Biometrics

Prof. Phalguni Gupta

Department of Computer Science and Engineering

Indian Institute of Technology, Kanpur

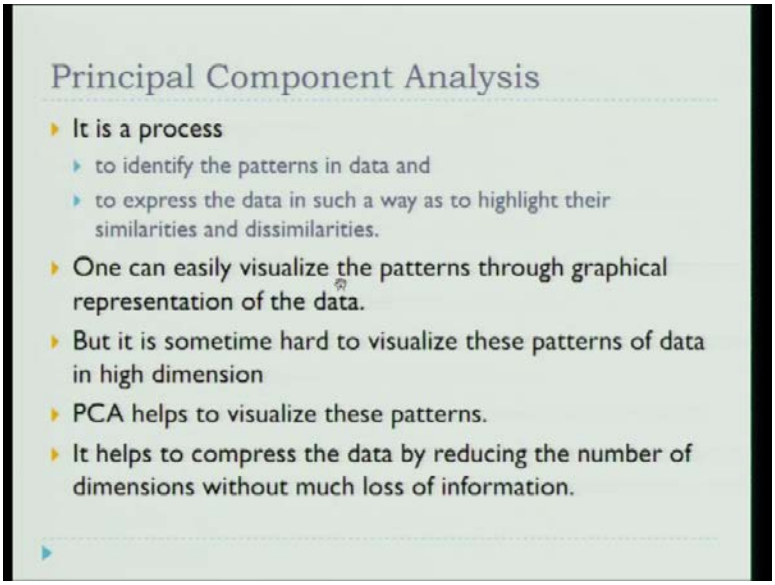
Lecture No. # 09

Principal Component Analysis

There principal component analysis which we like to cover today that will be very useful for us, while we will be designing the biometric system. That physical component analysis, it has the two major roles. First role is that you know it gives you the indication about the patterns of the data, and second thing that sometimes some information or the patterns of the data, they are not useful. They are close related, so why to consider those patterns? You can easily suppress them.

So basically, a principal component analysis will be useful, when you want to reduce the dimension, and by the word reduce the dimension it means that number of feature elements, you like to consider will be less. So, that is the basic.

(Refer Slide Time: 01:20)



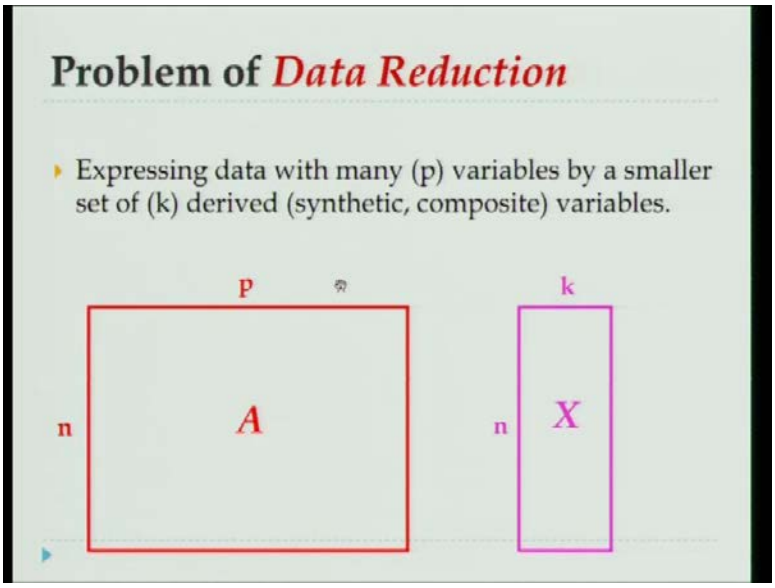
So generally, what to do by seeing the data, you like to see what is the pattern in that right. Are they maintaining certain relation or there is no relation exists randomly distributed. So these patterns you need to know.

Now, once you need to know these patterns, you want to highlight these patterns; highlight the pattern means that how much similarity there exists and how much dissimilarity there exist that is the issue. If there exist too much dissimilarity along a pattern that means that will give you some features, because our aim is to get certain unique features against an object.

So, to get the unique features against the object that means, you want to know how much variability you have within a particular pattern, so that is the aim. So, the first one is that you want to find out whether there exist a pattern or not, second part is that if there exist a pattern then how much similarity or how much dissimilarity is there. If the similarity is too much similarity exists in one pattern that means that pattern is not useful parameter or characteristics to consider as a feature element.

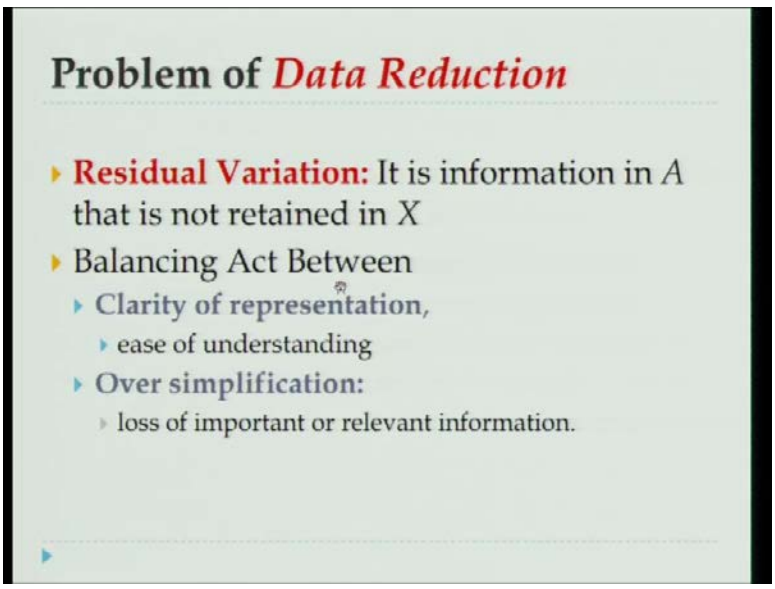
Now, this pattern you can visualize if you plot it. Now plotting is possible when you have one-dimensional, two-dimensional or three-dimensional, but more than that you will not be able to visualize whether there exist a pattern or not. So that the main problem is that you can think what for I should use the PCA, once I know the pattern, but pattern is visible only for the smaller dimension for larger dimension for this you have to take the help of PCA. PCA also as I told you, also help you that I want to reduce my dimension by discarding those features, which are having the most similar behavior in a pattern.

(Refer Slide Time: 03:36)



So, suppose I have the data of n elements each element containing the P variables of P variables, P variables means that P characteristics, P dimensions. So you have the P dimensions and you have the n observations, so this is your data set. Now through the PCA, you can reduce the data set into n cross K, where K contains top most patterns having the larger variance or variability that is the idea. So, what you can see that you have the original matrix A. Now a of size n cross P, I am reducing into K obviously you are losing something.

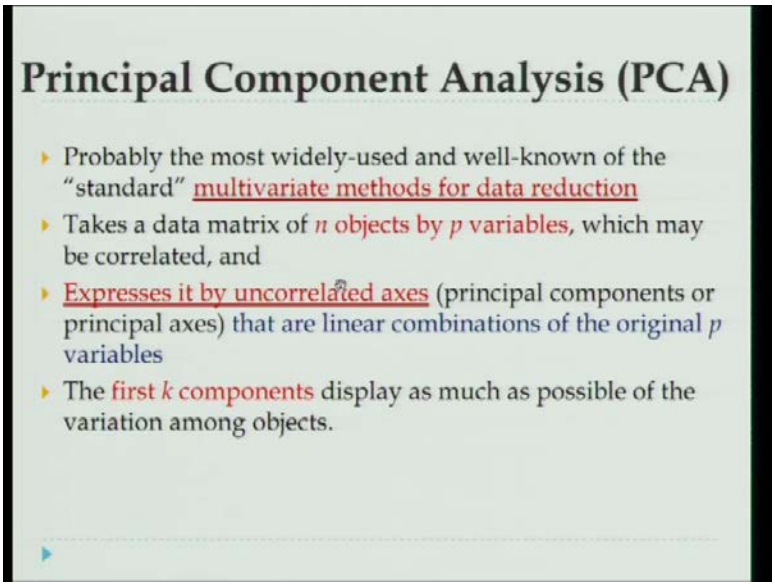
(Refer Slide Time: 04:35)



So, the term is residual variation. It is the information in A but that is not in your n cross K area that is known as residual variation. These residual variation should be as minimum as possible that is the aim now that gives you the indication that existed trade of relationship between the clarity of representation and oversimplification by the term. Clarity of representation means that whatever data you have you must be able to get proper representation.

So, if I have the matrix A is such n cross P obviously we will get the same matrix what about original a everything will be there now. if I reduce to n into K so, the information will be lost but you are simplifying by the process of oversimplification. You may lose certain information so there exist a trade of relationship between these two. So you have to choose such a K so that you get the maximum information but maximum reduction of size also that is the problem.

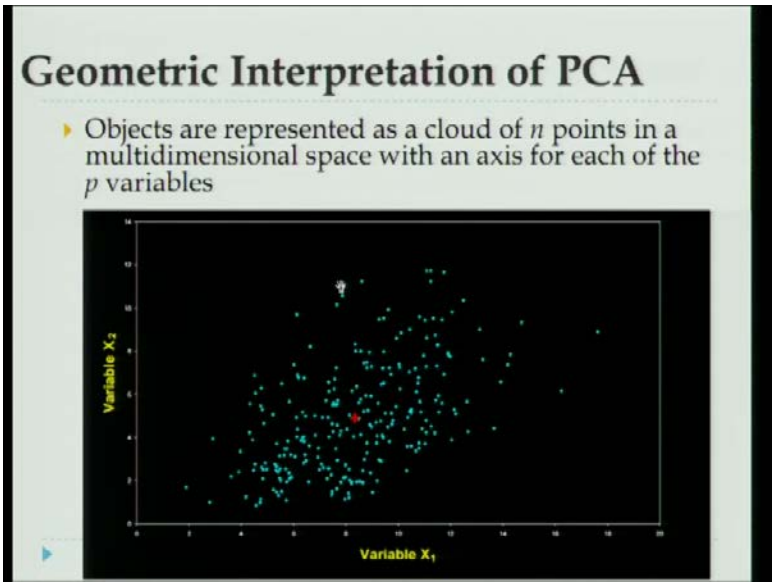
(Refer Slide Time: 05:53)



Now this is very well known method in statistics where you need to reduce the dimension to analyze it, because they have the multi dimensional things so need to reduce. Same thing will happen in our case also well, we will be covering our biometric system so it takes the data of size n cross p dimensions you have and they are correlated. That if you have one x 1 corresponding value will be there x 1 1 x 1 2 x 1 3 x 1 so corresponding values will be there but what we express it into the uncorrelated dimensions which is we want to plot it in the direction of principal axis and that has to be

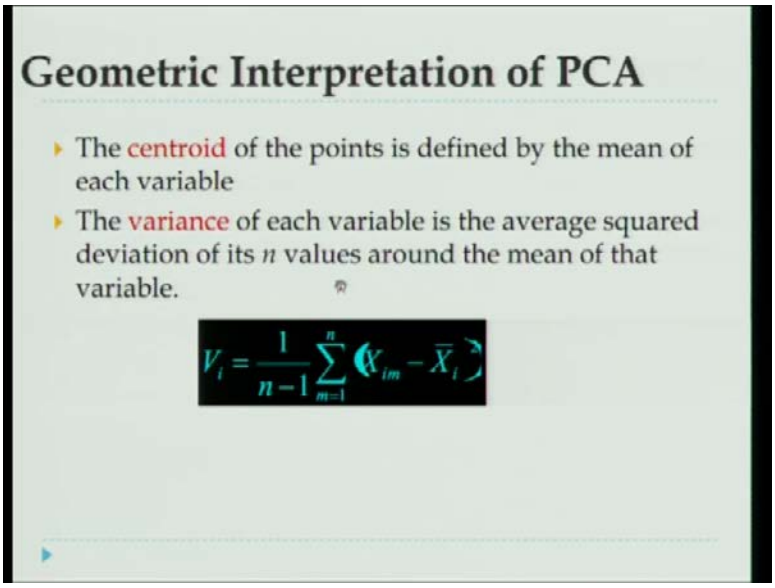
expressed in the form of linear combination of those variables, how to do it we will discuss this. And then we want to pick up the top k elements. Top k principal components those principal components you are selecting and having the largest variance k top variances.

(Refer Slide Time: 07:24)



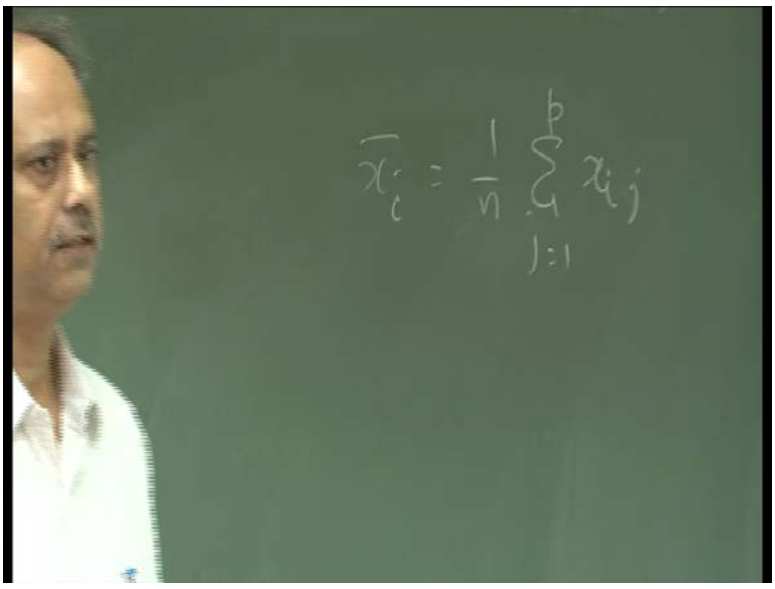
So, this is about the definition of principal components. This is the data you will have and what happens that you have the point cloud data. Point cloud data means that you have the point which is represented by p variables or p dimensions and here one is variable X 1 another is variable X 2 like that p dimensions are there this points are there.

(Refer Slide Time: 07:47)



And the term centroid by now in statistics you have heard centroid. Centroid is nothing but the mean of the variables each objects in all dimensions. So X i bar is equal to 1 upon summation.

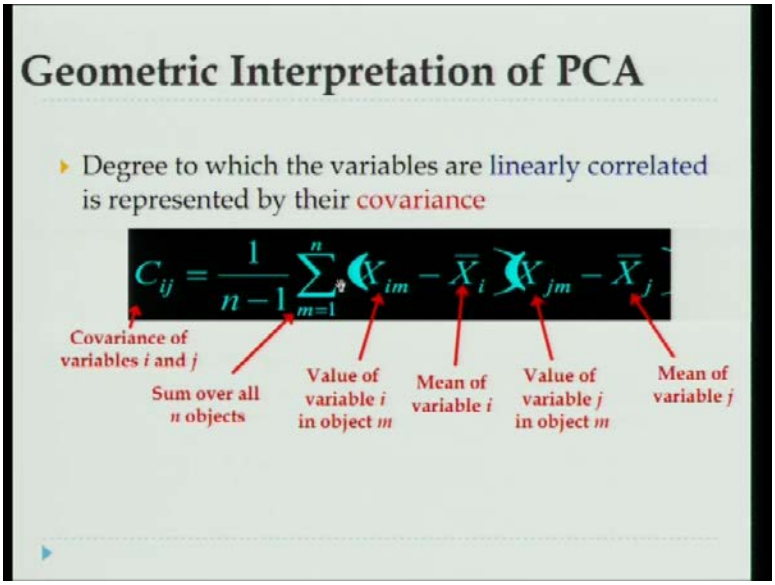
(Refer Slide Time: 08:08)



So x i bar is equals to 1 upon n summation over x ij j is 1 to p. So, centroid of the points defined by the mean of these variables and similarly. The term variance, variance is nothing but that sums of square of deviation of each variable from mean and then you are taking the average. You should have taken one by n but to get it why it is 1 by n minus 1

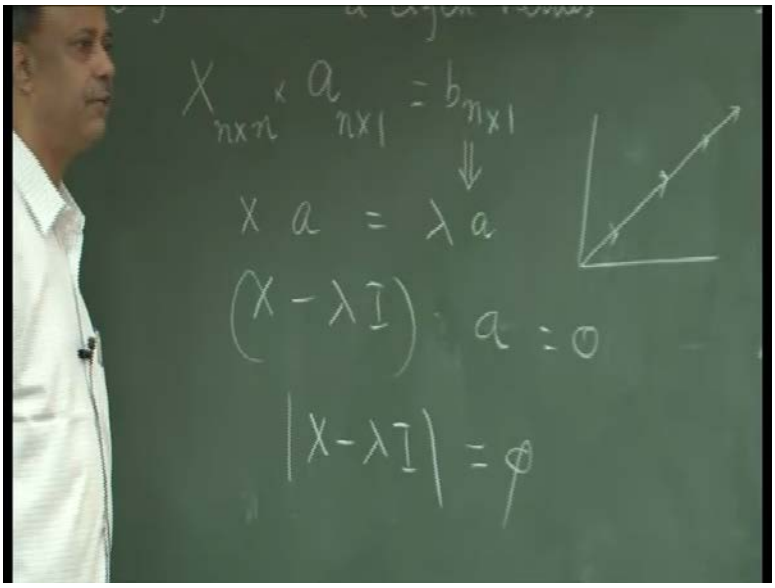
that n-eth term is degrees of freedom is n minus 1. So, we are dividing by n minus 1 sum of square and degrees of freedom are n minus 1 that is why we are normalizing by n minus 1.

(Refer Slide Time: 09:07)



Then next term you will be using the covariance. Covariance is nothing but you want to find out the degree of relation, correlation-ship in the linearly. Correlation-ship between the two variables so it is equals summation X im X i bar X jm minus X j bar and 1 by n minus 1. This C ij is the covariance variable relationship between i and j and this m is sum over all n objects and value of variables object. X i this is the average of this and similarly, the average with respect to the variable j.

(Refer Slide Time: 09:54)



So, now the term, what is the term? Term is Eigen vectors and Eigen values and its application you cannot forget, you may forget the formula remember that why I need that. This is required not only for biometric; you may require this thing for future also. These Eigen vectors and Eigen values are basically depend on the some certain transformation or it will give you certain another vector that is the thing.

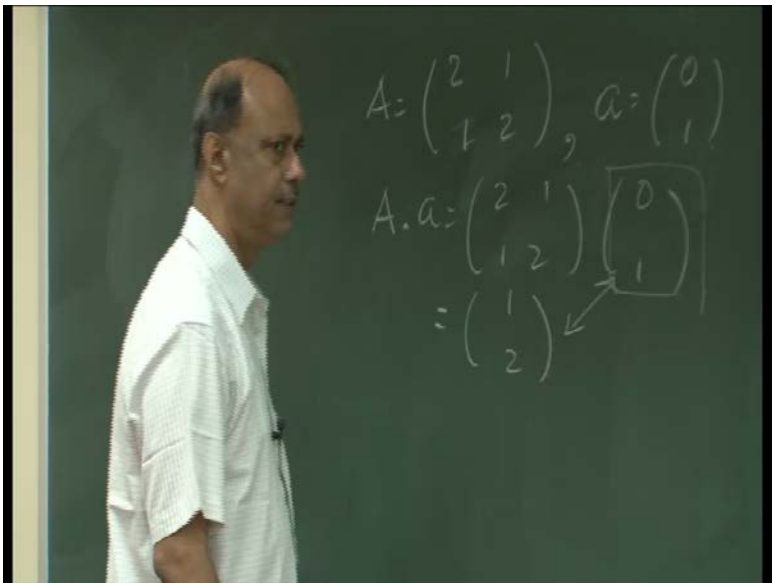
Suppose, you have n cross n observation matrix that whatever we have considered and we have n cross that n variables and n dimension. You have n cross n and you have a vector say a it is also n cross 1. So, if I perform this operation what I will get another vector. Some symbol is there no vector arrow mark and I assume that this is vector, because generally we write in bold letter but here there is no scope of writing. So, this is another vector. Now these two vectors are parallel when I can express, this is multiplication of some scalar multiplication of a. So, these two are parallel when this can be expressed as a scalar multiplication of a or one can be expressed in terms of the other one by multiplying by something some real number.

So, that means what X a can be expressed as lambda a then you can tell these two vectors are parallel and if you can express this where lambda is some real then this a is known as Eigen vector, what it means basically a point is here and this is origin and it gives the direction like this. Now lambda is a scalar so it will give you the increase form or here, it is giving only the direction of elements with respect to the origin. That is the vector and

what is this lambda? Lambda gives you what it means once you transform this a to this bigger one what is the deviation? What is the dispersion? What is the variance? Once you move this one to another transformation then there will be a dispersion variation. What is the variance with reference to that pattern? That will give you the lambda.

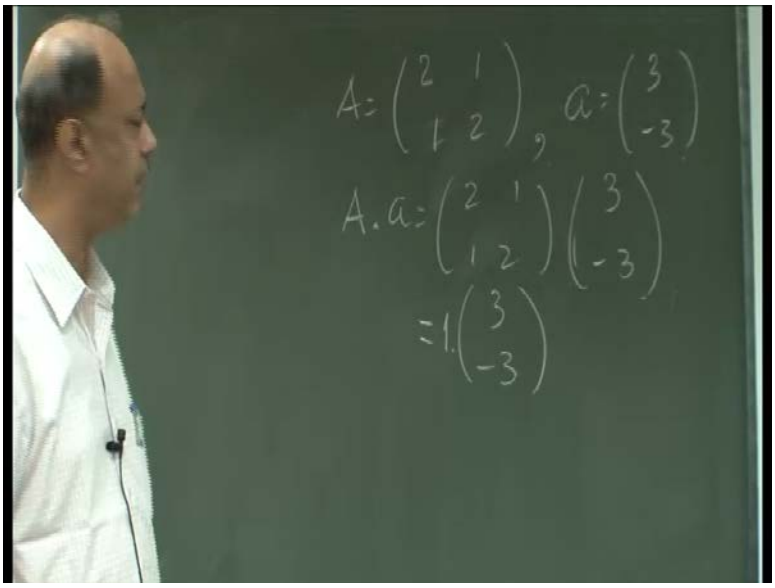
So, we will be looking for the larger value of lambda for our principal components. Largest value of lambda will help us to detect that yes. This is the principal axis so this can be written as you will be writing this is determinant of X minus a I lambda I equals to 0. That is the thing you have to do. Now you solve it you will be getting the values of lambda.

(Refer Slide Time: 14:08)



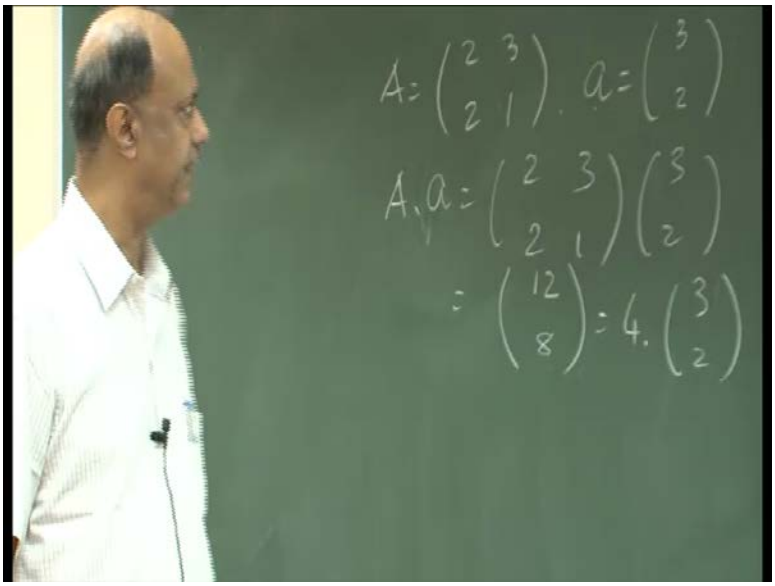
Let us consider two three small example so A dot a is a 2 1 1 2 0 1 and so it will become what 1 2 so this is observed that these two cannot be expressed in terms of this. So, this is not a this is not an Eigen vector.

(Refer Slide Time: 15:08)



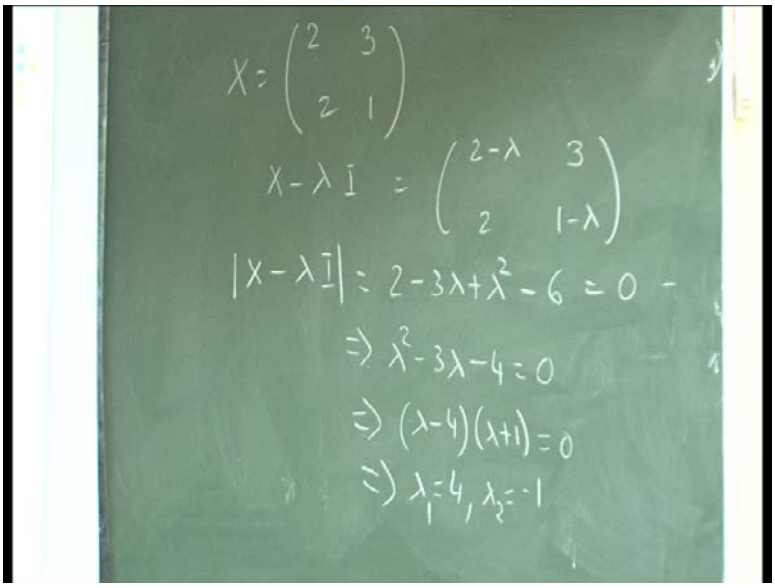
But if I consider three minus three so this becomes lambda is one and this x can be expressed a in the form of a so it is an Eigen vector.

(Refer Slide Time: 15:33)



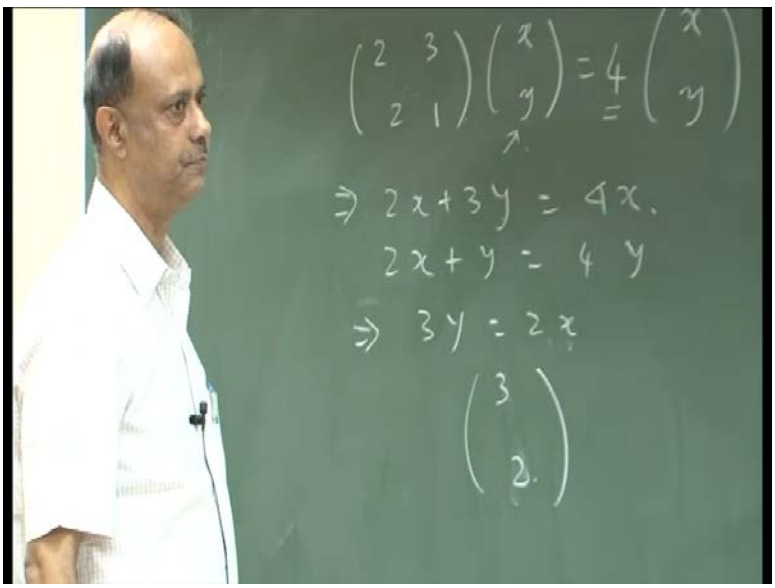
Now, tell me what should be a so that a is an Eigen vector. So, one possible could be 3 2, is it correct 1? So, we come later on how to find out or first this is a just trial and error method we are using. So, lambda is 4 and so 3 2 is your Eigen vector and now how to find out this, see here. I have just putted the value and I have told you that this is an Eigen vector.

(Refer Slide Time: 16:44)



Now in order to find it, you have to find out what is your X X is 2 3 2 1 and so you have to find out X minus lambda I is nothing but 2 minus lambda 3 2 1 minus lambda, because I is the identity matrix so you will be writing this. So, determinant of X minus lambda I 2 minus 3 lambda plus lambda square minus X, this has to be 0 lambda square minus 3 lambda minus 4 equals to 0. Lambda minus 4 lambda plus 1 will be equals to 0. So, this gives you lambda 1 equals to 4 and lambda 2 equals to minus 1 so you got the two Eigen values.

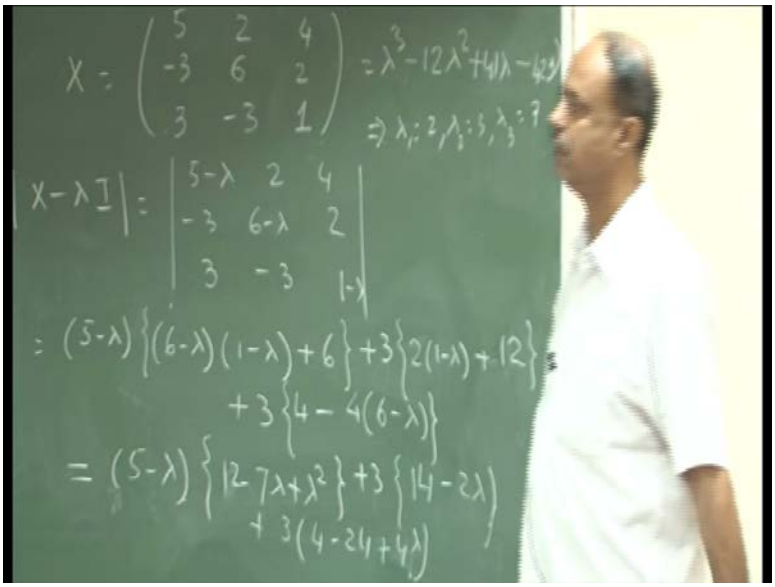
(Refer Slide Time: 18:00)



Now you have to find out what is your Eigen vector. Now you have 2 3 2 1 and x y is equals to 4 times of x y, see this is the things you have to find out. This Eigen you have obtained only this part lambda but not the Eigen vector you have to find out what is your Eigen vector.

So, this gives you 2x plus 3y is equals to 4x and 2x plus y equals to 4 y and this gives 3y equals to 2x. So, once you want to find out the principal axis so it will be a straight line. Y equals to x forms so this will be possible only when it is 3 and 2, 3 into 2 is 6 and 2 into 3. So, this is the only way you will be getting this line so this is your x y that is your Eigen vector.

(Refer Slide Time: 19:15)

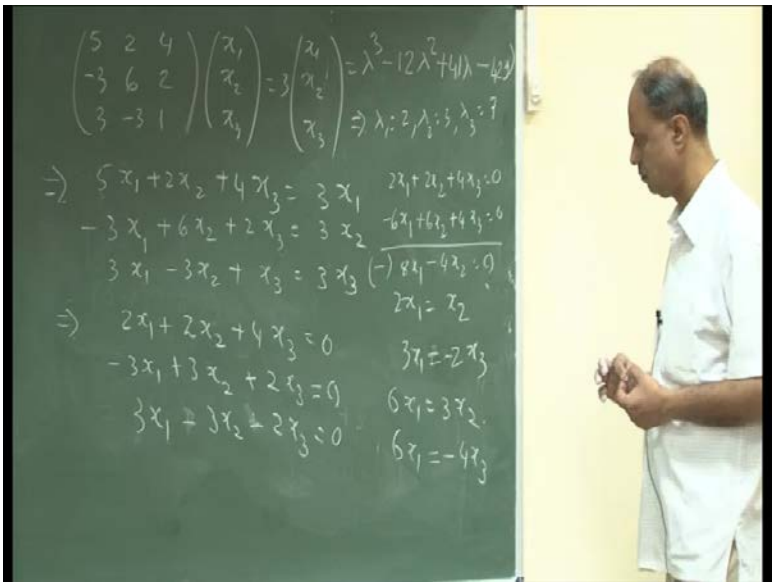


Suppose, I have X 5 2 4 minus 3 6 2 3 minus 3 1 and now can you find out what is this Eigen values and Eigen vectors. It is not essential that matrix should be square matrix but after multiplication you must get the square matrix and the minimum dimension will give you that many number of Eigen values. That means this size of this will give you the number of Eigen values. That means number of dimensions will give you the number of Eigen values. Here we have considered is the 3 cross 3 and now can you find out the first you find out the Eigen values and then you find what Eigen vectors are.

How many lambdas, you will be getting. First you tell me three so one of those should be at least correct. If one is correct remaining, two will be correct. We can assure you, is it this line first you check. This line if it is correct then you expand that is 5 minus lambda

it will be 6 minus 7 lambda plus lambda square plus 6 so it becomes 12 plus 3 times. What is the final equation, so you are getting lambda cube minus. What minus 12 lambda square then plus 41 lambda 42 and this will give you lambda 1 equals to 2 lambda 2 equals to 3 is it 2 3 and 7 you told lambda 3 is equals to 7. Hopefully it is correct.

(Refer Slide Time: 23:40)



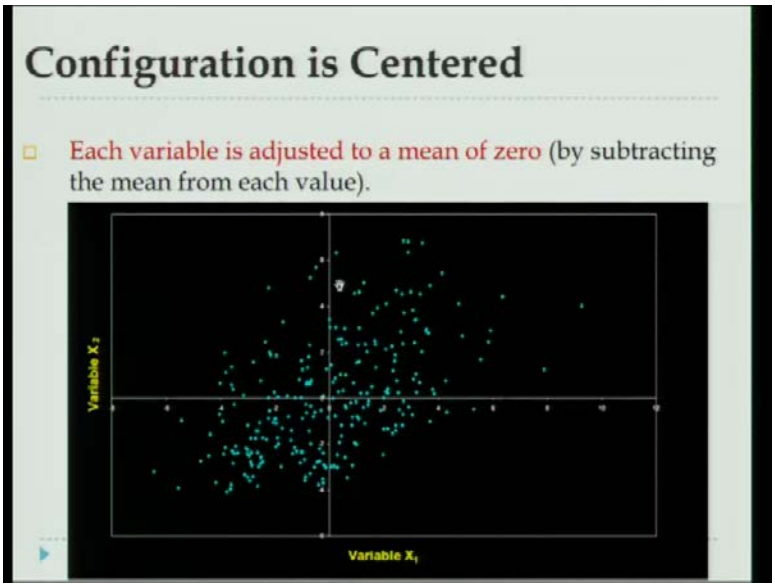
Now what is the Eigen vector? Find out one Eigen vector you take 3 lambda 2 equals to 3 and then you tell me the Eigen vector 5 2 4 6 2 was it 5 or 3 is this what about relation between x 1 x 3 or x 2 and x 3 3 3 x 1 plus x 3 2 x 3 2 4 minus 3 2 4 minus 3.

What is two x 1 2 x 2 4 how did you get 2 4.

Minus 3.

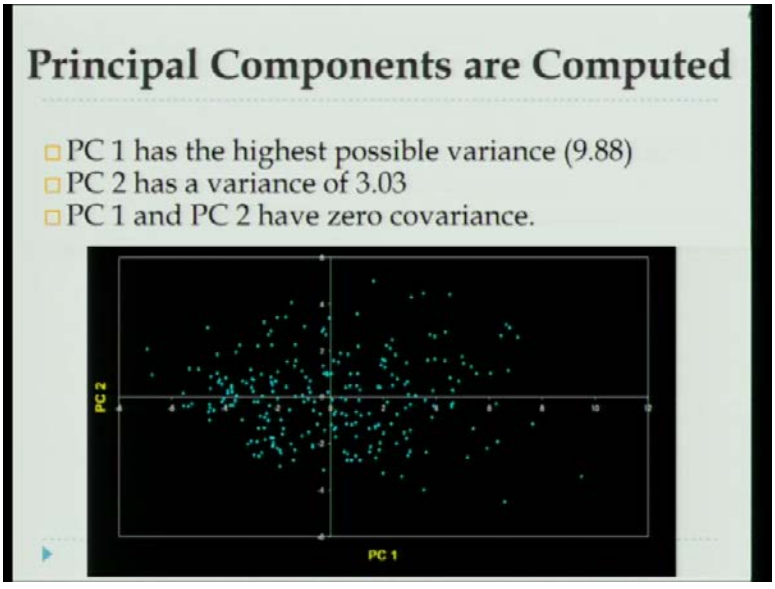
So, that 2 4 2 4 minus 3 so if you put this value you can find out the result is coming.

(Refer Slide Time: 29:56)



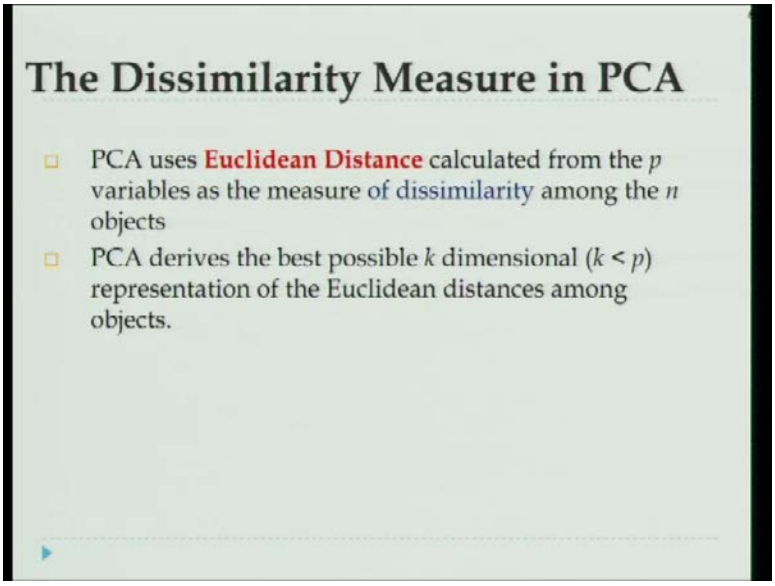
So, this is your background of Eigen values and Eigen vectors and that Eigen vector gives you the direction with respect to the principal axis and Eigen values will give you the variance with respect to the that pattern that principal axis has given the n dimensional, p dimensional elements of n observations. Your aim is to rotate the axis point to a new axis point that axis point is termed as which satisfies the following properties.

(Refer Slide Time: 30:36)



First one is that it should be ordered in such way that the principal axis has the maximum variance then the axis two. Which is the next maximum variance and you will be getting since it is a p dimension so you will be getting the p principal components. P principal axis but you will have to arrange in certain order increasing or decreasing order. Highest one will be your principal axis, one then next one and so on and also you have to select this axis in such a way that they are perpendicular to each other. The principal component I is perpendicular to principal component one two three and so one.

(Refer Slide Time: 31:18)



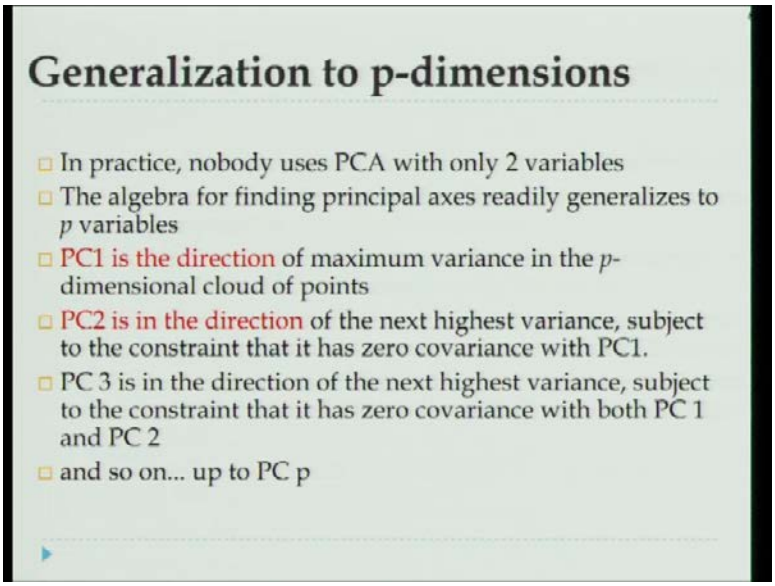
So principal axis should be perpendicular to so that is they are 0.

Now PC 2 to calculation competition is the important thing that PC 2 is also similar to whatever you obtain in the case of variance here also this is a user distance but its distance the measure of dissimilarity. How much it is deviating and based on this dissimilarity, you decide whether this component is getting the higher preference with the other one because the variable gives you the unit characteristics.

So, more the variance more the dissimilarity and you will be choosing it first and suppose you assume that it is a p dimensional and you observe that all p dimensional activities may not be there. Some case you will find that very closely similar you do not have to consider that one but you have to select in such a way that total dissimilarity values should be near 100 or some of the coverage some of the information's characteristics of the image, whatever you are considering within the k dimension should

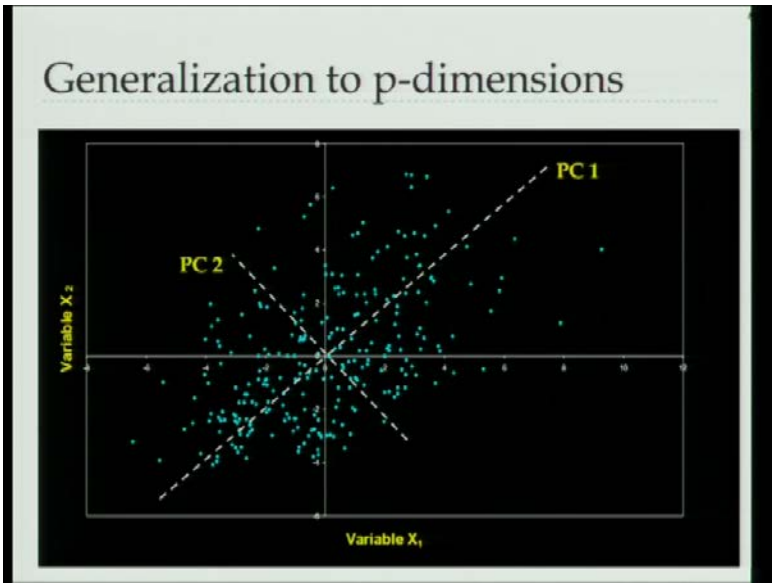
be as nearer as possible to the original image. That is the closely close representation or you have to have the original representation. You may not get the exact representation but as nearer as possible keeping as many as dimension reduces you can do on your image.

(Refer Slide Time: 33:05)



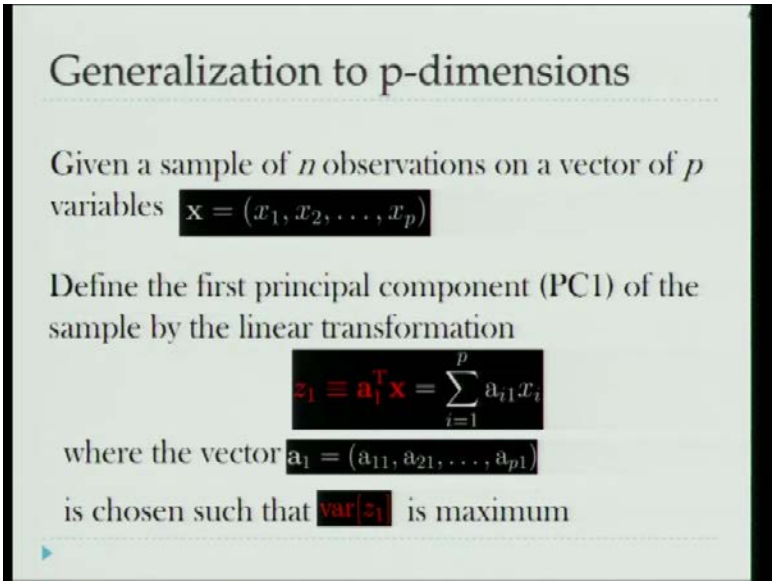
So, if we have the two variables and nobody does the PCA, just you find out the correlation coefficient. You can easily find out what is the pattern. You do not have to do that proper more than two dimension pattern for pattern re-cognize. Pattern analysis you need to do the principal component and as that principal means the direction of maximum variance to the p dimension and then you consider the PC 2 with the condition that this a perpendicular to PC 1 and so on.

(Refer Slide Time: 33:58)



So, in this image first you will be getting the direction this one and then with preference to PC 2 and then you rotate the image. So, it will maintain the property of all the things. How to obtain the corresponding coordinate of this value with preference to PC 1 We will be discussing later on.

(Refer Slide Time: 34:22)

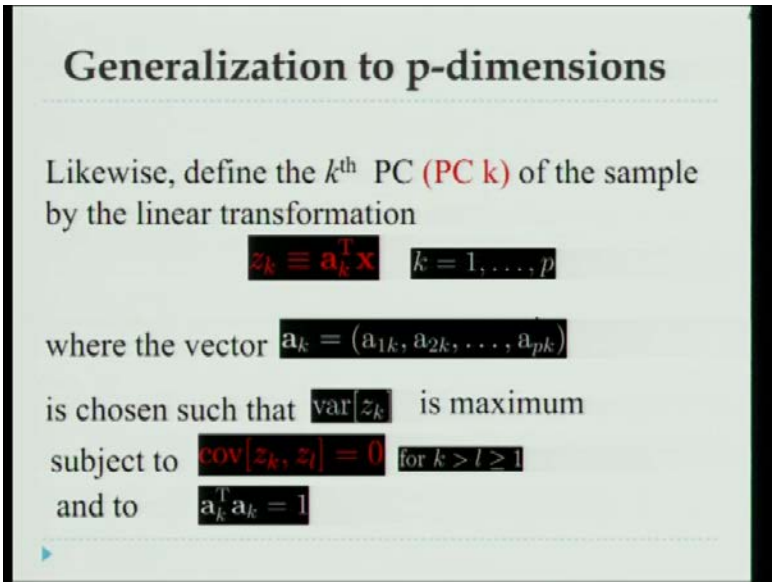


So, what is the formula? Or what is the algorithm? First let us find out how to find the maximum principal that is the first one then the second one third one and so on. Suppose you have the n observations X x 1 x 2 p dimensions this is a vector.

So first principal component is defined by z 1 equals to a 1 p transpose of a 1 X and which is nothing but summation over i 1 to p x 1 I a i. Now what is not known this you have to find out this Eigen vector? You have to find out such that the variance is maximum variance of z is maximum that is the problem. Your aim is to obtain the vector a in such way that variance of z 1 is maximum now can you recollect here. It was the three dimensional that is why we could solve this one but if it is a n dimensional how you could have been solve determinant finding is a very difficult problem. In n dimension what will be how can you solve and moreover do you know that PCA and other thing there readymade routines are available.

So, only thing conceptually it should be clear and other part is that if you just call PCA and you will get the results. So, tell me can you recollect how to find the determinant or find the values of solution of x minus lambda a equals to that 1 I want to find out the Eigen values. How can I find them, x minus lambda a equals to 0 we want to find the solution of that, there is no term singular value decomposition s v d. So, please and try to understand what s v d is? that is the only thing you do yourself because.

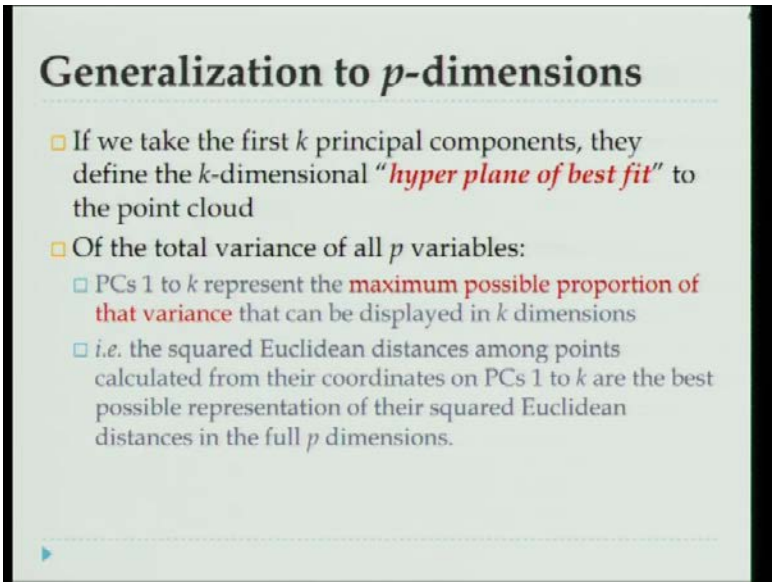
(Refer Slide Time: 36:42)



Now what happen the next if I want to find out the k-eth principal component? I will finding the z k same way a k T X, now you have vector and you have to select this vector with the condition that variance of z k is maximum but of course, must be less than z 1 z 2 z 3 z k minus 1 also the covariance between the z k and with the z 1 z 2 z 3 each of

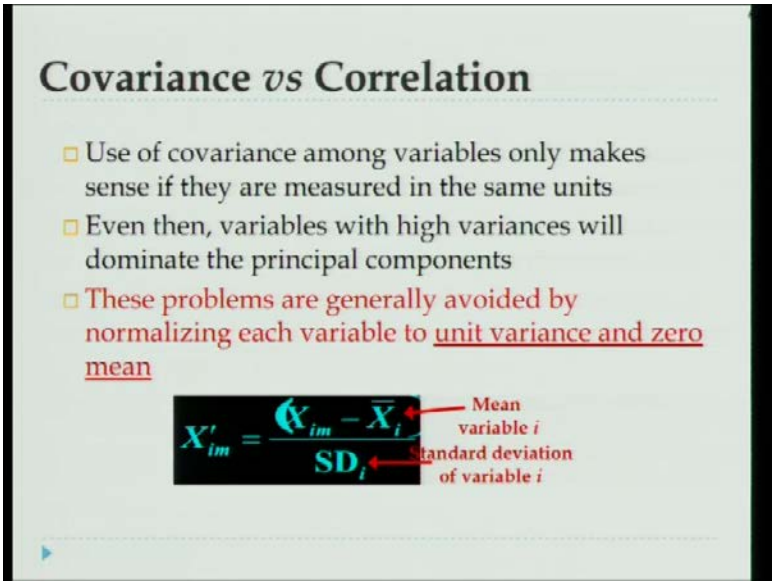
them that should be 0, because I told you that it should be perpendicular to every other axis and this is the condition of orthogonality so that should be maintained.

(Refer Slide Time: 38:01)



So, this part already I indirectly told that you are not selecting all the p value dimensional variables. We are selecting k of the k best of them on the hyper plane and these k piece parts of principal components 1 to k represents the maximum possible variations in the data and that is definition of nothing but we are considering the Euclidean distance.

(Refer Slide Time: 38:22)

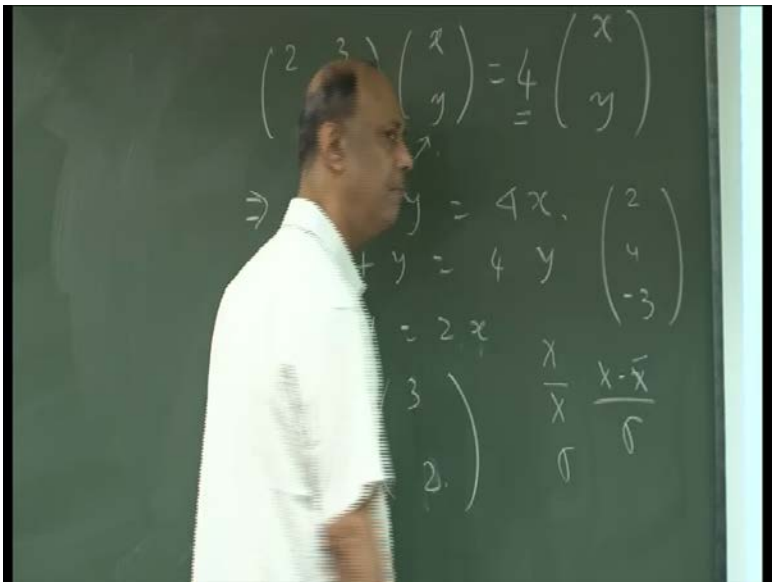


So, you could have used the simple covariance matrix on your data but there is a little problem, because their units the variables the p dimension units may not be the same. So, that is the issue if you know that the unit is same then directly you can obtain the covariance matrix.

Another thing is that you must ensure that the there is not too much dispersion not too much variance is not that high because the high variance will create or will have the positive impact towards the principal component. I have principal component so best way is that you normalize how to normalize it just X minus X bar divided by sigma.

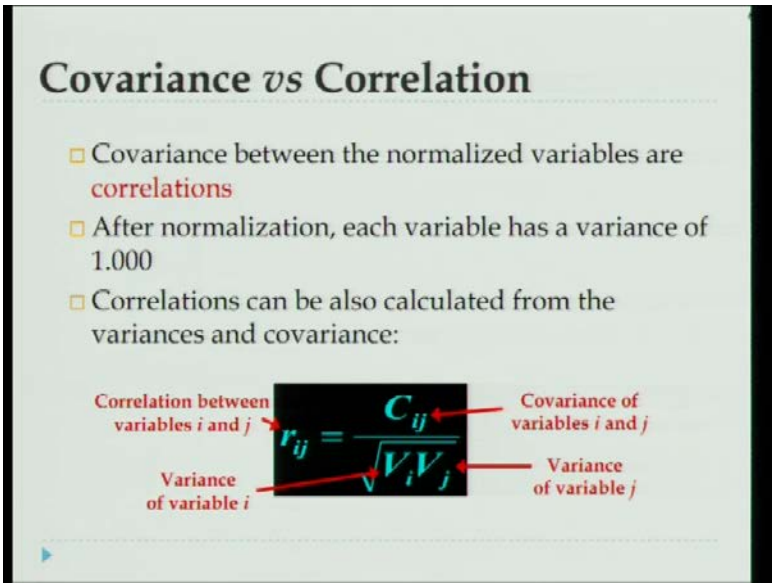
So, in that case X minus X bar divided by sigma if you do it then all the dimension your mean will be 0 and standard deviation will be one.

(Refer Slide Time: 39:41)



If you have the variable X and X bar is the mean and sigma is the standard deviation then X minus X bar by sigma obviously the normal distribution with 0 mean obviously the distribute the 0 mean and standard distribution one. So that is the thing so we generally normalize it to 0 mean and unit variance.

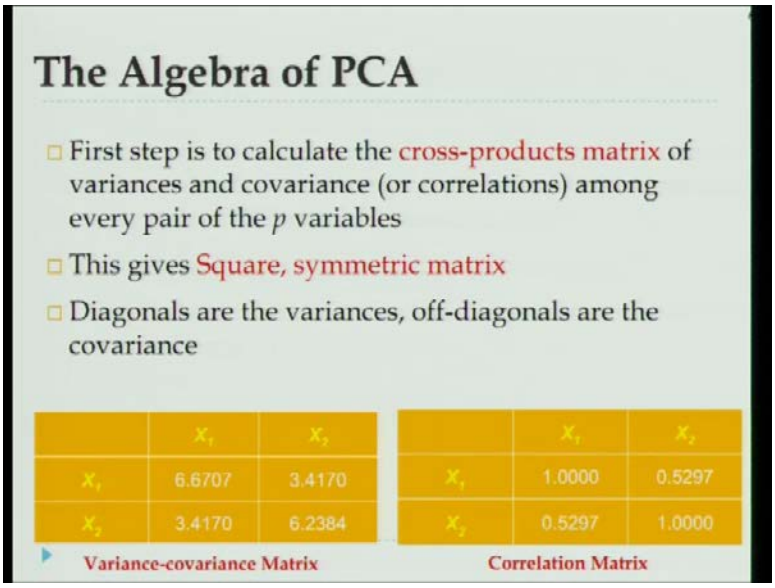
(Refer Slide Time: 39:58)



Now if you have the normalized elements or of subjects or objects of n dimension p dimensions then your covariance will directly and covariance will directly give you the correlation. Covariance because it is a normalized one X minus X bar divided by sigma so if I just covariance matrix will give you the information about whether there exists a correlation between any two variables and also that standard deviation, also each variable will have the standard variance one, because you have already divided by this sigma so variance will be always one.

Now if I have the covariance matrix that is C ij is the covariance between i and j then and if my variance is not normal that is not lying is not one then correlation matrix r ij correlation between i and j becomes C ij divided by square root of V i and V j .

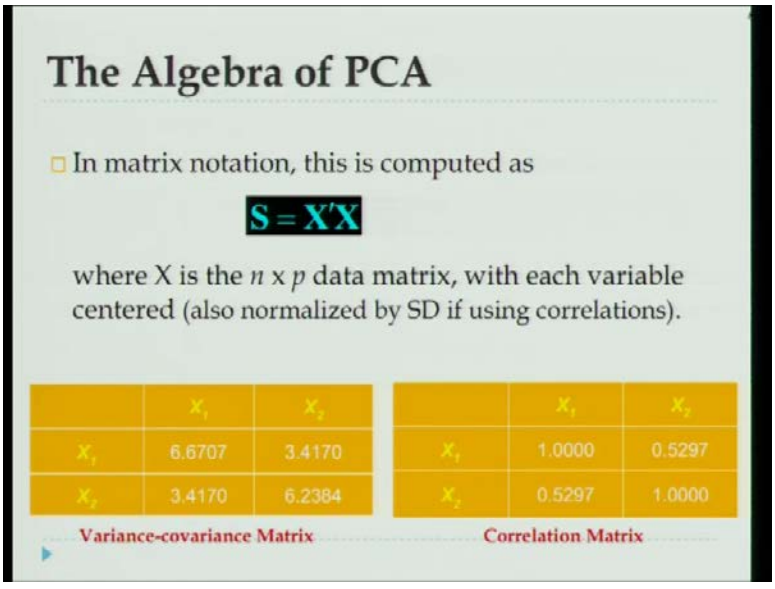
(Refer Slide Time: 41:32)



Now, this can be written that first you obtain the cross product of matrix of variance and covariance and if you have that cross cross product, what happen the diagonal elements. It will be the always variance and this will give you the correlation ship or variance covariance between X 1 and X 2. This is covariance between X 2 and X 1 and so on.

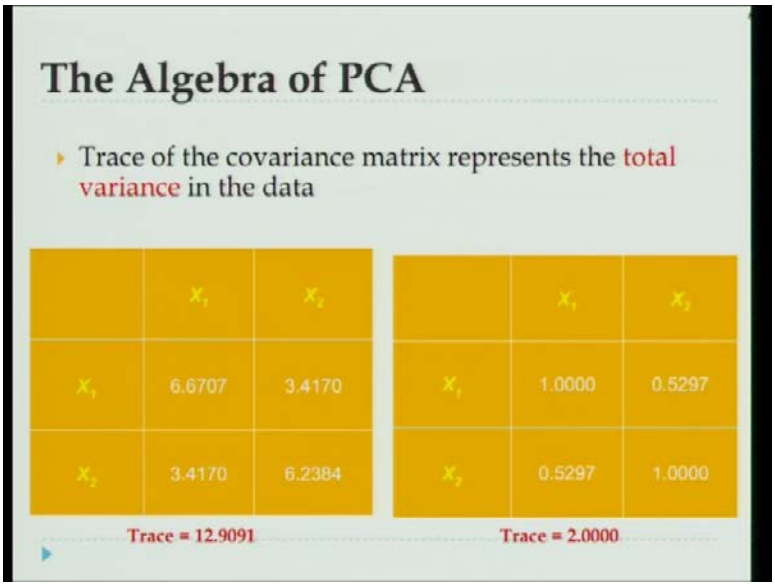
And this will become always a square matrix, because you are making x and x t the diagonal elements will give you the variance, and this will give you covariance and these are symmetry is there everything else clear.

(Refer Slide Time: 42:44)



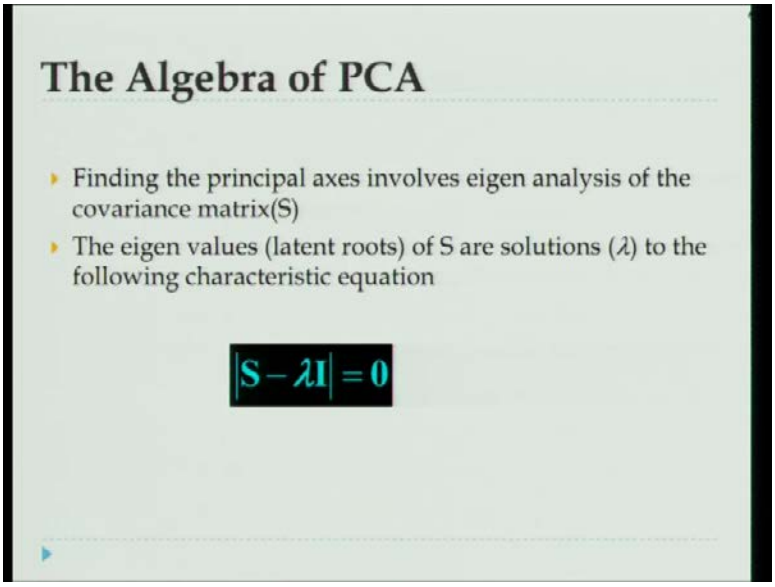
So, this can be in matrix form notation and this can be written as prime X, because that will give you the covariance matrix. Here X is a normalized elements and your size n cross p, otherwise you could have written x prime minus mu divided by sigma into. They could have written but this is normalized variables so you do not write those things.

(Refer Slide Time: 43:14)



Now remember that trace one trace of a matrix is sum of the diagonals elements.

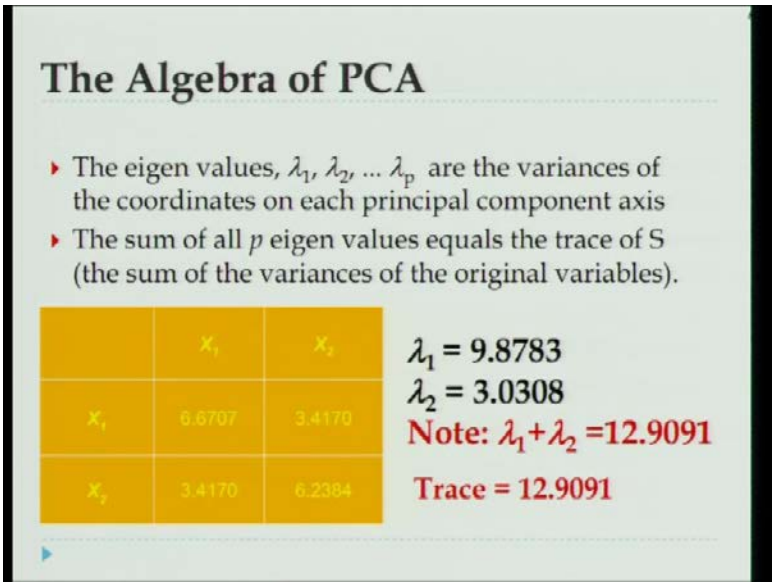
(Refer Slide Time: 43:34)



In that case it is a very useful property here or the relationship with the value of Eigen values. Now you have the here your notation x, and there you have written x that there is

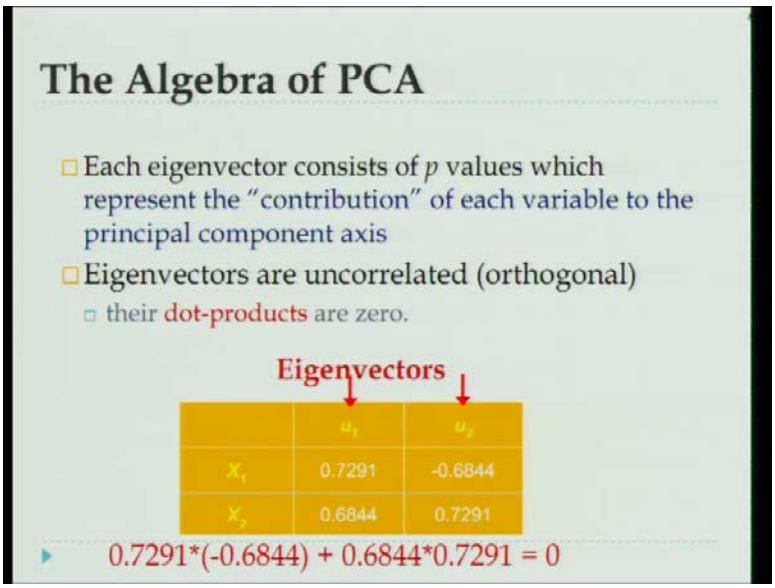
the determinant. You have to find out the solution of this characteristic equation and will be use the similar value technique to obtain the values of lambda.

(Refer Slide Time: 44:00)



And these Eigen values for this, one that Eigen values that will be coming lambda 1 equals to this and lambda 2 equal 0 3. That sum of these lambda values is equals to sum of the standard variables that this trace of this equals to trace of.

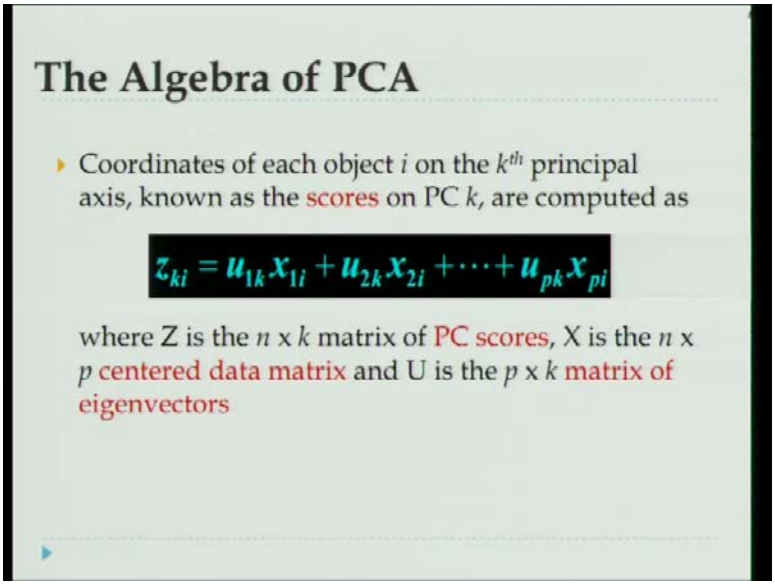
(Refer Slide Time: 44:29)



Now, what happen that you got Eigen vectors from these Eigen values? Now, each Eigen vector consist of how many variables, it represents the consist of p values. P values will

be there and it also gives you the contribution of each variable to the principal component, because that will provide you the principal axis and this for with reference to the original x y values; x values, it has certain contribution on it so Eigen vector consist of p values which represent the contribution of each variables to the principal axis and they are uncorrelated as their cross product. This into this plus, this into this will be 0 because they are uncorrelated.

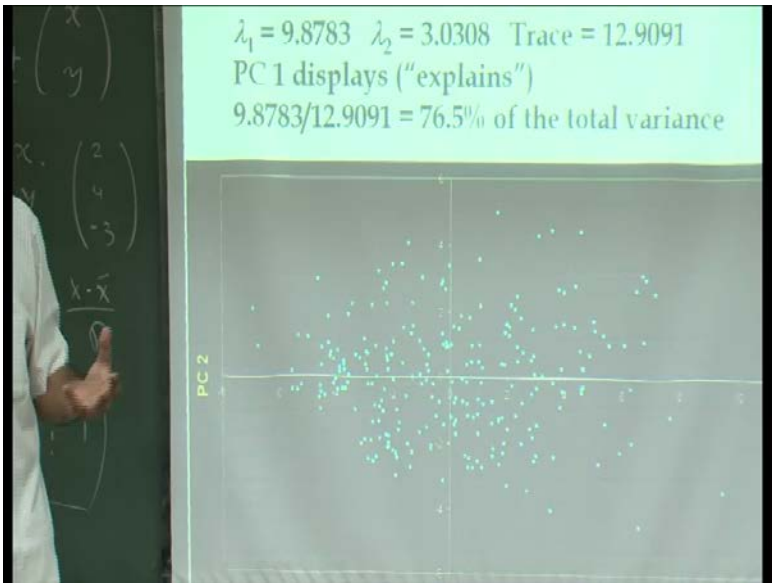
(Refer Slide Time: 45:25)



Now this x is given to you, the whole matrix original matrix and u is just now you obtained that Eigen. So, you need to know that what will be the mapping function for each x values into the principal axis. This is known as scope, so coordinates of each object on the k-eth principal component, which is known as the scope on PC k-eth principal component, which can be computed as z k i equals to this is known one k into x 1 I. These are also known u 2 k, because this is the vector Eigen vector and this is your original matrix.

Now, Eigen vector it has given you the p cross k matrix of vectors. How are you getting that one vector is of size p? There are k such vectors so the whole all this p, you have one vector of size. P then another vector of size and p like that there are k such vectors this form a matrix that is your Eigen matrix. So, you have X is of size n cross p and U is of size p cross k; so you will be getting z k i the corresponding value of on the with reference to the principal components.

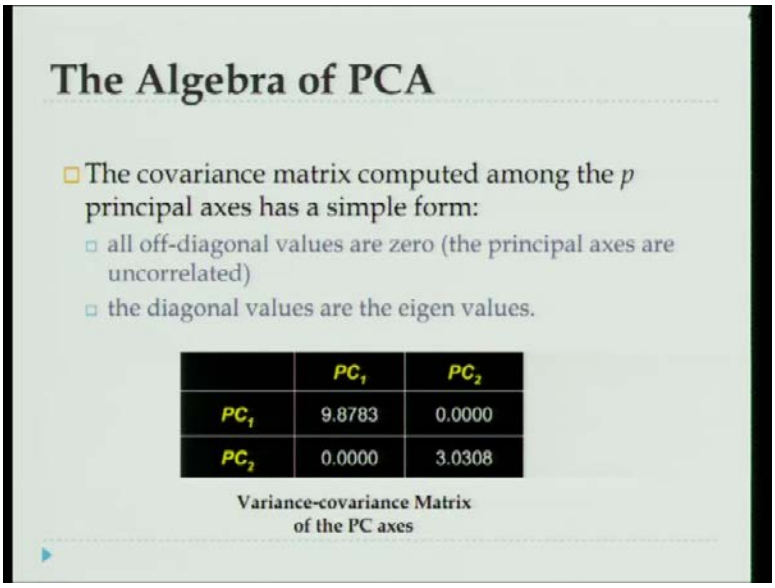
(Refer Slide Time: 47:36)



Now you know the lambda values, lambda is nothing but the eigen values lambda 1 and lambda 2. The trace is nothing but summation over lambda 1 plus and lambda 2, now what is the information you are getting with reference to PC 1. That is nothing but the difference or dispersion or variance or with respect to the first component. Eigen values divided by sum of all the lambda values that is the coverage or information you are getting with reference to the first principal component. That means that 9.8783 divided by 12 point this is 76.5 percent information are being carried out with respect to the first principal component.

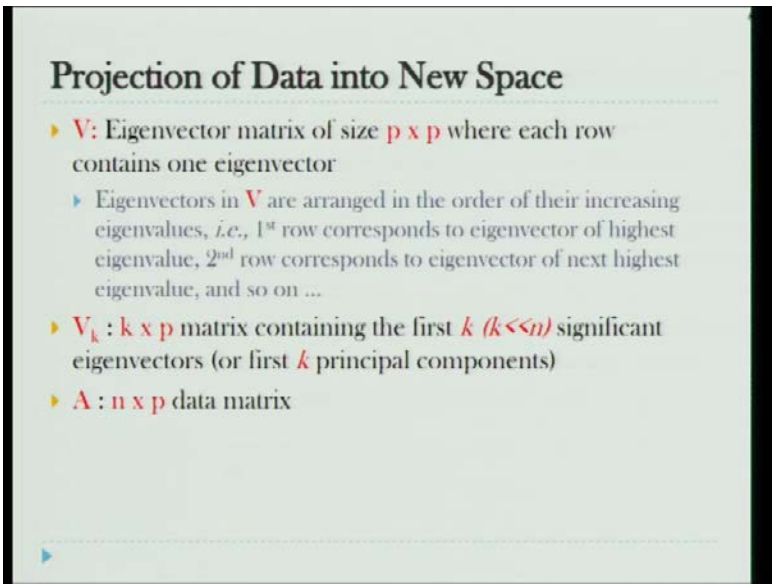
Now if I consider and second principal component, here since you have the two principal components. So, second will be the remaining of this element. It may so happen that there is a three principal components, and now you find out how much information, you are getting through lambda 1 divided by lambda 1 plus lambda 2 plus lambda 3, then lambda 2 by lambda 1 plus lambda 2 plus lambda 3. If you take the sum and you are finding that they are coming 99 percent that means the 99 percent of information you will be getting. If you consider only these two components and if you reduce the you do not consider the third one.

(Refer Slide Time: 49:13)



So, there is also again another correct all of them, because of this design problem. If I have that variance covariance matrix with respect to the principal component since they are orthogonal, they are uncorrelated. So, this diagonal these elements will be 0 only the diagonal elements will contain the variance and covariance part will be 0, because they are called not correlated.

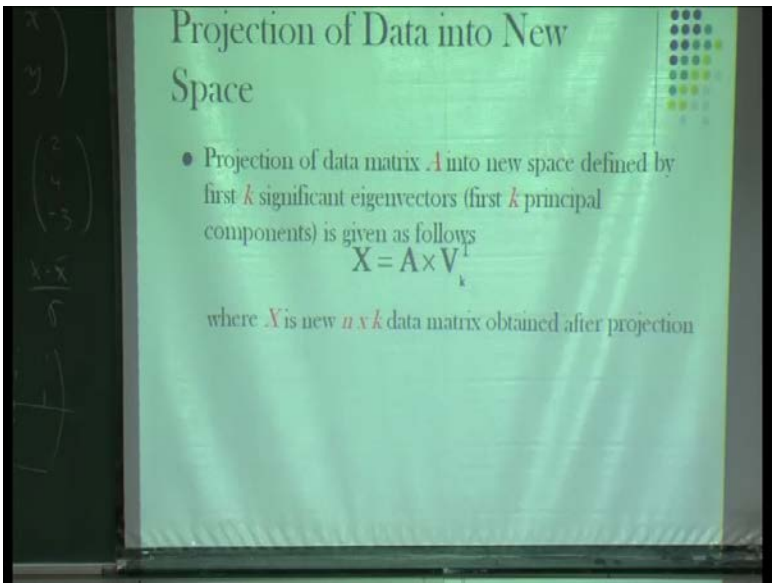
(Refer Slide Time: 49:38)



So, this is your variance covariance matrix. Now have to know the Eigen vectors you have to project it with reference to the principal axis. How it looks so you have these Eigen vector matrix of size P cross P with each row contain the Eigen vectors.

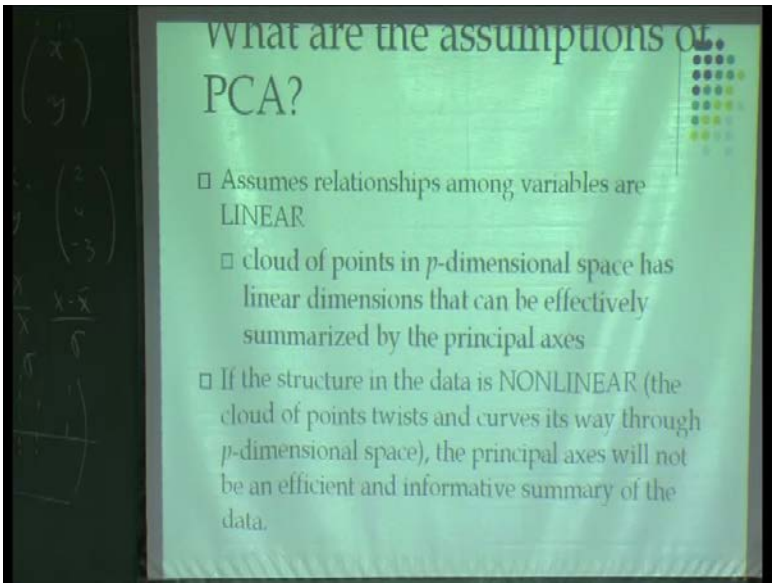
So, first row contain all the Eigen elements and the entire first row contains the large Eigen vector of the largest dispersion largest lambda value. Second row contains the all the Eigen values. Eigen all the elements of the Eigen vector having the second largest lambda values and so on V k is the sub part top k element top k rows. V k is the top k rows because do not want to consider the remaining ones. So, V k is the top k rows and a is your original matrix. N cross P so original matrix and V k transpose is your the top k rows of the Eigen matrix. So, if I multiply these two will be getting x which is the projected pattern.

(Refer Slide Time: 51:30)



Now, suppose I have the projected (( )) matrix and I want to see whether I can regenerate back or I can get back my A. So, of course you will be writing X V k transpose inverse should be equals to A. So V k transpose inverse is also V k, because V k v, it is relation is like this. That is the thing that you can always tell that this can be replaced by V. So, whatever value you are getting that need not be exactly a, because you have suppressed some of the information, but it is nearer to A. So your aim is to obtain such a V, such that after projection after regeneration I should get the near value of A. So, that k is important, that k is the require value.

(Refer Slide Time: 52:41)



Now, what assumptions you made? The most important assumption a is the relationship among the variables must be linear; otherwise, you cannot have this principal axis concept, because if it is non-linear, then this axis, principal axis, you cannot determine and everything has gone, whole idea will be gone. So, you must be able to express in terms of that something is equals to a into lambda plus something into lambda 2 plus something into like that.

So, this is the assumption in the principal relationship must be linear among the variables. So, any anything you want to ask on this. So we will be show, we will consider one example where face of once the way it has been used. PCA have been used intensively so we will consider that one while we will be considering that face biometrics.