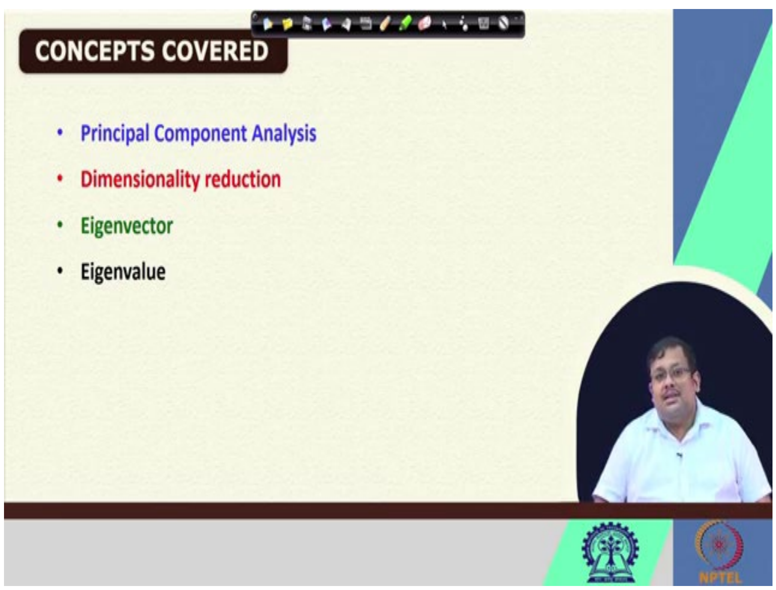
Machine Learning for Soil and Crop Management
Professor Somsubhra Chakraborty
Agricultural and Food Engineering Department
Indian Institute of Technology, Kharagpur
Lecture 11
Principal Component Analysis and Regression Applications in Agriculture

Welcome friends to this 11th lecture of NPTEL online certification course of Machine Learning for Soil and Crop Management. And today we are going to start week 3. And the topic of this week is principal component analysis and regression application in agriculture.

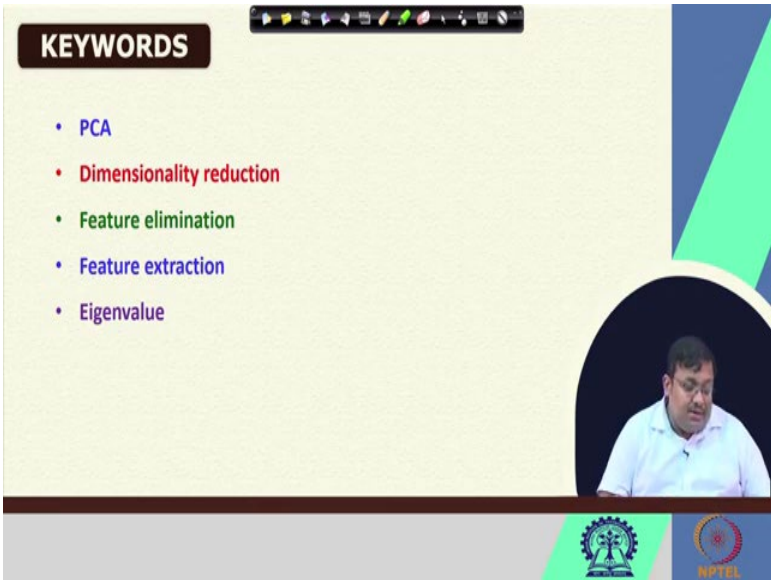
Although we are going to discuss principle component analysis and also different types of machine learning, prediction algorithms. We are going to confine our discussion on to soil and crop applications only. Although the applications for PCA and other prediction algorithms are widely variable and they can be used to any sector of agriculture.

(Refer Slide Time: 1:15)



So, this is the first lecture of week 3. And these are the concepts, which we are going to cover in this first lecture. We are going to first cover what is principle compound analysis. We are going to discuss, what is the benefit of principle component analysis and how principle component analysis can helps in dimensionality reduction. Also we are going to discuss different important terminologies of principle component analysis like eigenvectors, eigenvalue.

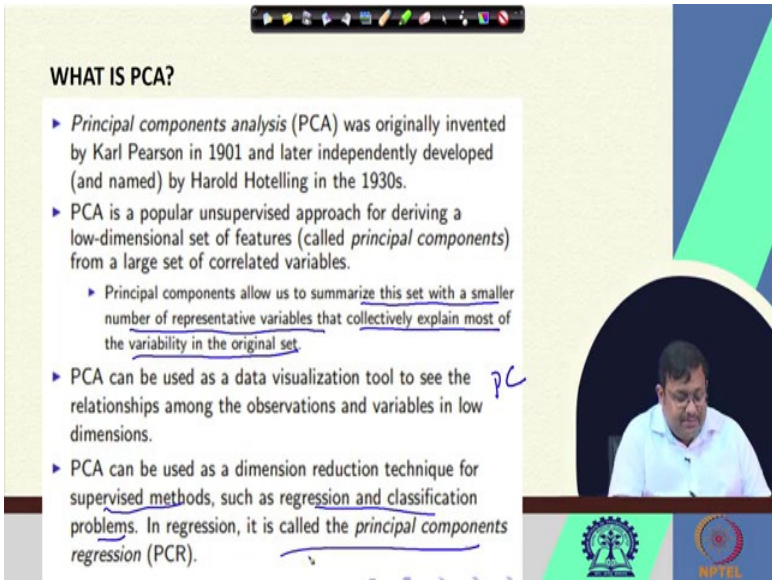
(Refer Slide Time: 1:45)



And also these are the important keywords, which are going to discuss today. First of all we are going to discuss what is PCA, a principal component analysis and dimensionality reduction. Then feature elimination, feature extraction. What is the difference between feature elimination and feature extraction?

And also we are going to discuss about how step by step you can do principle component analysis, in that we are going to also discuss in 1 of the important keyword that is a eigen value.

(Refer Slide Time: 2:26)



So, let us start with the PCA. PCA is the short form of principal component analysis and principal component analysis is a very important dimensionality reduction algorithm. As that in our previous lectures, we have discussed what is overfitting? Overfitting means when we include, when the model is too much trained on to the data, it can learns from the noise and it becomes so optimistic that it cannot generalize the unknown samples.

So, although calibration, or training, statistics for overfitted models are very good, but those models fails, fail measurably when we validate the results using the testing data set. So, couple of reason for over fitting is we have discussed, what are the reasons for overfitting, 1 of the major reason is a proper, improper designing of the experiment and inclusion of too many variables.

And also we have seen that when there is a multicollinearity, we have also defined what is multicollinearity. So, when there is a sufficient amount of multicollinearity, that also destroys the any multivariate prediction model, for example, multiple linear regression model, which is a parametric model. So, this type of problems, or pitfalls in a in MLR, or multiple linear regression, or any multivariate algorithms can be reduced by dimensionality reduction. What is dimensionality reduction?

What how many types of dimensional reduction, approaches are there we are going to discuss. But the major application of PCA is to decompose, or in other words to combine all the independent variables in the, in the data set to and give some and transform them into some new variables and then selectively remove some of those new variables we call them principal components, which are not so much informative.

Now, what are the pros and cons of this type of approaches we are going to also discuss, but at this point of time just remember that it is a dimensionality reduction, or reduction approach. So, let us start with the principle component analysis, the principal component analysis was originally invented by the scientist called Pearson in 1901 and later independently developed and named by Harold Hotelling in the year 1930s.

So, PCA is a very popular unsupervised approach for deriving a low dimensional set of features. What is dimensional? Dimension means when there are multiple variables, or features are there we call them high dimensional data, 1 of the example of high dimensional data is spectral data, which we are going to discuss in upcoming lectures.

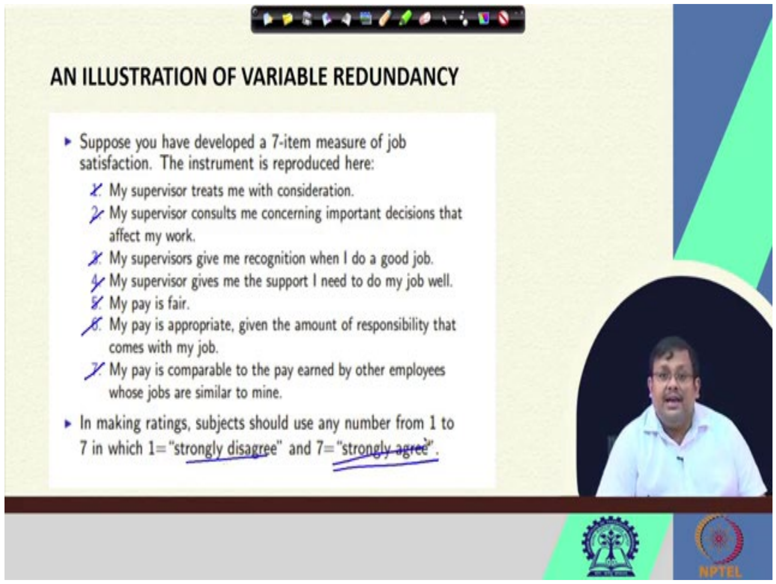
So, when they there are high dimensional data, PCA is a popular unsupervised approach for deriving a low dimensional set from this high dimensional data, from a large set of over correlated variables. If there is a multi-collinearity, how we can reduce that multi collinearity and get a small subset of features, that PCA generally execute.

Now, principal component allows us to summarize this, principal component allows us to summarize this set with a smaller number of representative variables, that collectively explain most of the variability in the original set. So, the idea is we decompose the high dimensional data into a smaller number of representative variables, we call them principal components, or PC, that collectively explain most of the variability in the originals data set.

So, how we execute that we are going to discuss. So, PCA can be used as a data visualization tool also to see the relationship among the observation and the variables in the low dimension. Sometime if you want to see any clustering pattern among the data set, we generally apply principle component analysis, because principle component analysis helps us to identify the cluster in the data set. And the similar observations are grouped together.

So, we are going to see also that. So, PCA can be used as a dimensionality reductionity, reduction technique for also supervised methods, such as regression and classification problems. So, in regression it is called the principal component regression, we are going to also discuss what is principal component regression.

(Refer Slide Time: 8:16)

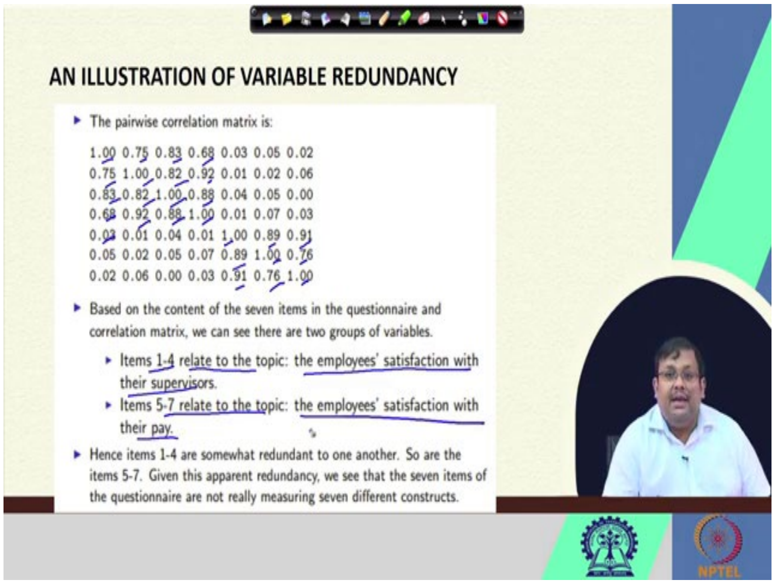


Now, let us consider a particular problem, suppose you have developed a 7- item measure of job satisfaction. And the instrument is reproduced here. So, basically there are 7 questions, or 7 points on to on which the employee has to give some kind of rating. So, first point is my supervisor treats me with consideration. Second is my supervisor concerns me with concerning important decision, that affect my work.

Third 1 is my supervisor give me recognition, supervisors give me recognition when I do a good job. Fourth 1 is my supervisor give me support, give me the support I need to do my job well. Fifth point is my pay is fair. Sixth point is my pay is appropriate given the amount of responsibility, that comes with my job. And seven point is my pay is comparable to the pay earned by other employees, whose jobs are similar to mine.

So, there are 7 points, so in making these ratings subjects, or employees should have, should use any number from 1 to 7, in which 1, it denotes the strongly agree and 7 denotes the, 1 is denotes strongly disagree, whereas 7 denotes the strongly agree. So, after we after the subjects read them these questions and we do a pairwise correlation for these 7, 7 points, we can see a pairwise correlation matrix.

(Refer Slide Time: 9:53)

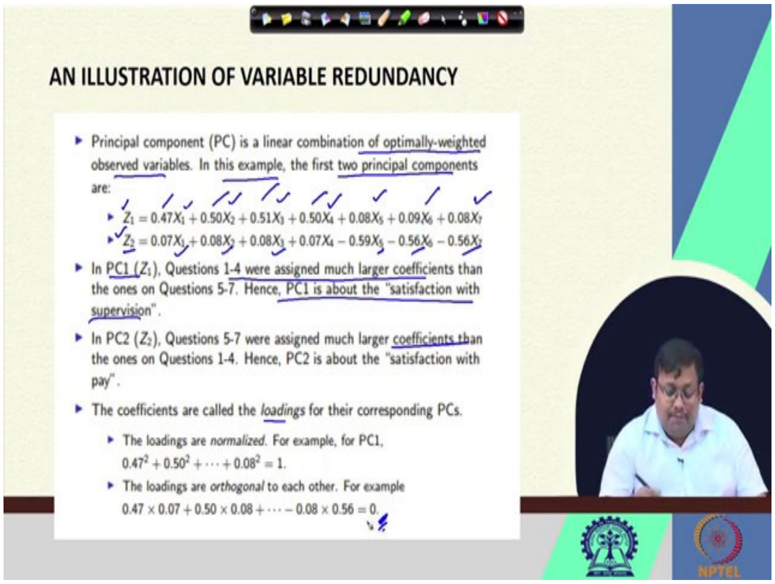


So, this is a pairwise correlation matrix and based on this pairwise correlation matrix, we can see that out of these 7 items in the questionnaire, there are clearly two groups, the first group is item 1 to 4, which related to the topic the employee satisfaction with the supervisor. So, you can see here first four columns or first four rows, you can see their correlation coefficient is quite high 1, 0.75, 0.83, 0.68, then these four.

So, we can see first four questions, or first four important points are really correlated with each other. And we can see the last three, questions, or points are also highly correlated among each other. So, the first 1 to 4 question related to the topic the employee satisfaction with the supervisor. However, item 5 to 7 related to the topic the employees satisfaction with their pay. So, hence item 1 to 4 are somewhat redundant to one another.

So, similarly several items 5 to 7 are also redundant to one another. Now, given this apparent redundancy, we can see the seven items of the questionnaire are not really measuring seven different constructs. So, there are some kind of multicollinearity available in this data set.

(Refer Slide Time: 11:50)



So, what is the remedy? So, principal component say in this condition we calculate the principal component, which is a linear combination of optimally weighted observed variable. So, in this example the first two principle components are suppose we calculate, if there are n number of variables we can calculate n principal components and here first 2 principal component are Z1 Z 2.

So, here you can see 0.47 X 1, 0.50 X 2, 0.51 X 3, 0.50 X 4, then 0.08 X 5, 0.09 X 6, 0.08 X 7. Similarly, Z 2, 0.07, 0.08, 0.0, so on so forth. So, you can see there the in principle component 1, which is denoted by Z 1, question 1 to 4 were assigned much larger coefficient. So, here you can see 1, 2, 3, 4, these are assigned much larger coefficient than the ones in question 5, 6, 7, their coefficient is small. Hence PC 1 is about the satisfaction with the supervision.

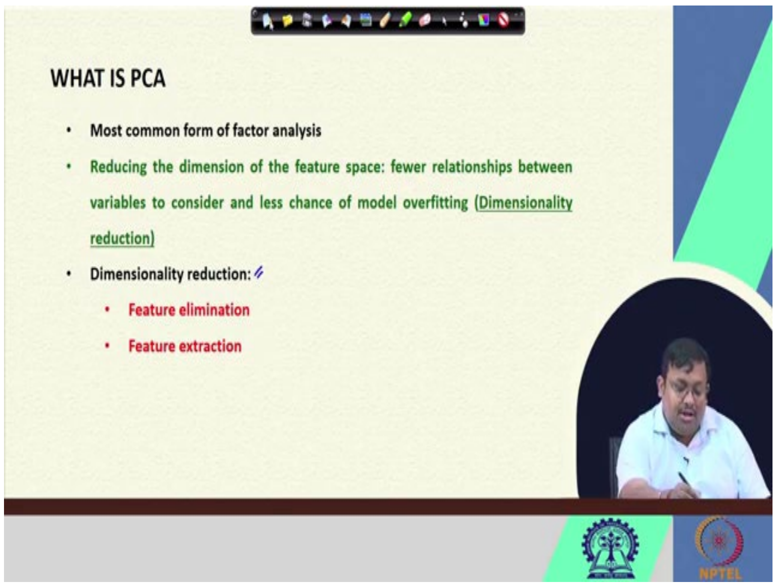
And conversely in PC 2 question 5 to 7 were assigned much larger 5 to 7 these three, 5, 6, 7, are given much larger coefficient than the question 1, 2, 3, and 4. Hence PC 2 is about the satisfaction with the pay. So, this is how we calculate the principal components and in principle components course and the coefficients here are known as the loadings.

So, here this coefficients are known as the loading for their corresponding PCs, corresponding PCs and the loadings are normalized, for example, for principle why we call it normalized?

Because if you take the square of the loading, it will always add to 1, and the loadings are orthogonal, that means they are, they do not have any interaction.

So, you can see when we multiply the loading with the, loading of the first principal component with the corresponding loading of the second principal component, we can and then sum up together we will get a total of 0. So, there is no interaction between the loadings also.

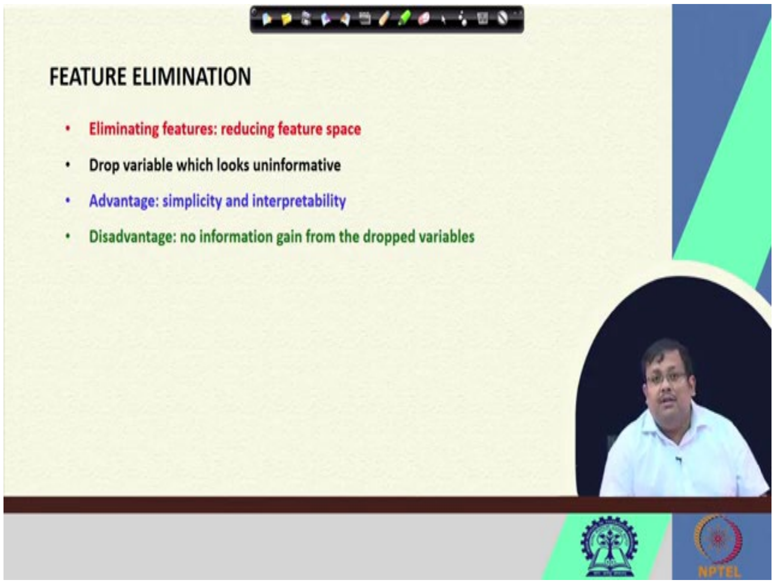
(Refer Slide Time: 14:16)



Now, what are the important features of PCA? The important features of PCA is there is a most common form of factor analysis. And it helps in reducing the dimension of the feature space, because it helps in the reduction of the feature space and because fewer relationship between variables to consider and less chance of model overfitting, when you reduce the dimension there is always chance less chance of overfitting.

Now, this operation is also known as dimensionality reduction and dimensionality reduction can be achieved by two methods, one is called feature elimination method, another is feature extraction method. So, we are going to see them one by one.

(Refer Slide Time: 15:07)

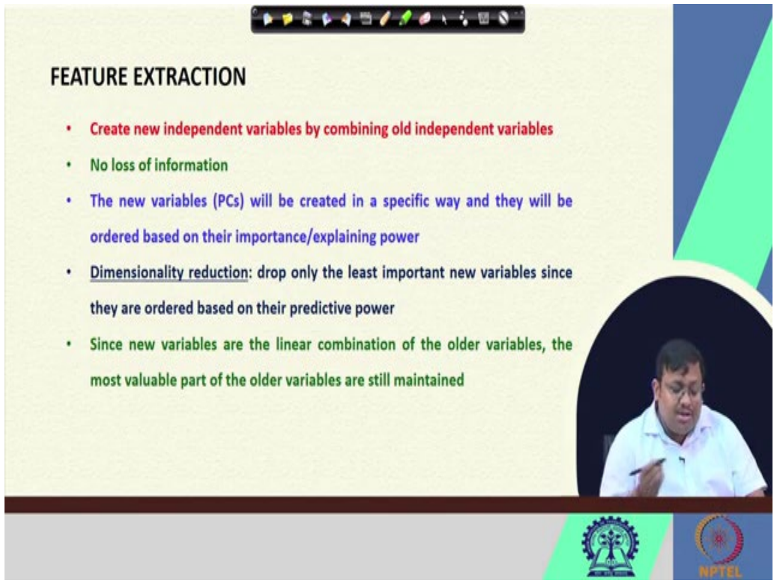


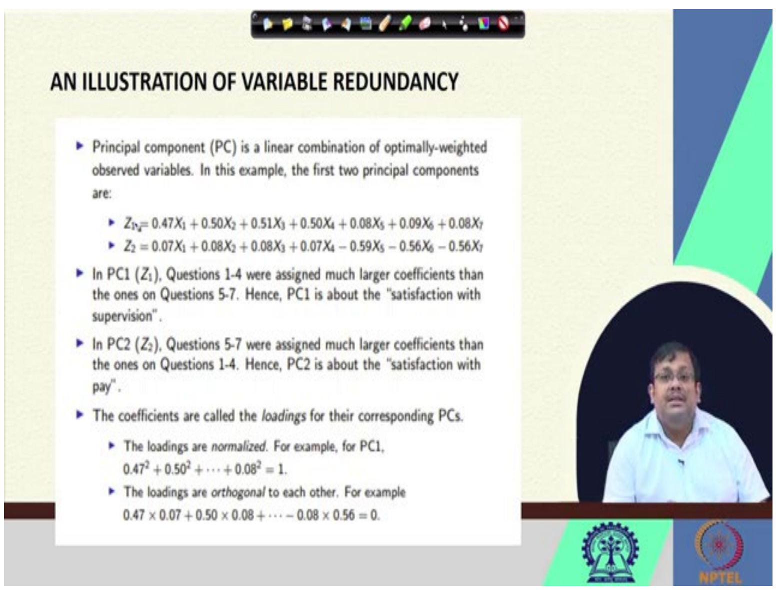
Let us see what is feature elimination method? Now, in the feature elimination method what happens? We eliminate the features and ultimately suppose there are 10 variables and we remove one feature as a whole or two features as a whole so that is called eliminating of the features. And ultimately it is reducing the feature space from 8, 10 variables to 8 variables.

So, in how we remove that? We remove those features which will which looks uninformative. Now, it has several advantage and disadvantage. The most important advantage of this feature elimination is it is simple and it is interpretable. What is disadvantage? Disadvantage is you do not get any information gain from the dropped variables.

So, there might be some amount of information, which is contained in those dropped variables, but we do not gain any information, when we remove from the feature space. So, this is the drawback of feature elimination.

(Refer Slide Time: 16:25)





Now, what is then feature extraction? Feature extraction means when we create in new independent variables by combining old independent variables, for example in that example you have seen that there are 7 questions, which were 7 independent variables, but we are combining them linearly to get this principle component 1 and principal component 2.

So, essentially we are converting the print the original features, or variables into some linear combination and converting them into some in into some new variables, new independent variables. So, in that way we are not losing any information, because remember while calculating these new independent variables, we are not removing, or dropping any variables, these I mean if you go back to this example, you can see that that this calculation of Z 1 all also consider not only X 1, X 2, X 3,and X 4, but also it consider X 5, X 6, and X 7.

So, how small the contribution may be but it is also counted while counting the principal component score. So, this is the benefit of feature, this is the benefit of feature extraction, because we are not losing any information by removing a feature completely from the data set. So, we are not losing any information, so the new way because new variables will be created in a specific way and they will be ordered based on their importance, or explaining power.

Now, what is principle component 1? Principle component 1, you generally see that when we combine these old independent variables into new independent variables and after calculating this new independent variables we order them. So, the principal component 1 will always explain the maximum variation in the data set followed by principal component 2 followed by principal component 3 up to principal component n.

So, this is the feature of principal component, so it the naming of the principal components follows the order of the explaining power, or variance explaining power of the principal components. So, in case of dimensionality reduction you can , so why we call it a dimensional reduction?

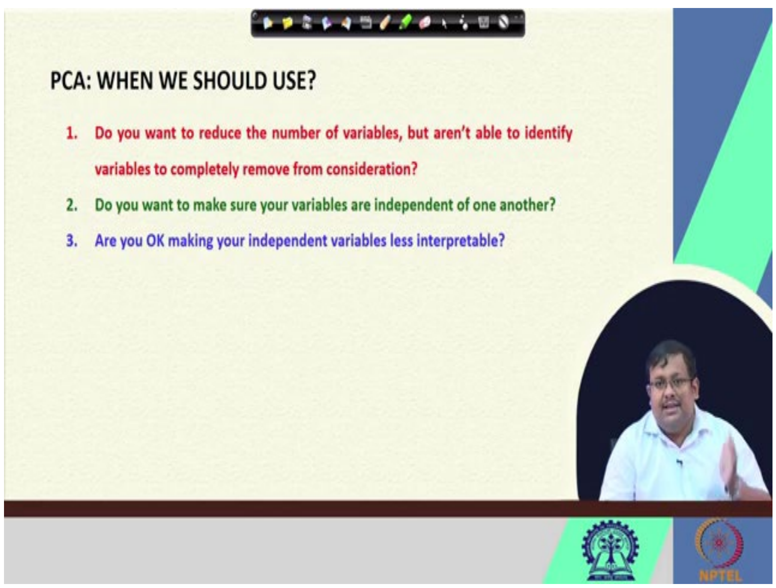
We are not removing a variable as a whole, but still we are calling it a dimensionality reduction, because we are dropping only the least important principle component, or new variables since they are ordered based on their predictive power. Since this principal component are ordered suppose n number of principal components are there and we have order them based on their predictive power.

So, we can selectively remove some of the principal components, which are having very less predictive power, by doing so we are reducing the dimension, because ultimately if there are n number of principal components and we are removing two principal component from the last. Then we are having actually n minus two principal components.

So, we are reducing the independent variables, but at the same time we are not losing the information important information, which you can get from all the important least important variables also. So, here it justifies that by doing the principle component analysis, you can do the dimensionality reduction, but at the same time you are not losing the information gain from any variable by entirely removing that variable, or feature from the data set.

Since new variables are linear combination of the older variables this most important valuable part of the older variables are still maintained in the PCA space, this is very important. We are still maintaining the important contribution from the older variable in the PCA space and we are removing only the unimportant principal components, whose predictive powers are not that much high, in that way we are we are conserving the information gain from the least important variable also, but at the same time we are reducing the dimension. So, that is why principal component analysis is a dimensionality reduction approach.

(Refer Slide Time: 21:56)



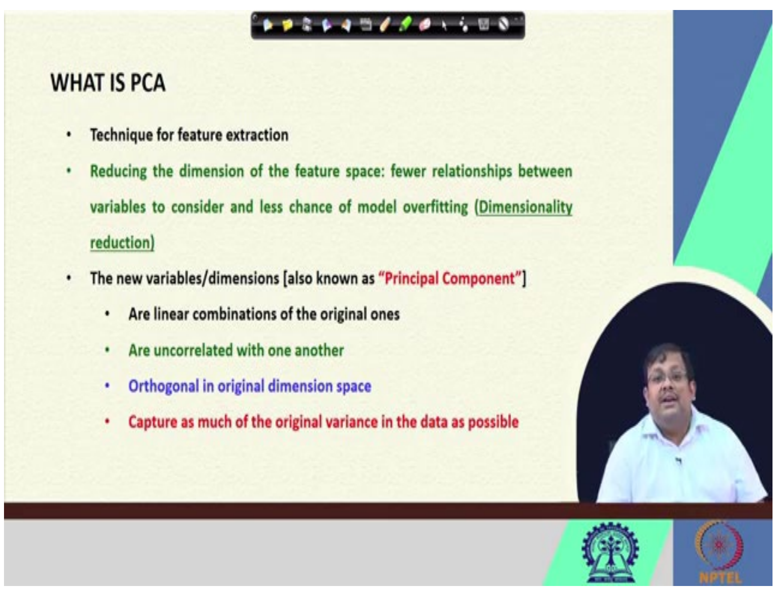
Now, when we should use the PCA? This is one of the important question. So, we should use the PCA, when we have 3, questions in our mind. First of all do you want to reduce the number of variables, but are not able to identify the variables to completely remove from consideration, you have a large dimension data, you want to reduce that dimension, you want to go for dimensionality reduction.

But you do not know which variables to remove completely from the feature space. So, in that way you go with the principal component analysis. Second important question do you want to make sure your variables are independent to one another, one of the major feature of principal component analysis is, in principle component analysis this principle components are orthogonally projected to each other.

So, if there are n number of principal components, they are projected in n dimension, which are orthogonal to each other. So, there is no interaction and they are independent to each other. So, if you want to make sure that your variables are independent to one another you can always go with the principal component analysis.

The third one are you ok making your independent variables less interpretable, remember when you are combining them together in a principal components the interpretability is not that very straight forward as compared to where we keep the variable as such. So, if you are ok, if you are having yes, for all these questions, then you can go with the principle component analysis.

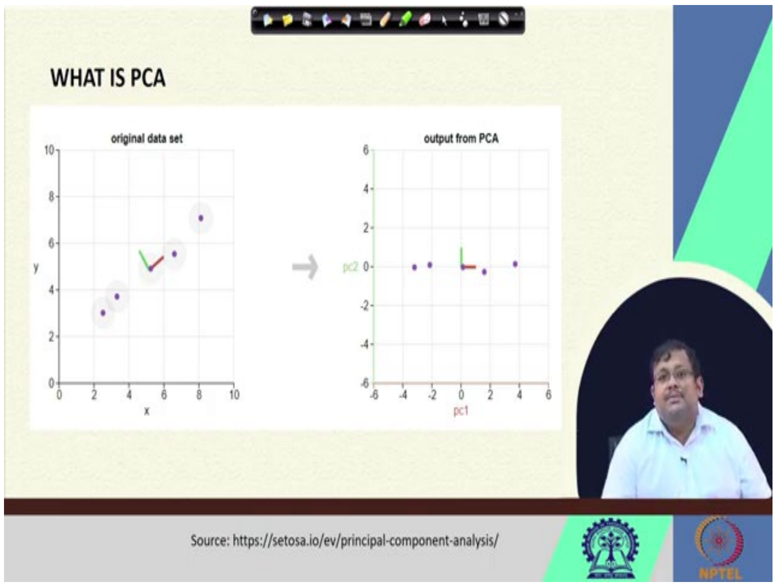
(Refer Slide Time: 23:55)



Now, what is PCA? Of course we know that it is a technique for feature extraction and it can helps in reducing the dimension of the feature space, fewer relationship between variables to consider and less chance of model overfitting, of course because when we are reducing the dimension and all as also ensuring that the features are not correlated, they are reducing the multicollinearity, they are also reducing the overfitting.

So, the new variables are known as the principal component are basically linear combination of the original ones and they are uncorrelated to one another and they are orthogonal in original dimension space, of course when they are orthogonally projected there, they will be uncorrelated to each other. And they can capture as much of the original variance in the data as possible. So, these are the features of principal components.

(Refer Slide Time: 24:51)



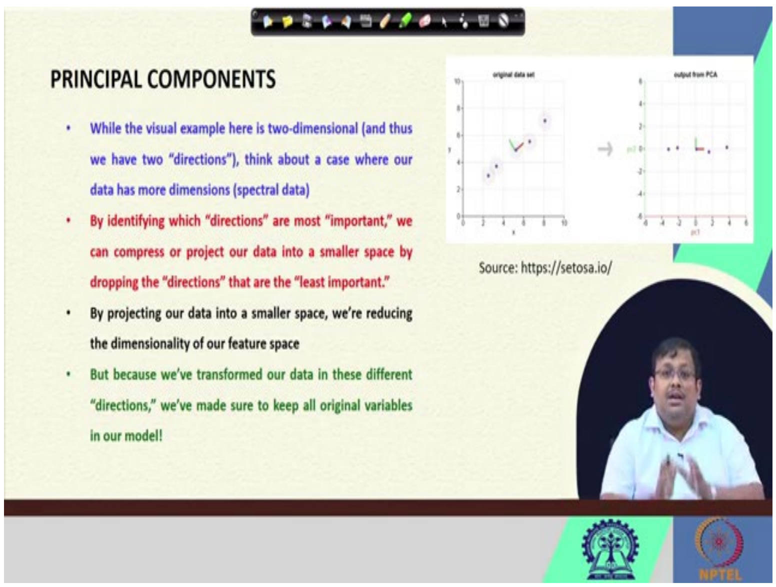
Let us see one example. Suppose here there are two variable X and Y, and we have got these points, which already have their X and Y value. So, in the principal components, if we imaginary draw a new set of coordinates, these are principle component 1, and principle component 2, by rotating the direction.

So, here you can see there are two direction one is red direction in the green direction in the data. So, by rotating the data in such a way, we can have a two new set of imaginary dimension, that are known as principal component 1 and principle component 2. And as I have mentioned that these two dimension will be orthogonal to each other.

So, you can see here this is principle component 1, and principle component 2. And why we call it a principal component 1? Because the maximum variance of the data set we can see in the principal component direction of the principal component 1. So, this direction is principal component 1 and the other direction is principle component 2, I hope now it is clear to you, what is principal component analysis.

So, basically it is a rotation of the coordinates in such a way, that we can get two coordinates where the data is showing the maximum explain variance, or maximum variance in the towards the dimension of principal component 1 followed by the other principle component.

(Refer Slide Time: 26:42)



So, while the visual example here is two dimensional and thus we have two dimensions, two directions, think about a case where our data has more dimension, when there is a spectral data I will show you in case of spectral data, there are thousands and thousands of variables. So, there are thousands of dimension, or directions. So, by identifying which directions are most important.

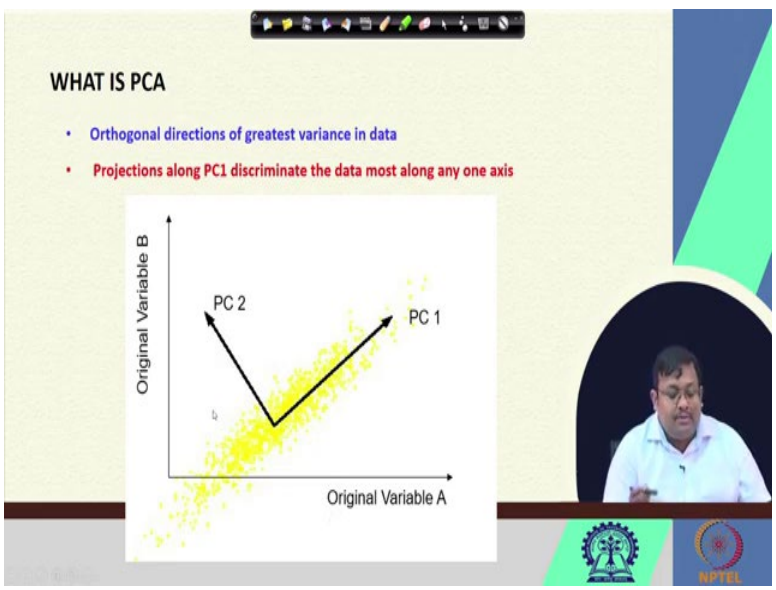
So, here in this example we can see this PC 1 direction is more important than PC 2 direction, because in the PC 1 along the PC 1 direction, we are getting higher variance. So, by identifying the direction, which directions are most important we can compress, or project our data into smaller space by dropping the direction that are least important.

So, what we will do? We will calculate the principal component analysis and then we will project the principal components into n dimension and we will see in which dimension the data is least variable. And then we will selectively remove those principal components to compress the data further.

So, by projecting our data into a smaller space, we are reducing the dimensionality of our feature space. But because we have transformed our data in this different direction, we have made sure we to keep all the original variables in the model, by remove by when we calculate this principle components and project them into different direction, we are not losing any information by removing the least important principle component.

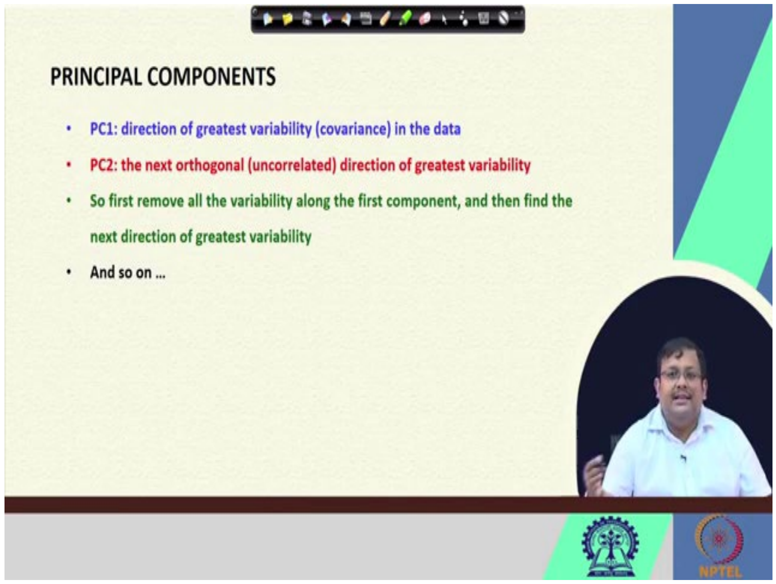
Because in the in the other principle component calculated principle components and project principal components, we have still maintained the important information, which is there in the least important variables, or we are keeping the all information required information in the which are present in the original variables in the model.

(Refer Slide Time: 28:52)



So, what is PCA? It is orthogonal direction in the greatest variation variance of the data you can see this is a principle component 1, principle component 2, original variables and principle original coordinates and this principal component coordinates. So, projection among PC 1 discriminate the data most along any 1 axis.

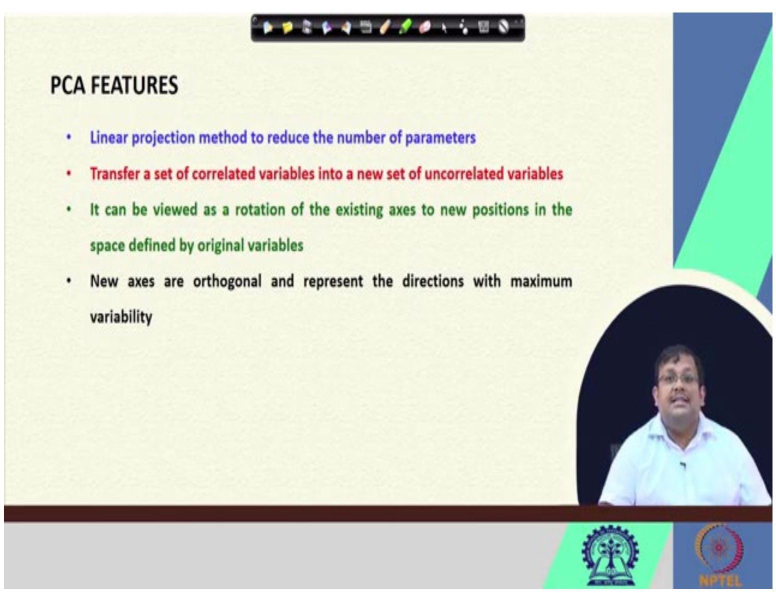
(Refer Slide Time: 29:18)



And principle component 1 is always a direction of greater variability, or covariance in the data. Followed by principle component 2, the next orthogonal, or uncorrelated direction of greatest variability greatest variability.

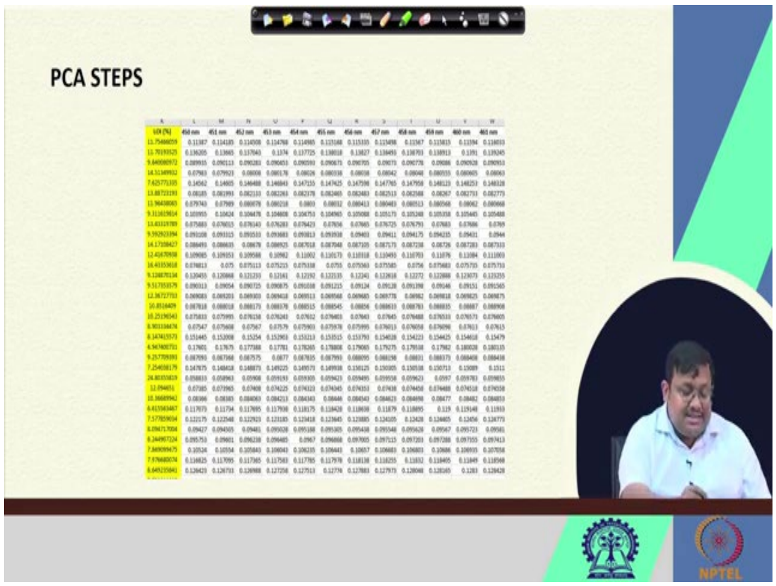
So, first remove all the variability along the first component and then find the next direction of the greatest variability and we can do in this fashion to ultimately gain the total n number of principal components. And then we can do selective removal of the principal components.

(Refer Slide Time: 29:53)



PCA features if you talk about the PC features, it is a linear projection method to reduce the number of parameters, it transfer a set of correlated variables into a new set of uncorrelated variables. It can be viewed as a rotation of the existing axis to new position in the space defined by the original variables. And new axis are orthogonal and represent the direction with maximum variability.

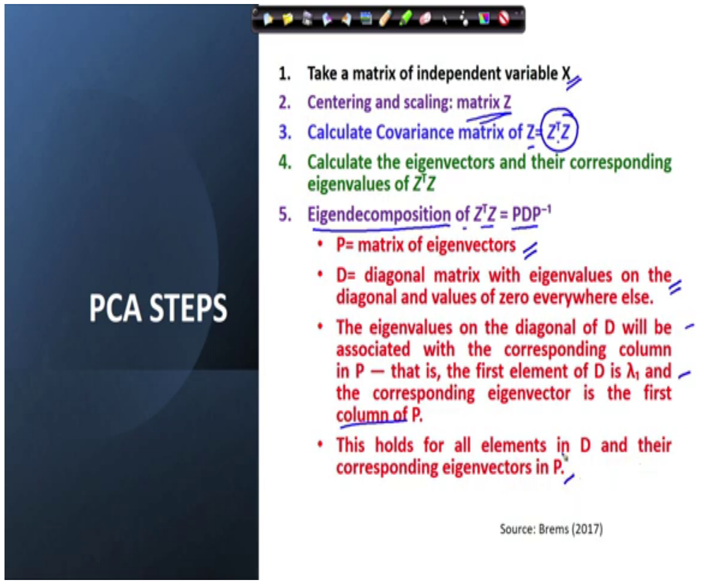
(Refer Slide Time: 30:19)

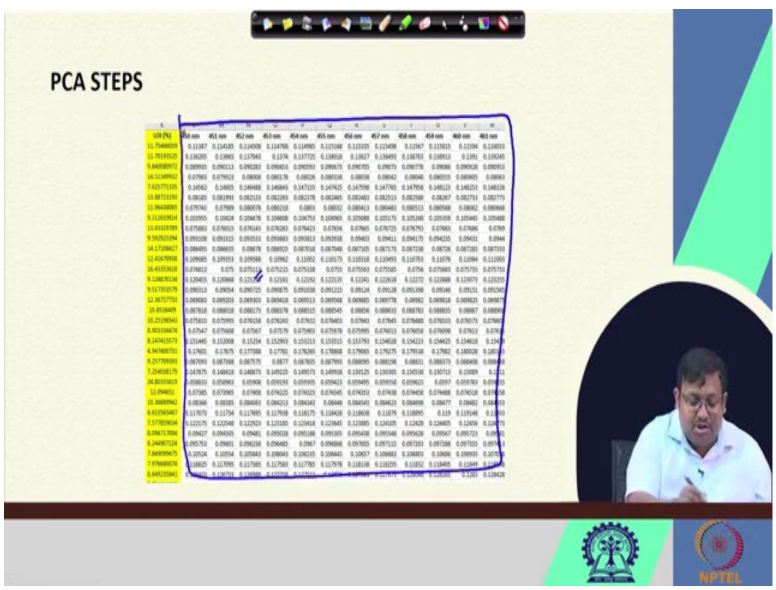


So, if you see this is a example, this is an example of a spectral data set, you can see how many variables are there. It is a snip of a whole spectral data set, the spectral data set can goes up to thousands and thousands of variable and this is a target parameter.

Suppose it is the loss on ignition organic matter in case of soil by Nelsons and Somers method. So, this is suppose in this data set this is a target variable and these are the independent variable.

(Refer Slide Time: 30:52)





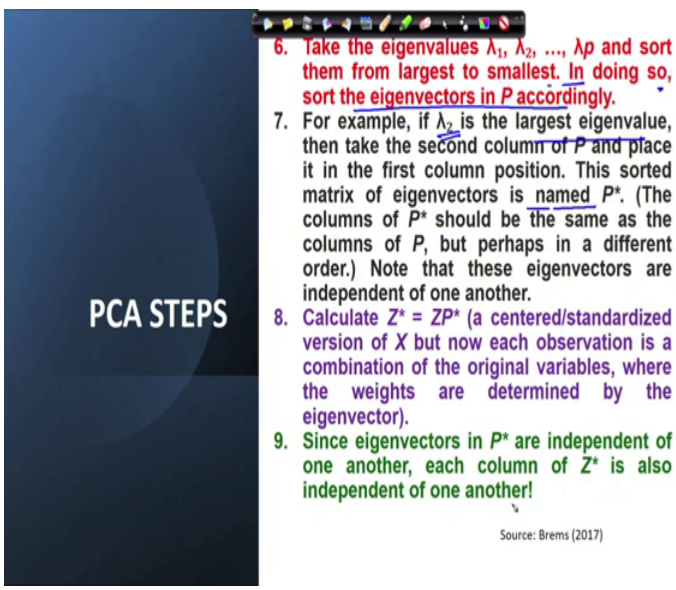
So, how to do the principle component analysis? These are the steps of principle component analysis, first we want to take a matrix of the independent variable. What is the independent variable? The independent variable is the is the spectral data. If we go back to our previous slide, here this spectral space is the independent variable, this is the independent variable.

So, we select this independent variable, we call it a, we call it, we know X, independent variable X. So, take a take a matrix of independent variable X, multiple X is there. So, we call it this matrix Z. And then we can calculate the covariance matrix of Z, which is basically multiplication of Z and Z transpose.

Now, in the fourth step we can calculate the eigenvectors and their corresponding eigen values of this covariance matrix. So, these are covariance matrix of Z. So, then we can calculate their eigenvectors and corresponding eigenvalues of this covariance matrix. And we do that by eigen decomposition. Eigen decomposition of Z transpose Z produce this PDP it can be represented by PDP inverse.

So, here P is the matrix of eigen vectors, D is the diagonal matrix with eigen values on the diagonal values of the diagonal values of 0 everywhere else and the eigen values on the diagonal of D will be associated with the corresponding column in P. So, that is the first element of D is lambda 1 and the corresponding eigen vector is the first column of P. And this holds for all the elements in D and their corresponding eigen vectors in P.

(Refer Slide Time: 33:01)



So, once we decomposed this covariance matrix into PDP inverse. Then what we do? We take the eigen values, these eigen values as I have already told you they are denoted by this lambda 1, lambda 2, lambda P, and sort them from largest to smallest. So, in doing so sort the eigen vectors in P accordingly. So, suppose if we find that this lambda 2 is the largest eigen values, among all these lambda 1, lambda 2, lambda P.

So, suppose we have found this lambda 2 is the largest eigen value. So, then we can take the second column of the P and place it is in the first column position, because we select the second column P, that is the second eigenvector and put it in the first position, because that is the most important. This sorted matrix of eigenvectors is now named as P star.

So, this columns of P star should be same as the columns of P, but perhaps in a different order, because here we are order these eigen vectors according to their importance, these eigen vectors are of course independent to each another. Then subsequently we calculate this Z star, which is Z into P star, which is a centered or standardized version of X.

But now each observation is a combination of the original variables, where the weights are determined by the eigen vector. So, once we calculate this P star, then if we multiply this P star with the original Z matrix, then we calculate this Z star. So, since eigen vectors in P star are independent of one another each column of this Z star is also independent of one another.

So, this is how we calculate using the matrix algebra, we can calculate these eigen values. And we can calculate this principal component analysis by based on their ordering their eigen vectors. And ordering of the eigen vectors based on the ordering of the eigen values.

So, this is how we calculate, this principal component analysis, principle components and this total analysis is known as principal component analysis, remember again these eigenvectors are orthogonal to each other and they are uncorrelated to each other.

(Refer Slide Time: 35:25)

So, these are some of the references, I hope that you have got some information, new information regarding PCA and how to calculate the PCA. Let us wrap up our lecture here. And in the next lecture we will continue from here and we will see how to select the important number, important number of eigen values, important principle components and their subsequent analysis. We will also discuss the principal component regression and their applications. So, thank you let us meet in our next lecture.