Applied Multivariate Analysis

Prof. Amit Mitra

Prof. Sharmishtha Mitra

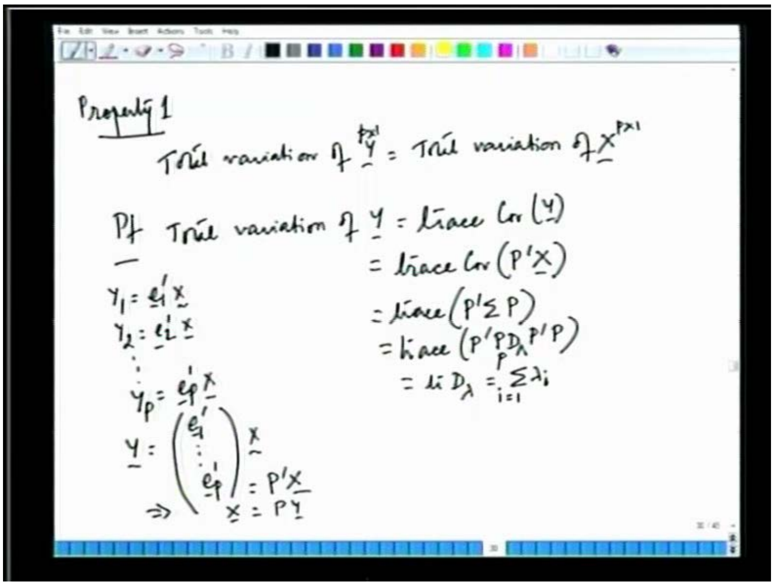
Department of Mathematics and Statistics
Indian Institute of Technology, Kanpur

Lecture No. # 24

Principal Component Analysis

After we have considered construction of PCs, the principal components, which are in fact very special type of linear combinations of the original variables; we said that we are going to check the properties that we had stated in the very begin in the desirable properties of the principal components. The first one of them was that the total variability present in the original data should be equal to the total variability in principal components.

(Refer Slide Time: 00:49)



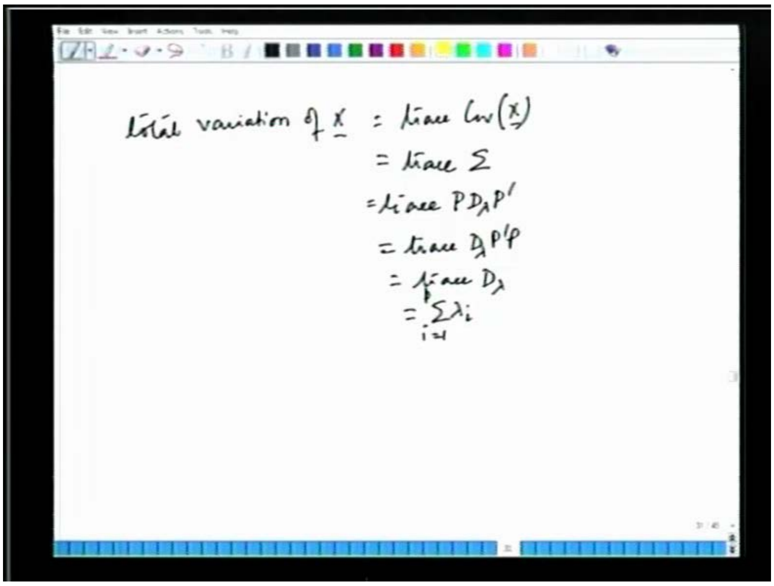
So, let us start with our first property today that is property 1, we have total variation of Y that is of the principle components is equal to total variation of X. Now, at this point, we are still talking about the P dimensional vector of principal components, we have considered as many elements of this principal of or we have considered as many principal components as they are are the originally variables. And we have also defined to total variation as the some of the diagonal elements of the corresponding of variance covariance matrix.

So, let us say that by have total variation of Y, what is this? By my definition of total variation of Y, it should be the trace of the variance covariance matrix of Y. Now we know that this is nothing but trace of trace of the variance covariance matrix of Y; and what is Y? Well Y is nothing but you know that Y is we have constructed the principal components in this way; Y is the linear combination; the first linear combination e 1 X, Y 2 is e 2 X, and similarly Y t is nothing but e P X.

So, this is in fact, if I want to write the Y vector, I can actually represent the whole thing; this is my Y vector, the collection of these Y 1 Y 2 to Y P in the column form, and I have this as e 1 e 2 to e P, all these are transpose, and I have this as x. So, this is, if you recall that this is nothing but our P transpose matrix, because this the P matrix is found, the column of the P matrix are nothing but the orthonormal eigen vectors e 1 to e 2 to e P. So, this is nothing but P X, which also gives me prevalently that X is equal to P Y, but I can handle this situation this relation, so this is trace of nothing but covariance of P transpose X. Why do I do this, because I know, what is the variance covariance matrix of X.

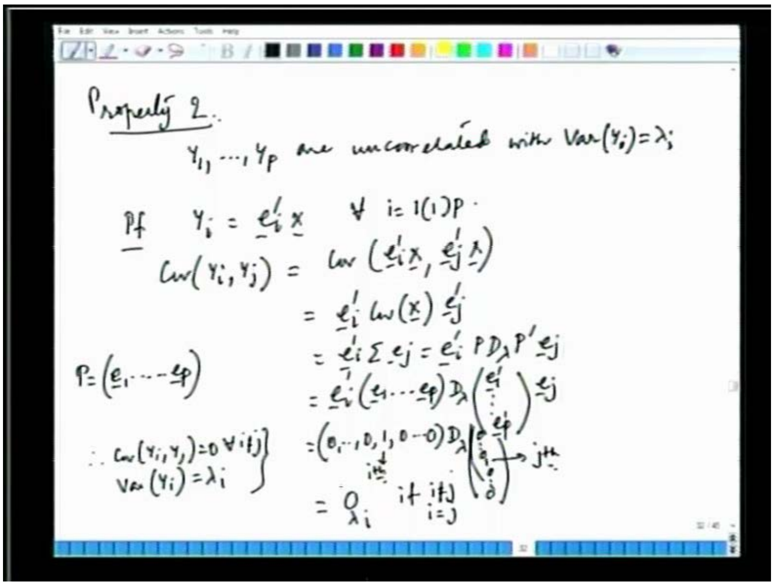
And trace of covariance of this matrix is trace of P transpose sigma P, because I have assumed, I have taken that sigma is the variance covariance matrix of the random variables X. And now, I use the spectral decomposition of sigma, in terms of PD lambda, so this is PD lambda P prime we know, what are these matrices already? And since P is orthogonal matrix, its columns are made up of orthonormal eigen vectors, this is nothing but trace of D lambda; and trace of D lambda is nothing but sum of lambda i(s), this is i from 1 to P.

(Refer Slide Time: 04:31)



And what is total variation of X? Well total variation of X is… Total variation represented by the original data that is X, well by definition it is trace of the covariance matrix of X, which is trace of sigma; and trace of sigma is nothing but trace of P D lambda P prime. Again using the spectral decomposition of sigma, and I am using the fact the trace of A B is trace of B A. So, this is nothing but trace of D lambda P prime P. Now P prime P now P prime P being an a identity matrix I will have to consider only trace of D lambda and which is again equal to some of lambda I from 1 to P giving me the Property of the first property that total variation of Y is in fact equal to total variation of X.

(Refer Slide Time: 05:30)

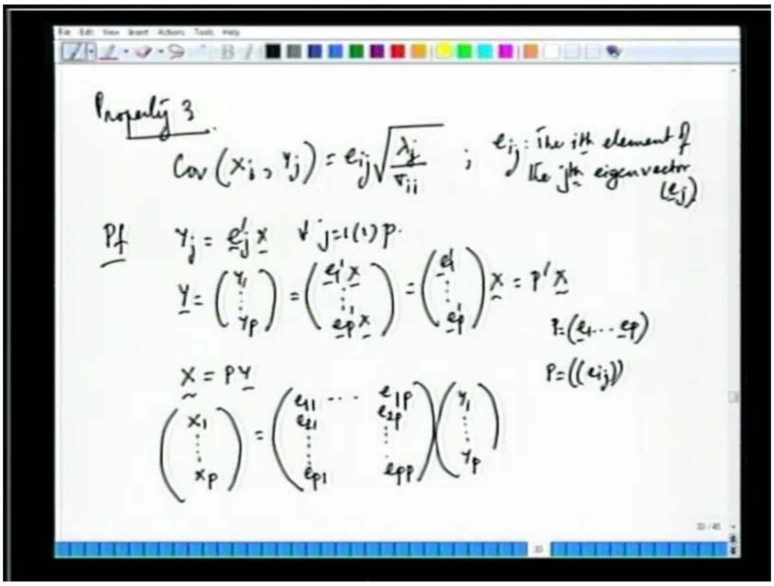


The next one was we said that the principle component Y 1, Y 2, Y P, they should be uncorrelated. So, let us see this, the principal components so constructed, they are boundary uncorrelated, so we have Y 1, Y 2 to Y P are uncorrelated. We combine one more thing with this with variance of Y i, the i eth principal component is the i largest eigen value of the covariance matrix of X. So, we have Y 1, Y 2 to Y P are uncorrelated with variance of, with variance of Y i being lambda i, this is again very simple, because we have the i eth principle component Y i is nothing but e i transpose X; and this is true for all i from 1 to 2 p. So, that if I really have to consider, what is covariance matrix between Y i and Y j, where it is nothing but covariance between e i prime X and e j prime X.

So, this is in fact e i prime covariance of X and e j prime, and this covariance X is nothing but our sigma matrix, the matrix which is very important to us for calculation of eigen values and eigen vectors, and we use this spectral decomposed representation of this matrix to write e i prime with P D lambda P transpose and e j right, where the columns of P matrix, these are nothing but e 1 e 2 to e P. Once I realize this, I can also realize very quickly that where if I write this as e i transpose, and here I have e 1 e 2 to e P, then I have a D lambda, and here I have transpose of this that is e 1 transpose to e p transpose, and e j here. Now since e i(s) are orthonormal eigen vectors, I am only going to have a 1 in the i eth position, when this e i transpose combines with the e i here.

So, that 1 in the i eth position and 0 elsewhere D lambda, and then again I have… Now this is important, because this is in the i eth position, whereas this is in the j eth position. So, they are never matching, and I am as a result, I have a 0, this is j eth, and as a result, I have 0, if well if i is not equal to j, and well if it is equal to i is equal to j. So, 1 is going match with 1 here, and resultant element that is going to remain is this i eth element of D lambda and that is going to be lambda i, if i is equal to j. So, in 1 go, I can proof, I have been able to proof that this not only proofs that covariance of which implies that covariance of Y i Y j is 0, for all i not equal to j and variance of Y i, which is covariance between Y i and Y with itself, this is equal to lambda i.

(Refer Slide Time: 09:20)



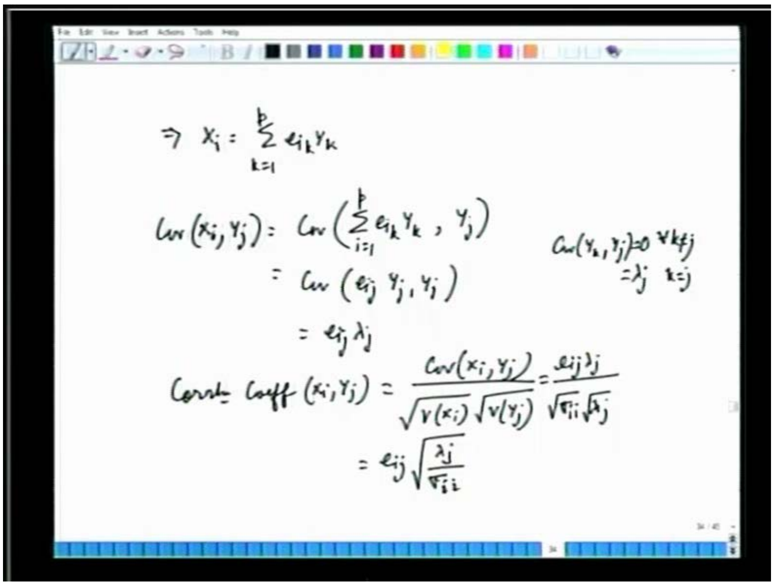
The next property that we talk of is property 3. Well, we know covariance between X i X j is sigma i j; the i j eth element of the sigma matrix have constructed simple components in such a ways such that the covariance between Y i and Y j that is the covariance between the i eth principal component, and the j eth principal component is 0, they are uncorrelated. Now, what about covariance between the original variable and principal components? Say I consider covariance between X i, any X i are not just the Y i, but say any Y j. So, let us look into that, because this relation, this covariance or this correlation coefficient that I am going to obtain from here is going to help me a lot in data interpretation.

So, this is the next Property that we are talking of that is covariance between X i that is the i eth original variable, and the j eth principal component Y j is e ij root of lambda i by sigma ii. Now I know, what is lambda i? It is the variance of the i eth principal component, and sigma ii is the variance of the original variable x, but what is e i j, where e i j is the j eth sorry the i eth element of the j eigen vector that is e j right. This lambda is lambda j 0 i, because I am considering y j here, so this is lambda j, variance of Y j, this is sigma i i variance of X i, and e i j is coming from the j eth eigen vector e j, it is the i eth element of that eigen vector. So our, again we start from the form of the principal component Y j, let us try to express it in terms of X, then we are able to use the known results, and Y j is nothing but e j transpose X. This is true for all j from 1 to P.

And then I have to consider, so well I know Y is again Y 1 to Y P, and these are nothing but e 1 X to e p X right. So, this is nothing but e 1 transpose to e P transpose combine with X, and this is P transpose X. So, I have Y is nothing but P transpose X; P being orthogonal, I have X is equal to P Y. And then if I have to consider, well X is this like Y, it is also P dimensional, I have X 1 X 2 to X P; and I simply have the P matrix here; if you recall the P matrix is nothing but e 1 e 2 to e P. And now I am saying that I am denoting the i eth element of the j eth eigen vector by e i j.

So, I can open up this all these vectors right, so for e 1, if I have to write the first element of the first eigen vector, I will write e 11, then the second element of the second this column vector, I am going to write e 2 1. So, this is exactly how I write up matrix P, if P is nothing but, and then P is nothing but represented by this e i j, and I have e 11 e 21 in this way up to e p 1, and finally, the first element of the P eth vector, P eth eigen vector, which is e 1P, then the second element of e P, which is e 2P, and then I have the last element of the P element of the P vector giving me e PP, and then this is Y 1 to Y P right.

(Refer Slide Time: 14:01)



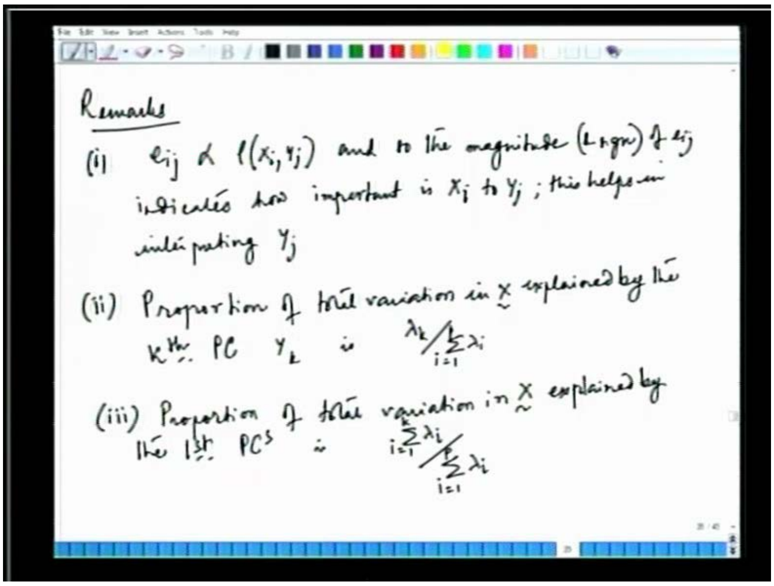
So, this gives me a new representation of X; I can say that X i, if I consider the i eth element of the X vector, it is nothing but sum of e ik Y k, and k the sum is over the common dimension obviously, it is P. So, now if I have to consider covariance between X i and Y j; well, I will have to consider covariance between X i, I am using this form of X i, this is sum over i, and for Y j, I am simply taking Y j. Now, I can do this, because I know what already what is the covariance between covariance between any Y k and Y j. So, it is simple for me to handle this. Now I know that covariance between any Y k and Y j, where k is not equal to, j is equal to 0. So, if I open up this sum, and consider the covariance with Y j, the only term that will make a difference in the in this covariance will make a contribution rather in this differences the P eth term that is when Y P come terms with the coefficient e iP; others are all equal to 0, because I am considering covariance between Y k and Y j, and k is not equal to i.

So, this is basically nothing but covariance between e ij Y j, and then sorry this is going to give me… So, this is going to give me covariance between e ij and e ik, that is the only part that that is going to remain, and I have e ij with Y j and Y j rest of this are not making any contribution, and covariance between Y j and Y j is of course, lambda j and e i j remains as such giving me e ij lambda j, because I know that covariance between Y k and Y j is 0 for all k not equal to j, and is lambda j, if k is equal to j. So, I am using this relation.

And then I have, I can also consider the correlation coefficient correlation coefficient between X i and Y j; this is covariance X i Y j by root variance of X i times root variance of y j, I know all this; so, I can immediately right this as e ij lambda j, and this is root of sigma ii, and this is root of lambda j. So, this is giving me e ij with root lambda j by sigma ii. Let us go back to the statement, we had stated that now we have to make a small correction again here; this this is not actually the covariance, but in fact the correlation coefficient. The covariance being equal to e ij lambda j; so if I consider the correlation coefficient between X i and Y j, then only I get this whole thing.

So from here, it is clear that this element e ij, it plays a very significance role in the correlation coefficient between X i and Y j right, because its sign is going to determine the sign of the correlation coefficient, I have the rest of it is ratio of two standard deviation always positive. So, its sign as well as its magnitude are very important for me to interpret the relationship between the original variable and the principal components. If there are some more parameters, which can be now checked from whatever we have obtained, and let us list down few important of those parameters, which are very important, which are very useful for data interpretation part.

(Refer Slide Time: 18:32)



The first one is I have say we can note down this as some remark, so some notes. So, the first one is the one that we have just written that is e ij that is the i eth element of the j eth eigen vector, this is proportional to correlation coefficient between X i and Y j; and so the magnitude of e ij and sign also of e ij indicates, how important is X i to Y j; and this helps in interpreting, because after all these principal components that we have constructed, these are not some direct observable variables. These are linear combinations of a number of several variables. So to give, we have to give some interpretation in terms of the physical variables, and if we have certain relationship like this, it becomes little easier for us to give the interpretation as to how this are related. So, this helps in interpreting Y j.

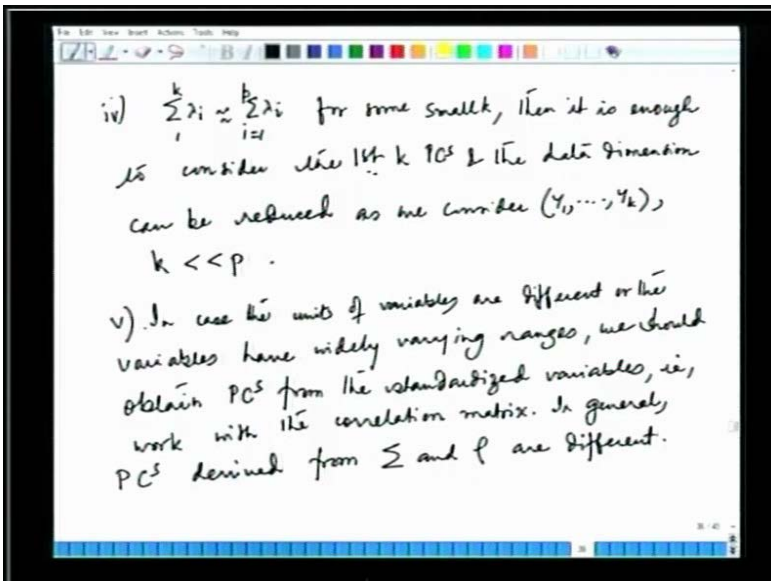
The second one is the proportion; now if you recall, we have now talked about one property of the principal components, which we said is very important that is though the total variation in X is equal to total variation in Y, ultimately to make the whole excise of fruitful one, I should have the total variation in Y, it should be explained by a very few number of total variation of a fewer number of the principal components that is you had recall we had written some thing like total variation in Y is approximately equal to the variation in Y 1 to Y 2 to Y k, where k is a number, which is much smaller than the true dimension P.

Now, that we have not checked, because that is that is going to come out from actual data only when we have data, we can actually see that how this what is the effect or what is the extent to which this has been this approximation is effective. But before that, I can also say that the proportional of total variability or proportion of total variation in X explained by the k eth principal component right. What is a total variation in X? It is sum of, sum the eigen values, and the variation explained by the k principal component is nothing but its variance, and that is equal to lambda k. So, this the proportion of total variation in X explained by the k eth PC Y k is equal to lambda k by sum of lambda i, i from 1 to P.

And what do you want? Ideally, we would like to have this lambda 1 by summation lambda k, a very large number. This should be as close to 1 as possible, then should come lambda 2 by sum of lambda i(s) on in this way. As a value of k increases, we would like to have this factor to be smaller and smaller, so that I can stop at a small value of k, and say that if I consider these many principal components, it is good enough for me to explain the total variability present in the data.

Then again it is not necessary that I have to consider the principal component in isolation that I I should say that one explains the the highest variability, so I stop at 1, you know, I can consider 1 and 2, two-dimensional is also good enough for me. So, in that case, I have to look at the total variability explained by the first principal component, and the second sample component. So, in this way, that is another one more parameter that is of importance to me is proportion of total variation in X, explained by the first k principal components that is that is some of lambda i, i from 1 to k and this is the total variation some of lambda i from 1 to P all of them.

(Refer Slide Time: 23:54)



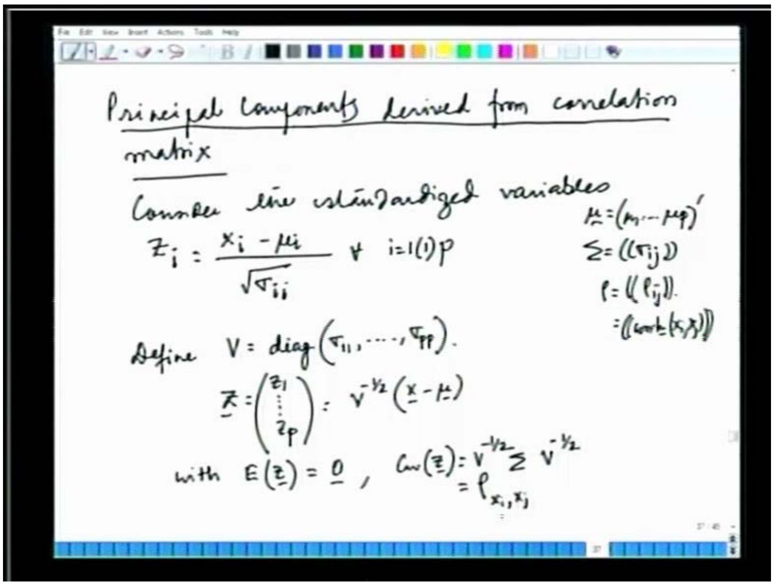
And then comes the important point that if I have sum of lambda i, i from 1 to k is a approximately equal to some of lambda i from 1 to P. Then I will say, and that two for some small k, smaller the better, for some small k, then it is enough to consider the first k principal components, and the data dimension and the data dimension can be reduced k that was our goal, primary goal; the data dimension can be reduced as we consider a part of y that is we consider now the k dimensional vector y 1 y 2 to y k, k is much smaller then P. Now one thing is we have to be careful about this is number 4, after the first three points, one thing we have to careful about we are saying that e ij is a very important factor for me to explain the important, the association between X i and y i j

Now, if I have the data, e ij is essentially the i eth element of the j eigen vector. Now if I have the data, which are in different units of measurement or data which have very different and wide ranges of variation. Then if I consider the data as it is, and then focus on these e ij(s) and other measures, then the comparison would not be truly correct. What we have to do in such a situation is we have to try and make the measurements unit free and so that we get the true picture of association. So this point also to be noted down, before we actually do the excise of that. So, standardizing the variable; so in case of, in case the units of variables, units of measurements of the variables, so the units of variables are different or the variables, which is bound to happen in practical situations, we shall we shall consider variables from different fields, which we have, which are bound to have different units of measurements, and they can also have totally different and very high range of variation. So, this we should and are the variables have widely varying ranges we should obtain principal components from the standardized variables.

So, standardizing the variables means, essentially we will have to divide the we have to first take the difference of this variables from the corresponding means and divide by the corresponding standard deviations. Now initially, when we working with the plane, the raw data; our matrix of importance to us was the variance covariance matrix of X; that is the sigma matrix. Now, you can easily realize that if we standardize the variables, the whole matrix of importance to us is now the correlation coefficient matrix of the variables. So now, we have to the everything is getting replaced by the correlation coefficient matrix rho, which which work initially we were worked with the sigma, the covariance matrix sigma.

Now, our matrix of importance is the correlation coefficient matrix rho. So, this is obtain from standardize variables; and we have to consider the that is work, we can put it like this, work with the correlation matrix, instead of the covariance matrix. Now, question remains that will the principal components be same, if we work with the sigma on the rho matrices? Now answer is no. In general the PC(s) derived from sigma and rho are different right.

(Refer Slide Time: 29:29)



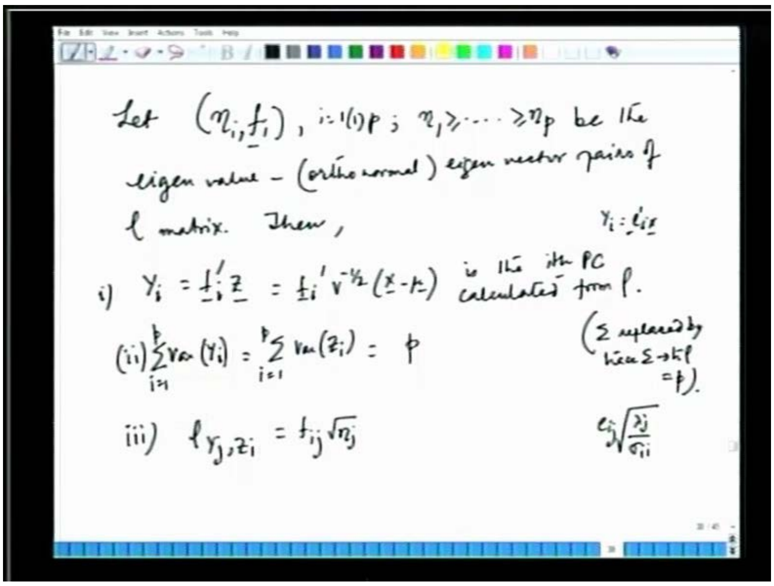
And now we consider calculation of principal components from the correlation matrix. So, I have principal components derived from correlation matrix. We really do not have do not have to do much here, we will see that how a simple transformation helps us and we will get all the results immediately from the once that we have really proved. So, first is we consider the standardize variables that is the first step; and use a very general notations set for those. So, I say that Z i is nothing but X i minus mu i that is the mean of the i eth random variable, we had earlier say that the mean vector; well if you have to write the covariance of the term covariance with in terms of a expectation etcetera, we will have to use mu being the mean vector mu 1 to mu P.

So, we have and sigma, the variance covariance matrixes sigma i j right. So, this is X i minus mu i by root over sigma ii, for all the i from 1 to P; there are P variables X 1 to X P, I define P standardize variables Z 1 to Z P. And now I just define a new matrix V, define V which is nothing but diagonal of the sigma, diagonal elements of the sigma matrix, so this is diagonal sigma 11 to sigma pp right. And then well I have z, how can I write this Z vectors, Z vector is nothing but Z 1 to Z p right. In terms of the X vector, if I have to do it, well simply, it simply that now that I have defined what is V; and the diagonal elements are all positives. So, V is a positive definite matrix, I considered the inverse square root of V, by speckle decomposition, by actually spectral decomposition of sigma, this is good enough to give me this also, V square root inverse of V, and then rest of it is nothing but X minus mu.

So, I combine all these standardize variables in this Z vector, in this fashion using X(s) mu(s) and these elements sigma ii root sigma ii coming into the picture in this way. So, now, with that I have Z is, this. So, what is expectation Z as expected, because I have standardized this and got z. So, this is nothing but the null vector, because expectation X is mu, and what is variance covariance matrix of z? Well, this is V minus half, covariance of X minus mu which is same as covariance of X, which is sigma, and then this is V minus half again. And what is that? Well, if sigma is the variance covariance matrix or the dispersion matrix of something of X right, and you have these are getting getting divided by this, the ij eth element of sigma matrix, what by what is it getting divided? Well, it is getting divided by this the element here that is it is getting multiplied here by sigma ii raise to the power minus half, and on the any other side its sigma j j raise to the power minus half.

So, in effect I am getting nothing but the correlation coefficients rho of X ij right. So, rho is nothing but rho ij, and what is rho ij? It is the correlation coefficient matrix of X ij or this as rho ij, the matrix of correlation of X i and X j. Well, this was the matrix of covariance between X i and X j. So, you may write it X i X j as a subscript, but there is no need actually. Just like we have sigma as covariance between X i i j and rho is the correlation matrix between X i and X j

(Refer Slide Time: 34:33)

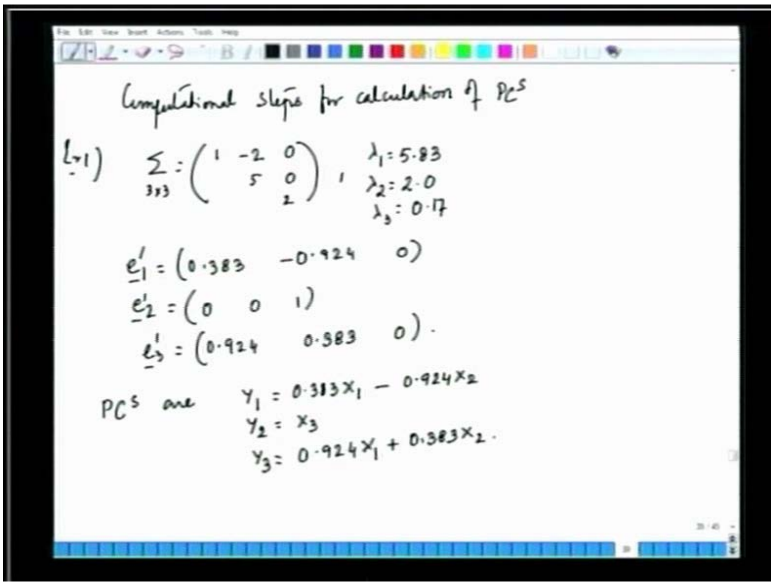


And then if I consider the eigen value and orthonormal eigen vector that is I say there I had lambda i e i, here let me have eta i and f i, i from 1 to P. The same way as we had considered the lambdas, I have lambda eta 1 greater than equal to up to eta P be the eigen value and orthonormal eigen vector pairs of rho known; instead of sigma, I had earlier lambda i and e i eigen value eigen vector pair of sigma; I do not calculate the eigen values and corresponding Eigen vectors for sigma. Now instead I calculate it for the correlation matrix rho, and then I get this is pairs of rho matrix, then then what is the i eth principal component? Let me use the same notation Y i for its principal component, and then what is the i eth principal component? Well, originally if you recall I had Y i is e i times X e i transpose X.

Then now using this notation I should have this as if I transpose Z right. And this is nothing but f i transpose v inverse X minus mu. So, this is the i eth principal component calculated from rho right. And if I have to consider the variance of this; well I consider the variance of some of the variance of first one, the first point to be noted is this. And then I have the total variation in variation in y that is sum of variance of Y i from 1 to P, there are P of them as before, and then this is nothing but, because I had variance of Y i equal to variance of X i. So, similarly I should have this variance of Y i this new Y i, their variance, sum of their variances should be equal to variances of Z i, i from 1 to P, and what is this? What is this matrix, actually this is giving me the correlation coefficient matrix, and then what are the diagonal elements, their diagonal elements are 1. So, some of them are actually equal to P. So sigma, everywhere sigma replace by rho. Trace of sigma, now I have trace of rho, which is equal to P right.

And another point that to be noted is, what is the covariance between Y j and Z i? So, as we had considered Y j and X i, well, there it was e ij if you recall it was e ij lambda j by sigma i . So, here it is going to be f ij, f ij right; with lambda j is now being replaced by eta j, and sigma ii is now actually equal to 1, because I am considering the rho matrix, instead of the sigma matrix. So, we will continue with the calculation from a data, calculation of principal components from data after this. Now we consider the computational steps required for the calculation of the principal components. First we consider a situation, where we are calculating eigen value eigen vectors from the, from the sigma matrix that is the covariance matrix. And next we will take up correlation matrixes, well.

(Refer Slide Time: 38:50)



So, the first example that we considered here, this is the computational steps for calculation of PC(s) calculation of construction of PC(s). Now that we know the theory behind this, we actually know that the this important information, the important data that we required is nothing but the variance covariance matrix of the original variables sigma. So from here, we can calculate the eigen values, eigen vectors and the principal components also. So, that is what, that is what matters to me, and then I have the first example say the example 1, where I have the sigma matrix given to me as 1 minus 2 0 5 0 2, it is a symmetric matrix obviously, I know that the data dimension is 3, I have a 3 by 3 square matrix, which is symmetric and which is positive definite also right.

So, I calculate eigen values from here, the three eigen values, and I denote them as the first one is the highest one in this way. So, I calculate the three eigen values, and then I put them as lambda 1 is equal to 5.83 lambda 2, the second one in magnitude is 2 and lambda 3 is 0.17, all positive in decreasing order 5.832 and 0.17, I have to consider the corresponding, calculate the corresponding orthonormal eigen vector. First I get the eigen vectors, and then normalize them to get the orthonormal eigen vectors, and these are e 1 is the eigen vector, the orthonormal Eigen vector corresponding to lambda 1 equal to 5.83 is this one, its 0.383 minus 0.924 and then 0.

The second orthonormal eigen vector that is the eigen vector corresponding to the second highest eigen value, which is equal to 2, this is equal to 0 0 1 and e 3; similarly is 0.924

0.383 and 0. If I know this, I know all the three PCs, the PCs are well, the first one is Y 1, it is e 1 transpose X. So, I know what is what are the co-efficient with each of the X variables for the y 1 for the first principal components? So, that is 0.383 X 1 minus 0.924 X 2 that is all, because of the third coefficient is 0. The second principal component is identically equal to the third variable, the original variable X 3; and the third principal component is 0.924 plus this X 1 plus 0.383 X 1 X 2, that is the third principal component. And if that is so, I am going to now see that how important these principal components are; how much of the total variation they can explain? So, this is the the the five characteristic that we are talked about.

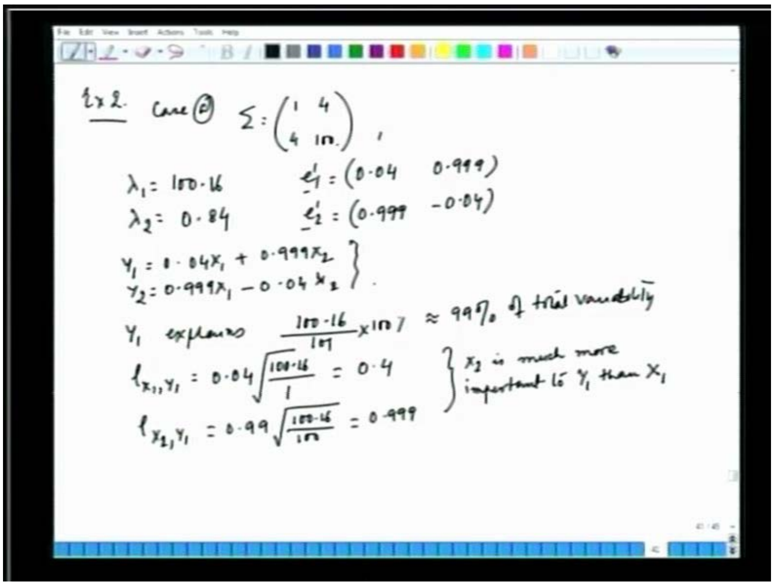
So, this tells me that I can reduce the data dimension from 3 to 2, because I see that the first 2 principal component, they are explaining 98 percent of the total variation, and this is the reduction of dimension by even 1 degree helps me a lot I can project this data or an X Y plane, and I can see various characteristics of this multivariate data right. And what about the covariance coefficient I have rho of X I, Y k; this is if you recall we have use a notation Y j, and this was e i j, and I have root of lambda j by sigma i i.

So, that I have rho of X 1 with Y 1 is 0.383, that is I am considering e i j, and considering the j th Eigen vector; that is the first eigen vector, and the first element of that that is the 0.583, and I consider the the first eigen value, because j is equal to once. So, that is 5.83, and variance of the first that is the first element of the sigma matrix, and this is equal to 0.25; this tells me about the relative importance between X 1, and Y 1. Similarly, if I consider X 2 and Y 1, what I get is that is k remains 1 here, but i is 2.

So, I have thus first Eigen vector, but the second element from 8, and that is minus 0.924 the first Eigen value, and the second element second diagonal element of the sigma matrix giving me 5, and that is equal to minus 0.998. I can see that there is a direct relationship between X 1 and Y 1, the first variable and the first principal component, whereas there is a negative relationship between X 2 and Y 1, but nevertheless importance’s are really high, there high in magnitude one is 0.925, another is minus 0.998. What about the third one X 3 with Y 1 will we have to consider the third element of the first eigen vector, and that is equal to 0 .

So, there is no degree of a association between the first principal component, and the third original variable. So, in this way we can, if we 1, 2 go to the second principal component, if you stock at the first principal component we need not go to the relation coefficient between the X j's X i's, and y 2, but if we want to retain y 2 we should ideally, because the only when the first 2 are combine, we have 98 percent of variability explained. And then one can also consider consider the calculate the correlations between X j's - the different X j with Y 2.

(Refer Slide Time: 47:20)

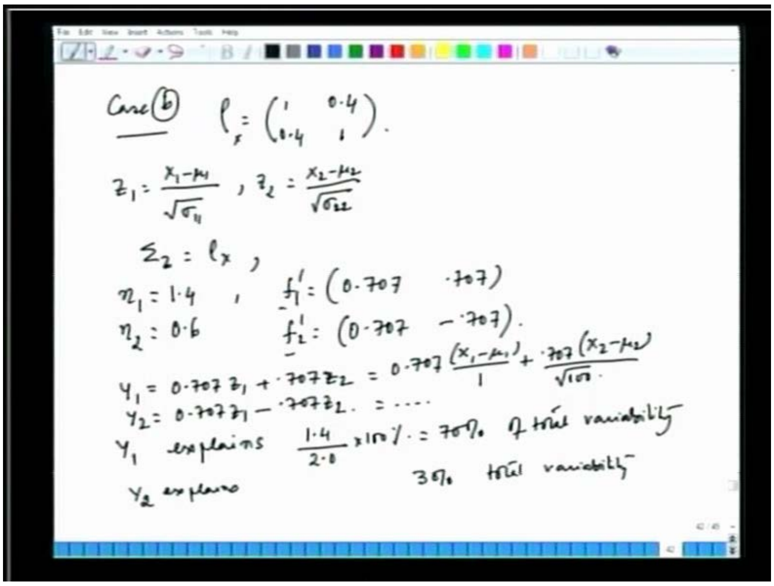


Let us take up the second example. And here we work with the, we will take a smaller matrix and work with both that is the dispersion as well as the correlation matrix. So, here first take consider this is case a, I have sigma this equal to 1 4 4 and 100; that is my variance covariance matrix. And my correlation coefficient matrix we can write it later on, when we consider case b. And what we have from here is the eigen values from this sigma matrix, I have two Eigen values lambda 1 and lambda 2, and these are 100.16 and 0.84. The corresponding orthonormal eigen vectors are e 1, this is 0.04 0.999 and e 2 is 0.999 minus 0.04. So, then I consider Y 1 and Y 2; Y 1 is nothing but I have to consider e 1 prime x, so which is 0.0 4 X 1 plus 0.999 X 2, e 2 is 0.99 X 1 this the reverse relationship, and then minus with a negative sign here, and minus 0.04 X 2 right.

So, then Y 1 explains, how much of the total variability it explains? The first eigen value lambda 1 by the some of it, which hardly makes a different, because I have 100.16 and 0.84 added to it, making this as 101, and this times 100 percent that is approximately equal to 99 percent of the total variability, 99 percent of total variability. And Y 2 explains a rest of it. So no need to consider, no need to go to Y 2; if I consider Y 2, there is no reduction in the data dimension, and if I consider Y 2 now, this explain 100 percent. But at least from two dimensions, I have come down to one dimension that is the purpose, whether it really helps mean data dimension reduction. And then I consider the correlation coefficients also I have rho X 1 Y 1, this is important, and this is only the eigen vector the first eigen vector is coming into the picture, this is the e i j part that is e 1 1 here, and then I have the variance of the first, the standard division of the first eigen of the first principal component that is 100.16 by sigma i i sigma 11, that is 1. And this is giving me 0.0 4; 0.4 actual not 0. 0 4; and what about the correlation coefficient between Y 1 X 2; well this sigma now going to be 0.9 9 it is the first eigen value 100.16, but now it is the second variance sigma 22 and this is equal to 0.99 right

So, this terms may that X to is much more important the degree of association of Y 1 the only principal component that we are considering is much higher with second variable then with the first variable perfectly this no problem with that. So, X 2 is much more important to X Y 1 that an X 1. Now how, what is the true degree of this importance is a because of the much higher value of this second variance. So, look into questions like this, what we should do is actual standardize the variables and work with the correlation coefficient matrix rather than the covariance matrix. So, that is what we are going to do.

(Refer Slide Time: 51:51)

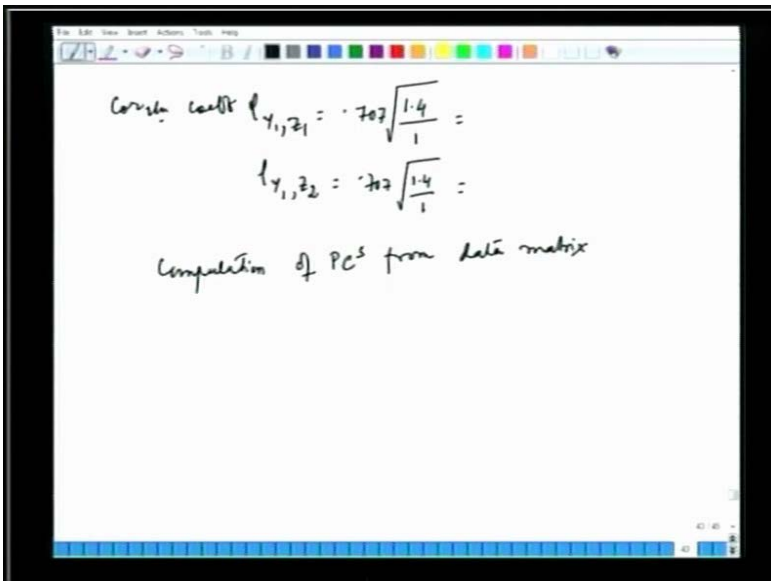


Now we consider case b part, where we have the correlation coefficient matrix calculated from the covariance matrix, just by that defining the V inverse, square root inverse of V matrix, and then we can easily get the correlation coefficient matrix, and that is equal to 1.4, I have and then what is this is 0.4 and 1 right. That is we are considering the standardized variables that is the standardize variables Z 1 is nothing but X 1 minus mu 1 by sigma 11 root of this. And Z 2 is nothing but X 2 minus mu 2 by root of sigma 2 2. And then sigma Z matrix is actually the rho X matrix right. This is rho x, but I can also have the sigma Z matrix, and then this is nothing but this matrix, what I am getting here; 1 0.4 0.4 and 1 right.

And then I calculate the eigen values eta 1 and eta 2 from this rho matrix, and eta 1 turns out 2 be 1.4 eta 2 is 0.6 right. And the eigen vectors corresponding orthonormal eigen vectors f 1 prime and f 2 prime also have to be calculated; this is 0.707 0.707 and f 2 is 0.707 with minus 0.707 right. And then we have Y 1, let us see that is the first thing that we going to check Y 1 explains; well what is Y 1 and Y 2; that is trivial. We need not write, but let us write here also Y 1 is nothing but if you want to write in terms of Z 1 and Z 2, I should write it ideally as 0.707 Z 1 plus 0.707 Z 2, these are to be replaced by X 1 and X 2 and Y 2 is 0.707 Z 1 minus 0.707 Z 2. And Y 1 explains, and ideally then further you can write this as 0.707 X 1 minus mu 1 by the element of the first first element of the sigma matrix that is 1, and you have 0.707 X 2 minus mu 2 by second diagonal element that is root 100 and in this way right.

So, Y 1 explains what percentage of total variability it explains? Its own eigen value, its own variance, which is 1.4 divided by the total variability, which is sum of eta 1 and eta 2 that is 2 times 100 percent and that is equal to 70 percent of total variability. And if I consider Y 2 part of it, it explains if the rest of it right, because Y 1 and Y 2 will together consider the total variability. Now, that is what we see here the difference here, if you work with the covariance matrix, and the correlation matrix, when we had worked with the covariance matrix, we saw that Y 1, the principal component can calculated from the covariance matrix explained 99 percent of the total variability, but here when we consider the correlation matrix, we see that Y 1 explains only about 70 percent of the total variation.

(Refer Slide Time: 56:23)



And let us look into the correlation coefficient part also, where we have correlation between if I have to consider the correlation coefficient of this is rho Y 1 with X 1. Well, I have to consider the first element of the first vector of the f vector that is 0.707. And then we have to consider the root of the first root of the first eigen value calculated from the correlation matrix that is going to be eta 1, which is root of 1.4; and then the variance, the first variance of X 1, this is well, this is actually equal to Z 1 right, and the variance here is nothing but 1. We can calculate this to get what is the correlation coefficient; and similarly I can have what is correlation coefficient between Y 1 and Z 2 right? So, this is going to be the e ij that is I am going to use the first element of the second vector, that is again 0.707. I will consider root of 4 and 1 here.

The interesting thing to see here is that Y 1 to Z 1 and Y 1 to Z 2, the degree of association is same, because I have the same value coming here, because of the fact that e 11 right; e 11 is equal to e 12, and I have the same eta i value coming here, and Z coming now from the correlation matrix have equal variances, which is equal to 1. So, that is precisely equal rho of Y 1 Z 1 is equal to rho of Y 1 and Z 2. A next example will be computation of PCs from data matrix; in these examples, what we had seen was that I did not have the data matrix, I only had the population variance covariance matrix given to me, that was good enough for me, I calculated the eigen values, eigen vectors, and then the principal components, and then I also talked about how much of the variabilities are getting, is getting explained by the principal component and so on and so forth.

But in reality, what we are going to have with going to have with us actually, the data, the the information on the multivariate data variables, and we are going to calculate our… We are going to calculate the whole sample covariance matrix from the data; and proceed for calculation of eigen value and eigen vector from that variance covariance matrix.