Multivariate Procedures with R

Prof. Shalabh

Department of Mathematics and Statistics

IIT Kanpur

Week – 11

Lecture – 52

Principle Component Analysis: Concepts and Theoretical Setup

Hello friends, welcome to the course Multivariate Procedure with R. So, in this lecture we are going to begin a new topic which is Principle Component Analysis, shortly called as PCA. So, the first question arises what is PCA or the Principle Component Analysis. So, we know that whenever we are working in real data usually if the process is complicated then there are many variables which are going to affect the outcome of the process. Some of the variables will be important that who are contributing more to explain the variability in the model and some variable may be contributing very less. And based on that we can say that we can have two types of variables which are one which are important and another not so important.

So, importance here is in terms of the variability that means those variables which are contributing more in explaining the variation in the data. So, the question here is how to find out these types of variables. So, this Principle Component Analysis is a way out, it is a statistical methodology which helps us and it explain us how we can obtain such type of variable in such a way such that the total variability of the system remains the same. So, it will try to find out another set of variables suppose we have some variable say x1, x2, xp.

So, it will try to find out another set of variables say y1, y2, yp which are some special types of linear combination of x1, x2, xp. But the beauty of such transformation will be that whatsoever it will be the variability due to x1, x2, xp will be the same as the variability due to the y1, y2, yp. However, those x1, x2, xp they may not be independent, but this y1, y2, yp will be independent. And because of this feature the new set of variables y1, y2, yp can be used in different places like as in solving the problem of

multicollinearity in the regression analysis. And these y1, y2, yp they are actually called as principle components.

So, in this lecture I am going to explain you the basic concepts, the basic definition I will be using very minimum mathematics, but there is a one good thing that I will try to take here sometime to explain you the basic fundamentals what is happening behind the principle component analysis. But the solution for the real-life data is very simple means in a simple word I can say you simply have to find out the eigen values and eigen vectors that is all and that you will see that how I am going to implement it. So, in this lecture I will try to take a very small data of say three variables and then based on that I will try to show you that how you can create the principle components and then in the next lecture I will try to take a data set where I will try to implement the R software to do the principle component analysis. So, let us begin this lecture and try to understand the basic concepts of principle component analysis. Okay, so now if you try to see we are going to talk about the concepts and theoretical setup of principle component analysis in this lecture.

So, the first question comes here what is this principle component analysis or popularly called as PCA. Suppose there are a large number of variable which are affecting the outcome of an experiment and the question is how many variables should be included or discarded and before beginning the experiment actually it is unknown to us that how many variables should be included and so that is why more than the required number of variables are usually included. But when we are trying to increase the number of variables then it has its own complications. For example, the number of elements in the mean vector and covariance matrix will also increase. For example, if I say if in our data vector say here x there are 10 variables that is p = 10, p is indicating the number of variables.

Then the mean vector will have 10 elements, covariance matrix have 10 into 10 that is 100 elements, the number of unique variances and covariances will be p into p + 1 by2 which is 55 and the number of unique covariance will be the total number of elements 55 - 10 variance - p that is 45. But if you try to increase this number of variable from here say this p = 10, to p = 15 then mean vector will have 15 elements, covariance matrix will have 15 into 15 that is 225 elements, number of unique variances and covariance will become 120 and number of unique covariance will become 105. So, if you try to compare this number 10 with 45, 120 with 55 you can see here this number is increasing a lot. So, once this number increases then different types of complexities also enters and then they try to create some other type of issues. So, one way to handle the situation is to consider the normalized linear combination of the random variable such that the total variance is not lost and we simply say in simple words that the structure of the variability in the data

is not lost that is the total variance remains the same and actually these normalized linear combinations are the principal components.

Why are we calling them as normalized? So, those linear combinations are called normalized linear combination who have the sum of squares of the coefficient as 1. Suppose if I say here a1 x1 + a2 x2 and suppose if a1 square + a2 square = 1 then I would say that a1 x1 + a2 x2 is normalized linear combination of x1 and x2. So, the principal components are the normalized linear combination of the random variables which have some special properties in terms of variances. So, remember variability is our main concern when we are trying to find out the principal components. So, in some way what are we doing when we are trying to find out the principal component analysis? We have a set of variable x1, x2, x3 and from there we are going to find new set of variable y1, y2, yp.

So, we are essentially trying to find out a new coordinate system. So, in effect transforming the original vector variable to the vector of principal component amounts to a rotation of the coordinate axis to a new coordinate system that has some inherent statistical properties. For example, the new coordinate system due to y's will also have the same variability as the older coordinate system based on x's, x1, x2, xp. So, in practice the total variation present is accounted only for a few principal component. What does this mean? So, when you are having say x1, x2, xp, p variables and you are finding out a new coordinate system which has new variable y1, y2, yp these are the principal component and then you will see that we try to find out what is the contribution of this y1, y2, yp in explaining the variability and so we will try to pick up some of the principal components who will take the majority of the shear of the total variation.

So, what is the basic purpose of principal component analysis? It is a method of reducing the dimensionality of the data without affecting the total variation of the data. The linear combination with large variances are found. For example, if you have 10 variables then it may be possible that only 6 variables out of 10 are sufficient and take care of the last part of the variance. So, why should I consider all the 10 variables? 10 variables suppose they are taking care of 100% variability and suppose 6 variables are taking care of the suppose 95-97% of variability. So, why should I spend my time, energy, money etc. in collecting the data on say on the remaining 4 variables. And this also gives up a solution to the multicollinearity. So, multicollinearity in the setup of multiple linear regression analysis arises when there exists exact relationship or a type of relationship among the independent variable. If you remember we had considered this model y = x beta + epsilon where x has the variables x1, x2, x3. And we had assumed that rank of x matrix = p which is full column rank matrix.

But suppose if there exists some linear dependencies among x1, x2, x3 then the rank of x will not be a full column rank matrix and then we will not be able to find out the ordinary least square estimator and it will create its own problem. So, whenever there are linear dependencies in the among the independent variable this is called as a problem of multicollinearity. So, this principle component gives us a solution right because your x1, x2, x3 are not independent but y1, y2, yp which are the principle components they will be independent and they will help in solving the problem of multicollinearity and this technique is called as principle component regression. Principle component regression which is quite popular but anyway I am not going to consider it here but I thought I must inform you. So, in practice yeah sometime the interpretation of the principle component is difficult but the operations are simple and they reduce the variability up to a large extent right.

So, now we try to consider the setup to create the basis for the principle component. So, suppose x is a random vector with mean vector 0 and covariance matrix sigma. So, without loss of generality we assume that the variables are measured from their respective mean otherwise you can take this mean 0 to be here mu but anyway we are because we are more concerned about the structure of the covariance matrix so the actual distribution of x is not important for us whether it has mean 0 or not. Otherwise without any problem I can take the mean vector mu also here which is non-zero. On the other hand if we additionally assume that x follows multivariate normal distribution then it has more meaning which can be given to the principle components because and then the calculation also becomes simpler and anyway after obtaining the principle component we also try to conduct the test of hypothesis for the principle component for which we will need this assumption.

Although I am not going to consider here the test of hypothesis related to the principle components and yeah that also gives us some sort of information but that is valid only when we try to assume the multivariate normal distribution for x vector. So, now means if I try to think algebraically the principle component are particular combination of the p random variables x1, x2, xp and geometrically these linear combination represent the collection of a new coordinate system which is obtained by rotating the original system x1, x2, xp as the coordinate axis and the new axis represent the direction with maximum variability and provides a simpler and more parsimonious description of the covariance structure. So, now let me try to explain you in a little bit mathematical way. So, suppose this random vector x is of order $p \times 1$ and it has p variables here like this and it has a $p \times p$ symmetric positive definite covariance matrix sigma. Now suppose the eigenvalues or I can call them as characteristic root alternatively both are the same thing of this sigma

matrix are say lambda1, lambda2, lambda p because sigma is the positive definite matrix so it will have p characteristic roots or say eigenvalues and suppose we try to order them like as lambda1 is the maximum value out of or lambda1, lambda2, lambda p and lambda p is the minimum value of among lambda1, lambda2, lambda p.

So, that we are assuming otherwise what will happen that you will obtain here lambda1, lambda2, lambda 3 then you will make here say lambda bracket 1 as maximum value of lambda1, lambda2, lambda 3 then lambda2 will make the second largest and lambda 3 will be the minimum value, but in order to make the expression more simpler I said of using this notation of this order statistics I am simply using here lambda1, lambda2, lambda p. So, now we try to consider here p linear combination of x1, x2, xp. Suppose y1 is indicated as say L1 transpose X where L1 is the vector with the elements L11, L21, Lp1 and similarly here y2 is another linear combination L2 transpose X and similarly here yp is the pth linear combination Lp transpose X which has here this type of element. So, these are the simple linear combinations. So, because we have assumed that x has a covariance matrix sigma so we can find out the covariance matrix of here L transpose X.

$$Y_1 = \underline{l}_1'\underline{X} = l_{11}X_1 + l_{21}X_2 + \cdots + l_{p1}X_p$$
$$Y_2 = \underline{l}_2'\underline{X} = l_{12}X_1 + l_{22}X_2 + \cdots + l_{p2}X_p$$
$$\cdots \quad \cdots \quad \cdots$$
$$Y_p = \underline{l}_p'\underline{X} = l_{1p}X_1 + l_{2p}X_2 + \cdots + l_{pp}X_p$$

So, from there I can say that variance of yi is Li transpose sigma Li and covariance between yi and yk is Li sigma Lk where i goes from 1 to p, k goes from 1 to p, and in the covariance $i \neq k$. So, the principle component are those uncorrelated linear combination of y1, y2, yp whose variances are as large as possible, very important statement. I have incorporated here two aspects, uncorrelated linear combination and variances are as large as possible. So, if you try to see here that the first principle component which we are going to obtain is the linear combination with the maximum variance. So, because from here you will have here variance of y1, variance of y2 up to here variance of yp.

$$Var(Y_i) = \underline{l}_i'\Sigma\underline{l}_i, i = 1, 2, \ldots, p,$$
$$Cov(Y_i, Y_k) = \underline{l}_i'\Sigma\underline{l}_k, i \neq k = 1, 2, \ldots, p$$

So, out of that, that linear combination which has got the maximum variance that linear combination will be the first principle component. And its variance is suppose variance of y1 = L1 sigma L1 transpose sigma L1. So, now the problem is this, now this variance can be changed. So, essentially this variance L1 transpose sigma L1 can be increased by multiplying any L1 by some constant. So, to eliminate this indeterminacy what we can do? We can restrict our attention to the coefficients vectors of unit length.

That means these are the coefficient vectors of this L1 sigma L1 are something like L1. So, we would like to have the coefficient vector whose length = 1 that is L1 transpose L1 = 1 or it is something like here L1i square i goes from 1 to p should be = 1. So, now after understanding this thing we can define the first principle component as the linear combination L1 transpose X that maximizes the variance of L1 transpose X subject to the condition L1 transpose L1 = 1. So, that is why I call this L1 transpose X as the normalized linear combination. So, we have obtained the first principle component.

Now we have to obtain the second principle component. So, that is again going to be another linear combination say for example L2 transpose X but now I am putting here one more condition that whatsoever be the second largest variance. Corresponding to that whichever linear combination is there that will be called as second principle component. So, the second principle component L2 transpose X that maximize the variance of L2 transpose X subject to the condition that L2 transpose L2 = 1 that means L2 transpose X is normalized. So, that we did in the case of first principle component also but now we are adding here one more condition that the covariance between L1 transpose X and L2 transpose X = 0.

$$Var(\underline{l}_2'\underline{X}) \text{ subject to } \underline{l}_2'\underline{l}_2 = 1 \text{ and}$$
$$Cov(\underline{l}_1'\underline{X}, \underline{l}_2'\underline{X}) = 0.$$

That means this y1 and y2 are uncorrelated. And similarly, if I want to find out here the third principle component so that is going to be something like L3 transpose X which has got the maximum variance with the condition that L3 transpose L3 = 1 and this L3 transpose X is uncorrelated with L1 transpose X and L2 transpose X. So, this third principle component will be uncorrelated with the first and second principle component. Similarly for the fourth principle component that is going to have the similar condition that its variance has to be maximize subject to the sum of squares of the coefficient should be = 1 and the fourth principle component should be uncorrelated of the first three principle components. So, in general if I try to say the i-th principal component is Li transpose X which is obtained such that it maximizes the variance of Li transpose X subject to the condition that Li transpose Li = 1 and covariance of Li transpose X and Lk transpose X is 0 for all i less than k.

$$Var(\underline{l}_i'\underline{X}) \text{ subject to } \underline{l}_i'\underline{l}_i = 1 \text{ and}$$
$$Cov(\underline{l}_i'\underline{X}, \underline{l}_k'\underline{X}) = 0 \text{ for } i < k.$$

So, this is how we can do it here. So, now it is very easy to find out. So, now I have given you the basic definition of this principle component. Now I try to do here something in which I am simply going to find out the characteristic root of the sigma matrix and the corresponding characteristic vectors. So, I am going to find out the

eigenvalues and eigenvectors of the covariance matrix sigma and then I will try to associate them with the principle component.

So, let us try to understand how. So, we have assumed that sigma is a p × p symmetric positive definite matrix of a random vector X like this because it is a positive definite matrix so all its eigenvalues are also going to be positive. Suppose the eigenvalues or the characteristics roots of sigma be lambda1, lambda2, lambda p such that I have arranged them as lambda1 is greater than = lambda2 and greater than = lambda p. So, lambda1 is the maximum value among all lambda1, lambda2, lambda p and lambda p is the minimum value among all lambda1, lambda2, lambda p. And for each of this lambda1, lambda2, lambda p we can obtain the eigenvectors also which are called as characteristic vectors. And suppose the eigenvectors of lambda1 is indicated by E 1, eigenvector of eigenvalue lambda2 is indicated by say E 2 and so on.

So, lambda1, lambda2, lambda p have the eigenvectors e1, e2, ep. Now if you try to understand what I try to do here. Here you have understood what is my here ith principal component which is the normalized linear combination of x1, x2, xp such that variance of Li transpose X is maximum and y1, y2, yi-1 they are uncorrelated with y i. So, now if you try to see what I am doing. The ith principal component is given by the linear combination yi = ei transpose X.

The i^{th} principle component is given by linear combination

$$Y_i = \underline{e}_i'\underline{X} = e_1X_1 + e_2X_2 + \cdots + e_pX_p, i = 1, 2, \ldots, p$$

with $Var(Y_i) = \underline{e}_i'\Sigma\underline{e}_i = \lambda_i, i = 1, 2, \ldots, p$

and $Cov(Y_i, Y_k) = \underline{e}_i'\Sigma\underline{e}_k = 0, \ i \neq k.$

What is here ei? ei is the ith eigenvector. So, now if you try to see here I can simply write here yi = ei transpose X which case something like e1 x1 + e2 x2 + ep xp, i goes from 1 to p. Yeah, I am just indicating here otherwise there will be your here i also here but anyway it does not make any difference because I will be using this straightforward terminology for ei transpose X. This yi has got a variance ei transpose sigma ei.

And now this is here same as lambda i. What is here lambda i? Lambda i is the i-th eigenvalue, right? Lambda i. And the covariance between yi and yk is ei transpose sigma ek which = 0 for i ≠ k. So, if you try to see what you have done here, you have found the eigenvalues and eigenvectors of the sigma matrix and you have used them in the construction of the i-th principal component yi. The coefficients are going to be determined by the values of eigenvectors and the variance of that yi is going to be

indicated by the value of eigenvalue that is lambda i. So, now what you have to do? You simply have to look at lambda1, lambda2, lambda p which are the eigenvalues of sigma matrix and then you have to see whichever is the maximum that will be the variance of the first principal component.

Now corresponding to that lambda you try to find out the corresponding eigenvector and that eigenvector will help you in the construction of i-th principal component, right? So, in case if I mean some lambda i's are equal given the choice of ei's and hence yi are not unique but we are not worried because in practice it will really happen. Anyway, we have a solution in the theory also. So, now if you try to see we have the first condition that whenever we are trying to find out a new coordinate system with y1, y2, yp their covariance structure should remain the same. So, you can see here that we have found here p principal component y1, y2, here yp which are something like y i = here ei transpose X. So, then the total variance if you try to see which is the trace of sigma is simply the sum of the elements on the diagonal elements sigma11 + sigma22 + sigma pp.

$$\boldsymbol{tr(\Sigma) = \sigma_{11} + \sigma_{22} + \cdots \sigma_{pp} = \sum_{i=1}^{p} Var(X_i) = \lambda_1 + \lambda_2 + \cdots + \lambda_p = \sum_{i=1}^{p} Var(Y_i)}$$

And now if you try to see I am saying that lambda i is the variance of yi. So, this sum of sigma11, sigma22, sigma pp which is nothing but your sum of the variance of xi = lambda1 + lambda2 + lambda p and this is the same as the sum of the variance of yi. So, this sum of all lambda i's is giving us the value of the total variance in the x1, x2, xp that is sigma matrix. So, consequently what will happen? The proportion of total variance due to which is explained by the kth principal component becomes here lambda k upon sum of all lambda i's.

$$\boldsymbol{e_{ki} \propto \text{Correlation}(Y_i, X_k) = \frac{e_{ki}\sqrt{\lambda_i}}{\sqrt{\sigma_{kk}}}, i, k = 1, 2, \ldots, p}$$

This will be the proportion. And the magnitude of the means kth element in the ei vector that is eki here measures the importance of the kth variable to the i-th principal component irrespective of all other variables. So, in particular what will happen this eki is proportional to the correlation between yi and xk and this can be shown that this is equal to eki into square root of lambda i divided by square root of sigma kk where i and k goes from 1 to p. So, you can see here what we are trying to do. You have got here all lambda1, lambda2, lambda p they may have different values. So, you are trying to see what is the proportion which is contributed by each of this lambda i and this can be obtained by this proportion, individual lambda divided by the total values of lambdas.

So, this is how you can find out the principal component. So, instead of using it directly into the R software and where we can find out this principal component directly, let me try to take here a very simple example to explain you how are we going to construct this

principal component. So, suppose there is a covariance matrix of three variables say x1, x2, x3 and it is given here by this matrix sigma and my random variables are x1, x2, x3. So, I try to create this sigma in the form of a matrix in the R software. Well, I am not going to use here the command, direct command to find out the principal component analysis but I will try to create them manually and for intermediate calculation I will try to use the R software.

So, this is your here sigma matrix. Now if I try to find out the eigenvalues and eigenvectors of sigma then it can be obtained by the command eigen that we have done if you remember in the beginning of the course. So, now this eigenvalues lambda1, lambda2, lambda p they are coming out to be here like this. These are the value of lambdas. Now you have to identify which is the lambda1, which is lambda2 and which is lambda 3. So, you try to see that which is the maximum value out of these three.

So, you can see here this is the maximum value. So, let me call it here as a lambda1. Then the second largest value here is this one lambda2 which = 2 and the minimum value out of these three values of lambda1, lambda2, lambda 3 is 0.171 which is my here lambda 3. Now corresponding to this lambda1 I can find out here the characteristic vector or say eigenvector. And similarly corresponding to the lambda2 means I can find out here this characteristic vector and corresponding to this lambda 3 I can find out here this characteristic vector.

Now using this characteristic vector, I can very easily write down the principal component. So, first principal component here is y1 = e 1 transpose X. So, you can see here I am writing here - 0.38 this here. Then the second element 0.92 here 0.00 actually not 0 but there is some very small value. So, this is 0.00 into x1, x2, x3. So, this will come out to be - 0.38 x1 + 0.92 x2. And similarly, here in the second case second principal component y2 = e 2 x1. I am using here these 0 0 1 here like a 0 0 1 and then x1, x2, x3 which is only here x3. And third principal component which is here these are three values 0.92, 0.38 and 0.00. I am taking only the two values after the decimal point and x1, x2, x3.

So, it will come out to be 0.92 x1 + 0.38 x2. So, you can see here that this is my here the first principal component which has got the variance of y1 = here 5.82. And similarly for the second and third principal component also. So, that is what I am trying to show you here that whether it matches with the outcome or not. So, if you try to see here this is my here first principal component. So, I am writing here first principal component and I try to find out its variance of y1. So, using the elementary statistical methods that it will be means variance of ax you know this is A square variance of x and variance of ax + by = a

square variance of x + b square variance of y and + twice of covariance between x y and then multiplied by here a b like twice of a b or I can write down here say + twice of a b into covariance of X Y.

Using this formula I have written here like this expression and if you try to substitute the values from the diagonal elements of the sigma matrix of variance of x1, x2, x3 and their covariances from the off-diagonal elements this value is coming out to be 5.83. And if you try to compare this is nothing but your this value. So, I have shown you that even if you try to compute manually the variance of the first principal component from the sigma matrix that is the same as the eigenvalue which has got the largest value among the three eigenvalues. Similarly, you can also find here the variance of y2 and variance of y3 which are the variance nothing but your lambda2 and lambda 3 you can see here and sigma is here by here like this.

So, I have used these values here in the computation. Now if you try to verify whether the total variance structure remain the same or not. So, if you try to see some of the variances of x i which are you here sigma11, sigma22, sigma33 this is sigma11, this is sigma22 and this is sigma33. So, this is 1 + 5 + 2, 8 and if you try to use this variances and you try to find out the sum of lambda1 + lambda2 + lambda 3 here this will come out to be here 8. So, you can see here it is not a very difficult thing to do and now based on that you can find out here that what is the proportion of the variance which is accounted by the first, second and third principal component.

So, the proportion of total variance accounted by first principal component is lambda1 divided by lambda1 + lambda2 + lambda 3 which is which is found here 5.3 by 8 which is 0.73 which is equivalent to 73 percent. Similarly, the proportion of total variance accounted by second principal component is lambda2 upon lambda1 + lambda2 + lambda 3 which is 2 upon 8 which is equivalent to 25 percent. And similarly, the proportion of total variance is variance accounted by the third principal component is lambda 3 divided by lambda1 + lambda2 + lambda 3 which is 0.17 upon 8 which is equivalent to 2 percent. So, you can see here that if you try to see the first principal component is taking care of 73 percent of the total variance and second principal component is taking care of 25 percent of the variation. So, if you try to add them together then all together, they are going to take care of the 98 percent of the total variability in the system.

And the third principal component is taking care only of the 2 percent of the variabilities in the system. So, you can see here that in this case instead of using x1, x2, x3 if I try to

use only the first two principal component y1 and y2 then possibly I am getting the accuracy of 98 percent of the variability being taken care. So, I can think of ignoring third principal component and instead of working x1, x2 and x3 we will prefer to work only with two principal component. Well, there is going to be a small loss of information but that is acceptable and if you try to replicate this thing for a bigger setup then instead of having 15 variables you can have only 8 or 9 variables for example. Now regarding the correlation coefficient between y1 and x1 which is indicated by rho y1 x1 which I have computed here this is something like 0.925.

The correlation coefficient between y1 and x2 is rho y1 x2 using the same formula that I explained you I have computed it to be here - 0.998. So, if you try to see here and then yeah means that also if you try to find out here rho y2 and x1 that means second principal component with respect to x1 and x2 their correlation coefficient is 0 and the correlation coefficient of y2 with x3 that is expected it is here coming out to be here 1 that was expected. So, all other correlations can be can be ignores in the third component is not so important.

So, now you can understand what I am trying to say. So, if you try to see here both this correlation coefficients, they are pretty high. So, x1 and x2 individually about equally are equally important to the first principal component right and in the second case x1 and x2 are more so important to the y2. And if you try to see here the correlation of y2 with x3 it = here 1 that was expected right. Now the next question comes here that I have told you these things on the basis of the some known covariance matrix but in practice this sample covariance matrix is unknown.

So, how you can obtain this principal component. So, we are understanding here now the maximum likelihood estimation of the principal components. So, suppose x1, x2, xn be a random sample of say size small n each variable each observation is taken on p random variable from multivariate normal population with mean vector mu and covariance matrix sigma. Sigma has p eigenvalue lambda1 lambda2 lambda p lambda1 is the maximum and lambda p is the minimum value out of lambda1 lambda2 lambda p and their corresponding eigenvectors are here suppose e1, e2, ep. Now I am not giving you here the proof but I am giving you here the final result that a set of maximum likelihood estimator of lambda1 lambda2 lambda p and e1, e2, ep they are the roots lambda1 hat greater than = lambda2 hat greater than or = lambda p hat of this equation sigma hat - lambda hat ip = 0 and a set of corresponding vectors e1 hat, e2 hat, ep hat they satisfy this

characteristic equation sigma hat - lambda i hat ip into ei hat = 0 with the constraint that ei hat transpose ei hat = 1.

Let $\underline{x}_1, \underline{x}_2, \ldots, \underline{x}_n$ $(n > p)$ be a random sample from $N\left(\underline{\mu}, \Sigma\right)$ where

Σ matrix has p eigen values $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_p \geq 0$ with

corresponding eigen vectors $\underline{e}_1, \underline{e}_2, \ldots, \underline{e}_p$.

Then a set of maximum likelihood estimators of $\lambda_1, \lambda_2, \ldots, \lambda_p$ and

$\underline{e}_1, \underline{e}_2, \ldots, \underline{e}_p$ consists of the roots $\hat{\lambda}_1 \geq \hat{\lambda}_2 \geq \cdots \geq \hat{\lambda}_p$ of $|\hat{\Sigma} - \hat{\lambda} I_p| = 0$

and a set of corresponding vectors $\underline{\hat{e}}_1, \underline{\hat{e}}_2, \ldots, \underline{\hat{e}}_p$ satisfying

$(\hat{\Sigma} - \hat{\lambda}_i I_p)\underline{\hat{e}}_i = 0$ with $\underline{\hat{e}}_i' \, \underline{\hat{e}}_i = 1$

where $\hat{\Sigma}$ is maximum likelihood estimator of Σ.

And this sigma hat is the maximum likelihood estimator of sigma. So, if you try to understand what I am trying to say here you have here normal mu sigma population you have a from here you obtain here a sample say x1, x2, xn then you try to obtain here sigma hat as a maximum likelihood estimator then now you try to take this covariance matrix estimator and try to find out its characteristic roots. Then whatsoever be the value of here lambda which are required to be used here lambda1 lambda2 lambda p hat they will satisfy this equation now you obtain here lambda1 lambda2 lambda p hat on the basis of given cut of data and then you try to find out the characteristic root using this equation lambda i hat ip into say ei hat = 0.

So, this will give you the ith characteristic vector or say eigen vector and then you have now all e1, e2, ep and then you can constitute your principle components. So, now we come to an end to this lecture and you can see here it was basically a conceptual lecture where I was trying to explain you the concept of principle component analysis and you can see that after understanding the whole set up the solution comes out to be very simple that whatsoever be the covariance matrix or the sample covariance matrix try to find out their characteristic roots and from there you try to find out the characteristic vectors and then whatsoever be the values in the characteristic vector they are going to help in constructing the linear combination which are your principle component and the value of the characteristic roots or the eigen values they are indicating the variances of the principle components. So, you can see here that the solution is pretty simple very straight forward but the main thing is that how are you going to interpret it because now y1, y2, y p they are the linear combinations of x1, x2, x p.

So, now many times people say that okay this x1, x2, x p they may have different units so in that case they prefer to scale the data first such that all the observations on x1, x2, x p they become unit free and then they try to conduct this principle component analysis

but anyway in real data finding the principle component on say unscaled x1, x2, x p and scaled x1, x2, x p may give you a different result. So, then you have to see that how are you going to decide which one to choose. We have different types of test of hypothesis based on which we can take a call we can also try to see which of the components are more close to the real-life data which are representing it in a better way then we try to choose it. But here basic objective is that we try to understand what are principle component, what they are interpret, what is their interpretation and how they are going to help us in reducing the dimensionality of the data. So, principle component analysis is a very important technique for the reducing the dimensionality of the data and that is why it is also called as dimensionality reduction technique.

There are some other techniques also but principle component is one among those. So, you try to understand these concepts and in the next lecture I will try to take an example and I will try to conduct the principle component analysis in the R software. But to understand their outcome it is very important for you that you understand these basics so that you can correlate the outcome of the software with these concepts. So, you try to practice it, try to understand it and I will see you in the next lecture till then goodbye. Thank you.