boundary towards one of the class means okay so the amount of drift of this decision boundary will be depending among tow factor one is log prior another is the variance of the covariance matrix good so moving on to the next case which is an arbitrary covariance matrix but still keep it identical for all class okay.

That means we cannot now just simply say that we have a diagonal σ the 0 of diagonal terms you cannot say that any more we do have non zero of diagonal terms as well as artery σ means that we do not have equal variance on the diagonal so we have $\sigma 1^2$ $\sigma 2^2$ and so on up to $\sigma\ d^2$ and σij typically will be non zero in general okay in such cases but the covariance matrix is same for the class when you write this expression

(Refer Slide Time: 11:56)



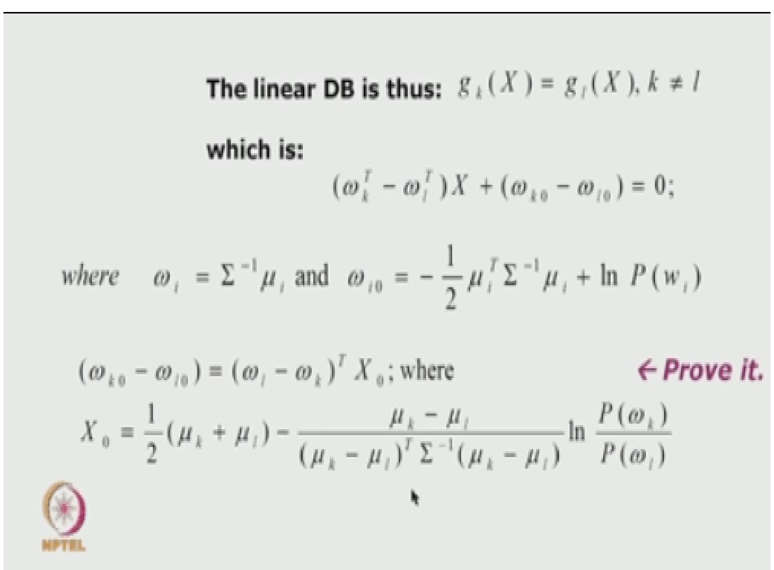
When you write the expression for the discriminant function for a particular class gi(x) yi can notice here that the class mean term μi available here as the case earlier and then we have the log prior term as well but we do not have the sub to the covariance why do not we have it here because this σ is identical for all class so if you have this expression and we can remove the class element quadratic term which we have done so far.

That means the corresponding σ inverse x^2 so we will basically left with the linear terms here okay in fact this is the linear terms of function of x and we have two constant terms which de0onds on the corresponding covariance matrices in fact for the individual discriminate function for a particular class we have σi is here you have the log prior term but you do not have this subscript for the covariance the reason is the same it is identical for all class and this gives us.

The linearized model for the discriminant function of course have to write the decision boundary where is shown which we will be depended in a very much as a same this term is return form here this particular multiplied by x so the corresponding weight vector is this maximum of course the transpose sign is anywhere and the corresponding with other two terms which are here are all coming at the base term okay.

So this is a possible to linearized as long as you have an arbitrary σ which is identical for all class again I will mention that in the some time back as well in the earlier class with unrealistic expansion what you are making an assumption is that the variance of color of a certain flower is same as the fruit if fruits and flowers are 2 different class okay the weights vary by the similar amount and so and so forth okay, so this what the expressions of mu gi(x).

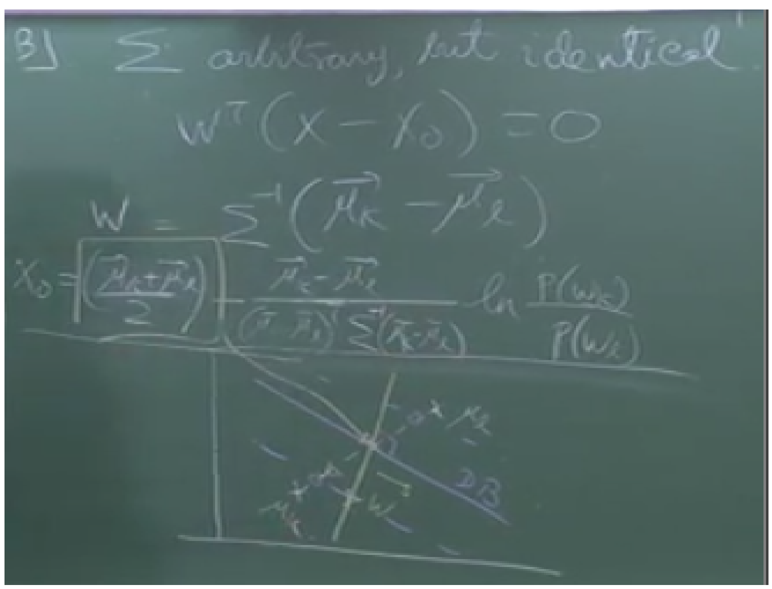
(Refer Slide Time: 13:46)



And the linear decision correspondingly using the same as we are done earlier is given by this particular function where k and L are 2 different class well you can imagine terms of I and j also if you like but these are 2 different class here and so you will have where each of the wi's from the previous slide is a covariance matrix multiplied by the class P or the class main mean P multiplied by the inverse of the covariance matrix and this the corresponding bayes term.

And we substitute it here this is what in fact you should be able to rewrite this particular term difference or the 2 bayes as function of the weight vector here okay and again I leave this as an excursive for you we had written this earlier for the case when we had a diagonal σ you can see that the expressions actually A is the same and we are going to write this again on the board and if you remember that is this is midpoint than the 2 class means – this factor is a same except the σ^2 is replaced by the inverse of the covariance matrix okay that is what you have here and what you have is basically the this is the vector which show that you are moving the decision boundary moves along the line joining the two class means okay.

(Refer Slide Time: 15:02)



So what it is means is covariance matrix is now arbitrary it is no longer diagonal like the case A but it is identical for all classes it that is what the word identical means here and the corresponding expression is the same overall expression does not change it is still linear decision boundary and where the W will be given as inverse of the covariance matrix divided by this okay and the corresponding X_0 and it will become space there the transpose should be here correct to multiplied by the log.

Okay so you can now compare the two expressions here okay this is what in the case when just diagonal very strictly the numerator remain the same the numerator change by the inverse of the covariance matrix the small correction what is the small correction here, okay so now it leads the same okay the inverse of the covariance matrix will give you $1/\sigma^2$ if it is diagonal and that goes to numerator here, so case a is simplified version of the case B okay and if you want to draw the diagram here.

In a similar manner like the one so I am not drawing the Iso contour lines will have slides coming up for that so I am just for the sake of simplicity say this is my class mean k and that is my another class mean μl for the two classes K_0 = L and you have let us first draw this just the numerator drawn without the covariance matrix which is μk = μl so what will happen is look I will put this dot okay so this is basically my μk – μl this particular term here which was the same as this.

Where just as scale factor earlier so this is the same vector which we have drawn for the simplest case A is same as numerator now okay without these term okay is so that the dash line with arrow basically gives this term and of this vector multiplied by the inverse of the covariance matrix we know that the covariance matrix is symmetric matrix okay symmetric matrix so I leave it as an exercise for you to visualize that this will actually create a short of an fine transform.

Okay for the corresponding vector so if this is of course an orthogonal matrix then that could have maintain the rotation okay but this is and also D dimensional rotation combines with an translation or okay so this basically means an a fine transformation in some sense you can visualize that the W now which was actually here okay not to scale of course we will what will now happen is that the W will get till 10 okay and it will probably give you a vector let say would be some like that.

Okay so that is my W the amount of tilt of this W with respect to the line during the class means will actually depended on this the covariance matrix which is arbitrary here but you can say this is some sort of tilt one, now what will happen to this point is the same effect which we had earlier because in general if this term vanishes again the term could vanish if these two class as are same then we are talking about the point is passing through, so if you looking at this particular term.

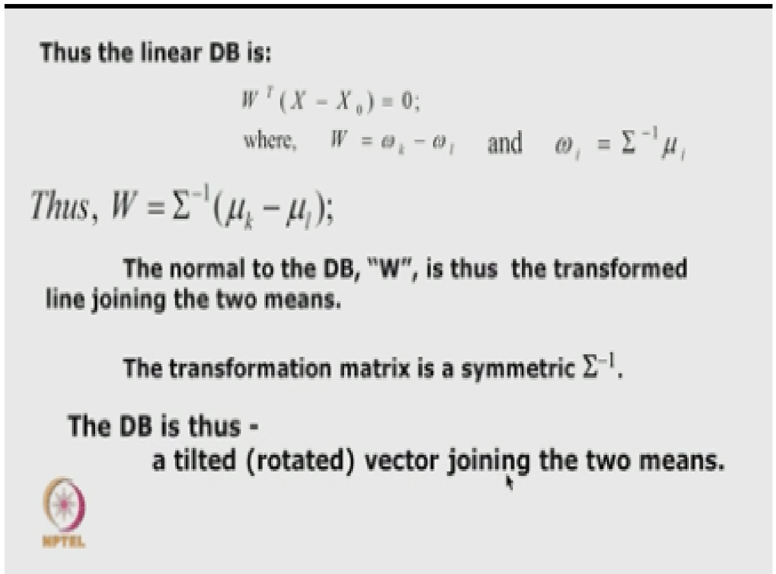
Here again and let us say that this particular point is this midpoint which is the mod point between the two class means so that is going to remain the same and what will happen is one thing you must be keep in mind always that this W vector is normal to the linear decision boundary and if it is normal to the linear decision model I have put that as a – sign so I am going to keep that consistently here may be I am going to use a slightly different color okay so this is what could be your DB.

The linear DB which is orthogonal to the W okay passing through the midpoint in the two class means if this term is equal to 0 which is same concept which you discussed earlier and if this term non zero is what will happen is bound to move along this vector because you have a now vector dictate by this so it is going to move along this what will happen is you can draw different types of decision boundaries here all of them will ne orthogonal to the W or strictly speaking W will be normal to the DB.

All these DB is which will appear due to different values of these particular term which will actually shift this point along the line joining the two class means, so at this part by shift from here to here is where your X0 will occur okay so that these are potentially some other points where this value might but at that point also the normal will still be the same W it does not change this does not change depending upon this particular term the class pairs is not affecting the W so the normal is basically the same this boundary just parallel to that and it keeps shift up and down.

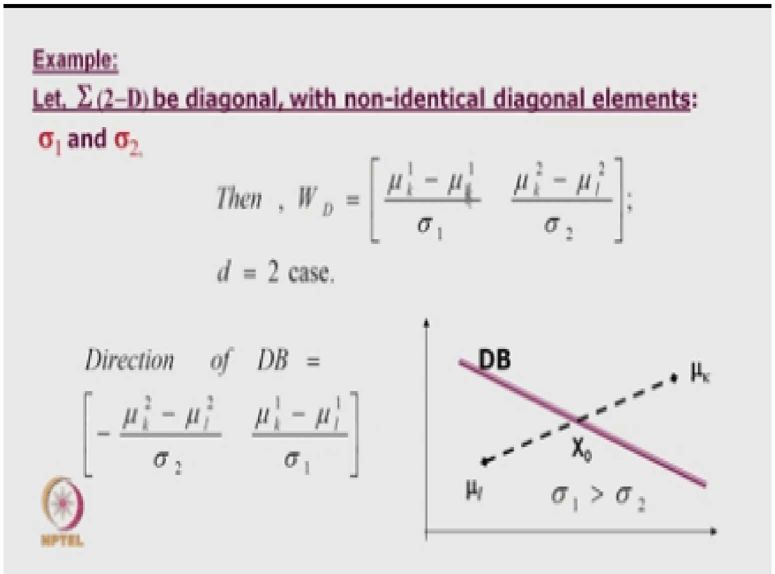
Here long this depending upon okay so okay you can just look at some of the examples of so this the expression if you look back one second that we writing in this particular form the linear DB is basically inverse μk – μl.

(Refer Slide Time: 22:37)



That this is how why it will get you know okay this is the expression of the weight vector and so it is just the transform line generally the two means so these is the line generally two means μk – μl as it transformed by correspond symmetric remember the inverse the σ or the corresponding covariance matrix is symmetric the inverse is also symmetric and the such can be considered as treated a rotated vector joining the two means.

(Refer Slide Time: 23:02)

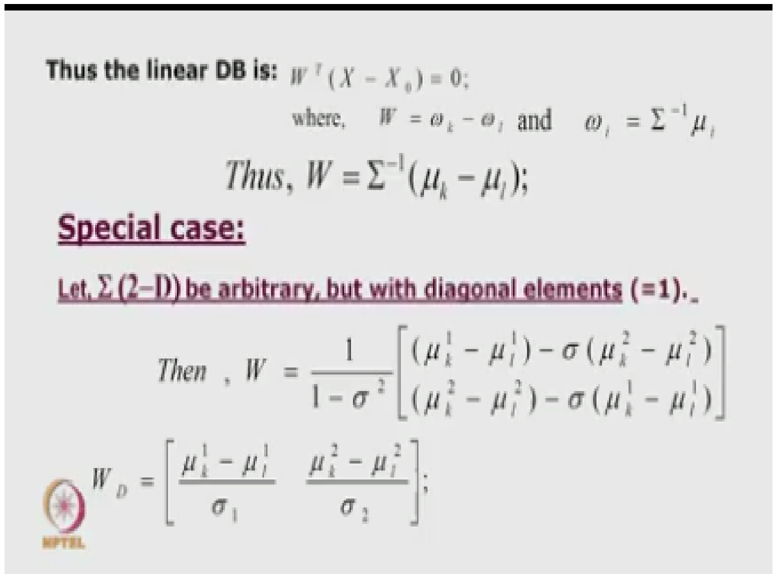


I have a simple example of this in two dimension that means if the covariance matrix is a diagonal term with non-identical diagonal elements remember this is not arbitrary still diagonal but for the sake of analysis we are just taken diagonal matrix but with unequal elements so this is somewhere between case A and B, remember case A we had diagonal and but identical elements here it is identical over classes arbitrary so this somewhere in the middle this diagonal but we have non diagonal, diagonal elements.

None sorry, non identical diagonal elements as given here okay, so the corresponding wet vector will be this I will leave it to you as an exercise and the direction of the corresponding DB will be this, because this is the w which will not be the line on the two means so the w will be a vector which will tell this is the vector join the two class means so we have a titled vector which is orthogonal to DB and this is an example of the direction vector of the decision boundary which is going to be normal to w, okay.

So these two are orthogonal vectors okay, so the line joining sorry, the decision boundaries marked here in the graph which is basically this vector.

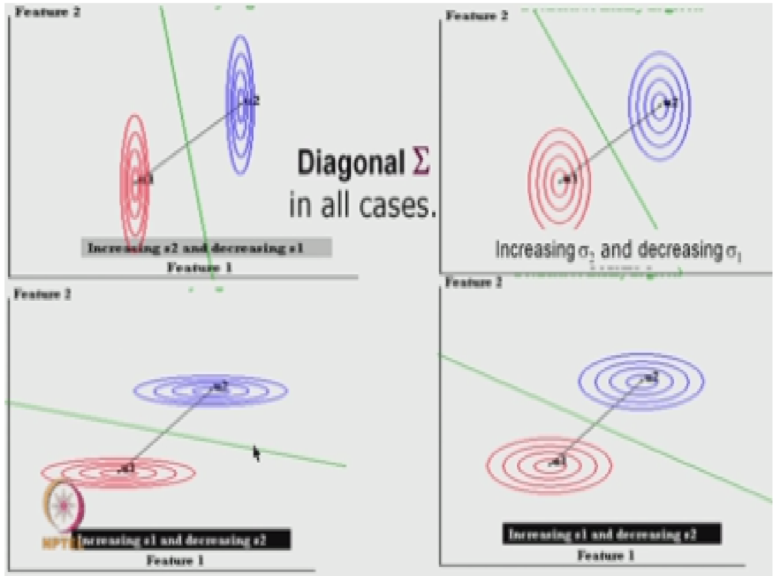
(Refer Slide Time: 24:18)



Okay, so just to yeah this is special case again carrying on with another simplest of the most simplest example but all diagonal elements equal to 1, if this is arbitrary so now what we are talking about is a special case where the diagonal elements whether the covariance matrix is arbitrary but the variance terms on the diagonal equal to 1 okay, so the σ^2, $\sigma2^2$, $\sigma3^2$ along the diagonal are equal to 1 you have an off diagonal terms.

Since it is a 2D case which is simple to 2x2 matrix so it is one along the diagonal and the off diagonal terms are equal to σ, okay so in such particular in such a case you should be able to obtain the value I leave this as an exercise for you to derive the equation for the case when W= this D indicates that it is a diagonal elements are equal to 1, okay that this has to does not indicate the dimension it indicate the case when diagonal elements and you can derive this very easily. So this is the σ^{-1} which you have and the corresponding terms are given here, okay.

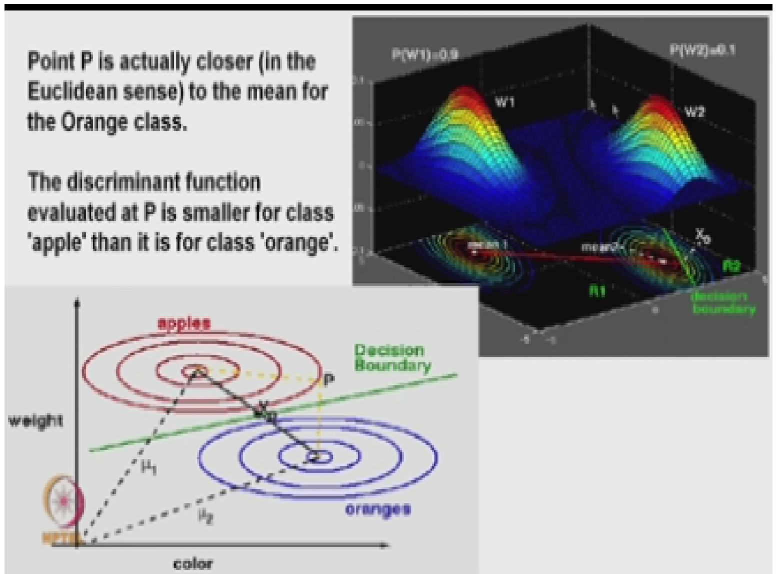
(Refer Slide Time: 25:31)



So these are certain examples of cases of arbitrary σ's which are diagonal in all cases with increasing feature 2 verses feature 1 so you can see that the decision boundary is no longer orthogonal to the line joining the two class means here it is almost diagonal but still the variance along the y direction or feature 2 is more than 1 so you can see that it is no longer a sphere it is becoming elliptical and it is all tilted.

It is titled more here in this particular case because the variances much, much more higher this is the case when the variance along x is higher than the variance along y okay, and the DB is not orthogonal this would have been the orthogonal line you on the two means so this is a tilted case here this is also tilted as well, okay so these two are not orthogonal. Remember this figure is in two dimension.

(Refer Slide Time: 26:33)



We will have a look at this figure where this is obtain from an e-document from the web where we have these two class means let us say that indicates apples and oranges and you can see here that the variance for the apple along the color dimensional what are the two features selected here weight and color are the two features for apples and oranges the variance of color for apples is large as well as those for oranges where as the weight vectors have a less degree of variance.

So if you take the line given the class two class means here and the decision boundary would have been orthogonal by a line joining as shown by them moving cursor at this particular point so the line would have been here if we had circles as my isotropic lines of one of the distances from the center, but in fact it is not so what this point P means is that this point P may appear larger by equivalent distance from this class mean for apple compared to those orange you can see that.

This distance is much more than the distance of P to the midpoint for oranges but based on margin of distance criteria these are equal lengths, so this is say D=1,2,3 and 4 the same case 1,2,3 and 4 so in some sense that this is the distance to the force of the, so the discriminate function evaluated P is smaller for class apples that is for class orange Y because this is the value, this distance is D=4 for oranges, this is D=4 let us say for apples.

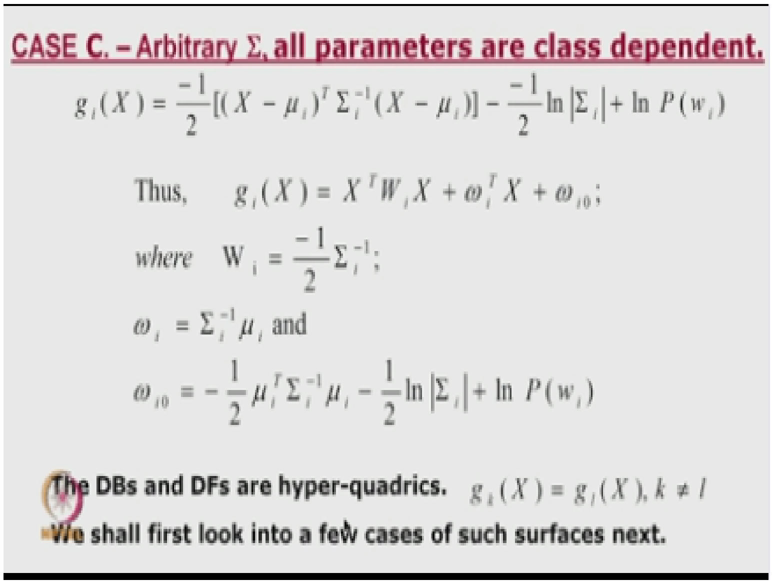
So this distance here is much than a value 4 okay, so if you observe this very carefully that I repeat again that the distance of this point P in the equilibrium says is much more from the class mean than for the mean corresponding to oranges. However if you take main of this factor that is the covariance matrix into account then this distance is much more than this distance here. in fact this is at the same distance as yet so, you can see the why the decision boundaries getting tilted because this is the points where my main distance from the two mean should be identical.

These are the points where the means are identical okay, and if you observe the same for in 3D figure with the corresponding class conditional priors class concern probabilities are shown here and this is the case where the decision boundary are shifted away from the midpoint joining the two class means towards one of the other class means because the class for again for class 1 is much higher than the class prior for the other class which is point 1.

So this is so large that it has not only move towards the class mean but gone further away, so we can see the effect of the class means and the tilt in the decision boundary due to an arbitrary covariance matrix which can take place if it is not diagonal, as long as it was diagonal and it is identical elements we had always the constrain that the W wet vector was the class, was the line joining the two class mean and the decision boundary was like a perpendicular bisector.

But not any more when we have arbitrary covariance matrix the decision boundary will be a tilted vector there the corresponding normal to that also will be the tilted vector which is the vector joining the two class means., so we have seen that.

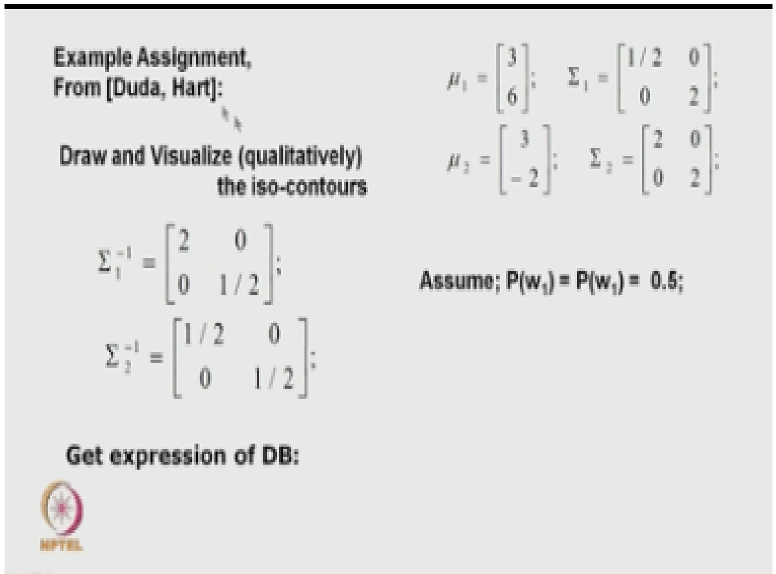
(Refer Slide Time: 29:52)



We will move to the last clays where you have an arbitrary σ and all parameters now are class dependent okay, that means this is different from the class B, the class B also we had an arbitrary covariance matrix but you had σ to be identical over our class now if you take flowers and fruits, so flowers will have a σ 1 as a co variants matrix fruits will have something else okay. let us look at the expressions this is the expression for the discriminate function and now you have to put back the index here which we could avoid earlier for the cases a and b.

And when we open up and write this expression in this particular form as a short of a non linear discriminate function for the first time you may having that then the corresponding weight vector is yeah this is a quadratic term which we cannot eliminating in more because there is subscript I which is just the inverse of the co variants matrix this is just a factor which is used and then other terms are same these two terms is what we already had in our previous of course with minor variants here because earlier this was arbitrary as well but the subscript was missing earlier.

So in general the discriminate function based on this expression of gi and the decision boundaries which will be obtained by taking the difference of gi for k and l all what are call in general hip[per quadrics and the decision boundary can be obtain by using this expression.
(Refer Slide Time: 31:27)



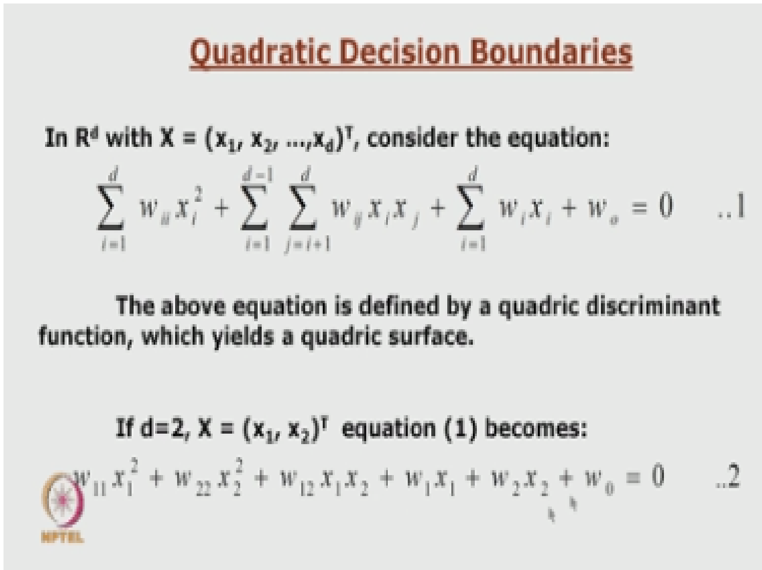
So just look at a few cases of the a simple case this is an assignment which I am leaving for you which I have picked up from the book by Duda, Hart on pattern recognition so these are the corresponding class means for this is the two class problems so this is two class means and these are the corresponding co variants matrix well we have kept them identical sorry we have kept them diagonal we have kept the co variants matrix diagonal but they are not same for both the classes okay.

So draft diagonal dimension are 0 alright for the seek of simplicity, so that you can work it as a home exercise using pen and paper but the main important fact is they are not identical okay, so I leave it an exercise that you can compute the co variants matrix inverse from the first one as this it is varies because it s diagonal one so it be one by the element so this is a very simple extension of that we talked about this in the discussion on linear algebra and vector spaces when we had seen the inverse of the diagonal matrix is just the elements of this.

Assume similar class priors and I leave this an exercise for you to write the expression of the decision boundary using these two class means and these two co variants matrix what is the formula we will use will go back this is the individual expression for gi so use this expression of gi take gk = gl and take expression of you just need the co variants matrix you just it the class means and the class pair they are all given the next slide.

Here is what you have for the class priors here of the class means here the co variants matrix use the previous one this is solved out problem in the book by Duda Hart they can refer to that for the answer.

(Refer Slide Time: 33:08)



Quadratic decision boundaries it is very nice to note that in d dimension in general it has an expression of this nature we will come to those expression very soon for the case c and those is responsible that two set of terms which are responsible for quadratic the terms this is the linear term and in a special case when the dimension is true you can actually open up and write this expression this particular form.

How may terms we have 1, 2, 3, 4, 5, 6, you can count for the equally to the two the number of terms here will be equal to 6.

(Refer Slide Time: 33:44)

$$\sum_{i=1}^{d} w_{ii} x_i^2 + \sum_{i=1}^{d-1} \sum_{j=i+1}^{d} w_{ij} x_i x_j + \sum_{i=1}^{d} w_i x_i + \omega_o = 0 \quad ..1$$

$$\overline{X}^T W \overline{X} + w^T \overline{X} + \omega_o = 0 \quad ..3$$

In equation 3, the symmetric part of matrix W, contributes to the Quadratic terms. Equation 3 generally defines a hyperhyperboloidal surface.

If W = I/0, we get a hyper-spheres/planes.

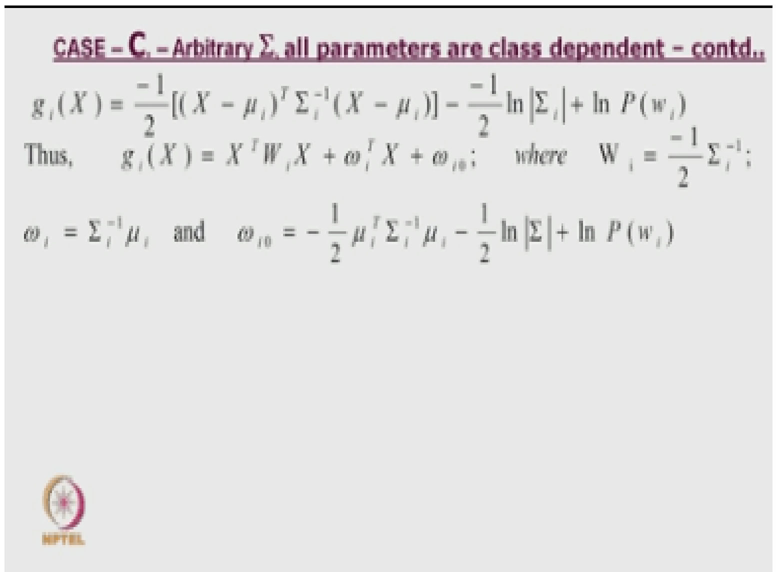
For a generic case when the dimension d is equal to this is the expression which we have so these are the bio variant terms the bio variants co efficient and the individual presences in the above equation total number of parameter which you have is this why you have 2d you have d + d here. The d term here d is here so that is 2d + 1 here okay and if you look at the number of terms here is basically a summation I leave it an exercise for you that what you will get.

So total number of term is given her or you can write it in terms this so you have to organizes them in this matrix because this is expression we got from the non linear decision boundary when we put gk = gl or the expression of gi itself was in this particular form when we are talking about an arbitrary sigma not identical for all classes. If we absorb this expression here in equation number 3 the w has d terms okay, which w? This one this matrix is a linear vector to continuing d so this is one wij is a d cos d matrix, and what we need to look basically to match the terms is that you need to duplicate so it is a symmetric matrix w here.

And the off diagonal terms which is d2 – d is obtained from this matrix so this wi are these elements which sits in the diagonal of the capital w here and the wij are duplicated by making half each and the sit in the half diagonal terms of w so this is how you construct this expression of the quadratic surface in d dimension or hipper quadratic surface in d dimension from an arbitrary polynomial nature of the expression is given here in d dimension this is geometric this is a vector representation.

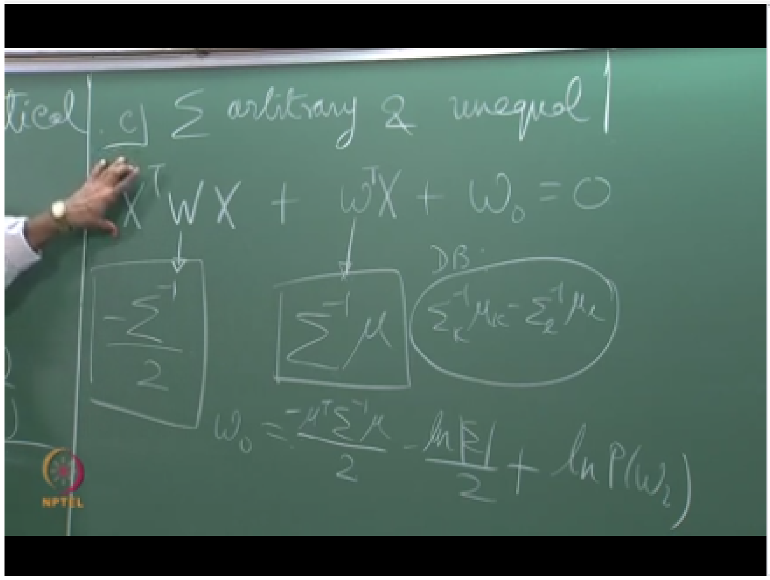
So in this equation which we had our gi this is the symmetric part of the w which contributes the quadratic terms and it generally it is called a sort of hipper or hipper poodle surface is high dimension three or higher dimension is what you will get of course if w is equal to an ordinary matrix then you can get fears or hipper fears if w is equal to null matrix you just have a linear decision boundary then you have lines planes and hyper planes.

(Refer Slide Time: 36:00)



Before we move ahead with some examples.

(Refer Slide Time: 36:11)



So what you have for case Σ is arbitrary and an equal that means non identical for all classes what was case a this was diagonal this simplest case of course the most simplest case which you are not return in the board was the case when this was an Andretti matrix okay and then the this we can consider two classes back so when this is again you consider as a special case of a when this is an identity matrix for we get actually Euclidian distances from the class means here you have this to be diagonal and identical class b this is arbitrary but identical over all classes okay Σ is arbitrary un equal over all classes.

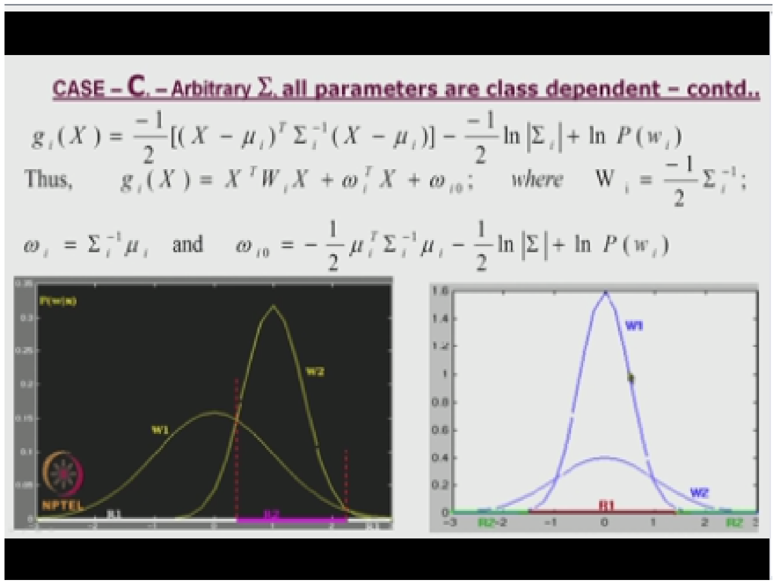
So now we have the generic case where you have unequal diagonal elements arbitrary of diagonal elements and they are not similar over all classes that mean the features over fruits and from those of flowers over something else say animal in some terms of image signal if you are talking about recognizing patterns of classifying okay. So in this particular case the expression so this is I am ignoring the I is there a + or – W^T so you need to have a T and for a arbitrary i^{th} class it is the discriminate function you have to put an I here as well as here.

This one is co variance matrix inverse and that is all/ 2 – that is what is your quadratic term, linear term here, you will be co variance matrix inverse x the ∑ okay and correspondingly W0 will be is this log prior – ½ there is – in okay. So as you see here that these are special cases of that in fact you should be able to map especially for the leaner terms, so if you look at this is basically this they have written in terms boundary that you can do because what you will have here is a W k – l here you will have σi μkl.

So I will just write one particular term so if you look at the decision boundary point of view, so you will have this, this is what it will look like because you are talking about gk – gl = 0 so this is what you will have, of course this is the co variance matrix are identical then you can go and write like this, we can take out that is the common factor and the same things hold good here as well.
The log prior term which looks different well in the true sense term this as no meaning in case of the quadratic because first of all the what you are going to have as lines surfaces so this is the linear form of this particular one in fact, so at least the weight factor is the one which you can simplify from the linear term and there is the quadratic term which is coming out from here and sitting here.

So in general so if you have scenarios where typically let us say this is μk and μl you will have these diagrams drawn on the board, so it is possible that you may have something like this and you may have then something like that arbitrary co variance matrix inverse, it is very unlikely that it going to be linear okay because it seems there is an effect of both are unequal first of all, the diagonal terms are not same and there are off diagonal terms as well, so it is possible it will be something like that.

(Refer Slide Time: 42:12)



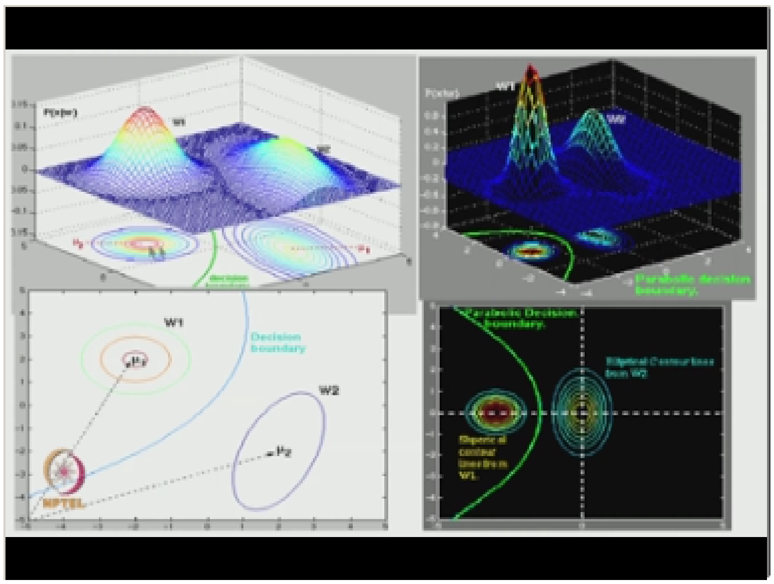
So this is the case which we are observing in 1d and remember the distance boundary in 1d is a still a point okay and if it is point in all other cases like the simplest case of A, B the weights two parts for class a and b or a and l or class 1 and 2 what do you think will be a non linear decision boundary in 1d okay, it should be a point but it may be more than one point, think of an intersection what is the point?

Intersection of two straight lines, I will give a geometric interpretation a point is intersection of two straight lines a typical is example intersection of weight vector there is an boundary, think now the intersection of curve and a straight lines non linearity because of course it is non d and think of a non linear curve which actually represents, so it will intersect the state at two points. So you can observe interestingly here you look at the 2 distributions in 1d what is the co variance matrix here, basically you have 1 variance term.

And this is the two class means w1, that means class variance for class 1 is much more than class 2 the prior also have the role to play which is not given here, so you can see here that look at the region this is the region because the class condition prior for class 2 is more than class 1 look at this you need to assign it to a class wherever thee class condition probability is more than 1 particular class.

So here 2 wins here class 1 wins here also class1 as a larger probability function distribution then the class 1, so this region belongs to R1 in between you have the region r1, this is a case of completely over lap class to the right with two class means are identical lot of errors you are going to make internal average on 50% or more in classification but look at here to here may be the red mark is not that great enough so there can be effect of class pad, so from here to here you have the condition density function or PDF of class 1 more than W. So this region belong to r1 but here you on both sides you have region r2 which means so you have two points as additions.

(Refer Slide Time: 45:13)

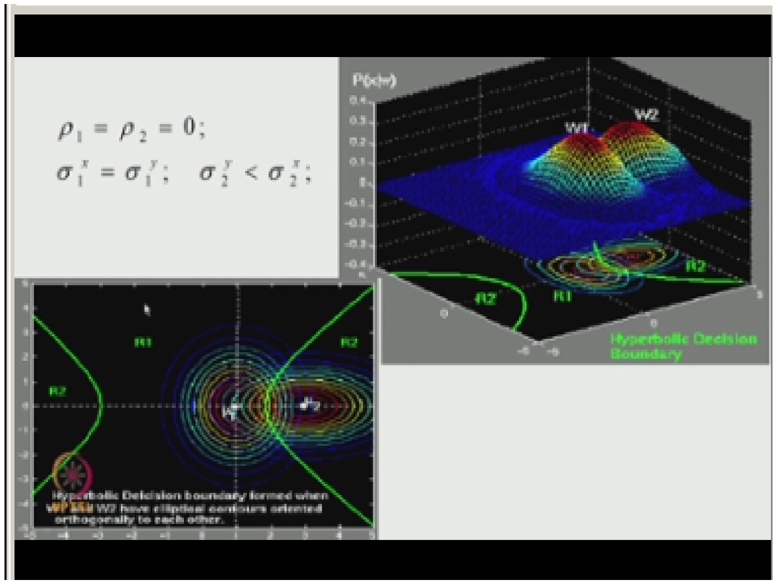


These are showing examples look at the figure of the bottom this is similar almost whatever is drawing so you have class means for two classes here and you have while this seems to be diagonal cohesive matrix this is x and y because these are not worrying take cautions these are just ellipse as you can see here in 3 d the corresponding is concurrent lines is given here and look at the parameter boundary here for the two cohesive matches.

This particular space there is an another example of Gaussian distribution functions the two class means look at the two is now you have case of oriented one of them is oriented and the other is not so it is possible that you have half diagonal terms which is 0 here for class one here you have the half diagonal term which are non 0 for class two because it is titled this is just oriented this is sorry I should be corrected.

This is called asymmetric Gaussian as an ellipse both ellipse mind as a control lines are ellipse but this is a asymmetric Gaussian this is a Gaussian oriented Gaussian sort of these if you look at these two both are ellipses okay and but they are not tiled okay they are just ellipse one of the major axis so in this particular cases also you will have oriented decision of boundary has given here you will not have non linear boundary.

(Refer Slide Time: 46:47)



So let us look at this example of non linear decision boundaries of a two class problem in two dimension the Gaussian function show the corresponding distribution function of the two different classes which appears like the mixture of Gaussians which belong two different classes you look at the bottom of the signal lines are shown two class means the covariance matrix is strictly diagonal with half diagonal terms equal to 0.

Which is 2x2 matrix since it is in 2 d let us look at the case for the class one the corresponding diagonal terms are same giving us circular signal lines which would have been perching 3 d or hyper spheres in dimension let us look at the distribution of the cohesion matrix as given for the class two and here the variances along the x direction or the first dimension is more than that of the second indicate the class id which is class two.

So this is for the class one where the variances are same here in case of class two the variances along x2 is more than standard deviation strictly speaking we can take them to variances well I taken this square x is both on y so this is nice scenario where you can see that you have two discriminating the decision boundary which create what are called a decision of the hyper polar decision boundary in 2 d they create a hyper polar surface or hyperbolic surfaces in 3 d.

As dimensions as well okay so if you here in between you have the class are r1 but if you here or there you need to assign for class r2 okay so this is the g1 this is the corresponding g2 g1-g2 will give you this short again simple case the variance for class one in x direction which is this one is same as that of y for the second class that I repeat again the variance along the diagonal terms for class 1 along x direction is same as for the second class in y direction.

So assuming this is x is the first direction and y is the second direction look at the other one here e variance along y for the first class is same as the variance along x for the second class the half diagonal terms at 0 indicated it is the diagonal covariance matrix and look at the means the they are shifted along x by this thing so basically if you look at the positions of the class means the differences in x and y directions have to be same to create what are basically called sort of decision boundaries which are asymmetric cases of ellipses okay.

And all these individual iso- lines are elliptical for both class one and class two but they are in different direction that means they are oriented orthogonal to each other and look at the two decision boundary which is the same which we had earlier for the case of ellipse and so we are in these region class I and here assign to class two okay so this a case of special case of hyperbolic decision boundary when this true corresponding boundary is come very close.

And the decision boundary which you observe here is a special case of an hyperbola called the rectangular hyperbola so here the boundaries appear to be linear but it is a very special geometrical case is of hyperbolic non linear decision boundaries which into this particular form of two intersecting lines or Para intersecting lines which is called rectangular hyperbola as we stop here the discussion on discrete function bored.

From the concepts of the based decision rule base theorem and decision rule the base theorem and the density function which gave us the manner decision criteria and from there we saw two crisis which provides us with the linear decision boundaries the last case the arbitrary sigma for different classes.

They are not the same anymore class wise provides non linear boundaries in general that is what we have we showed the samples into 2 d and we have to visualize now things will happen in 3 dimension and high dimension thank you very much.

Online Video Editing /Post Production

K.R.Mahendra Babu
Soju Francis
S.Pradeepa
S.Subash

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

Copyrights Reserved