Deep Learning
Prof. Prabir Kumar Biswas
Department of Electronics and Electrical Communication Engineering
Indian Institute of Technology, Kharagpur

Lecture – 08
Discriminant Function – III

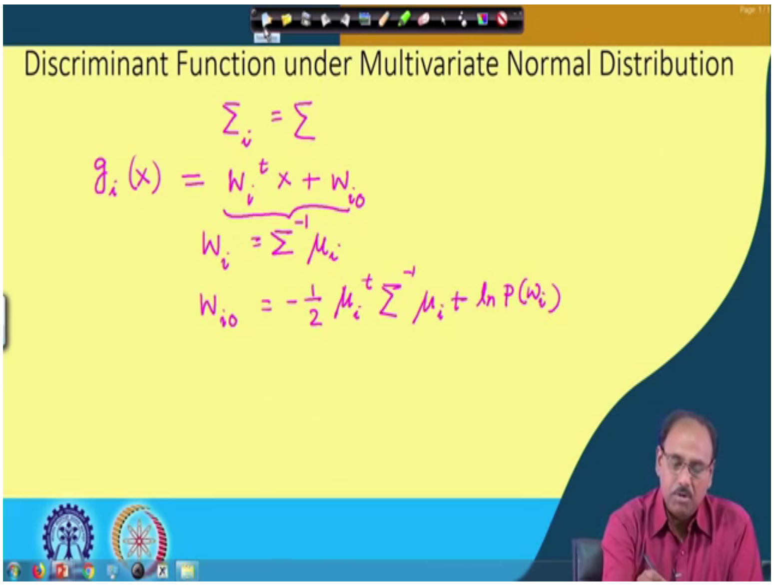
Hello welcome to the NPTEL online certification course on Deep Learning. We are discussing about the Discriminant Function and the decision boundary among different classes. So, in the previous class we have considered two simple cases or the covariance matrices of the different classes they are same. And, in one of the case we have assumed that the covariance matrix is of the form sigma square i, where sigma is the variance of all the components of the vectors. And, i is an unit matrix which indicates that the covariance matrix is diagonal, all the off diagonal elements are 0 and all the diagonal elements are same.

And, that is the case of distribution of points or the points are spherically distributed or hyper spherically distributed. And, in the second case we have assumed that sigma or the covariance matrix of all the classes are same, but the covariance matrix need not be of the form sigma square i; that means, I also have off diagonal elements which are non-zero. And, this is a case of distribution of points where the vectors are distributed in ellipsoidal fashion or hyper ellipsoidal fashion.

(Refer Slide Time: 01:56)

So, what we have discussed in the previous class is the decision boundary under various cases of covariance matrices and we are going to continue with the same discussion in this class with few more examples.

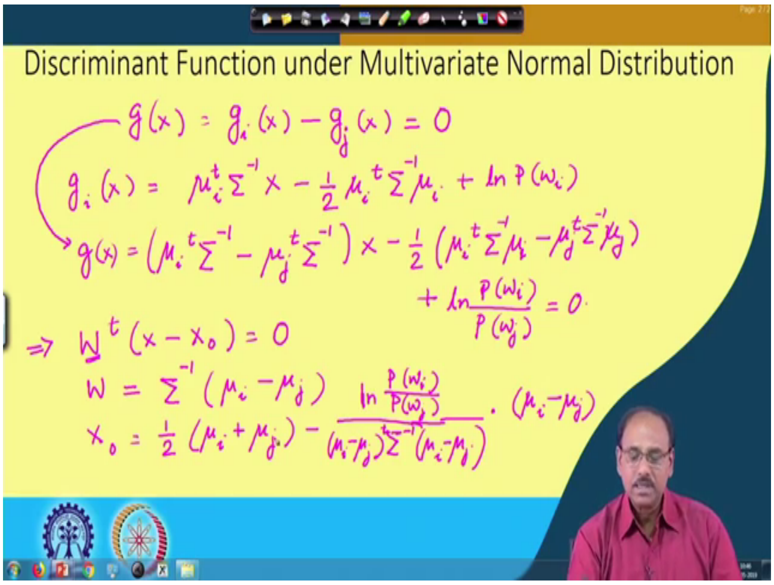
(Refer Slide Time: 02:11)



So, where we stopped in the previous lecture is we have assumed the case where the covariance matrix sigma i is equal to sigma ; that means, the covariance matrix of all the classes are same. But, here the covariance matrices need not be diagonal only and given this case we have in the previous class computed that g i X or the covariance the distant function of the ith class is given by W i transpose X plus W i naught. Where, we have seen that this W i is of the form sigma inverse mu i and W i naught was simply given by minus half mu i transpose sigma inverse mu i plus log of P of omega i, where P of omega i is the a priori probability of class omega i.

So, given this again as we have done in the other case I can compute the decision yeah. So, here you find that because this g i X is of the form W i transpose X plus W i naught; again this discriminant function is a linear function because I do not have any quadratic term in X in this expression. So, the discriminate function here again is a linear function. So, given such a discriminant functions as we have computed in the previous case, I can compute the decision boundary between the 2 classes omega i and omega j. So, what will be the decision boundary?

(Refer Slide Time: 04:20)



The decision boundary here again will be given by g X is equal to g i X minus g j X and because, on the boundary g i X and g j X they are equal so, I will have g i X minus g j X equal to 0. And, you find that what was our expression for g i X? g i X was of the form sigma i mu i sigma mu i transpose sigma inverse X minus half mu i transpose sigma inverse mu i plus log of P of omega i. Similarly, when I compute g j X, g j X will be mu j transpose sigma inverse X minus half of mu j transpose sigma inverse mu j plus log of P of omega j.

So, if I simply subtract this g i X minus g j X and equate that to 0 my expression will simply become, when I compute here g X which is nothing, but g i X minus g j X equal to 0. That simply becomes mu i transpose sigma inverse minus mu j transpose sigma inverse X of this minus half of mu i transpose sigma inverse mu i minus mu j transpose sigma inverse mu j plus log of P omega i upon P omega j that will be equal to 0 ok.

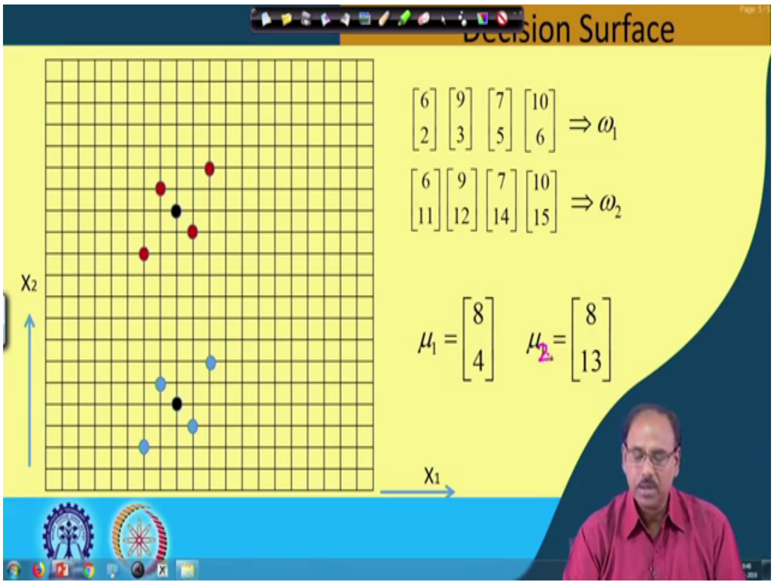
And, if you simplify in the same way that you have done earlier, you find that this will lead to an equation of the form W transpose X minus X naught which is equal to 0. Where, this W in this case will be given by sigma inverse mu i minus mu j and X naught will be given by half of mu i plus mu j minus 1 upon mu i minus mu j transpose sigma inverse mu i minus mu j. Here it will be log of P omega i upon P omega j into mu i minus mu j.

So, what does it indicate? In indicates as before that because the equation of the boundary, the decision boundary between the classes omega 1 and omega g is of the form W transpose X minus X naught equal to 0. And, if you remember this form is similar to what we have obtained in the previous case where, the covariance matrix was a diagonal matrix for all the classes. But what is the difference? In the previous case this W was simply mu i minus mu j; that means, W was the vector drawn from mu i to mu j, but in this case W is sigma inverse mu i and j mu i minus mu j.

So, which means that W is no longer in the direction of the vector from mu i to mu j, but the direction of W depends upon the covariance matrix sigma, because the expression is sigma inverse mu i minus mu j. X naught as before if I assume P omega i and P omega j to be equal that is the classes to be equally a priori, in that case sigma naught as before becomes half of mu i plus mu j ok. So, this expression simply becomes again my decision surface, the decision boundary between the classes omega i and omega j is orthogonal to W, but unlike in the previous case it is not orthogonal to the line joining mu i and mu j right.

So, it is orthogonal to W, but under the situation that P omega i and P omega j to be equal, the line becomes a bisector of the line. The decision surface becomes a bisector of the line joining mu i and mu j, because when P omega i and P omega j they are equal X naught is half of mu i plus mu j. So, it is halfway between mu i and mu j. So, my decision surface is a bisector of the line joining mu i and mu j, but it may not be an orthogonal bisector because, W is no longer in the direction of the line joining mu i and mu j in general. So, let us see that what will be the nature of the decision surface in this particular case.

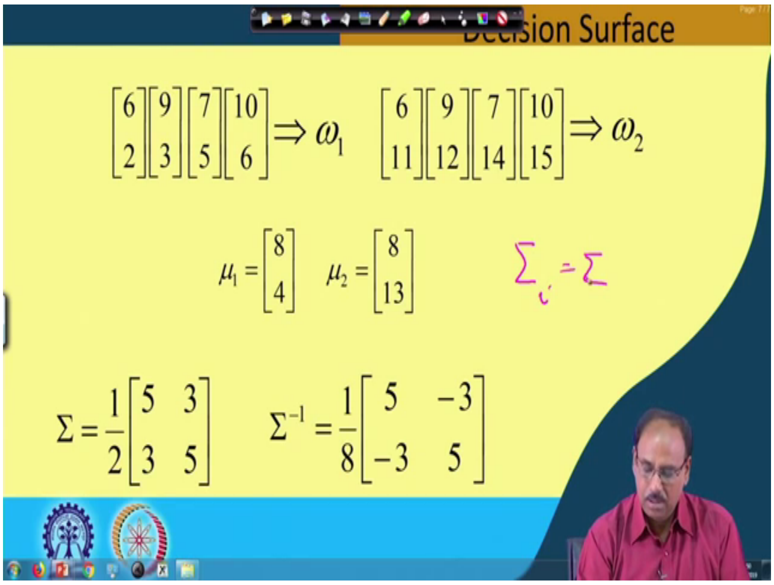
(Refer Slide Time: 10:56)



So, again for that I take a number of examples, number of feature vectors from class omega 1 and from class omega 2. So, in this case the feature vectors that I am taking from class omega 1 are all these feature vectors which are marked in blue. So, these are the feature vectors that I am taking from class omega 1 which are 6 2 9 3 7 5 and 10 6. Similarly, all these feature vectors which are in red they are taken from class omega 2, the feature vectors are 6 11 9 12 7 14 and 10 15. So, here you find that unlike in the previous case what the state of feature vectors are spherically distributed, in this case they are elliptically distributed; they are not spherical distribution anymore

So, given these feature vectors now let us see that how we can find out the decision surface. So, for these two sets of feature vectors I compute the mean vectors mu 1 for class omega 1 and also I compute mu 2 for class omega 2. Again there is a mistake, the second one here this mu 1 this actually should be mu 2, there is not mu 1, but it is mu 2. So, mu 1 is equal to 8 4 and mu 2 is equal to 8 13, they are the mean vectors of the feature vectors taken from class omega 1 and class omega 2.

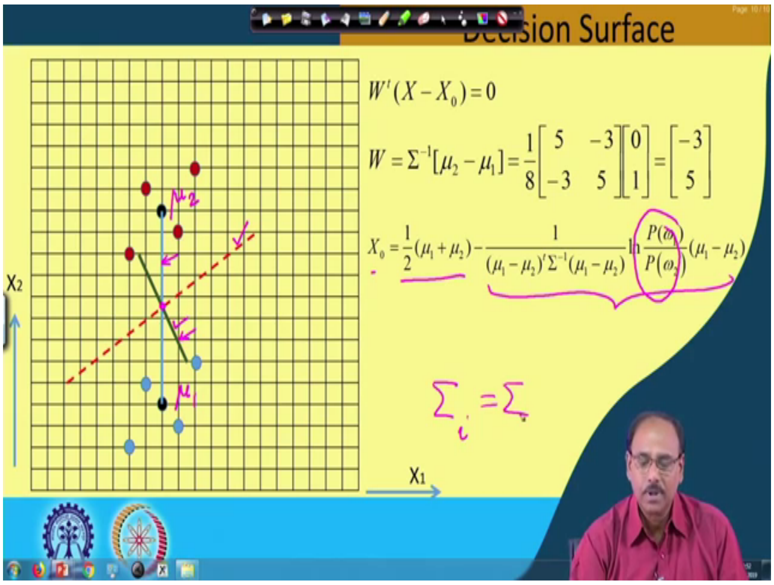
(Refer Slide Time: 12:51)



So, once I have this then again as before I can compute the covariance matrices. So, as we have computed in the previous lecture, in the same manner if I compute the covariance matrix you will find that a covariance matrix for both the classes omega 1 and omega 2 will be half into 5 3 3 5. So, this is the covariance matrix that you get for both the classes omega 1 and omega 2. And so, this is a case where at my off diagonal elements they are not a 0 anymore. So, unlike in the previous case where, I had the covariance matrix which was completely a diagonal matrix and all the diagonal elements were equal; in this case I have off diagonal elements which are non-zero.

So that means, the feature components the different components are not statistically independent anymore. However, for both the classes I have the same covariance matrix that is half 5 3 3 5. So, this comes under the case sigma i is equal to sigma. So, once I have this covariance matrix, I can compute sigma inverse that is the covariance matrix inverse ok.

(Refer Slide Time: 14:15)

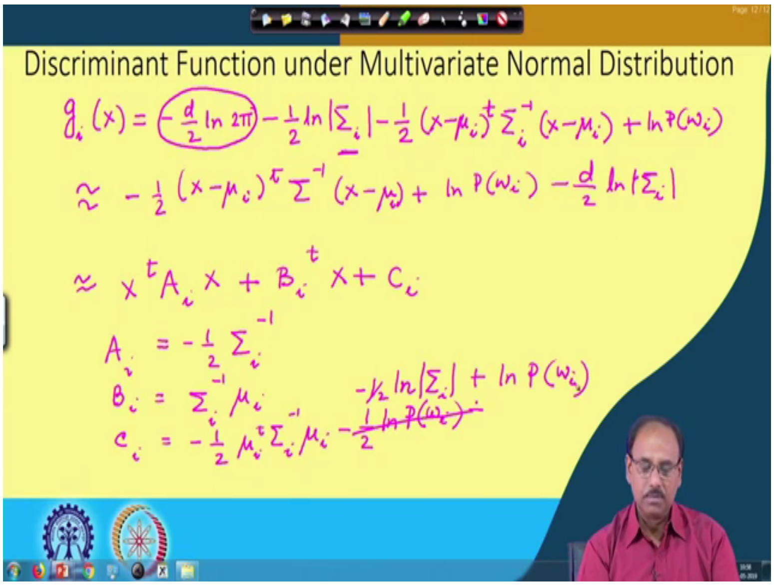


And, using this as we have seen earlier that I can compute the decision surface and the decision surface is given by W transpose X minus X naught equal to 0 where, W is sigma inverse mu 2 minus mu 1. And, you remember the sigma inverse that we have computed was 1 by 8 5 minus 3 minus 3 5 and mu 2 minus mu 1 is nothing, but 0 1 so, I get W which is minus 3 5. So, that simply says that the W is in the direction of minus 3 5, W will be in this particular direction. So, here you find that this dark green line, this line represents in the direction of W ok.

And, given this I can also compute what is X naught as X naught is given by minus X naught is given by half mu 1 plus mu 2 minus this. And, under the situation if I assume that P omega 1 and P omega 2 to be equal log of P omega 1 upon P omega 2 that will be equal to 0; that means, this term will be cancelled giving X naught to be half of mu 1 plus mu 2. So, that simply says that our decision boundary which will be orthogonal to sigma inverse mu 2 minus mu 1 or mu 1 minus mu 2 and it will pass through the midpoint between mu 1 and mu 2. So, this dark blue line is the line joining mu 1 and mu 2, dark a green line is sigma inverse mu 1 and [vocalized- noise] mu 1 minus mu 2.

And, this dotted red line is the line which is orthogonal to this dark green line and it passes through X naught where, X naught is half of mu 1 plus mu 2. So, this is mu 1 and this is mu 2 ok. So, here you find that the decision boundary in this case is also linear, but the decision boundary is no longer orthogonal to the line joining mu 1 and mu 2. So, this was our second case where sigma i was equal to sigma. Now, let us consider that what will be the discriminant function and the decision boundary in the most general case.

(Refer Slide Time: 16:57)



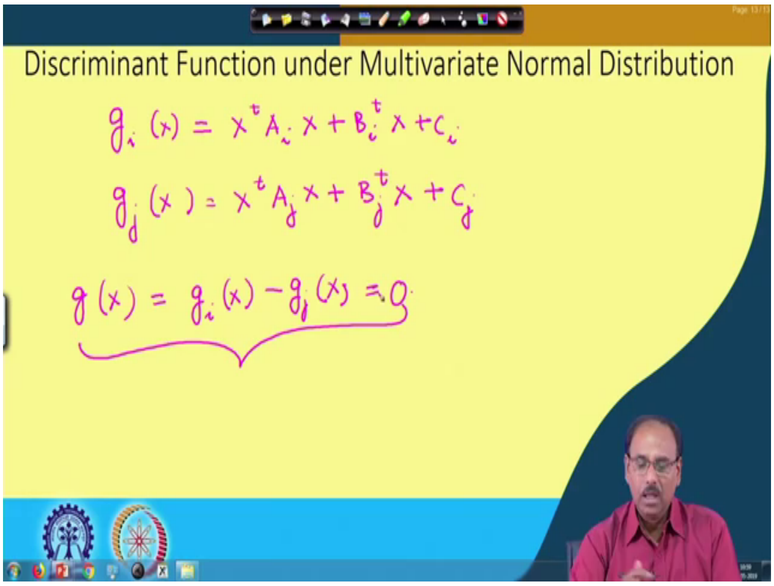
Where, I have sigma i to be the most general one, that is for different classes I can have different sigma y. It is possible that for some of the classes the sigma i will be orthogonal, but some of the some other classes sigma i can be diagonal for, but for some other classes it may not be diagonal. So, every class or the sample vectors taken from every class they have their own covariance matrix sigma i. So, to see that what will be the nature of the discriminant function in such case, let us go back to the original g i X where, we have seen that g i X was minus d by 2 log of 2 pi minus half log of determinant sigma i minus half X minus mu i transpose sigma i inverse X minus mu i plus log of P omega i.

So, here you find that as before I can remove this term from g i X because this is same for all values of i whereas, in the earlier cases we had ignored this term as well half log determinant sigma i because, sigma i was same for all the classes. Now, sigma i is not same for all the classes it is different for different classes so, I cannot ignore this coming more. So, that gives me that g i X will be simply minus half X minus mu i transpose sigma inverse X minus mu i plus log of P of omega i minus d by 2 log of mod of determinant sigma i right.

So, given this you will find that this expression is of the form X transpose A i X plus B i transpose X plus C i. I can simplify this expression in this form X transpose A i X plus B i transpose X plus C i. And, now find that because of the presence of that term X transpose X I will have if the components are X 1 X 2 X 3 and so on of the feature vector X; I will have terms X 1 square, I will have terms X 2 square, I will have term X 3 square, I will have term X 1 into X 2 X 1 into X 3 X 2 into X 3 X 2 into X 5 and so on. So, that leads to a situation that the discriminant function g i X does not remain linear anymore, but it becomes a quadratic discriminant function because of the presence of quadratic terms.

And, over here this A i it will be simply minus half sigma i inverse and B i will be sigma i inverse mu i and C i will be minus half mu i transpose sigma i inverse mu i minus half log of P of omega i plus sorry, this is minus half log of determinant sigma i plus log of P of omega i. So, C i will be minus half mu i transpose sigma inverse mu i minus half log of determinant sigma i plus log of P of omega i. So, my discriminant function becomes a non-linear or quadratic discriminant function.

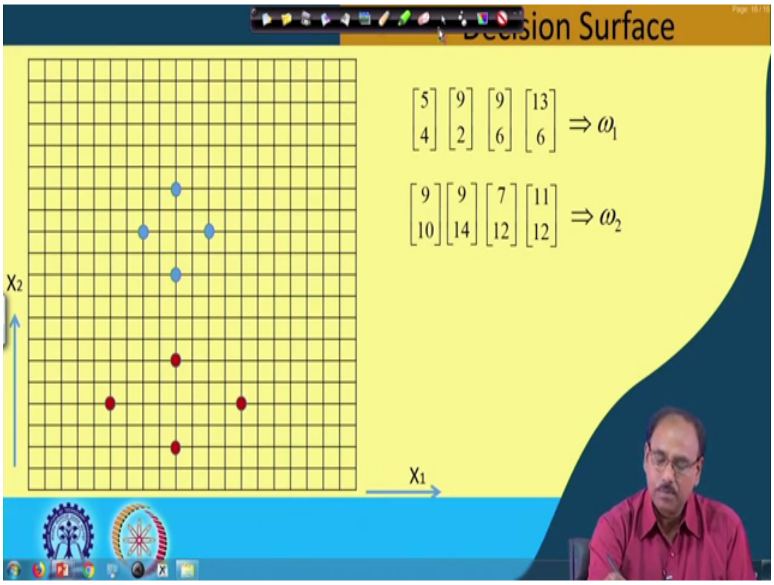
(Refer Slide Time: 22:15)



And, when I have such non-linear discriminant functions for to find out the decision boundary between 2 classes omega i and omega j. Now, it is difficult to simplify as we have done in the previous case unlike in the previous case. So now, what I have to do is I have to find out what is g i X which will be as we have seen, it will be X transpose A i X plus B i transpose X plus C i. Similarly, I have to compute what is g j X which will be X transpose A j X plus B j transpose X plus C j.

And, to compute g X which is the decision boundary between the 2 classes I have to compute g j X which is nothing, but g i X minus g j X and I have to equate this to equal to 0. So, whatever expression I get that will be the decision boundary between the classes omega i and omega j. So, now let us see what will be the decision boundary in this kind of scenario.

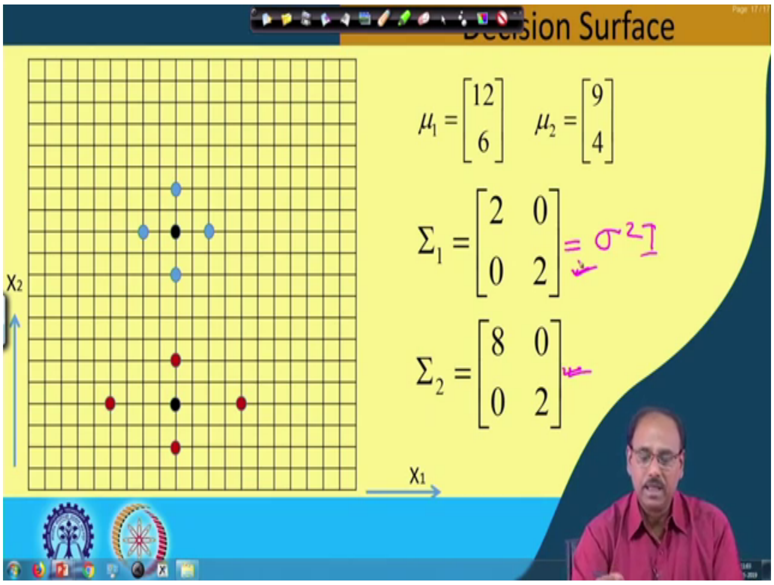
(Refer Slide Time: 23:33)



. So, for this again I take a set of feature vectors. So, a set of feature vectors say 9 10 9 14 7 12 and 11 12 which are taken from class omega 2, that is this one. And, I have a set of feature vectors 5 4 9 2 9 6 and 13 6 which are taken from class omega 1 and the feature vectors belonging to class omega 1 at these points right.

So, as before as we have done before given these feature vectors I can compute what is mu 1 and what is mu 2 that is the mean of feature vectors taken from class omega 1 and mean of feature vectors taken from class omega 2. And, I can also compute the covariance matrix sigma 1 and sigma 2 for this two different classes.

(Refer Slide Time: 24:34)

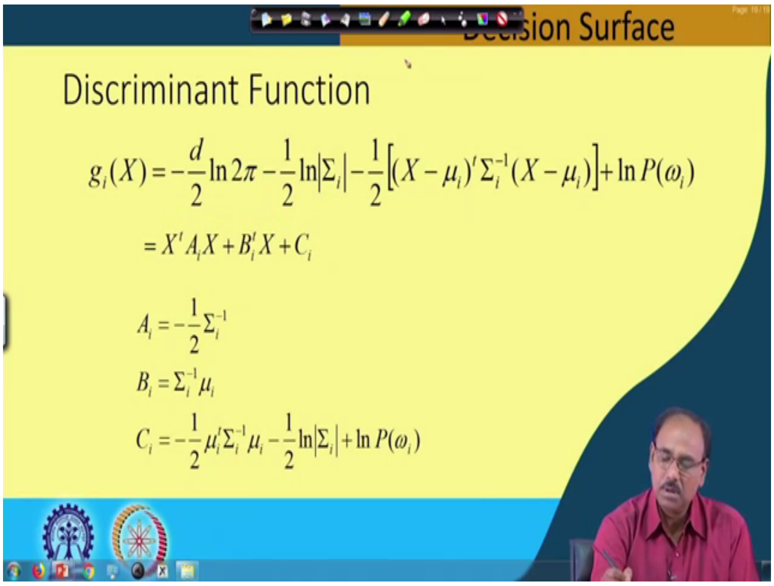


So, what here my mu 1 that is the mean of the feature vectors from class omega 1 will be 12 6 and mu 2 will be 9 4 and the covariance matrix sigma 1 is 2 0 0 2. So, you find that this 2 0 0 2 is the for variance matrix for these feature vectors, I think as we have doing our doing earlier this sigma 2 and sigma 1 should actually be sigma 2; anyway that does not matter much. And, for the other class your sigma or the covariance matrix is 8 0 0 2 which is the covariance matrix over here right.

So, you find that in the first case, in this case the points are spiritually distributed and here the covariance matrix is of the form 2 0 0 2 which is nothing, but of the form sigma square right. So, the points are spherically distributed whereas, in the second case while my covariance matrix sigma 2 is 8 0 0 2, it is elliptical distributed. And, because the variance of the component X 1 component is more than the variance of the X 2 component so; obviously, the spread in X 1 direction is more than the spread in X 2 direction and that makes it elliptical. So, given this to a set of feature points which are vectors now as I find that the covariance matrix for the 2 classes are different.

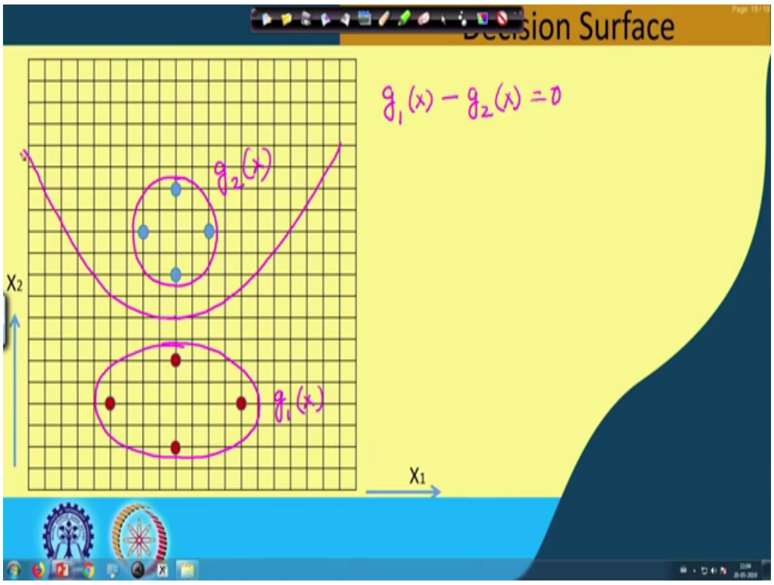
So, my discriminant function for the 2 classes for the second class, where covariance matrix is 8 0 0 2 will be a quadratic one whereas, the discriminant function for the first case where, sigma 1 that is this one because it is of the form sigma square i the the discriminant function for this class will be linear. So, over here my discriminant function will be quadratic for this the discriminant function will be linear. So, to find out the decision surface between the 2 classes over here I simply have to make g i X minus g j X and equate that to 0. And, after equating that to 0 whatever I get that becomes the decision surface.

(Refer Slide Time: 27:07)



You find that as we have done before that this is the expression of the covariance of the discriminant function or the covariance matrix is not of the form i square i right. And, the discriminant function is a quadratic discriminant function and given this particular situation if I try to find out that what will be that decision boundary between the 2 classes.

(Refer Slide Time: 27:30)



So, as we said that for this set of points the distant function will be linear one, for this set of points the discriminant function will be a quadratic one. And, if I if I call it say g 1 X and if I call it g 2 X the discriminate functions the equation of the boundary will be given by g 1 X minus g 2 X that will be equal to 0. And, if you plot the decision boundary you will find that the decision boundary will have a shape something like this which is not a linear decision boundary anymore that we have obtained earlier. But, in this case the decision boundary will be a non-linear decision boundary or a quadratic decision boundary well.

So, with this I come to the end of this lecture. So, in this lecture and the previous two lectures what we have tried to do is, we have tried to find out the discriminant functions of different classes assuming that the distribution of points is multivariate normal distribution. And, there we have taken three different cases, in the first case we have assumed that the covariance matrices for all the classes are same and they are of the form sigma square i.

And, we which case we have obtained the discriminant functions to be linear and not only that the decision boundary between these different classes they are also linear. And, under the situation when the a priori probabilities are same we have seen that decision boundary is orthogonal bisector of the line joining mu 1 and mu 2.

In the second case we have assumed that the covariance matrices for different classes are same, but they may not be of simple forms in sigma square i. In that case also the discriminate functions we have found to be linear, decision boundary was also linear. Decision boundary was a bisector of the line joining mu 1 and mu 2 under the situation when a priori probabilities are same, but the decision boundary in general was not orthogonal to the line joining me 1 and mu 2 because, of the presence of the term sigma inverse. So, that direction of the decision boundary in that case depends upon the variance matrix.

And, the third case was the more general case, where we assumed that the covariance matrix of all the classes are different. And, in which case we have found that the discriminant function is not linear anymore, the discriminant function is a quadratic discriminant function. And, given such quadratic the decision boundary to be the discriminant function to be a quadratic one using that, if I try to find out the separating boundary between the 2 classes; the separating boundary in general is quadratic, it is not linear anymore. I will stop here today.

Thank you very much.