Indian Institute of Technology Madras
Presents

NPTEL
NATIONAL PROGRAMME ON TECHNOLOGY ENHANCED LEARNING

Pattern Recognition

Module 02

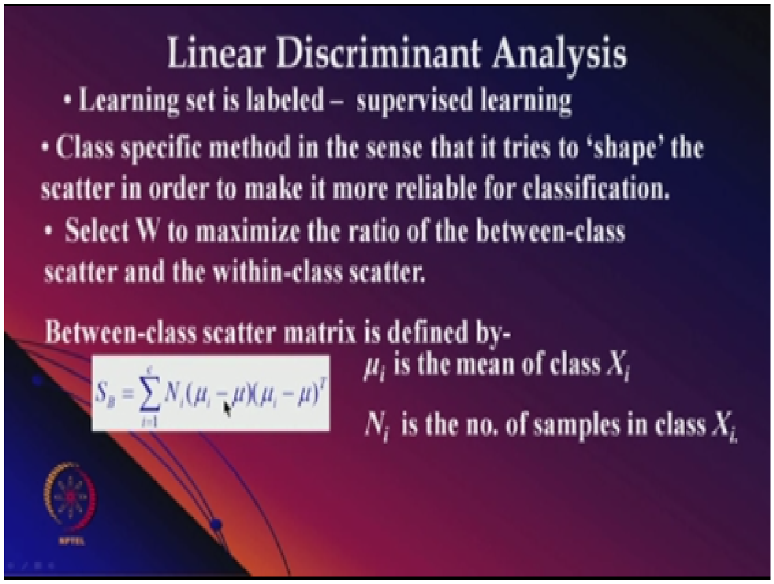
Lecture 14

Fisher's LDA

Prof. Sukhendu Das
Department of CSSE, IIT Madras

In this last section we were discussing concept of supervised method of pattern classification which is based on the fishers linear discriminate analysis or LDA it is sometimes called FLD as well in the pattern recognition literature and we are just introduced 2 different matrices the one is within class scatter matrix and there is between class scatter matrix we will look at our expressions one more time and then look at see some important properties and expressions of LDA and we will winde up this class with a few examples which can be hand out with you okay.

(Refer Slide Time: 01:08)



So let us look back what we are saying is LDA is a method of supervise learning it is one of the classifiers which needs set of training samples for learning a set of parameters in this case the parameters are SW and SB the within class scatter and the between class scatter expletively the learning set is labeled unlike the PCA where we had unlabeled data and that was also called unsupervised learning so it is a class specific method in sense it tries to shape the scatter in order to make it more reliable for classification.

Remember in the case of PCA we are interested in trying to find out the directions in which the maximum scatter of the entire data exists that is what principle components analysis or PCA does in case of LDA we want to maximize class separability okay that means in some scene the between class scatter if you take a 2 class problems we want to find a certain direction of the data within the data certain dimension in which the distance between the 2 class means or that two clusters of the 2 class become large or become more.

And in the same direction the within class scatter becomes less okay if you recollect the animation slide which we had long ago in earlier class when we were trying to distinguish between classification verse clustering we also way say that it is better and easier for a classifier to perform better if the between class distance or the distance between 2 clusters centers or the 2 class themselves is very large and the clusters are very compact.

So if you can find such dimension or set of dimensions what can be considered also as sub space okay with respect to the origin higher dimension of the data if you can find sub space where the directions point out that we going to have or is it satisfied with satisfies a constrained that the inter class distance is very large and the inter cluster distance is very small that is what is trying to achieve okay so that is what is the meaning of the sentence which you see now that the class specific method it tries to shape the scatter in order to make more reliable or better fort classification okay.

So this is accomplished with the help of trying to find out a weight matrix w which maximizes the ratio of between class scatter SB and with class scatter SW where the terms we will define them again for your ease of understanding that this is the between class scatter SB if you at the expression here the NI is the number of samples per the for particular class Xi let us say the class label is Xi μi is the mean for a particular class Xi and the μ is the overall μ of the data okay.

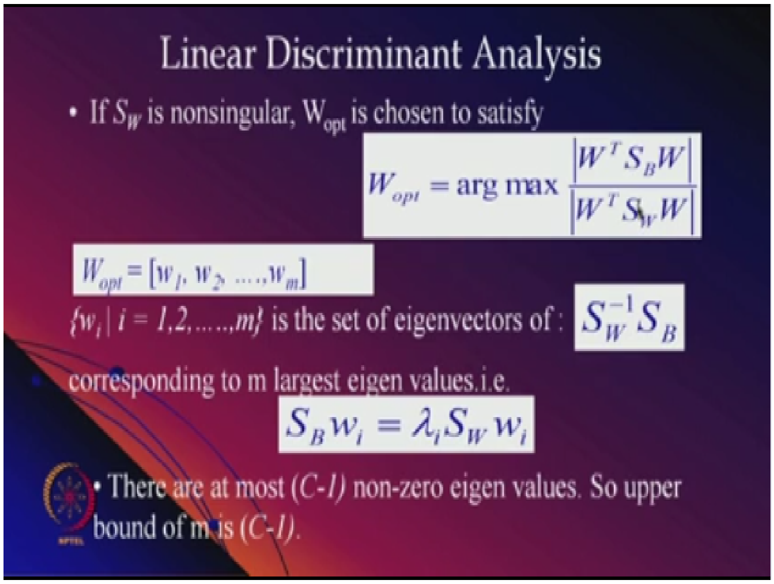
So it is something so if you look at Py- B μ it is basically that the individual process centers are normalized with respect to the mean okay of the entire data set and summation over Ni will actually give the number of samples but a particular classes are in weight age C or c here ware whatever you see is the number of overall number of classes okay so you need to sum this over al class.

That is the between class scatter let us look the expression once again for SW which is the with class scatter matrix you need to sum it over all class all right but you sum it over now this expression is similar to the PCA it is an outer product of the samples with respect to μi that in case of PCA where the overall data mean here which was μ the same μ which you see here on the left hand side would have occurred here in the case of PCA but in case of LDA the you have the mean subtracted from the data which is the individual class means.

Okay the μi is the mean of the class Xi which you have to take that and take the outer product you sum it over all the samples for particular class in fact Xk belongs to the samples set Xi that means basically Ni the summation will go over Ni number of times and the over c number of class okay so the total number of summation term which will be having is basically N multiplied by c SW and SB have the same dimension as the scatter matrix we had for PCA the dimension is the same.

But the matrices themselves are little bit different okay if you look at the expressions both are in some sense the outer product put one of them is computed with respect to the means only the other is computed with respect to the samples means subtracted with respect to the class specific means okay the PCA we subtracted the overall determine I am repeating again here you're subtracting the class means.

(Refer Slide Time: 06:09)



Okay so SW and SB is what you have okay and nest so what we are trying to do is find out an optimal W and we will just talk a little bit later on that what happens if SW is singular but assuming that the within class scatter matrix is nonsingular you try to find out an optimal value of W which maximizes this expression okay so we try to find out a W which maximizes expression and in the process of doing so you get an optimal W which can written in terms of a set of Eigen vectors as give here m sort of Eigen vectors.

and they are the Eigen vectors of this particular matrix which is SW^{-1} x SB in fact what you are doing here is trying to find out the m largest Eigen vectors of this characteristic equation which is the if you think of SW^{-1} x SB as an overall scatter matrix S then you are actually trying to find out the Eigen vectors and Eigen values λ is the corresponding Eigen values of this particular matrix okay and this is the reason why it is essential that the within class scatter matrix is a nonsingular because you need to obtain it is inverse of that matrix then multiply it with SB and then find the corresponding Eigen values and Eigen vectors.

This is the basic approach for FLD or LDA correct and I do that you need to find pretty inverse of the matrix SW the question comes is SW always singular we will have a look at it very soon I will just give key points with respect to some properties of within class scatter matrix SW. So there are actually at the most C+1 – non zero Eigen values in the λ so if you look at what is called the Eigen spectrum of this particular matrix the upper bound of m here is basically number class – 1.

So the restriction on the number of non zero Eigen values and the corresponding Eigen vectors for the W will actually depend on the characteristic or properties of SW weather it is singular or not these are some of main criteria which we need to follow.

SW is singular if the total number of samples $N < $ Dimension D remember this total number of samples not the so we can say that this is the average number of samples for class multiplied by the total number of class okay that will give you that total number of samples N it is rank is at most N- C okay where C is the number of class so you cam see that the number of samples here is going to dictate as very important say on the singularity or rank of SW.

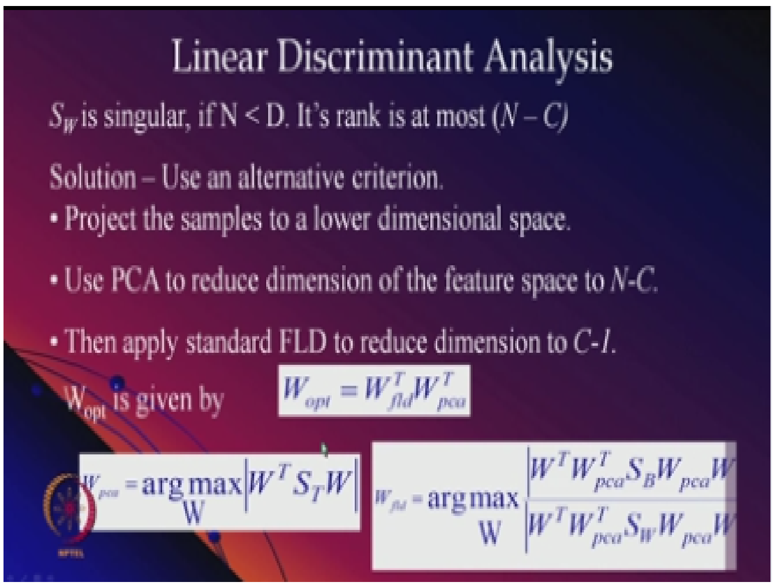
Typically if the if you have sufficient number of samples you do not need worry about the singularity of SW if you sufficient number of samples very large okay but there are certain applications where there are a death of the number of samples both for a particular class and for all the data samples but together or all the class put together a typical example is the case of face recognition problem, in the case of face recognition the dimension it typically may go to a few millions.

Okay it is very large okay in fact the resolution or size of the image N^2 if you take as not this N put if you take this 640/4 at team age 640 x 4 it will be the dimension of the problem weather you are doing PC or LDA now you will not have those many samples per individual or total number of cases available in your data base.

If you take let us say an example of a very large data base it may have even a few 1000 individuals you may have around 10 let us say 10 to 100 at the most samples per particular class in fact there are certain situations where you have just a few samples may be just about half does not 10 samples per particular, for a particular class in this case the class is a particular subject.

So in such a case definitely what will happen is the number of samples n will be less than the dimension, if you look at the slide.

(Refer Slide Time: 10:29)



This constrained will not be satisfied in all applications or situations of pattern recognition problems, all problem this may not satisfy in certain situations if the number of samples is less than the dimension or the dimension is much larger than the number of samples this becomes singular, okay and it is rank is at the most N-C. We will move at with this assumption that the rank is N-C for the time being we looking into this problem.

However, if the number of samples are very large and it exceeds the number of dimensions in certain cases which it is possible yes, in such cases you do not need to worry about the singularity of SW, but there are certain situations where you need to worry the number of samples are less typically those are the cases which are called the SS problem which is called the small samples size problem and the number of samples being very large per class or for all the classes put together that again is less than number of dimensions in such a case you cannot invert SW.

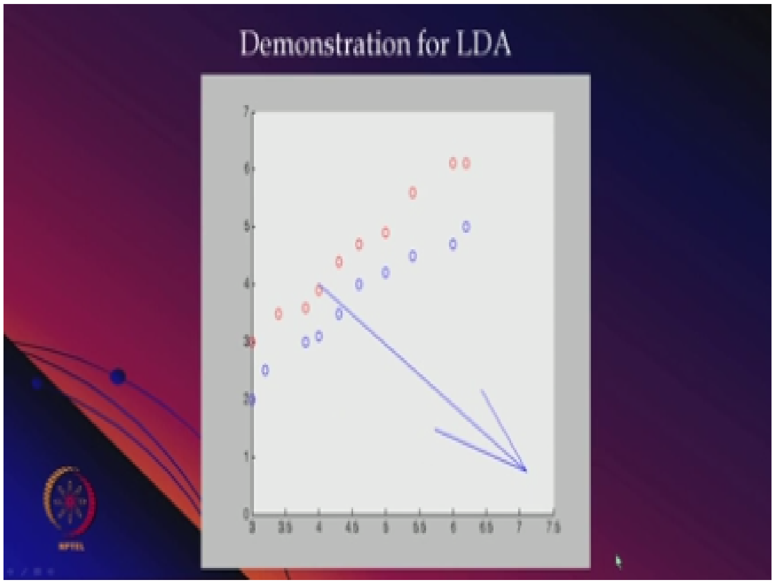
Okay, it is not a full rank matrix it rank is restricted by the number of class in number of samples. Solution to this such problems if you have single in SW is the following, project the samples to low dimensional space and to do that we use our method of PCA which we have studied in the last class to reduce the dimensions from the feature space from the original dimension D to a dimension mxc.

And one that is possible to apply the standard Fisher Linear Discriminate criteria or FLD to the reduce dimension C-1, okay. So in such a case when you are applying PCA which reduce the dimension that W optimal value of W can be visualized as that you have done a PCA earlier and the corresponding W which you would have obtain by the FLD criteria.

And the PCA criteria for getting the W for the PCA is given by this where S_T is the original scatter matrix. Remember we had an expression earlier which said that the scatter matrix S_T is the sum of SW+SP, so ST is SW+SP and W FLD can be written as an expression you can see that this is the similar expression to what you are we got earlier expect that you have the PCA done before the LDA.

So that corresponding matrix comes and sits as a pre and a post multiplication with respect to SW, so if you do this method under the condition that N is less than D SW is a singular you want to reduce this dimensionality of the space go to a sub space where you can reduce the dimension corresponding to the rank of expected rank of the matrix. Then you do a PCA first and then do an FLD to reduce the dimension and people typically take C-1 dimensions for the FLD.

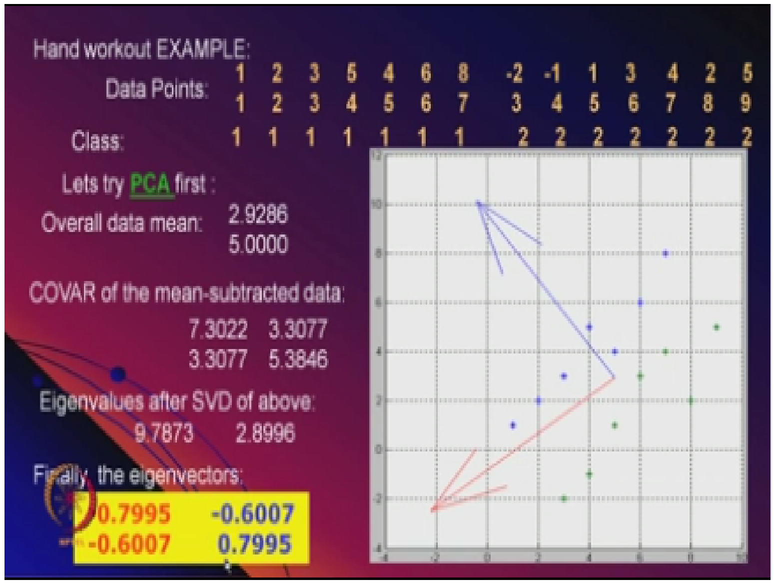
(Refer Slide Time: 13:19)



Okay, so let us look at some hand worked out example this is a very synthetic diagram drawn with respect to two data sample points and we had this similar diagram in the case of PCA where the PCA would have given a direction along the maximum scatter or variance of the data samples which have been along the direction orthogonal to this vector okay, so that along this line along the data samples as this arrow indicates that would have been what the PCA directs.

But the LDA will give a direction in which you are expected to have the maximum scatter that means now if you project the samples point they expect to the red points to be in the left hand side the blue points on the right hand side. Although you may not have separability in this case because data samples are very nearby alright but this is the direction in which you will have maximum separability or the condition of $SW^{-1}xSP$ is maximized the correspondingly whatever with S^T.

(Refer Slide Time: 14:16)



Let us take this hand worked out example it is easy to do this calculations in a sheet of paper but of course you can keep your calculators ready if you want, so let us take this example of data sample points in 2D way we are taking an example into 2D because it is easy to visualize so these are set of x,y points then the top you have set of x coordinates and the bottom you have set of y coordinates you have them for two different classes so the class 1 is the class level at the bottom, so for the class 1 you have 7 points, for the class 2 you also have 7 points and the data samples in the x,y two dimensional space we look something like this.

Where you will have the data samples to the left given by this blue points here and the data samples to the right as identified by class number 2 will be given by the set of green points. So let us try PCA first before trying LDA, we will try LDA also by using the class information but let us shut off the class in 1.

Let us say I give you a scatter so forget the color of this code, color code for these two different classes let us take all of these as a set of points as set of black markers let us say forget the color and let us try PCA that means you have just taken the data samples and ignoring the class level and we do that this is the overall data mean for all the samples, okay which can be averaged out by taking the average of all x values which you will get as this.

Average of all the y values is what you will get as 5 okay, and this will help you to compute the overall covariance of the mean subtracted data from this that means what you need to do is subtract from each of the sample points the mean and then do an X^T to actually obtain the covariance matrix it shall be a 2x2 matrix as given here. Because it is a two dimensional data so that is what this is size or dimension of your ST or covariance matrix or scatter matrix.

The corresponding Eigen values obtained after a singular valued decomposition of this matrix is basically this, which basically shows that along the first dimension you have quite a bit of scatter along the second dimension you have then you know maybe much a less than the first dimension. What are the corresponding Eigenvectors for the first dimension this is the dimension of the Eigen vector we will draw that in this diagram very soon.

You will have it up and this is the second Eigen vector, what do you expect along the first Eigen vector you should expected to have a large class scattered because the Eigen value is very large along the second dimension you will also have a scatter but it would not be that scattered or that stretch as with respect to the first Eigen value okay.

So if you look at the diagrams here the Eigen vector shown here is given by the same color along this particular direction so it basically says that along this direction you have the maximum scatter of the data which is quite obvious if you look into this the overall data seems to form an elliptical pattern in terms of a scatter and this seems to be the major axis of that particular scatter.

What is the other dimension, which is given here so this is given by also the blue color corresponding the same color code I have used for displaying the values of the, components of the Eigen vector as well as the vector shown in 2D so you can see that along this dimension you have very less scatter of the data samples which is probably almost 1/3th or even less than these scatter which you have along that.

So what PCA is giving you are two different directions okay, as we are discussed in the earlier class along the first direction you have a very large scatter along the second dimension given by the PCA you also have a scatter but it would not be as large as the first dimension. How much of the scatter exist in the second or even third if it exist in this case of course we are just discussed a hand worked out example in two dimension.

But if you take a N dimensional problem you will get a scatter matrix of dimension N corresponding N Eigen values and N set of Eigen vectors. So along the first Eigen vector, second Eigen vector, third Eigen vector and so on you will have as diminishing scatter, the scatter will be going less how much is it happens down will actually depend on what is called the Eigen spectrum or the distribution of the Eigen values along the diagonal of that diagonal matrix.

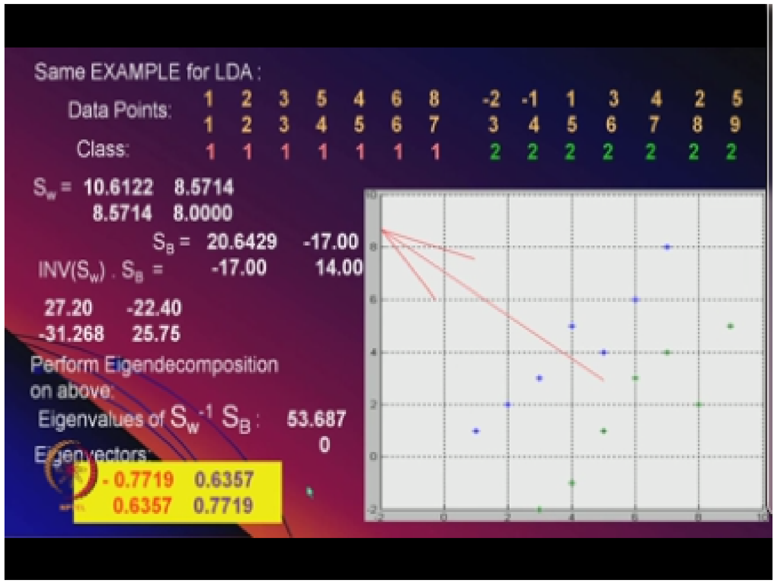
Okay, so in this case is a two dimensional problem if you look here that this scatter along the first is much larger than the second which is expected and the ratio will probably tell us how much more is the scatter along the first dimension here or the first principal component that is the way one should say precisely, this is the first principal component this is the second principal component you could have a third if the dimension was three.

And the corresponding Eigen vector it has given here and the corresponding Eigen values are given here, this is what PCA will do. But as you can see here that if you project the data samples along the first Eigen dimension you will not have separability in fact is the second dimension of the PCA or the second principal Eigen vector which will give you a separability.

So PCA should not be use for classification in general, we talked about this earlier it is typically used for dimensional reduction for data representation okay, are trying to find out a sub space where you can have maximum scatter up to that point is it okay. As for example you have just in today's class that PCA supersedes LDA comes before LDA because you want to reduce the dimension in certain cases when the SW the within class scatter matrix is singular, okay.

So PCA can come before the LDA to reduce the dimension that is one of the applications, if you want ot actually obtain class separability and if of course class labels available you better try an LDA rather than the PCA, so let us use this example in the next slide and move ahead.

(Refer Slide Time: 20:08)



And try to use this class information which is now level with 2 show that the data on the left side belong to the class1, the data on the right side belong to the class 2 this is the distribution which we had for class 1 and class 2 in the previous slide as well where we use the sample for PCA without the class information, now using LDA, now using class information we will perform LDA on this dimension on this data.

Look at the Sw okay, so we have given the expression earlier but I am just are giving you the values, so you can use the entire data set to class problem 2d to compute within class scatter 2 + 2 within between class scatter matrix is given here. The inverse of Sw in this case it is non singular okay because you have enough number of data samples to dimension of the problem, os it will not be singular here.

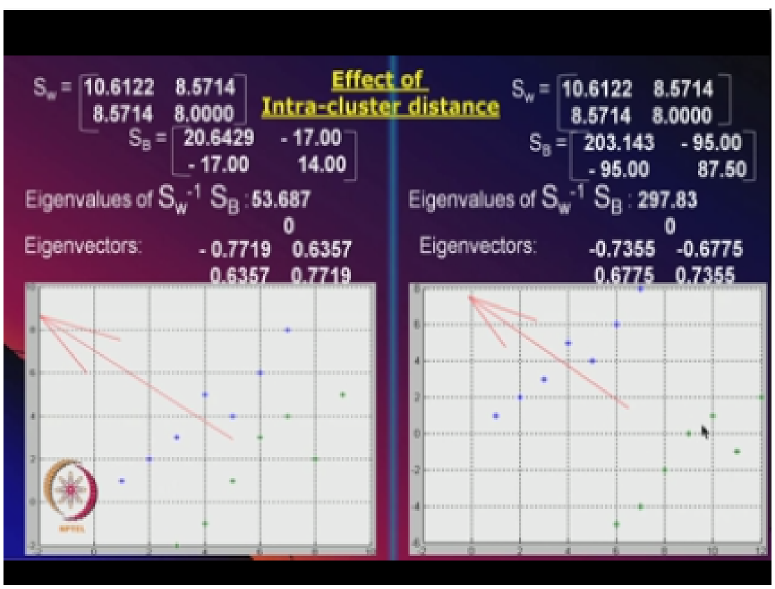
So inverse of this matrix I will leave it as exercise to do that it is simple 2 x 2 matrix multiplied with Sb okay, os if you look at the inverse of Sw which is non singular matrix in this case as number of data samples is large compare to the dimension this is what you get. So what is the corresponding I can get the composition of this okay? Eigen values of Sw inverse are given by this, so what does this basically tell you, that along the 1^{st} Eigen vector you will have a large scatter.

You will this scatter is 0 the separability is 0 of the data and the corresponding Eigen vector is given in, so let us draw them on the data, so this is the 1^{st} corresponding Eigen vector given by LDA remember PCA gave us the direction which was along the maximum direction of the scatter, LDA is given along the direction of maximum separability this is the direction as given in the red. The corresponding color as we use also to show you the direction what about the, I am not shown you the other direction but let me tell you this direction will be orthogonal to this.

So it will be along this direction in which is also similar to the direction given by the 1st principle component of the PCA let us go back to have a look, if you look at the PCA Eigen vector here and if you invert the sign it is 0.8 and 0.6 which will give you the direction towards the upper right and it same as the 2nd Eigen vector direction almost similar it is not same it is not identical but along the similar direction you have a scatter which is 0.

So I am not drawing the second Eigen vector the main reason being that you are not expected to have any separability or the criteria is null so in this particular case we will take only the 1st dimension to predict the data to have the maximum separability.

(Refer Slide Time: 23:15)



Let us look at the effect of the inter cluster distance on the Eigen vectors and the Eigen values as an example. Now let us assume that these on the left hand side you have the within cluster scatter and between cluster scatter matrix the data which you have just discussed in the previous example okay these are same values okay, you can check your notes that these will be the same values obtained from the data.

If you go back you can see here this is what Sw and Sv is for this particular data, so you have kept that on the left hand side, on the right hand side I give you the same Sw but different Sb and I will tell you the mechanism by which I obtain this, what I did was or what you can do is? You need to actually separate the 2 data out okay.

The data sample points which you have seen in the previous slide you need to bring in more separability will show that with the diagram. So if you do that basically give it a shift along the direction of not the maximum scatter but the maximum depth on separability okay so that can be done and I will show what I am mean by that. When you do this keeping the within the class separability same within class scatter same we consider this is change or increase in the corresponding values of the between class scatter.

What this basically means I have not changed the distribution of the individual data samples but taken the two class bin and the separated them a part okay I repeat again, kept the inter cluster that means within class scatter of each particular set of class samples I have not change this scatter of the data points but I have change the class means. Let us say one of the class means if you change to bring in more separability this is what the result is going to happen.

That means you will have a larger between class scatter matrix and the same as W we will go and do Eigen for both the net results is following look in Eigen spectrum of the original data we will compare with this whatever we have now okay. So I do not make the 2nd component of the Eigen as non 0 still keep it 0 but there is increase in the γ1 which is basically the corresponding first Eigen value along the first principle component obtain by LDA let us look at the corresponding Eigen vectors.

Well there is not much of a change okay change is what you will see in the 2nd or the 3rd decimal places of the corresponding Eigen vectors, so the directions more or less and the magnitudes have remain the same let us look at the data okay. this on the left hand side what you see is are the data points which are given to you earlier in the previous slide as the previous example and that the corresponding component is given by 1st Eigen vectors.

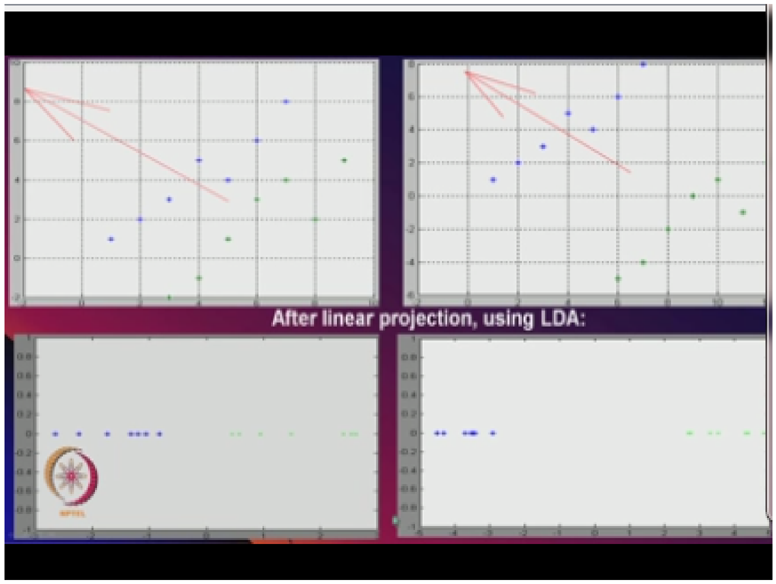
We will never consider the 2nd Eigen vector because the corresponding Eigen value is 0, so this is the 1st Eigen vector, this is same case here minor change in the value which respect to the Eigen vector here, so it is almost the same but look at the separability, this is separately which I brought into the data actually have a large value for the both the diagonal of the diagonal terms between the class scatter matrix compare to the data that which we had earlier and the net result is larger.

You can see that the separability of the data between classes getting deflected in the corresponding Eigen value, so this is lesson so in some sense you have maximum separability along the 1st dimension the less you go to the 2nd and so on and after some point of time you may not have any separability on certain dimensions which are you know beyond C – 1 the number of classes = c here = true.

So c – 1 = 1 so you will have only 1 dimensions so you will have only 1d in which you will have separability and you will not have any separability after c- 1 Eigen vectors that is 1 c -1 is non 0 and 2 = 0 but if you bring in more separability of the data it gets deflected in the corresponding Eigen values, it will shoot up if the class separability is more or the inter class within class scatter also gets reduced.

This example shows that between class is the spread or the distance is more diagonal values are going higher the same will happen within class scatter matrix goes down that means the data samples are much close to the mean themselves for a particular class that also can happen.

(Refer Slide Time: 28:10)



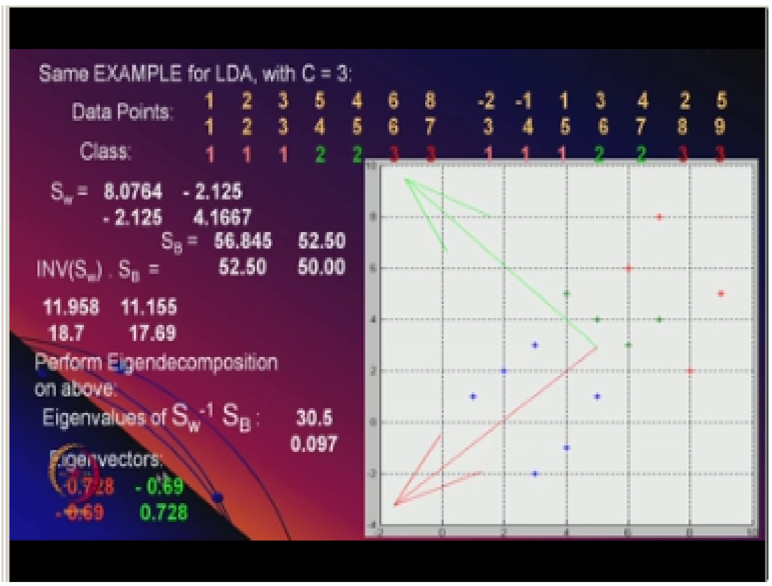
So this is the example which shows that if I take the data points and project it along this direction well there is a little bit of scale problem here which the direction is not shown problem but in way it is more or less accurate here this is generated by a small program. So if you do this you can see the separability here that means all the blue points are projected to left hand side, so what you are seeing as the x axis here is the axis along the direction of the principle Eigen vector given by LDA.

So all the blue points will pay from left hand side as it is here all these green points are projected here and there is some degree of separability between these 2 data samples okay. Look at this particular data and if I project now this on the corresponding Eigen vectors you see the separability now, corresponding principle Eigen vectors 1st principle Eigen vectors given by LDA, so if you projects this blue points they all form a very strong cluster here.

The corresponding green points from another cluster here when you compare the separability on the right hand side which is respect to this, not only the separability is larger but the within class scatter, you look at the scatter of the blue samples here compare them with the scatter here, look the scatter of green points, look at the scatter here. So both the examples show that between class scatter as increased and within scatter class as come down.

Remember of course one must be careful when comparing these 2 plots remember one thing here that they are not to the same scale, now this is of -3 to +3 so it is basically in a range of 6 pixels or 6 units in dimensions here the dimension is so in a scale of 10 the separability here if you say will be up to 5 units in terms of length it is a huge pressure and this sub pressure is about two units more than one so this separation is also increased and within class scatter reparability also come down these are the two lessons and that get deflected in the Eigen values not in the corresponding Eigen vectors.

(Refer Slide Time: 30:38)

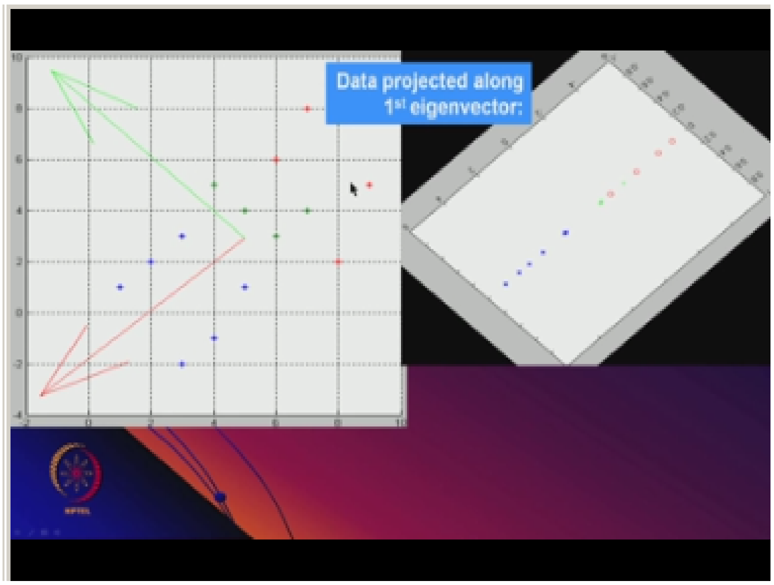


To wind up let us take an example where I increase the number of classes to three okay so what I have done the distance distribution of points is same you have set of points in blue corresponding to class 1 set of points in green corresponding to class 2 and four points in red corresponding to class 3 okay we will quickly go through SW and SB since the number of classes is 3 you can expect that you will have seperabilty in two dimension c-1=2.

When you perform Eigen decomposition $SW^{-1}xSB$ you have both the Eigen values which are non zero now and the corresponding Eigen vector are given here so you have some small amount of seperabilty in the other dimension which is second dimension but look at the first dimension here you have high degree of seperabilty.

Which is reflected in the corresponding Eigen values this is the first degree of separation so you can see this is the class one, this is class two and class three so off course this is the best possible dimension in which you can project the samples to have a fear degree of seperabilty you may not have a seperabilty around the dimension so that is why you the less value of degree of Eigen value.

(Refer Slide Time: 32:03)

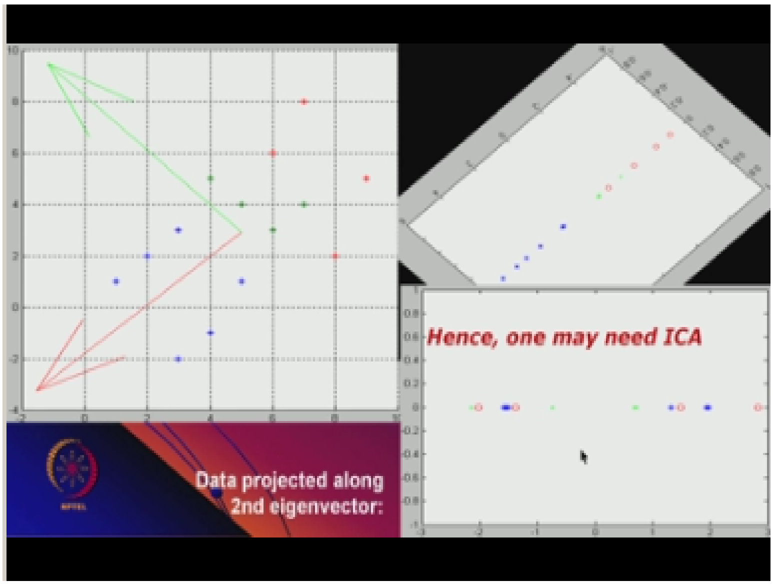


When you project this sample here in the first Eigen vector this is what you have okay that means again I have kept the inclination of this direction same that means so take this blue point project it here take the green point project on this take the red point project it here this is what you have okay if you tile it and show here this is the seperabilty you have a good amount of separability with respect to the blue class as compared to the green.

And the red points there is some degree of overlap here between the red and green points which you cannot probably over come so the red point here is the blue point projected here is original point at appoints here in two dimensional space they all project here so this is fake degree of seperabilty in the first Eigen vector obtained by LDA the same may not be true in the second Eigen vector you can see a huge degree of overlap.

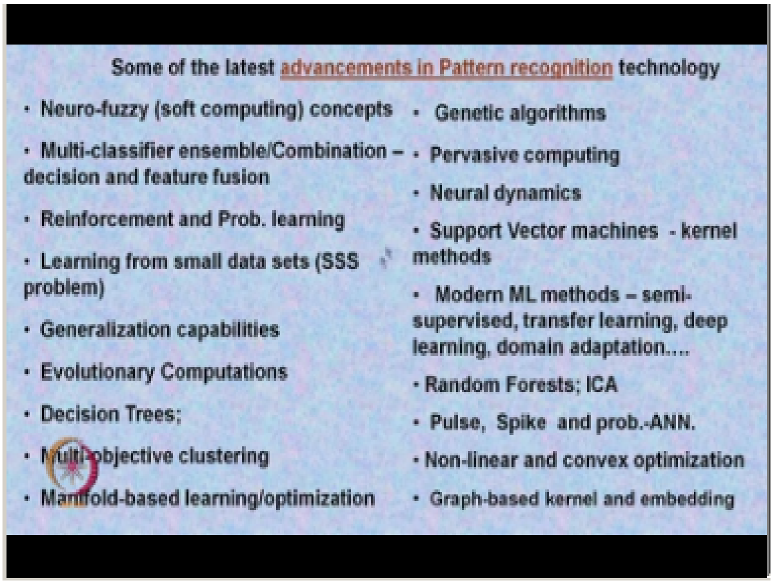
This is along the second Eigen vector as given by the green arrow along the second Eigen vector when you over lap you can see a scatter which is overlapping in all three different classes so this is another example if possible to hand out using a calculator where you can find that the Eigen spectrum or set of the Eigen values are obtained by SVD is going to give you the degree of severability between classes okay.

(Refer Slide Time: 33:19)



So in such cases one is to do while the other methods are supervise learning which is called independent component analysis one can even try that is to separate data to wind up this class I have a put forward a list of

(Refer Slide Time: 33:34)



Some may not be extensive exhaustive but some adventurous are the technology over the last decade or some advent in the field of pattern recognition which is taken place over the last decay and I am just going to name them may not discuss at all any one of them in the detail as talked about by professor murthy and myself that is a course for the beginners in the field of pattern recognition.

And the hope you get encouraged in this field of study from these lectures read books and read much more advanced topics which have come out in several other books as well as advanced literature getting published in which conferences and journals so I will look at them but let us look at the list of techniques which we may not have covered and they are some reason advanced people adopts of computing methods based on nuke technique we talked a little bit of perception.

And new networks you combine that with reasoning century you could architecture for class discriminately rich to in fact these are the methods which are used often to discriminate between classes which are overlapping some of them do not Gaussian distribution and so and so far multi classifier ensemble combination which involves both decision and feature fusion this actually talks about trying to use all set of different classifiers together to form and if a decision the classifier themselves different.

In terms of architecture or they could be trained with different number of samples there are lots of theories of this people also work on the reinforcement and probalastic learning there are which try to handles small data size problem so what is called as small sample size in the field of new network and pattern recognition it can work on generalization capabilities of new networks and pattern recognition algorithms people also work on revolution computation which is very near it is a sub soft computing.

There are methods based on decision trees multi objective clustering and most of the clustering algorithms which you have discussed tries to minimize one criteria if let say take k means it tries to reduce some sort of aquarium distance with respect to the mean but there are but clustering algorithms which actually tries to minimize one critic or try to minimize one criteria with by keeping some other constraint.

So those are some examples of multi objective clustering very recently people are also working on man force based learning and optimization where people take ideas from the areas of the differential geometry and topology people who are genetic algorithms pervasive computing neural dynamics they support vector machines and kernel method modern machine learning methods have contributed a lot in the field of pattern recognition.

In fact there are certain areas in which machine learning and pattern recognition overlap quite a lot where you have to discuss as pattern recognition some of the terms and methods are semi supervise learning transfer learning deep learning domain learning these situations very important because it is possible that you may not have lot number of samples in certain data set but there is an oxalises layer in set which has huge number of data set.

And samples so you could change your classifier which such samples and try to perform we cannot train directly on another data set and train a another data set and tested another one so there are methods in which you try to transfer the information fro0m one domain to other there are methods on other methods of transfer learning and simple supervise which actually you have the human intervention between which tells you that some of the classifier decision made during testing are correct or not.

And you try to readjust the vertices of the particular classifier carrying on there are methods on ransom forest independent component analysis well I will put one minute sentence here remember PCA LDA and then you have ICA people talk about these settings almost together PCA works without class levels it gives you the scatter along the maximum direction okay LDA supervise learning which tries to give you.

This scatter along the direction which you have maximum seperabilty between classes the difference is PCA takes the entire data into account and gives you maximum scatter where and LDA gives you maximum reparability we know that criteria $SW^{-1}x$ that is criteria which tries to maximum and maximize and it you need class labels for that n hence LDA is supervised is unsupervised ICA relax one constraint for a LDA it each class levels.

It is a supervise method of learning but it relax the constraint that the second Eigen vector is normal to the first is normal to the first third is normal to the first and second and so on if you so there are certain data sets in which after you obtain the first principle Eigen vector let us say in which you are maximum separable it is not necessarily true always in all data sets that the second seperabilty will be orthogonal to the first that may not be the case.

So in such cases you need to find out the mechanism where you get the second component of this second maximum degree of separation which should be in a certain direction may not be necessarily orthogonal to the first one but it basically means if you looking at some ith principle component in LDA that may not be orthogonal to the rest previous i-1 principle Eigen vectors that is what ICA tries to do to overcome the restriction on LDA people have worked on pulse spike probabilistic network.

And there are methods of non linear and convex up to optimization related to both minify based learning related to modern methods of ML which people are trying out to solve complex problems in the field of pattern recognition graph based kernel and embedding are also some of the methods in structural pattern recognition and syntactic pattern recognition which we have not covered in this course 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