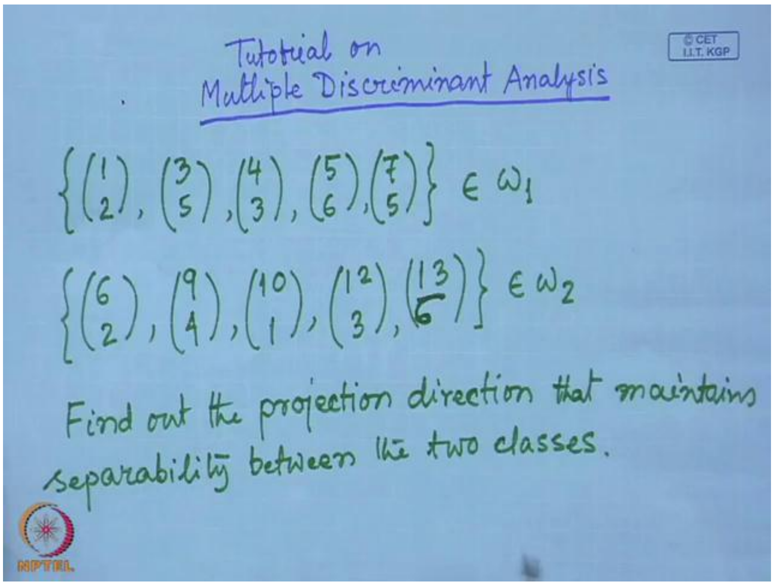
Pattern Recognition and Applications
Prof. P. K. Biswas
Department of Electronics and Electrical Communication Engineering
Indian Institute of Technology, Kharagpur

Lecture - 19
Multiple Discriminant Analysis (Tutorial)

Hello, so in the last class we have started a tutorial. We have taken up a problem of a few feature vectors belonging to class omega 1 and few feature vectors belonging to class omega 2. Actually, the problem that we were trying to solve is the dimensionality deduction problem that from a hard dimensional space if I project onto a low dimensional space. There in many cases the problem that we face that in hard dimensional space if the feature vectors belonging to different classes are linearly separable, when I project that into a lower dimensional space they may not remain linearly separable anymore and we have also discussed that the best way to project is onto the Eigen space. So, we have taken this particular problem in the previous class.

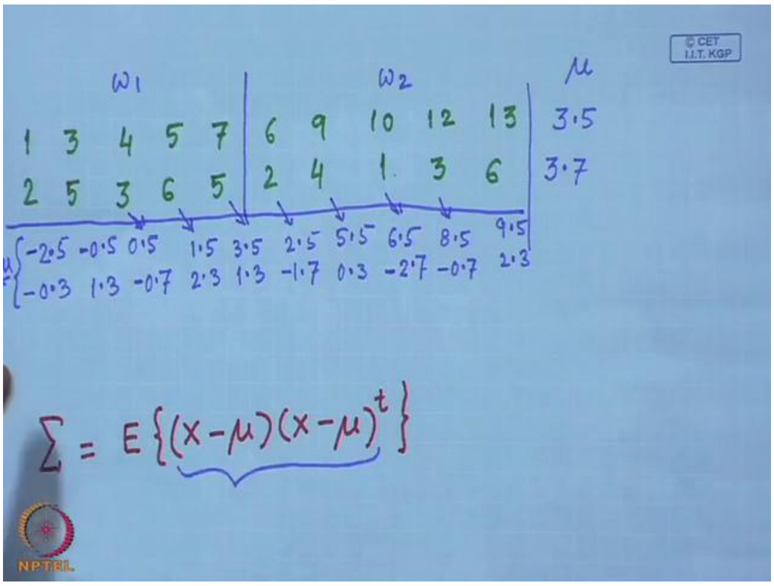
(Refer Slide Time: 01:28)



That there are 5 feature vectors which are taken from class omega 1 and there are 5 features which are taken from class omega 2. And we want to find out a projection direction for reduction of the dimensionality that maintains separability between the two classes that, is even after taking the projection onto a line, the projections of the vectors belonging to class omega 1 and the projection of the vectors belonging to class omega 2. These two states of projected points should still be linearly separable.

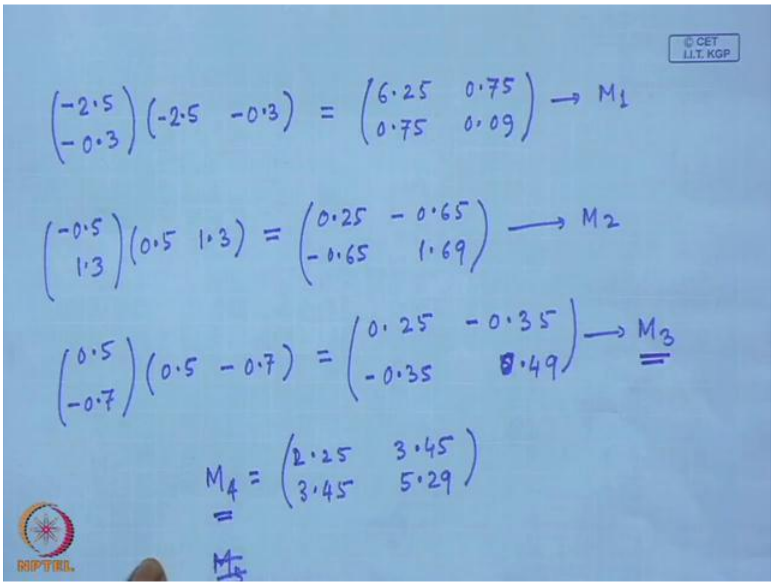
So, before solving this problem we have taken up the issue that, if I project for the best representation of the multi-dimensional data. And the for that we have seen earlier that the best way of projection is to project onto the Eigen vector. So, the first problem that we have tried to solve is what will happen to this projected point, if the projection is taken on an Eigen vector. So, for that let us briefly recapitulate what we have done.

(Refer Slide Time: 02:51)



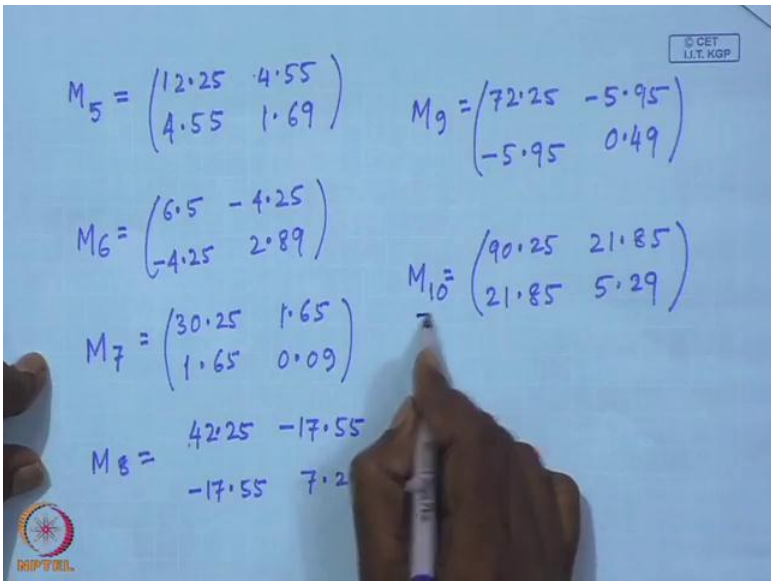
So, this is the inter set of feature vectors. And I want to project this feature vectors onto the Eigen vector. So, I have to find out what is Eigen vector. So, for this feature vector we have computed the min vector. And we know that the co-variance matrix the definition of the co-variance matrix is sigma is equal to expectation value of x minus mu into x minus mu transpose, where mu is this mean vector. So, for every vector we have computed x minus mu. So, these are x minus mu for different feature vectors. And once I have this x minus mu, for every individual vector the x minus mu we have computed x minus mu into x minus mu transpose giving me a number of matrices.

(Refer Slide Time: 03:46)



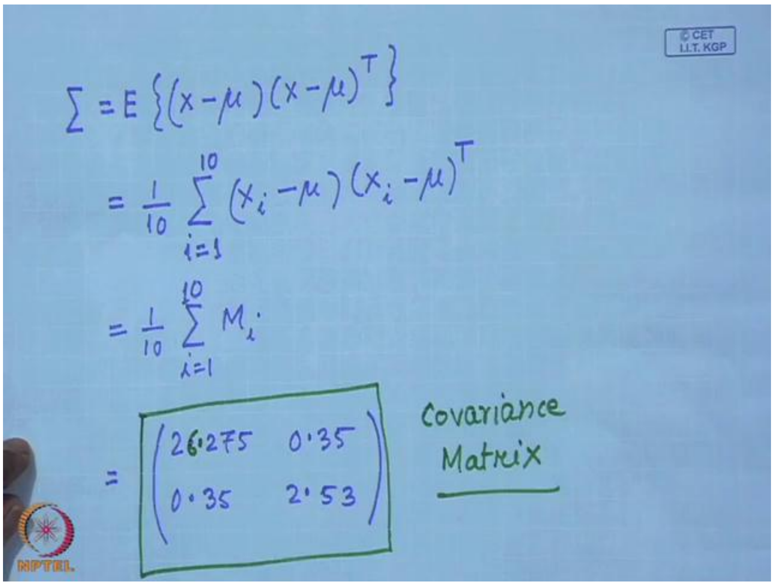
So, the first one x 1 minus mu into x minus mu, x 1 minus mu transpose that gives me a matrix m 1 x 2 minus mu into x 2 minus mu transpose gives me another matrix m 2. Similarly, x 3 minus mu into x 3 minus mu transpose gives me another matrix m 3. Similarly, I get m 4 and as I have total 10 number of feature vectors. So, I get 10 such matrices.

(Refer Slide Time: 04:16)



So, once I have 10 such matrix I can compute the covariance matrix which is nothing but expectation value of all this 10 matrices or effectively, these are the average or mean of all this 10 matrices, which gives me the co-variance matrix.

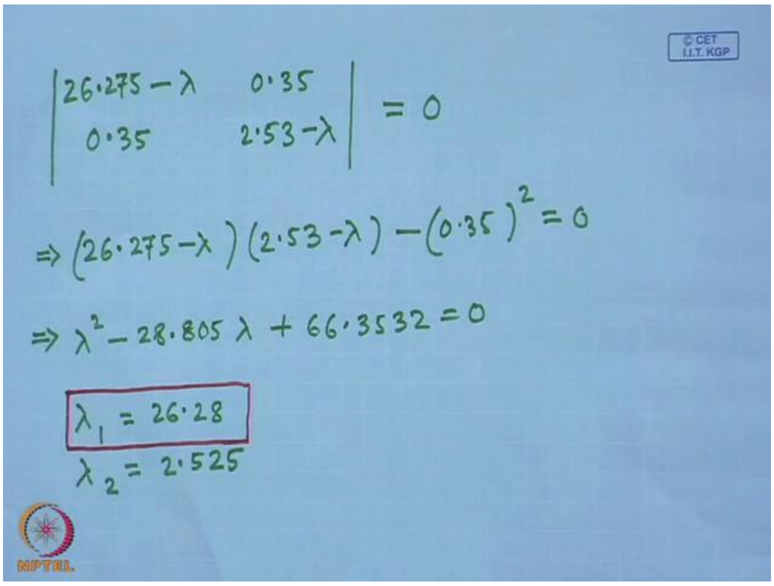
(Refer Slide Time: 04:45)



So, I compute the co-variance matrix by taking the mean of all these 10 matrices. And the co-variance matrix came out with this that is 26.275, .35, .35 and 2.53. Now, you note and interesting over here is. In the matrices that we have computed starting from m 1 to m 10, this was the matrix. You find that 1 2 and 2 1 is same, 1 2 and 2 1 is same, 1 2 and 2 1 is same. And obviously in the co-variance matrix also element, 1 2 and 2 1 is same, because this is the co-variance between the first element and the second element x 1 and x 2.

So, these two elements have to be. So, I get the co-variance matrix. And to get the Eigen vector of this co-variance matrix, I have to solve the Eigen value equation. So, first I have to find out the Eigen values of this co-variance matrix out of that I have to choose the maximum Eigen value. And find out the Eigen vector corresponding to the maximum Eigen value. So, to find out Eigen values of the co-variance matrix the equation is the procedure is well known you form a determinant of this form.

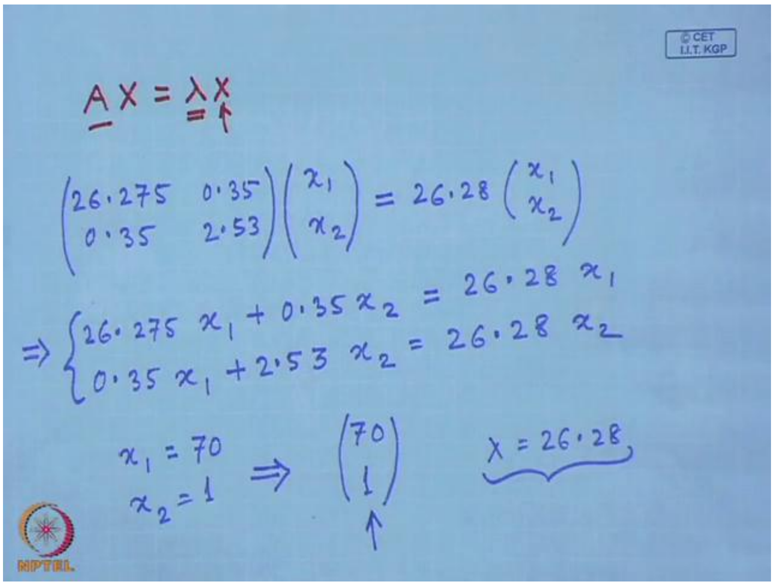
(Refer Slide Time: 06:17)



So, along the main diagonal if lambda is an Eigen value then along the main diagonal from all the elements you subtract lambda from a determinant and set this is determinant value equal to zero. So, I will get a number of equations. Solve those equations to find out the value of the lambda. So, that is what I get over here. I get a quadratic equation you find that it is an equation which is quadratic in terms of lambda. And because it is a quadratic equation so I will get 2 values of lambda to Eigen values.

So, I get lambda 1 which is 26.28 and I get lambda 2 which is 2.525. So, this co-variance matrix has two such Eigen values and out of this two Eigen values, I have to consider that Eigen value which is maximum. And corresponding to this Eigen value I have to find out the corresponding Eigen vector. So, obviously the matrices of size n by n I will have n number of Eigen values. So, it depends upon what is the dimensionality of the matrices.

(Refer Slide Time: 07:48)

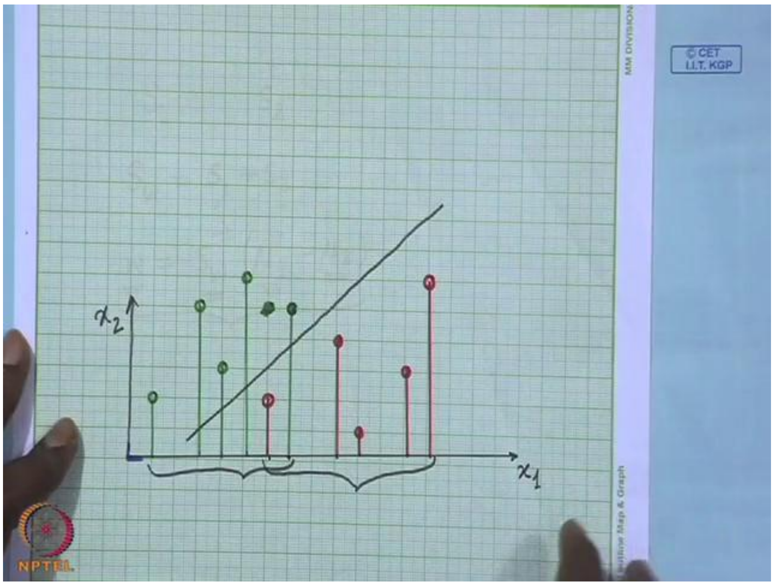


So, for this Eigen value I try to compute the Eigen vector and for that we have this well-known Eigen value equation that for matrix A, if x is the Eigen vector and lambda is the corresponding Eigen value then, this equation must be satisfied that is a x is equal to lambda x. In all case this matrices A is the co-variance matrix which is this right 26.275, 0.35, 0.35, 2.53. And this Eigen vector x is x 1 x 2 which will be equal to so I took the maximum Eigen value which is 26.28.

So, 26.28 x 1 x 2 I get two simultaneous equations. Solve this equation to get the values of x 1 and x 2. So, here you find that x 1 would be is equal to 70, x 2 would be equal to 1. If it try to get the value from the other one the values will be slightly different and the values that difference in values from these two equations is because of the error. In many cases the decimal values we have position we have ignored.

So, I get it later. So, because of the transition error the x 1 x 2 values that I get from here. And the x 1 x 2 values that I get from here, ideally this should be same, but I get a difference because of the transition error, but anyway the values will be very close instead of 70 here we may get 68 69 or something like that. So, this is my Eigen vector, Eigen vector is 71 then what we have done is…

(Refer Slide Time: 09:42)



We have plotted all the feature vectors onto a graph paper. And this is the Eigen vector 71 projected on all those feature vectors onto the Eigen vector and these are my projected points. And here as expected because we did not consider the separability of the feature vectors belonging to two different classes. We just wanted to find out projections directions that are a projections space that best represents the higher dimensional feature vectors into a lower dimensional feature vectors.

So, we said that it may not be possible to maintaining the separability between the classes. And exactly that is what has happened. You find that all this feature vectors belonging to class omega 2 their projected within this space. And all these feature vectors belonging to class omega 1, they are projected within this region and these 2 regions now get. So, though in the original 2 dimensional space the feature vectors belonging to two classes omega 1 and omega 2.

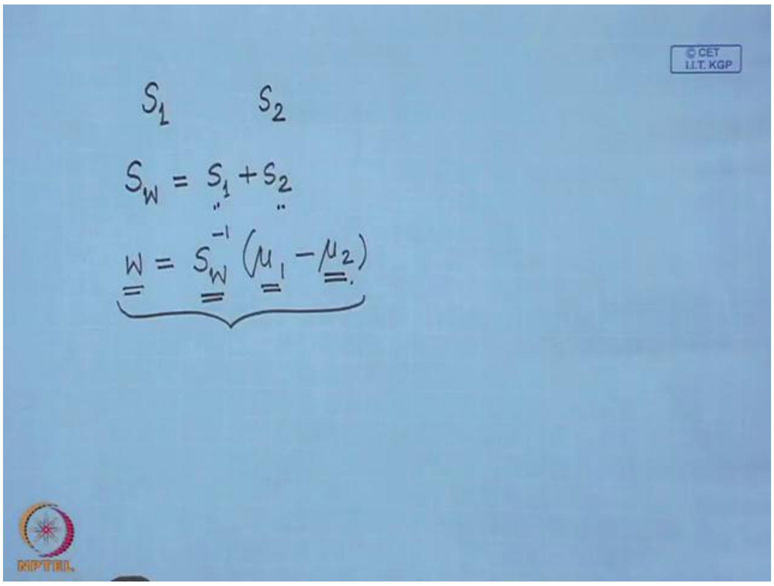
They were separable on the projected space they are no more separable. So, this is a problem that we face if we go for best way of representation or project the original set of data onto the Eigen space for reduction of dimensionality. So, to solve this problem what feature linear discriminator does is it tries to find out a projections space which maximizes the separability between the classes or separation between different classes. And at the same time it tries to form compact clusters of the feature vectors belonging to individual classes.

So, what happens if I have two points clouds belonging to two different classes. When I project into a lower dimensional space, I try to see that these two classes are wide apart and at the same time the feature vectors belonging to individual class they are very compact. So, if I can maintain this, I can maintain the separability between the classes even in the projected space or in the lower dimensional space. So, in order to do that I have to consider two things one is I have to minimize the within class scatter so that every class is very, very compact.

And at the same time I have to maximize the between class scatter that is the classes should be pushed well apart. So, if I do them simultaneously then I get a sort of criteria function. So, after optimization of the criteria function I get a solution vector. So, this solution vector gives me a vectored or a projection direction on which if I project higher dimensional feature vector even in the lower dimensional projected space the separability of the vectors can be vented. Of course, when I tell all about this my assumption is in the initial space in the original higher dimensional space that data points are separated, they are separable.

If they are not separable in the original higher dimensional space then whatever tricks I play I cannot guarantee that they will be separable in the lower dimensional projected space. So, that has to kept in mind that when I talk about separability after projecting onto a lower dimensional space. I always assume that in the original higher dimensional space the feature vectors are well separated. So, then only the other things are valued. So, by optimization of that criteria functions which minimizes the within class scatter and maximizes the between class scatter. I get a solution which is obvious form.

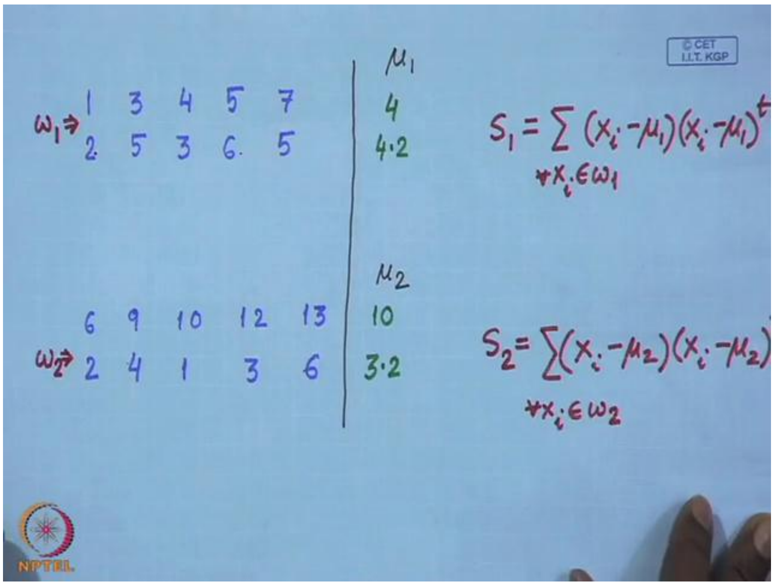
(Refer Slide Time: 14:17)



That is if S 1 is the within class scatter of the data points belonging to class omega 1 and S 2 is the scatter of the data points belonging to class omega 2. Then S w which is equal to S 1 plus S 2 this is called the total within class scatter, that is the sum of the scatters of the samples belonging to class omega 1 and the samples belonging to class omega 2. So, I take these two scatters individually then add this two I get total within class scatter which is S w mu 1 is the mean of the samples belonging to class omega 1 and mu 2 is the mean of the samples belonging to class omega 2.

So, once I have this then, the direction of the projection or the projection vector which maintains separability will be given by S w inverse into mu 1 minus mu two. So, this solution has been obtained by solving a generalized Eigen value expression. So, to illustrate this I will take up the same problem for which we have gone for projection onto the Eigen space. So, that is we have two sets of samples, two sets of feature vectors one set of feature vector belonging to class omega 1 the other set of feature vector belonging to class omega 2.

(Refer Slide Time: 16:05)



So, the feature vectors which belong to class omega 1 are 1 2 3 5 4 3 5 6 and 7 5. So, these are the 5 feature vectors which are taken from class omega 1. Similarly, 6 2 9 4 10 1 12 3 and 13 6. These are the 5 feature vectors which are taken from class omega 2, and the scatter of data points belonging to a particular class they are defined by this. It is the sum of x i minus mu 1 into x i minus mu 1 transpose where mu is the mean of the feature vectors belonging to class omega 1.

So, for every vector belonging to class omega 1, I have to compute every vector x belonging to class omega 1 I have to compute x minus mu or x minus mu 1 into x minus mu 1 transpose. So, when I have n number of feature vectors in belonging to class omega 1 I will have each of this x minus mu 1 into x minus mu 1 transpose gives me matrix. So, if I have n number of feature vectors in class omega 1, I will have n number of such matrices.

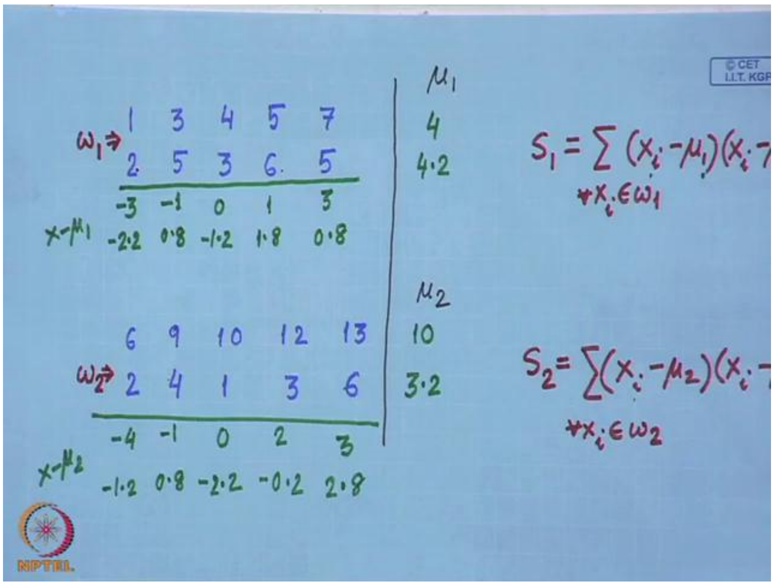
And sum of all those matrices gives me the scatter of the samples in class omega 1. Similarly, here for every vector belonging to class omega 2, x minus mu 2 into x minus mu 2 transpose that gives me 1 matrix. If there are m numbers of feature vectors in class omega 2, I will get m such matrices. And sum of all these m matrices gives me the scatter of the samples belonging to class omega 2. So, that is how I compute scatter for class 1 and I compute the scatter from for class 2 S 1 and S 2. And you find that there is this scatter and the co-variance matrix they are very, very similar.

If I normalize the scatter matrix by the number of samples that I have, I get a co-variance matrix because co-variance matrix is nothing but expectation value of this term, which is nothing but take the summation of all these matrices normalize by the number of data elements that number of matrices that I found I get a co-variance matrix. So, they give more or less same information in one case it is normalized in the other case it is not normalized.

So, what I do is I consider this feature vectors belonging to class omega 1 and I compute the mean of all this feature vectors because finally, I have to compute this term. So, I compute the mean of this feature vectors which is mu 1 and mu 1 in this case will be 4 4.2. And then considering these vectors from class omega 2, the mean of this feature vectors which is mu 2 is 10 3.2 which you can easily compute from here. Take the summations of all these components divide by 5 I get 10.

Take the summation of all these second components divide by 5 you get 3.2. And similarly, from here and the next 1 I have to compute is for each of the data points for each of this feature vectors I have to compute X minus mu 1. For each of the feature vectors over here I have to compute x minus mu two. So, that I can compute x minus mu 1 into X 1 minus mu 1 transpose. In the other case X minus mu 2 into x minus mu 2 transpose.

(Refer Slide Time: 20:28)

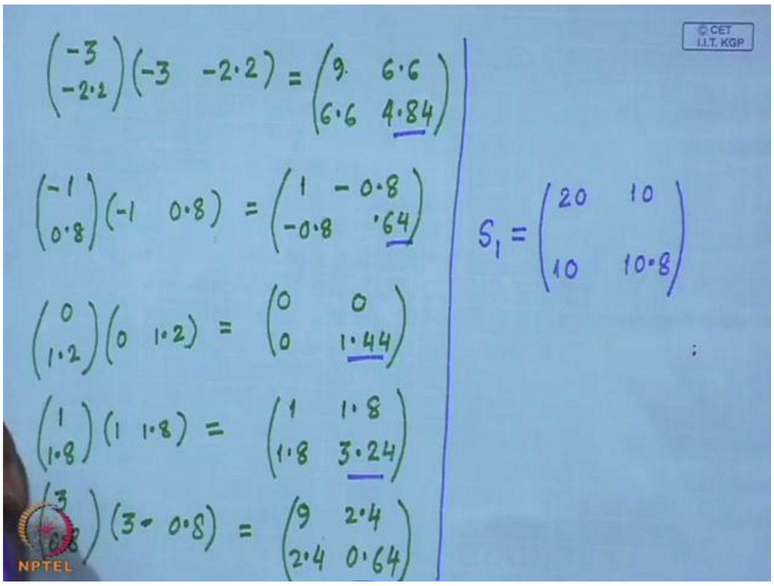


So, over here you find that for all the samples belonging to class omega 1, X minus mu 1 when I compute x minus mu one. So, x minus mu 1 over here it will be simply 1 minus 4 that is minus 3 for the first sample then 2 minus 4.2 so this will be minus 2.2. Similarly, over here for the second one it is 3 minus 4 that will be minus 1 5 minus 4.2 that will be 0.8. Similarly, over here this will be 0 minus 1.2. Here it will be 1 1.8 and here it will be 3 0.8. So, these are x minus mu 1 for individual vectors x. In the same manner I have to find out x minus mu 2 for individual vectors x belonging to class omega 2.

So, let us see what those values will be for the second one. So, here I have to compute x minus mu 2. So, the first one it will be 6 minus 10 which is minus 4, 2 minus 3.2 which is minus 1.2, 9 minus 10 which will be minus 1, 4 minus 3.2 is 0.8, 10 minus 10 it will be 0, 1 minus 3.2 it will be minus 2.2. Similarly, here 12 minus 10 will be 2, 3 minus 3.2 it will be minus 0.2, 13 minus 10 it is 3 6 minus 3.2 it will be 2.8.

So, I have got x minus mu for every individual x belonging to class omega 2. And Now, I have to go for computation of the scattered matrices S 1 and S two. So, for computation of S 1 I have to find out what is x minus mu 1 into x 1 x minus mu 1 transpose for every x minus mu 1 taken from the class omega 1. So, the first one is minus 3 minus 2.2.

(Refer Slide Time: 23:17)

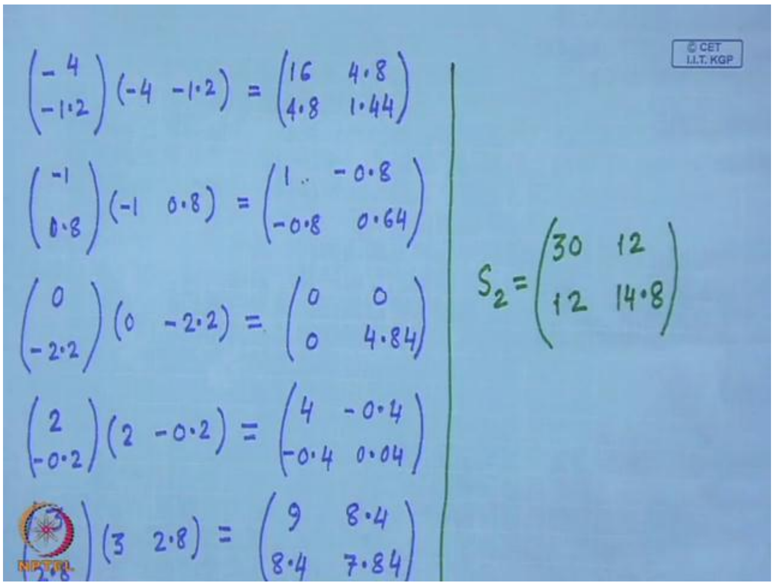


So, let us see that what will be this first one is minus 3 minus 2.2, this is x minus mu 1 into x minus mu 1 transpose will be minus 3 minus 2.2. And if you compute this value the matrix will be 9 6.6 6.6 then 4.84. This is one of the matrices the second one was minus 1 0.8. If you remember this it was minus 1 0.8. So, it is minus 1 0.8 the column vector into minus 1 0.8 row vector. It will be simply 1 minus 0.8 minus 0.8 and .64. For the third one it was 0 1.2 0 1.2.

So, if I compute this it will be 0 0 0 1.44. For the fourth one it is 1 1.8. So, I will have 1 1.8 column vector into 1 1.8 row vector. So, if I multiply this it will be 1 1.8 then, over here 1.8. And this will be 1.8 into 1.8 that will be 3.24. And then fourth one X minus mu 1 for the fourth one was 3 and .8. So, if I compute this 3 then 0.8 3 0.8 if I compute this it will be 9 2.4 2.4 and 0.64. So, these are the 5 matrices that I have 1 2 3 4 5. And from these 5 matrices I have to compute the scatter is 1 which is nothing but sum of all these 5 matrices.

And if I take the sum you find that the first component would be 9 plus 1 plus 1 plus 9. So, that is 18 plus 2 which is equal to 20. Similarly, this is 6.4 minus .8 plus .8 plus 2.4 6.6 minus .8 plus 1.8 plus 2.4. So, this will be simply 10 this term will also be equal to 10 and here I will have 10.8 which is nothing but sum of this plus this plus this plus this plus this which comes out to be 10.8. So, this is my scatter matrix S 1 Similarly, let us try to compute what will he the scatter matrix S 2. So, for computation of scatter matrix S 2 I have x minus mu 2 over here. For every x belonging to class omega 2 so that is given by this. So, I have to compute the scatter matrix S 2. So, the first one is minus 4 minus 1.2.

(Refer Slide Time: 28:21)



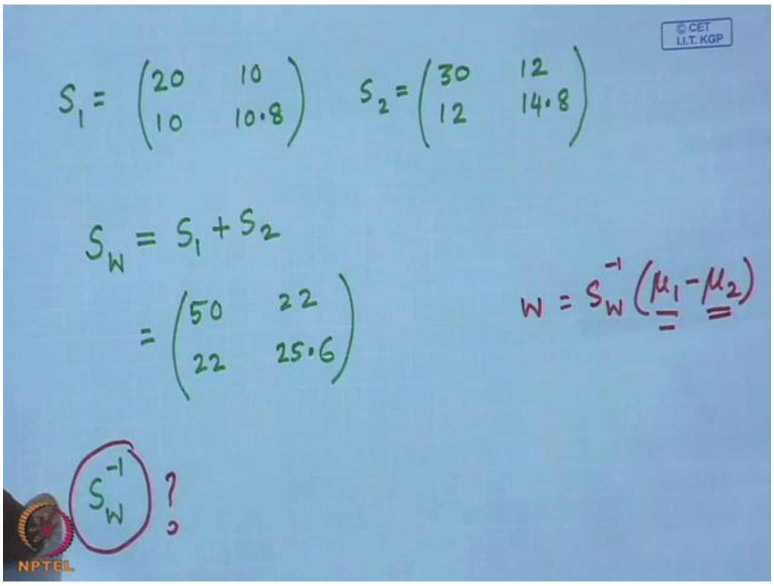
So, the first matrix that I get is minus 4 minus 1.2 into minus 4 minus 1.2. So, that simply becomes 16 4.8 4.8 and 1.44. So, this is what I get from the first vector. The second vector is minus 1 1.8. So, the matrix that I get from the second vector is minus 1 0.8. So, it is minus 1 0.8 into row vector minus 1 0.8. So, this matrix will be 1 minus 0.8 minus 0.8 0.64. For the third one which is 0 minus 2.2. So, it is 0 minus 2.2 into row vector 0 minus 2.2. So, this will simply be 0 0 0 then 4.84.

So, this is my third matrix. To compute the fourth matrix it is 2 minus 0.2. So, what I have is 2 minus 0.2 this is the column vector into 2 minus 0.2 the row vector. And if I multiply these two what I get is 4 minus 0.4 minus 0.4 and then 0.04, this is the fourth matrix. And then to get the fifth matrices I have to take this 1 3 2.8. So, this is 3 2.8 3 2.8, multiply these two what I get is 9 8.4 8.4 and then 2.8 into 2.8 it is 7.84. So, these are the 5 matrices that I get from the 5 samples belonging to class omega 2.

Now, sum of all these 5 matrices will give me the scatter S 2. So, if you take the sum scatter S 2 will be given by this. So, I have S 2 which is nothing but sum of all this 5 matrices. So, over here it will be 16 plus 1 17, 17 plus 2 is 21 plus 9 30. So, the first component will be 32 components 4.8 minus 0.8 minus 0.4 plus 8.4. So, this will be 12. Similarly, here it will be 12 and this component will be 1.44 plus 0.64 plus 4.84 plus 0.04 plus 7.84. So, this will be simply 14.8.

So, we find that I get two scattered matrices one for class omega 1 which is S 1 the other scatter matrix for class omega 2 which is S two. So, for class omega 1 the scatter matrix S 1 is given by this and for class omega 2 the scatter matrix S 2 is given by this. Now, from S 1 and S 2, I have to find out what is the total within class scatter that is S w which is nothing but S 1 plus S 2.

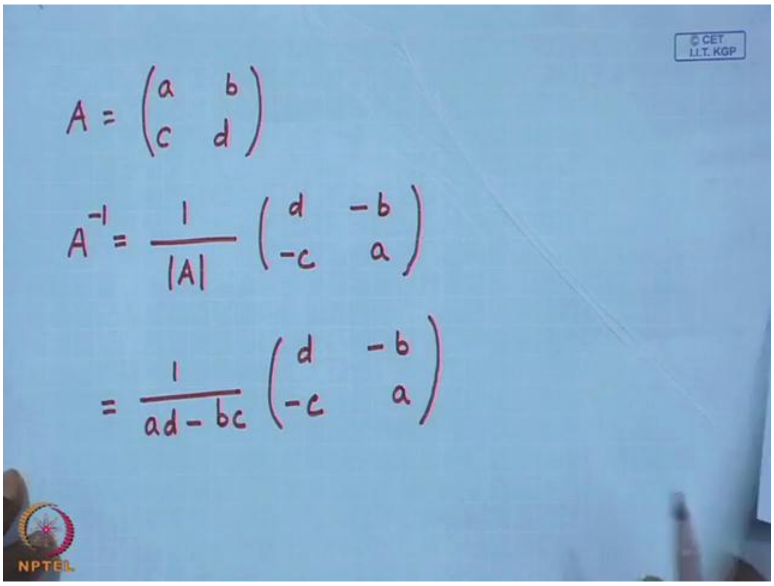
(Refer Slide Time: 33:43)



So, what I had is I had S 1 which is 20 10 10 10.8 and s 2 which was 30 12 12 and then 14.8. So, I have to compute the total within class scatter which is S w which is nothing but S 1 plus S 2. And this will be simply if I add this two matrices 50 22 22 and here 10 plus 8 plus 14 plus 8 that will be 25.6. So, this is my total within class scatter. And from this what I have to compute is S w inverse because finally, I have to compute the projection direction.

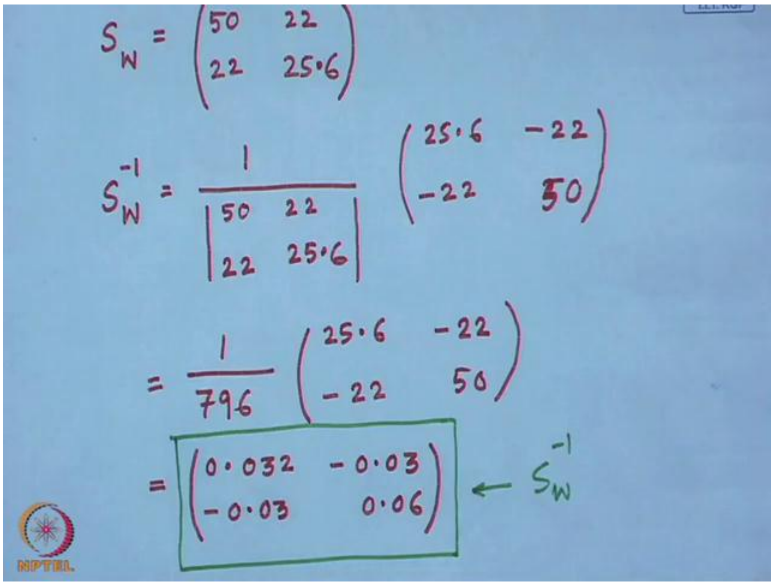
And the projection direction S w is given by S w inverse into mu 1 minus mu 2 where mu 1 is the mean of the samples belonging to class omega 1 and mu 2 is the mean of the samples belonging to class omega 2. So, from this total within class scatter is w, I have to compute what is S w inverse. So, this is what I need to compute. Now, for a 2 by 2 matrix computation of inverse is very simple. So, if I have a two dimensional matrices something like this.

(Refer Slide Time: 35:47)



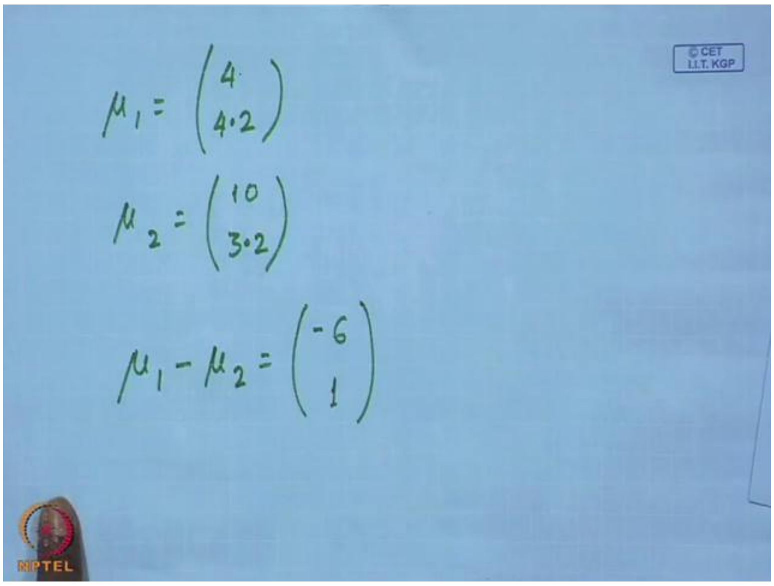
Say, matrix A which is given by a b c and d this is a 2 dimensional matrices. Then A inverse is simply given by 1 upon determinant of A. And then matrix is d minus b minus c a which is simply this determinant A is nothing but a d minus b c. So, which is simply 1 upon a d minus b c into d minus b minus c and then a. So, this is what is the inverse of A 2 by 2 matrices, where the 2 by 2 matrix is given by elements a b c d. So, simply using this expression when I know that my total within class scatter S w is given by 50 22 22 25.6.

(Refer Slide Time: 37:04)



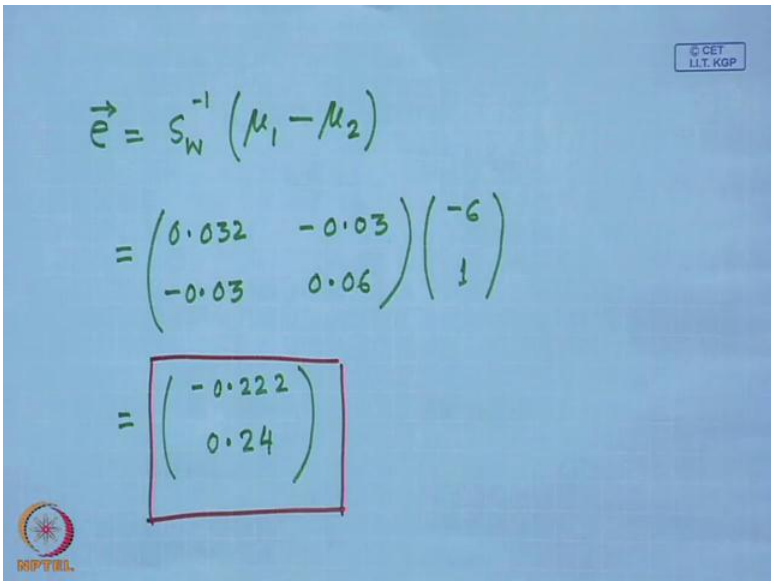
This is the total within class scatter I can very easily compute the inverse of the total within class scatter that is S w inverse. So, over here this is S w inverse will be given by 1 over determinant of S w that means determinant fifty 22 22 25.6. And this matrix will be 25.6 minus 22 minus 22 20. And this one will be 50. And if you compute this determinant you find that this determinant is 50 into 25.6 minus 22 into 22 which comes out to be 1 over 796 into 25.6 minus 22 minus 22 50. And if you complete this computation the inverse of the matrix will come out to be 0.032 minus 0.03 minus 0.03 0.06. So, S w inverse is simply this. So, this is the matrix which is important to me that is S w inverse. And the other quantity that I have to compute is mu 1 minus mu 2.

(Refer Slide Time: 39:38)



And we know from before that mu 1 was given by 4 4.2 and mu 2 was given by 10 3.2. So, these were the means of the samples belonging to class omega 1 and the mean of the samples belonging to class omega 2. So, I can easily find out that mu 1 minus mu 2 will be is equal to this minus this that is minus 6 this minus this which is equal to one.

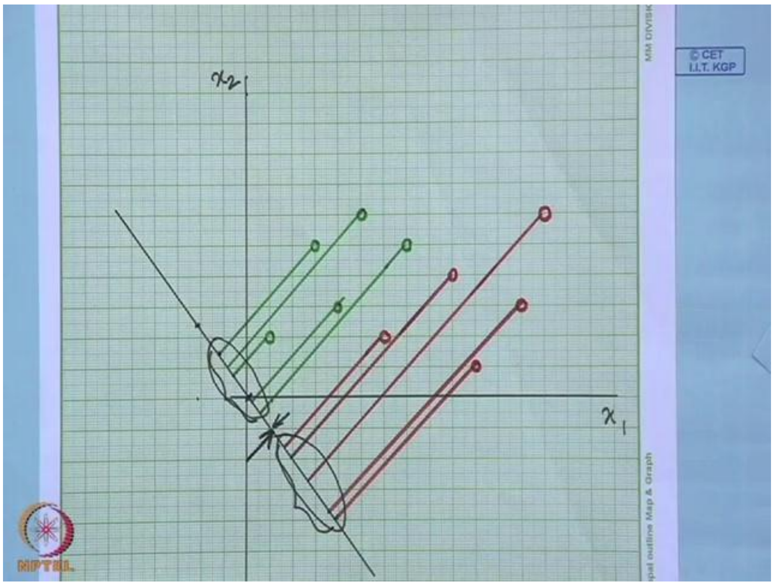
(Refer Slide Time: 40:30)



So, once I have this the projection direction is simply given by, if I write it as a vector e is simply S w inverse into mu 1 minus mu 2 that is the direction of projection where S w inverse. You remember from it is 0.032 minus 0.03 minus 0.03 0.06. Of course, we are not considering the higher decimal spaces. So, it is 0.032 minus 0.03 minus 0.03 0.06 into mu 1 minus mu 2 which is nothing but minus 6 1. So, you multiply this by the column vector minus 6 1.

And if you compute this you will find that this 1 will come out to be minus 0.222 and 0.24. So, the projection direction of the vector on which if I take the projection which will maintain the separability between the classes should be this 1, which is minus 0.222 and 0.24. So, again let us plot these points on a graph paper and take the projections on to this vector. And let us see what improvement we get at all if we get some improvement. So, I will take this graph paper. So, as before let me assume that so this is my origin. So, I will draw x axis and y axis over here or x 1 and x 2 axis.

(Refer Slide Time: 43:21)



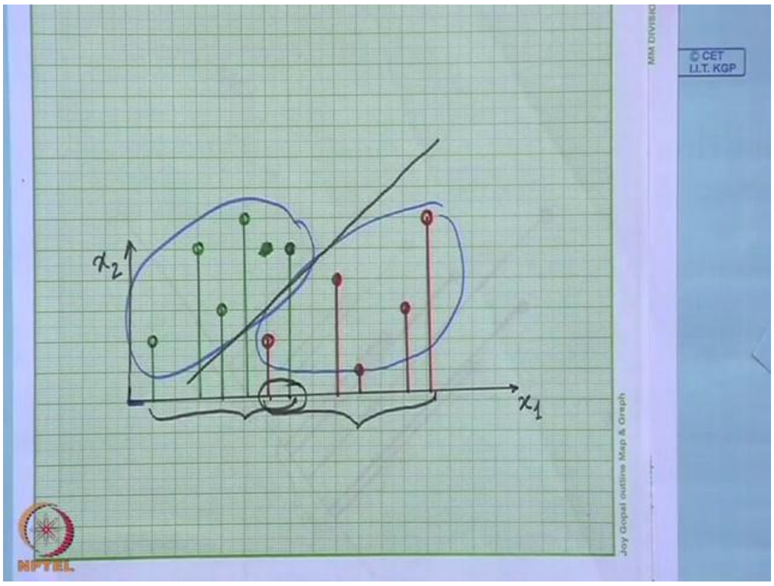
So, this is x 1 axis and this is x 2 axis. Let us put the vectors belonging to class omega 1. I will use this skin color for that, the vectors are 1 2 then 3 5. If you remember this vectors which we had used 1 2 3 5 4 3 5 6 and 7 5 from class omega 1, 6 2 9 4 10 1 12 3 and 13 6 from class omega 2. So, for class omega 1 it is 1 2 the next one is 3 5 next one is 4 3 that is over here the next one is 5 1 2 3 4 5 1 2 3 4 5 6 5 6 which is over here the next one is 7 5 1 2 3 4 5 6 7 1 2 3 4 5 7 5.

So, these are the vectors which are taken from class omega 1, similarly if I take the vectors from class omega 2, which is 6 2 9 4 10 1 12 3 and 13 6. So, I will take 6 2 1 2 3 4 5 6 6 2 is over here then 9 4 1 2 3 4 5 6 7 8 9 1 2 3 4 this is 9 4 then comes 10 1 10 1 then comes 12 3 12 1 2 3 and then thirteen 6 thirteen 1 2 3 4 5 6 13 6. So, these are the vectors from class omega 2. And my projections direction as we have computed is minus 0.222 and 0.24. So, I will put it as let us this point so minus 0.22 minus 0.24 minus 0.22 and plus 0.24.

So, because I am an interested only in the direction. So, even if I multiply this by then does not matter. So, I will take it as minus 2.2 multiplying this by 10 minus 2.2 and 2.4. So, I will take this point minus 2 or if minus 22 if I multiply this by 100 even that does not matter. So, I will make is minus 22 and then plus 24. So, minus 22 is somewhere over here plus 24 is over here. So, this is the point so my vector is, this is the direction of projection.

Now, let us project or take the perpendicular projection of all the feature vectors belonging to class omega 1 and the feature vector belonging to class omega 2 on to this projection direction. So, this will be, if I take perpendicular projection. So, these are the projections. Now, we find when I projected it the feature vectors belonging to class omega 1 onto this projection vector and the feature vectors belonging to class omega 2 onto the projection vector they are not mixed anymore there are well separated. So, over here this is the region in which the feature vectors belonging to class omega 1 they get projected and this is the region over which the feature vectors belonging to class omega 2 they are projected. And these 2 regions are well separated. So, the projected points are no more mixed.

(Refer Slide Time: 50:12)



Unlike in the previous case if you remember this particular figure at the feature vectors from 2 different classes they are mixed in this region. So, the projected points are no more separable, but if I maintain that constant that when I take try to take the projection directions I will maintain the separability I will try to maximize the separability between the classes or I will try to maximize the between class scatter. And I will try to minimize the within class scatter.

So, while doing so the solution vector of the projection direction that I get is this and here it is quite clear. This demonstrates that on this direction if I take the projections even in the projected space. The Feature vectors in different classes they are well separated. So, this solves one part of the problem that is happened able to reduce the dimensionality, but for classification application, I need to go further that is I have to design a classifier which works in this projected space.

So, how do I do that? Clearly, over here these are 1 dimensional variable all these projected points if I take the distance of these points from the center these are 1 dimensional variable. So, I get a set of 1 dimensional value or scalar values over here. I also get a set of scalar values over here. So, I take this set of scalar values to belong to one class and this set of scalar values that belonging to another class. And for classification purpose what I need to do is. I need to identify a point on this line which will separate between these two classes.

So, if this point follows a particular distribution I can find out what is the mean and variance of this set of points. I can also find out what is the mean and variance of this set of points. And from there I can identify what should be the threshold of this line. So, this is not a big problem, but it is a very simple problem. So, through what we have done is, we have reduced the problem from a higher dimensional space to a lower dimensional space and the lower dimensional space the problem is more manageable. So, I stop this lecture here.

Thank you.