Pattern Recognition and Applications

Prof. P. K. Biswas
Department of Electronics and Electrical Communication Engineering
Indian Institute of Technology, Kharagpur

Lecture - 18
Multiple Discriminant Analysis
(Tutorial)

Hello. So, for last two three classes, we have discussed about the problem of dimensionality of the feature vectors. We have seen that in case of pattern recognition, depending upon the complexity of the patterns that we want to classify or we want to recognize, we have to capture different types of features from the patterns. The number of features in some cases can be very large it can be two features, it can be three features it can be ten features, it can be hundred features or it can be even thousands of features.

So, we have seen through our analytical analysis, that as the dimension of the feature vector increases, the training complexity of the time required to train the classifiers also increases to a large extent. So, it is always better that if we can reduce the feature dimensions. So, we have said we have discussed earlier that one way to reduce the dimension of the feature vectors is to project the feature vectors into lower dimensional space.

And when I project a higher dimensional feature vector onto a lower dimensional space, then there is always a risk that if I have 2 different classes, say omega 1 and omega 2. Say higher dimensional space in n dimensional space, if the feature vectors belonging to omega 1, and the feature vectors belonging to omega 2, they are separable when I reduce the number of dimensions, it is quite possible that in the reduced dimension the feature vectors will not remain separable anymore.

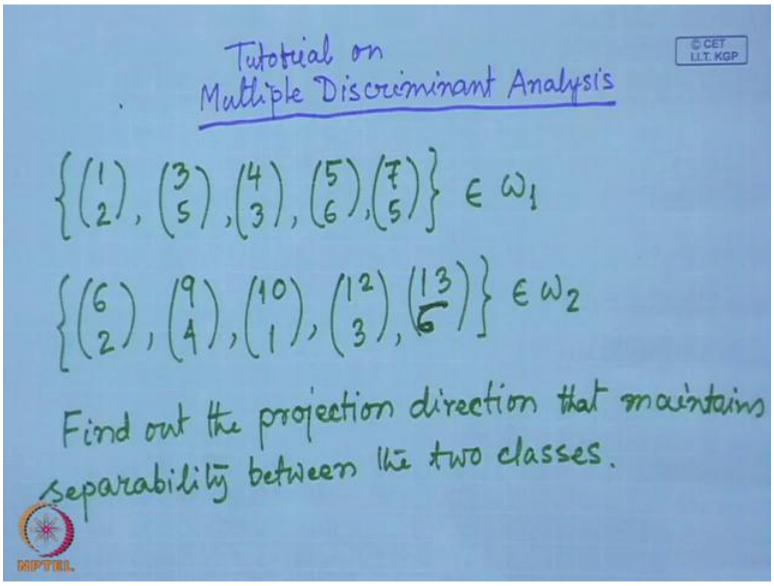
The other thing is that what should be the best projection direction or what is the best projection space. So, we have seen that. So, far as data representation is concerned in a lower dimensional space, the best way of projection is to project the higher dimensional feature vectors onto the Eigen vectors as computed from the feature vectors. We have also seen that if I simply project the higher dimensional feature vectors onto Eigen vectors or onto the Eigen space, then the separability among the feature vectors may be lost.

The Feature vectors belonging to different classes may not remain separable anymore. So, for that we have gone for features linear discriminator. So, in case of features linear discriminator, we have seen that we try to maximize the between classes scatter and at the same time we try to minimize the within class scatter. That means when you project the higher dimensional feature vectors onto a lower dimensional space, we try to find out the projection space in such a way that the clusters or the patterns belonging to different classes they will be as apart as possible. Simultaneously, the patterns belonging to the same class they will be as compact as possible.

That is what we mean by it tries to maximize the between class scatter and it tries to minimize ah the within class scatter. Accordingly we have computed different projection directions. So, in the first case when simply data representation is our aim onto a lower dimensional space the projection directions are actually the Eigen vectors. This simply Eigen vector projecting onto the Eigen vector may not maintain the separability. So, which is not very suitable for classification purpose because in case of classification it is not only data representation onto the lower dimensional space, but it is also the separability of the data belonging to different classes that is very important.

So, for classification our aim is that we want to find out the projection directions, which maximizes the between class scatter and it tries to minimize the within class scatter. So, today we will take up an example, we will try to solve a problem which will illustrate this fact that if I take projection onto an Eigen vector, the separability may not be maintained, but if I project onto another space following the features linear discriminator the separability of the data may be maintained.

(Refer Slide Time: 05:30)



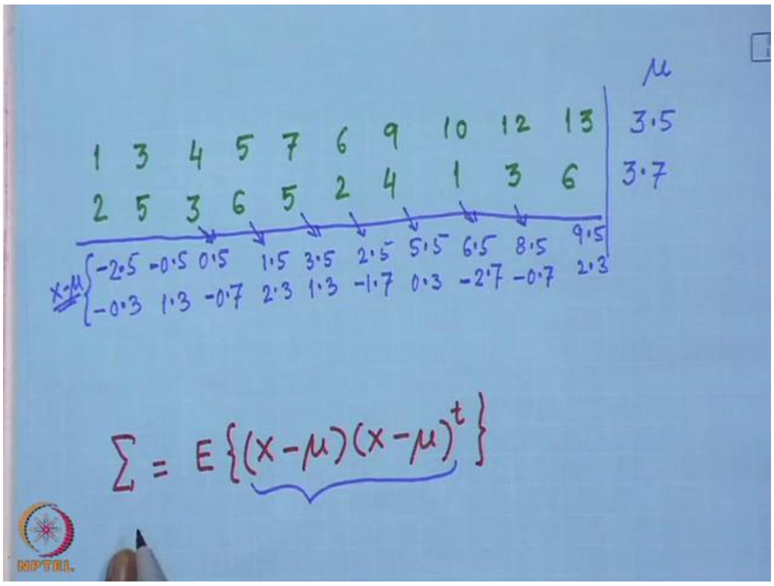
So, we are going to have today a tutorial on multiple discriminant analysis. So, we will take a number of feature vectors from 2 different classes, a set of feature vectors from class omega 1 another set of feature vectors from class omega 2. So, suppose we have been given some feature vectors, let us take them let feature vectors say, 1 2 then 3 5 then say 4 3 then say 5 6 and say 7 5. So, we are considering two dimensional feature vectors. So, suppose these are the five feature vectors which belong to class omega one.

So, these are the feature vectors which are taken from class omega 1. We have another set of feature vectors 6 2, let us take 9 4 then say 10 1 then say 12 3 and say 13 6. Say, these are the feature vectors which are taken from class omega 2. We want to find out the projection direction. So, that when these feature vectors are projected onto a lower dimensional space projected on to a line, we want to find the direction of the line. So, that when this feature vectors are projected onto that line the separability of the feature vectors between the two classes omega 1 and omega 2 that will be maintained.

So, what we want to do is find out the projection direction that maintains separability between the two classes. Of course, what we will do is before finding out such a projection direction which maintains the separability, we will find out the best projection direction. So, far as data representation into a lower dimensional spaces concerned and then we will solve this problem. We will find out the projection direction that will maintain the separability. So, that we can compare the performance of the two. So, that we can demonstrate that in one case the separability is maintained and in the other case the separability is not maintained. So, we have these four five feature vectors from class omega 1 and we have the 5 feature vectors from class omega 2 right.

So, first if I want to find out the projection direction for best representation, then what we have to do is we have to find out the Eigen vectors of the co-variance matrix of these data elements. So, let us see how we can find out the co-variance matrix first and from the co-variance matrix we have to find out the Eigen vector corresponding to the maximum Eigen value. So, first let us find out the co-variance matrix. So, to find out the best projection direction, we have to consider all these feature vectors together forgetting about their class belongings

(Refer Slide Time: 10:30)



So, we have the feature vectors 1 2 3 5 4 3 5 6 and 7 5 from class omega 1 and 6 2 9 4 10 1 then 12 3 and 13 6. These are the feature vectors from class omega 2 and considering all these feature vectors I have to find out co-variance matrix. So, to find the co-variance matrix first what I have to do is I have to find out the mean vector mu. This mean vector is nothing but mean of all these feature vectors. So, when I try to find out the mean of all these feature vectors, I have to find out the mean of the first component I also have to find out the mean of the second component. The mean of the first component and mean of the second component these two taken together gives you the mean vector.

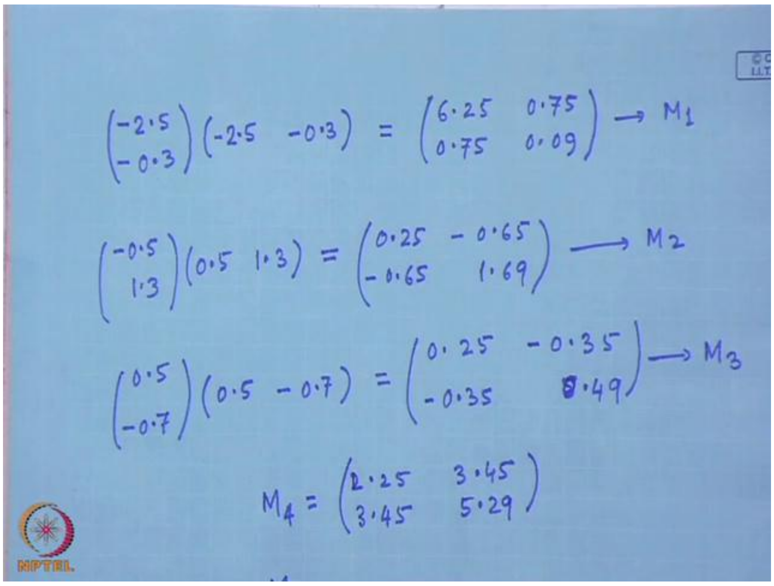
So, when I compute mean of all these ten feature vectors the first components of all this ten feature vectors this mean you can compute. This will come out to be something like say 3.5. The mean of all the second component of all the ten feature vectors that will come out to be something like 3.7. So, this is my mean vector mu. As you know that when I compute the co-variance matrix, the co-variance matrix is defined as sigma is equal to expectation value of x minus mu into x minus mu transpose, where this x where are individual feature vectors and this mu is the mean vector.

So, I have to compute x minus mu. So, from every feature vector x i have to subtract the mean vector which is mu. So, when I subtract the mean vector from every feature vector what I have is, let me put here I will put x minus mu for individual feature vectors. So, I will put x minus mu here. So, here we find that, when I subtract this mean vector mu from the first feature vector which is 1 2, this will simply be minus 2.5 minus 0.3. So, this is the first feature vector minus the mean vector. For the second feature vector it will be minus 0.5 and 1.3. For the third vector it will be 0.5 and minus 0.7. For the fourth vector it will be 1.5 and this will be 2.3, for this vector it will be 3.5 and 1.3. So, these are x minus mu for all the feature vectors belonging to class omega 1.

Now, coming to other set of feature vectors this is 6 minus 3.5 which will be 2.5 and 2 minus 3.7 that will be minus 1.7. Similarly, for this is it 9 minus 3.5 which will be 5.5 and this is 4 minus 3.7 which is 0.3. Similarly, for this it will be 10 minus 3.5 which is 6.5 and 1 minus 3.7 this is minus 2.7. For this vector it will be 12 minus 3.5 which will be 8.5 and this is 3 minus 3.7 obviously it is minus 0.7. For the last one it will be 9.5 and 2.3. So, this is these two row’s actually give me the feature vectors, x minus mu or x minus mu have been computed by subtracting the mean vector from the individual vectors.

Then for each of these vectors after subtraction of the mean vector I have to compute the x minus mu into x minus mu transpose. Then I have to take the average of all those matrices, which I get that gives me the expectation value of x minus mu into x minus mu transpose which is nothing but my co-variance matrix. So, if I do that for therefore the first one, you find that the first one is minus 2.5 and minus 0.3.

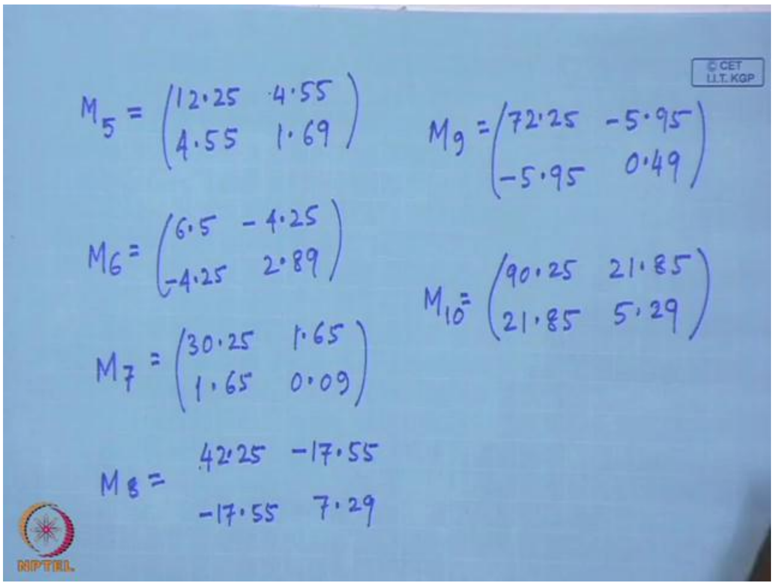
(Refer Slide Time: 17:15)



So, what I have to compute is minus 2.5 minus 0.3, this is my x minus mu and x minus mu transpose will be minus 2.5 minus 0.3. So, this is a column vector which is of dimension 2 by 1 and this is a row vector of dimension 1 by 2, obviously I can do matrix multiplication. The product matrix that the resultant matrix will be of dimensional 2 by 2. If I simply compute this you find that the value of this matrix will be 6.25, 0.75, 0.75 and then 0.09. So, this is one of the matrices similarly, for the second 1 which is minus 0.5, 1.3, 0.5, 1.3 when I do this multiplication this matrix would be 0.25 minus 0.65 minus 0.65. This one will be 1.69, for the third one which is 0.5 minus 0.7 into the row vector 0.5 minus 0.7, this will be simply 0.25 minus 0.35 minus 0.35. And this one would be 0.4 9.

So, this is my third matrix and this way if I compute all the matrices, once I have said how to compute individual matrices x minus mu into x minus mu transpose, this can be done for all the vectors x minus mu and I can get similar such matrices. So, as I have 10 different vectors I will have 10 such matrices. So, let us put those 10 matrices one by one. So, this is my matrix one m1, this is my matrix m2, this is my matrix m3. Similarly, m4 with the fourth vector, that will be simply 12 yes. So, that will be 2.25, 3.45 then 3.45 5.29.

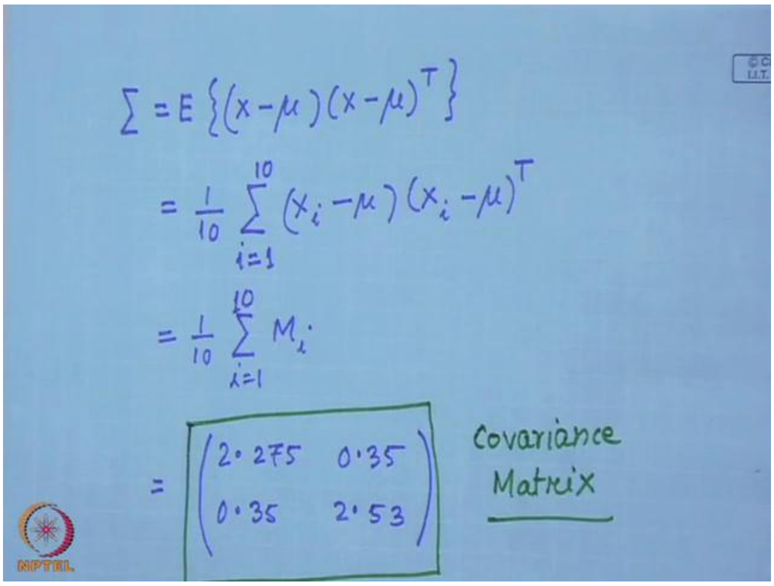
(Refer Slide Time: 20:58)



Fifth matrix will be 12.25, 4.55, the sixth matrix M 6 will be 6.5 minus 4.25 minus 4.25 2.89, M 7 seventh matrix will be 30.25, 1.65, 1.65 0.09, M 8 will be 42.25 minus 17.5 5 minus 17.55, 7.29, m9 will be 72.25 minus 5.95 minus 5.95, 0.49 and the last one, M 10 will be 90.25, 21.85, 21.85 and 5.29 right. So, these are the 10 matrices that we have. So, the first matrix is this, which is x 1 minus mu into x 1 minus mu transpose.

The second one x 2 minus mu into x 2 minus mu transpose that gives this, third one x 3 minus mu into x 3 minus mu transpose that gives this, the fourth x 4 minus mu into x 4 minus mu transpose that gives this. Continuing this way up to the tenth, x 10 minus mu into x 10 minus mu transpose that gives me this. So, I have this total of 10 matrices, now out of this 10 matrices when I combine them I add them and. Divide by the number of matrices that I have what I get is the expectation value

(Refer Slide Time: 24:10)

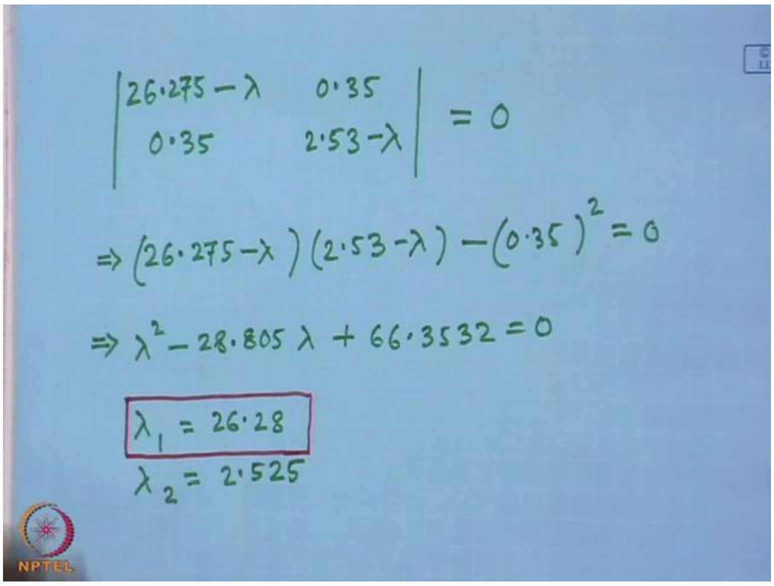


So, as per our definition, the co-variance matrix sigma is equal to expectation value of x minus mu into x minus mu transpose. Over here because we have 10 different vectors. So, this expression will simply be 1 upon 10 into summation. So, here I use this symbol for the summation operation x i is the i th vector minus mu into x i minus mu transpose. Where the summation has to be taken over i is equal to 1 to 10 as I have 10 different feature vectors and accordingly I have 10 different matrices. Each of this x i gives me a matrix mi the one that we have computed here, M 1 M 2 M 3 and so on. So, each of x i minus mu into x i minus mu transpose gives me a matrix M I. So, this expression will simply be 1 upon 10 into summation of M I, I varying from 1 to 10.

So, effectively what I have to do is, I have to add all this 10 matrices, M 1 to M 10 all these different matrices have to be added and then the matrix have to be the summation matrix have to be divided by 10. When I do that, you will find that this simply becomes 26.275, 0.35, 0.35 and 2.53. So, this is my co-variance matrix. So, this is what is the co-variance matrix, now the next step that I have to do is, I have to find out the Eigen vector corresponding to the maximum Eigen value of this co-variance matrices.

So, first let us try to find out what will be the Eigen value or the maximum Eigen value and then let us try to find out what will be the Eigen vector, corresponding to that maximum Eigen value. So, you all know, that if lambda is an Eigen value of this matrix then I can simply find out the Eigen value lambda by using this equation.

(Refer Slide Time: 27:58)

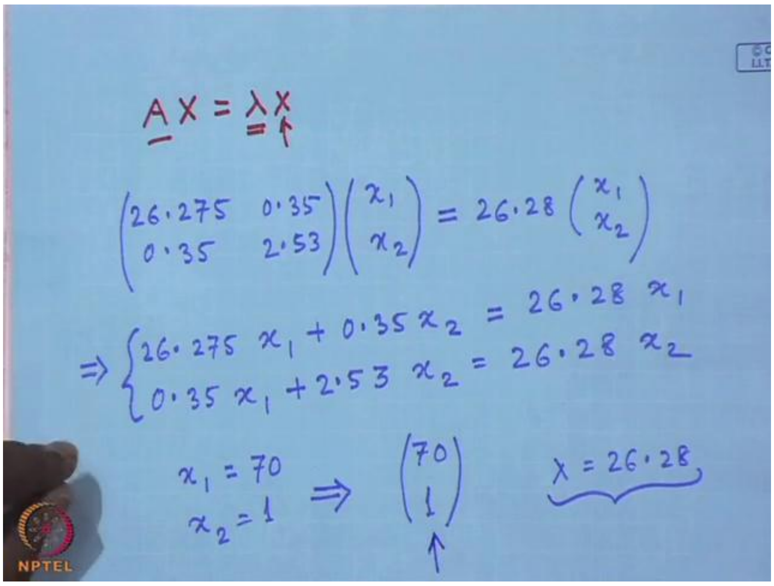


So, the equation that I have to use is 26.275 minus lambda, where lambda is an Eigen value 0.35, 0.35 and then 2 .5 3 minus lambda take this determinant and equate this determinant to 0. You solve this equation, I get the different Eigen values lambda. So, simply this equation leads to 26.275 minus lambda into 2.53 minus lambda minus 0.35 square. This is equal to 0, which is reduced to lambda square minus 28.805 lambda plus 66.3532 is equal to 0. So, here we find that, I get a quadratic equation in terms of lambda, well lambda is an Eigen value of the co-variance matrix.

As this equation is a quadratic equation. So, obviously the lambda will have two solutions, which indicates that this matrix this co-variance matrix will have two Eigen values. Corresponding to each of this Eigen values I have one Eigen vector. So, this co-variance matrix will have two Eigen vectors one corresponding to each Eigen value. So, when I solve this equation this is a quadratic equation the solution is very simple. You can find out that one of the solutions, lambda one will be 26.28 and the other solution lambda two will be 2.525.

So, these are the two Eigen values of my co-variance matrix. Out of these two I have to choose that particular Eigen value which is maximum. So, we find that this Eigen value lambda one which is 26.28, that is the maximum one. So, once I have this maximum Eigen value I have to find out what is the Eigen vector corresponding to the maximum Eigen value. You also know what is this Eigen equation, Eigen value equation which simply says that, if x is an Eigen vector of a matrix a, a is my matrix and suppose x is the Eigen vector then a x has to be is equal to lambda x.

(Refer Slide Time: 31:16)

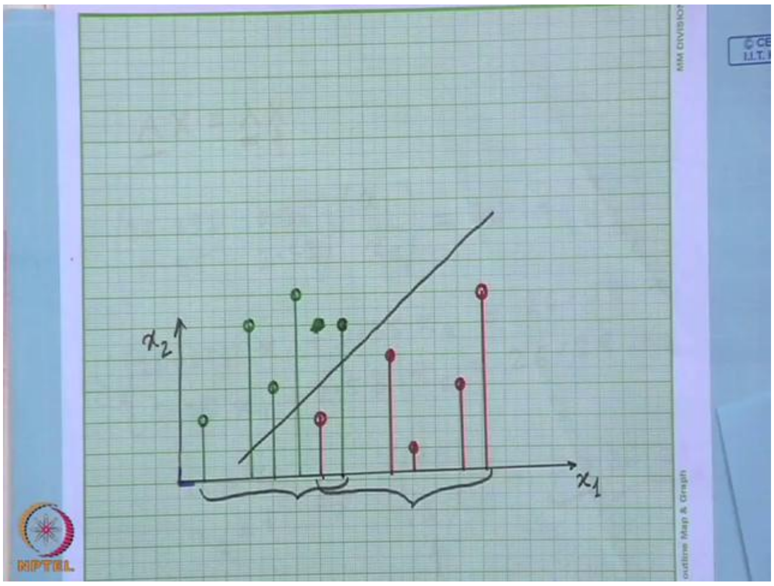


Where this lambda is the Eigen value and x is the Eigen vector of matrix a. So, following the same one here we can find out what is my Eigen vectors corresponding to this maximum Eigen value, lambda which is equal to 26.28. So, let us try to find out the Eigen vector corresponding to this maximum Eigen value. So, in order to do that we know that our co-variance matrix was given by 26.275, 0.35 then 0.35, 2.53. This was my co-variance matrix sigma and if I assume that the Eigen vector corresponding to the maximum Eigen value of this co-variance matrix is x. This being a two dimensional matrix which has got two components x 1 and x 2 right.

So, this has to be equal to lambda times x and where this lambda the maximum Eigen value is 26.28 . So, this has to be simply 26.28 times x 1 x 2. So, if I reduce this equations into linear equations I get two equations, one is 26.275 times x 1 plus 0.35 times x 2 is equal to 26.28 x 1 and the other 1 is 0.35 x 1 plus 2.53 x 2 is equal to 26.2 8 x 2. So, I get these two simultaneous equations. So, I have to find out the value of x 1 and the value of x 2 by solving these two linear equations, simultaneous equations. The solution also is very simple, if you solve this you will find that I will get x 1 is equal to something around 70 and x 2 will be something around 1.

So, that clearly indicates that 70, 1 is the Eigen vector of the co-variance matrix corresponding to the Eigen value lambda is equal to 26.28. So, corresponding to this Eigen value the Eigen vector is 70, 1. What we have to do is we have to project the feature vectors which are given onto this vector 70 1. Now, let us plot this points on a graph a paper to see that what kind of projection we get right. So, I will use a graph paper to this job. So, what we have is data session that we have the origin somewhere over here.

(Refer Slide Time: 35:24)



This is our origin, I will put this different feature vectors which are given onto this graph paper. Over there if you remember that we said these are the feature vectors 1 2 3 5 4 3 5 6 and 7 5 this first 5 feature vectors they belong to class omega 1 and the seconds 5 feature vectors 6 2 9 4 10 1 12 3 and 13 6 they belong to class omega 2. So, when I plot these feature vectors onto my graph paper I will plot them into two different colors. So, the first one I will use the green color. So, the first 1 is 1 2 , 1 2 that comes somewhere over here then 3 5, 3 5 is somewhere over here then 4 3, 4 3 is this, then comes 5 6 over here and then 7 5.

So, these are the feature vectors which belong to class omega 1. Now, let us plot the feature vectors that belong to class omega 2 and I put them into red color. So, the first one is 6 2 from here the feature vectors are 6 2 9 4 10 1 12 3 and 13 6. So, 6 2 1 2 3 4 5 6 2. So, this is one feature vector, this was 5 6 and the next one was 7 5 1 2 3 4 5 6 7. So, this one not here it will come here. So, 6 2 that comes over here then we have 9 4 7 8 9 4. This is another feature vector belonging to class omega 2. Then we have 10 1, this is another feature vector belonging to class omega 2. Then we have 12 3, this is another feature vector belonging to class omega 2.

Then we have 13 6 3 3 6 this is another feature vector belonging to class omega 2. So, these are the feature vectors belonging to two different classes. We have computed that our Eigen vector is 70, 1, what is that one. So, the Eigen vector was seventy one. So, I want to plot that Eigen vector which is 70,1. So, it will pass through a, somewhere over here. So, I will take a scale to get this direction of this Eigen vector which is this. So, this is my direction of the Eigen vector corresponding to the maximum Eigen value.

Now, if I take perpendicular projection onto this Eigen vector what I will have is, the different projections will be like this. Feature vector will be projected here, this feature vector will be projected here, this feature vector will be projected here. So, we find that as we said that all the feature vectors belonging to class omega 1, they are represented by green color and all the feature vectors belonging to class omega 2, they are represented by red color. When I project these feature vectors onto the Eigen vector corresponding to the maximum Eigen value.

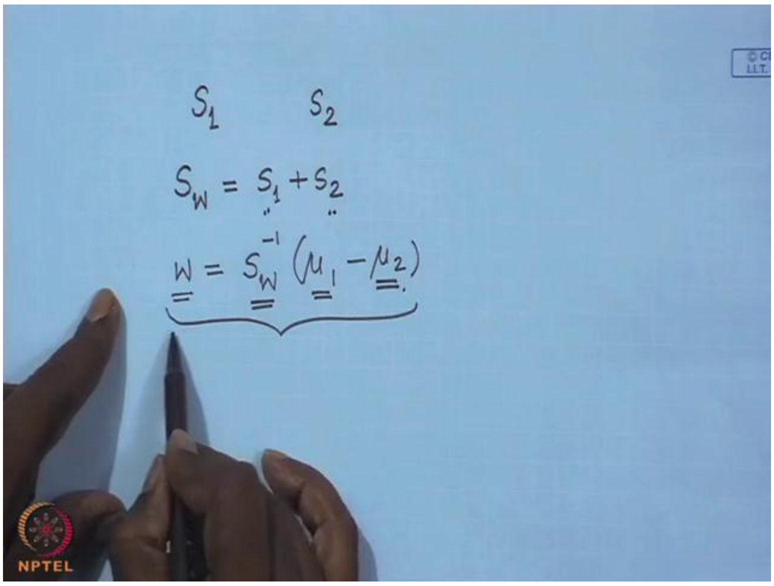
So, these are my projections. So, here we find that feature vectors belonging to class omega 1 they are projected. So, what we have used is, this is my axis x 1 and this was my axis x 2. So, the feature vectors belonging to class omega 1, they are projected onto a points in the line which is in this range, the feature vectors belonging to class omega 2 they are projected over here. So, it clearly says that when I take the projection onto the Eigen vector the projected points, the projected vectors are no more separable.

The projections from class omega 1 and the projections from class omega 2 they get intermixed which is quite obvious in this particular case. So, the feature vector from class omega 1 is over here, where as the feature vector from class omega 2 is over here. Whereas, the rest of the feature vectors which are projected onto the other side the rest of the feature vectors from class omega 1 they are projected onto the other side. So, I have a mixing of the feature vectors the projected feature vectors onto a lower dimensional space. Whereas, you can clearly say see that I can obviously draw a straight line over here. Separating the feature vectors in class omega 2 and the feature vectors belonging to class omega 1 by a linear straight line.

So, in our original space in the two dimensional space they were linearly separable, but when I take projections onto the Eigen vector they are no more linearly separable right. So, this a problem that we face when we project feature vectors from a higher dimensional space onto feature vectors onto a lower dimensional space. This is the problem which is addressed in the features linear discriminator, where we try to find out the projection direction which maintains the separability, which tries to maintain the separability by increasing the between class scatter and by decreasing by minimizing the within class scatter.

So, if you remember the features linear discriminator, now let us talk about the features linear discriminator. So, features linear discriminator tries to find out the mean of the individual classes. So, it considers the feature vectors belonging to class omega 1 and the feature vectors belonging to class omega 2, find finds out what is the mean of feature vectors in class omega 1 and what is the mean of the feature vectors in class omega 2. So, one is mu 1 other one is mu 2. Then it computes what is the within class scatter or what is scatter of the samples of the feature vectors in class omega 1 and what is the scatter of the samples in class omega 2. Then it defines what is called total within class scatter.

(Refer Slide Time: 45:35)

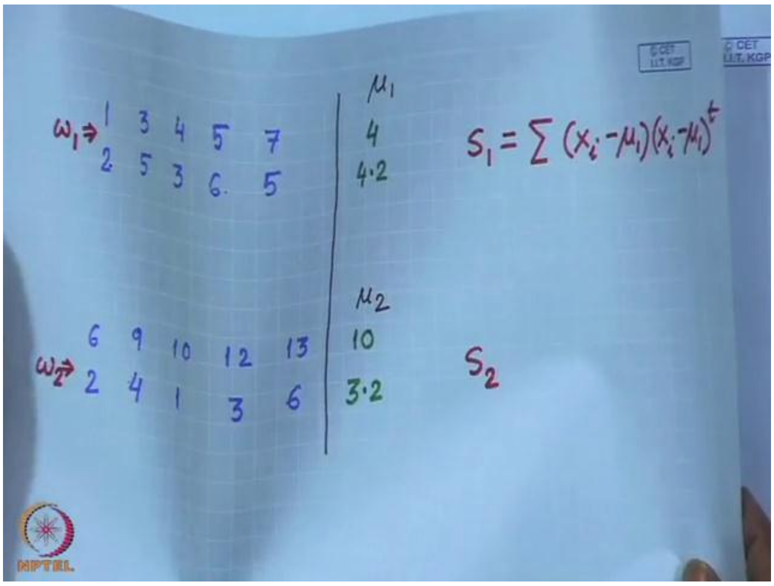


So, if you remember we have said that if S 1, is the scatter of the feature vectors in class omega 1 and S 2 is the scatter of the feature vectors in class omega 2. Then the total within class scatter is given by S w which is nothing but S 1 plus S 2. So, this is the total within class scatter. If mu 1 is the mean of the feature vectors in class omega 1 and mu 2 is the mean of the feature vectors in class omega 2, then when you discussed about the features linear discriminator, we have said the direction the projection direction is given by a vector w which is nothing but S w inverse into mu 1 minus mu 2. Where S w is the total within class scatter, mu 1 is the mean of the feature vectors in class omega 1 and mu 2 is the mean of the feature vectors belonging to class omega 2.

This expression was obtained by solving a generalized Eigen value equation and that Eigen value equation took care of maximization of the between class scatter. That means pushing the clusters so the classes as apart as possible. By minimizing the within class scatter that is S 1 or S 2 by making those classes as compact as possible. So, we try to make the feature vectors belonging to a single class very compact. Simultaneously, try to maintain the distance between two different classes. So, accordingly we had obtained an criteria function and by optimization of the of the criteria function, we could come to a solution something like this, which says that the projection direction which maintains the separability is given by S w inverse into mu 1 minus mu 2. Where S w is the total within class scatter.

So, if I have 2 different classes one is S 1 and other one is S 2, S 1 is the scatter of class one, S 2 is the scatter of class two. So, total within class scatter S w is given by S 1 plus S 2. So, let us try to see that by using this features linear discriminator approach, what projection direction we get and how does it differ from the Eigen direction. So, here again I will consider same set of feature vectors belonging to class omega 1 and omega 2. So, that I can compare the performance of projection onto a Eigen vector direction onto a Eigen vector and the projections onto the features discriminator. So, as we said that we have two sets of feature vectors two dimensional feature vectors 1 set taken from class omega 1 the other set taken from class omega 2.

(Refer Slide Time: 49:21)

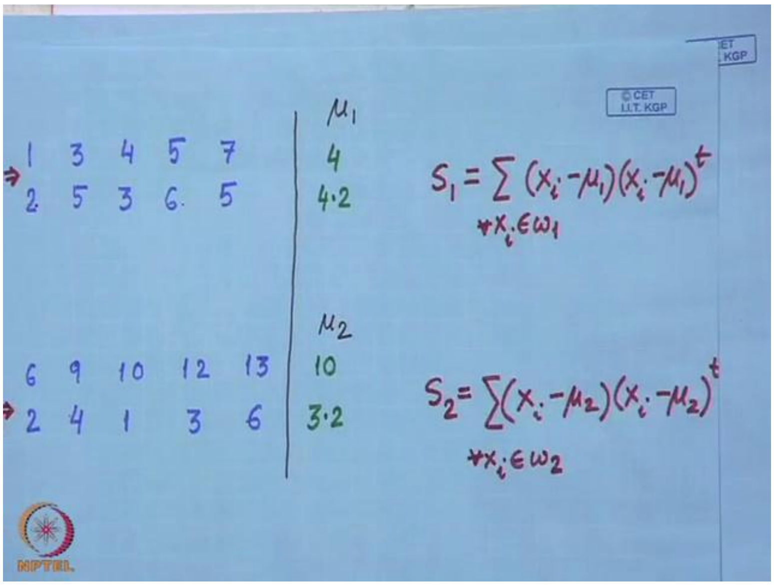


So, from class omega 1 the feature vectors we have taken are 1 2 3 5 4 3 5 6 7 5 . So, these are the feature vectors which are taken from class omega 1. So, each of them is a two dimensional vector 1 2 is a vector, 3 5 is a vector, 4 3 is a vector, 5 6 is a vector and 7 5 is a vector. Similarly, I had other set of feature vectors which are 6 2 9 4 10 1 12 3 and 13 6 right. These are the feature vectors which are taken from class omega 2. So, first what we have to do is, we have to compute the mean of the feature vectors from class omega 1 and the mean of the feature vectors from class omega 2. So, here I will compute mu 1 And here I will compute mu 2.

So, if you compute the mean of these feature vectors which is mu 1, you can easily find out that this is 4 plus 4 8, 8 plus 5, 13 plus 7 20, 20 divided by 5 as I have 5 different feature vectors. So, the first component would be 4 of mu 1. Similarly, some of this second components divided by 5 that will simply be 4.2. So, mu 1 the mean of the feature vectors belonging to class omega 1 is 4 4.2. Similarly, mu 2 the mean of the feature vectors belonging to class omega 2, that also I can compute from here.

So, mean of the feature vectors belonging to class omega 2 mu 2 will simply be this is 15 plus 10 25 37 divided by 5 that will be 10. So, that is the first component and in the same manner the second component you can compute 3.2. So, this is the mean vector of the feature vectors belonging to class omega 1, this is the mean vector of the feature vectors belonging to class omega 2. So, once I have these two means, then I have to compute the scatter of the feature vectors belonging to class omega 1 and scatter of the feature vectors belonging to class omega 2. That means I have to compute what is S 1 and what is S 2. S 1 is nothing but sum of x i minus mu 1 into x i minus mu 1 transpose.

(Refer Slide Time: 53:21)



For all x i belonging to class omega 1. So, you compute x i minus 1 into x i minus mu 1 transpose for all the feature vectors belonging to class omega 1. So, each of them gives me a matrix. So, I will get a number of matrices and because here I have 5 different feature vectors. So, I will have 5 different matrices, sum of all those matrices gives me scatter S1. Similarly, s 2 will be sum of x i minus mu 2 into x i minus mu 2 transpose. For all x i this summation has to be computed for all x i belonging to class omega 2. So, we have to compute these two. So, let me stop this lecture here, in the next lecture I will complete this problem. Then we will compare the performance of this and the projection onto the Eigen vector.

Thank you.