Data Analysis and Decision Making – I
Prof. Raghu Nandan Sengupta
Department of Industrial & Management Engineering
Indian Institute of Technology, Kanpur

Lecture – 37
Copula theory

Warm welcome to all my dear friends and students, a very good morning good afternoon good evening to all of you. Welcome to this Data Analysis and Decision Making One course on the NPTEL MOOC series. And as you know this is a 60 class or lectures NPTEL MOOC series, which will go on for 12 weeks that is total 30 hours and each week we have 5 lectures, each being of half an hour and we are in the 8th week which is the 37th class.

Now, if you remember we are discussing about copula function and that some and the mapping, which basically happens between any the cdf or any distribution and the univariate distribution case. Univariate means, the discrete case or the continuous case as that, we do I want to map a one mapping and find out the actual values of x, such that they can map on mapped onto the u space. So, I am using the words for the first time, but you will understand.

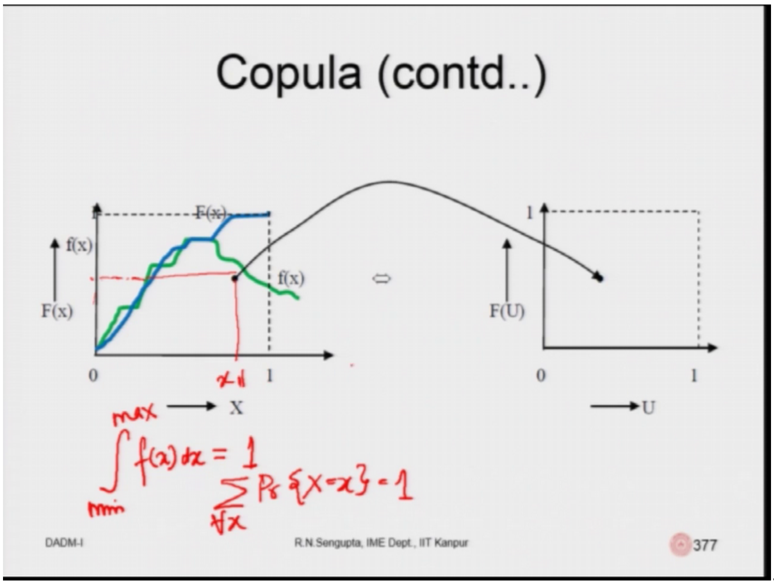
So, what we have if in a 2 dimensional case where we are plotting along the x axis, the random variable x consider that is only for the univariate case. And along the y axis, which I have the diagram which I have drawn in the 36th class and along the y axis we are basically plotting both the pdf or the pmf as well as the cdf of that function. And then, we also told that the main idea based on which we are doing it such that, we are able to map from the x plane to the u plane; x being the random variable which we want to find out for which we want to find out the copula function and u being the plane or the coordinate system for the uniform discrete case or you will consider the uniform continuous case.

The main fact why this can be done is because, the cdf function is between 0 and 1, maximum value being 1. The cdf function is monotonically increasing and all the properties of the probability distribution function of probability mass function holds for both the random variable x as well as for the random variable u. U being again I am repeating is the uniform continuous and distribution case.

Now, this idea can be also extended to the 3 dimension or the 2 dimension, we have already said 3 dimension 4 dimension and the higher dimension. Such that, if we have a the multivariate case for p number of random variables, we can find out the corresponding discrete, distribution or the continuous discrete, continuous a uniform distribution on the discrete uniform distribution, such that the mapping can be done.

Now, for considering the 2 dimension I will basically, first show the diagram again which I have drawn, I will again show it and then basically, go into the three dimension case that is, it will be easy for you to appreciate So, consider for the 2 dimension case, so on the screen on the left hand side.

(Refer Slide Time: 03:21)



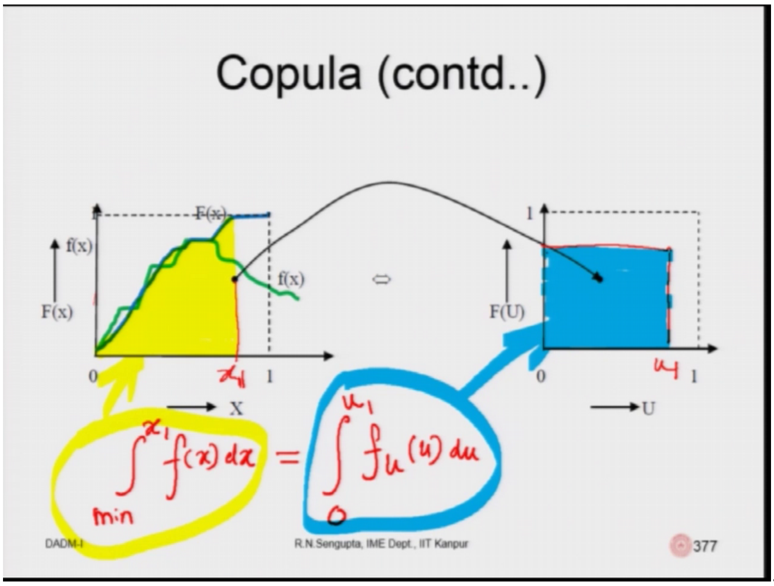
The left plate, you are basically the random variable x, we are measuring x along the x direction along the y direction you are basically, finding out the values of the pdf or the corresponding cdf.

So, consider this is a pdf function the continuous case. Now the green line which I have I have, drawn it very arbitrarily. So, the green line which I have drawn we basically, have the pdf values and consider it looks like a normal for the time being. And the blue line which I have drawn even though it is may not be exact, but I am trying to convey the picture. So, that is the cdf value.

So, one thing you notice, for the cdf value the minimum value is 0, where the pointer is and the maximum value; obviously, for the case would be 1. Now one value here which I am drawn, does not make sense, the actual value which I want to denote which as is a hidden behind this arrow is the one value which means, the pdf values if we add them up which gives the cdf, the maximum value is basically one because, the sum of the probabilities.

So, this is integration for the case summation means integration So, on the on the minimum value to the maximum value or if you have the this would be true. Now what you do is like this, you pick up a X value here, some x1 and its corresponding value is given for the pdf. Now consider this is x 1, which means let me erase it, so it will be easy.

(Refer Slide Time: 05:27)



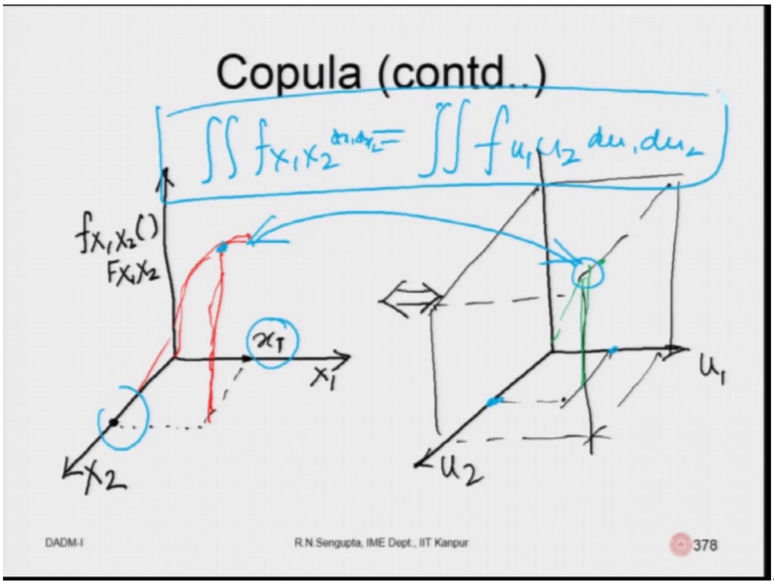
So, it will be from minimum to the x1 value f of x dx, so that would be the total area. this is This will be the total area.

So, this is the total area of the sum of the probabilities, this is the case. And that would be mapped on to the uniform a continuous case. So, if you want to map it, continuous f of u u du from uniform remember uniform, continuous case is between standardized 1 means is between 0 and 1. So, the minimum value is zero and this is the small value of u 1 which I want to have.

So, I will have sum, so this uniform case would basically be sum u 1. This is the u 1 value and the total cda value consider this is here, use wait right. So, this is the blue part, which is here and this is the yellow part, which is here. So, there is a 1 to 1 correspondence between the values of u 1, which is here and x 1 which is on this the actual for this univariate case and you can find out the u 1 and then do the calculations for the copulas accordingly.

In the three dimension case, what it becomes I will try to draw it and show it to you.

(Refer Slide Time: 08:37)



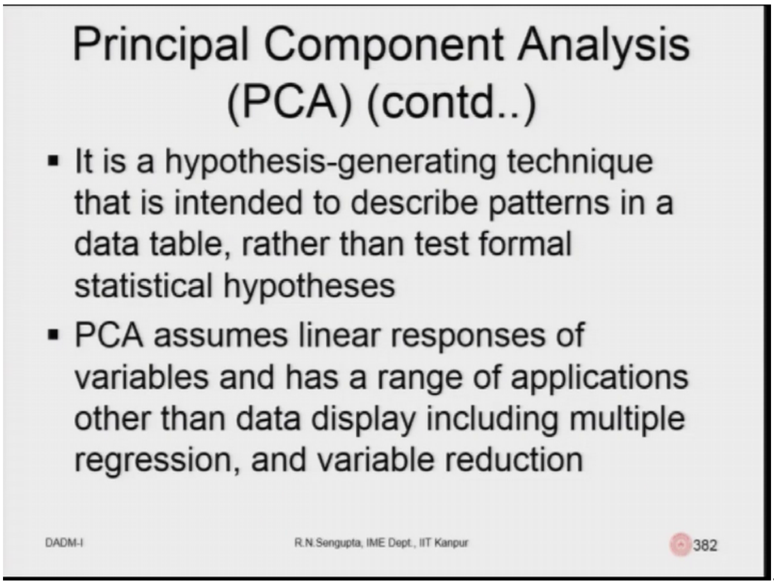
Let us draw multivariate case, you have x 1 being measured, x 2 being measured and this is the f x 1, x 2 which is the pdf joint distribution, along with that you have capital f, fx 1, fx 2 which is the cdf value So, if you pick up any 2 values x 1, x 2. so in the third three dimension, it will be some we use the red colour. So, this is some distribution which is in the three dimension case, some value which will be here.

So, now what we do is, this has to be mapped. Now I have the discrete case this is u 1, u 2. Now this will be in the continuous discrete case, where u 1 u 2 are between 0 and 1 it will be cube of dimension, 1 cross, 1 cross one. Maximum value 1, I am just drawing the cube first So, this would basically, have sum u 1 sum u 2 and the value, the discrete case would be somewhere here consider.

So, what you are doing is, you are mapping the value of these one, x 1, x 2 and the corresponding cdr value which you have, with u 2, u 1 and this such that, this will be possible So, here double summation of the total area, it will be the double summation of f of u 1, u 2, obviously here dx 1, dx 2, du 1, du 2 will come ok. So, based on that you can find on the copulas that has given, this is the idea which we have.

So, obviously that means that the cdf values is always has the same property for both the pdf, whatever the pdfs are. This I give a blank screen, if we needed to draw it anyway. Now, we will come to the copula concepts later on. So, this is the idea we will use it in some problems.

(Refer Slide Time: 11:51)



So now, we will consider the principal component analysis, which is a multivariate technique ordination technique used to display patterns in multivariate data.

So, we want to find out the components, which are principal based on which what is the maximum amount of information, which we can that is basically the idea. So, in the multivariate case, is there are different type of variables which have their effects. We want to basically, break down the effects in such a way that, we are able to in a in a simple way rank them from the maximum variability to the least variability such that, we can say with confidence. Confidence the word I am confidence I am using is not the confidence from the hypothesis testing, with we can see with confidence with a lot of information being there that which way set of variables are important and what is the overall effect on the on the dependent structure of the set of information, based on which we are trying to do the study.

Now, principal component analysis aims to graphically display the relative. So, we will I will draw a small graph or I have a small graph, which will give the better picture of this principal component analysis which is PCA. It basically, aims to graphically display the relative positions of data points in fewer dimensions; that means, lower the number of variability or such that, we have can have maximum amount of informations.

So, in a fewer dimensions, while retaining as much as a information possible and also to explore the relationship between the dependent variables and whether the dependent variables can be placed in such a way that, they are independent of each other. So, graphically when you are doing that, you will understand how it can be done, considering the independent structure. And that is a very simple concept which you already are all studied in class 10 or 12 in basic physics, basic mathematics, in basic concepts of sciences such that, the idea what we have there would can be brought down concept sheet on the graphical framework.

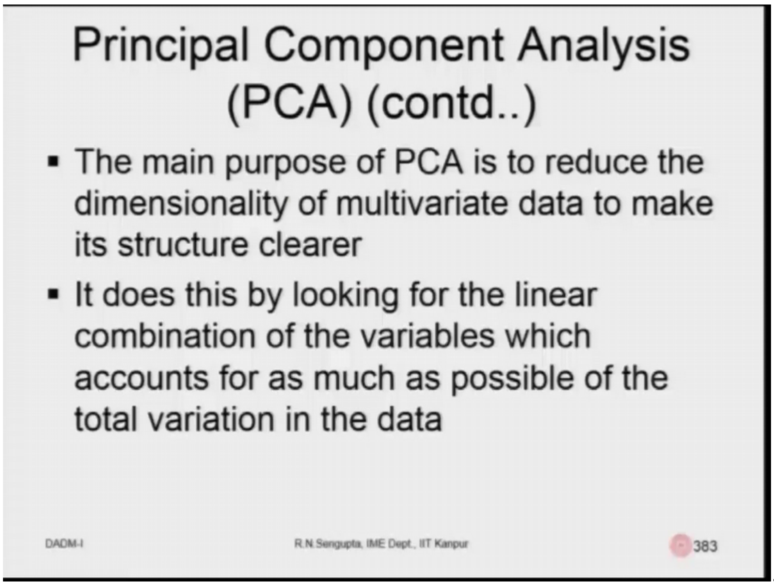
Not I am not talking about the calculation front will tackle it accordingly. Now in principle component analysis or PCA is an hypothesis generating technique, that is intended to describe the patterns in the data table and rather than test the formal statistical hypothesis. So, using the data we will find out plot, them on the graphical frame, try to find out what is the relationships and rank them from the maximum dependence to the minimum dependence and we would do not want to basically formally state the hypotheses as they are applicable.

PCA assumes linear response of variables; so obviously, we will have the relationship between the dependent and the independent variables. I should use the word the of independent variables in the sense there are more than one variable. So, to consider this x 1, x 2, x 3 till x p and you want to find out the relationship of each of them on the dependent structure. This is not nothing to do with a multiple linear regression, so we want to basically find out what are the ranking sort of thing and that means, maximum effect to the minimum effect.

In regression we give a corrective dependence structure between the x 1, x 2, x 3, till X p and the dependent 1 which is y. So, PCA assumes linear responses of variables and as a range and applications other than data display including multiple linear regression it can be utilized, but it is a little bit different. And variable reduction where you want to reduce the number of variables which are very important, in order to basically understand what that what the dependent structure can be.

The main purpose of PCA is to reduce the dimensionality that means, if there are p number of variables which have been giving, you want to find out that what are the minimum number of variables which gives us the maximum amount of dependent structure or the output; whether by the word output I means the dependent structure.

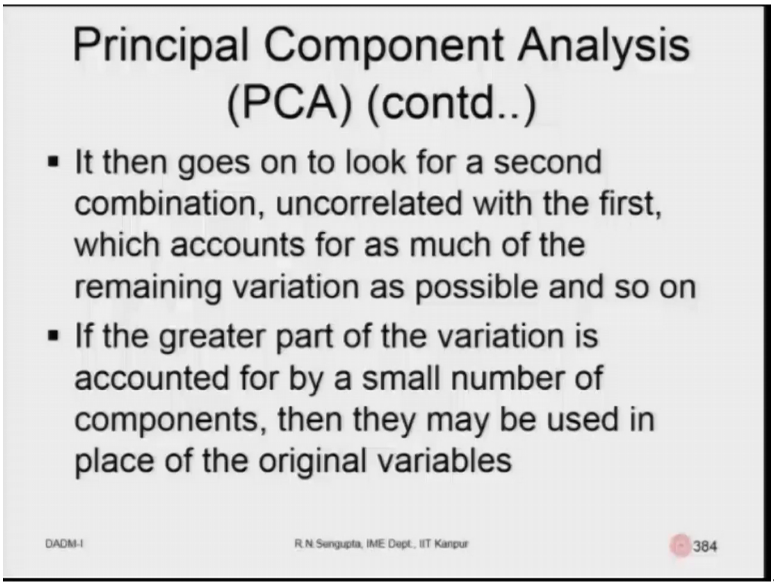
(Refer Slide Time: 16:01)



So, it is basically reduce the dimensionality of a multivariate data to make structure clearer and find out their dependence concept. It PCA basically does this by looking for the linear combination of the variables, like we want to combine the variables linearly, combine the variables linearly means those x 1 to X p which are there we want to basically, give them some weights such that the combination of them would give us the dependent structure stage by stage.

That means, we want to basically give the maximum amount of information, based on the first set of linear combination. So, once that linear combination is there we try to gather or glean the maximum dependence structure take it out. Then the rest of the structure or the dependence the information which is there, we basically propose a second set of linear structures between the same x 1 to X p and continue doing it in such a way that, we are able to glean the maximum amount of information the dependence structure point 1. And point number 2, when we are doing that, it basically formulate the depend the linear combination between the independent variables in such a way that, technically they are those sets are independent of each other. So, that is what I mentioned that I will try to basically portray it graphically, so, we will understand in much better way.

(Refer Slide Time: 17:21)



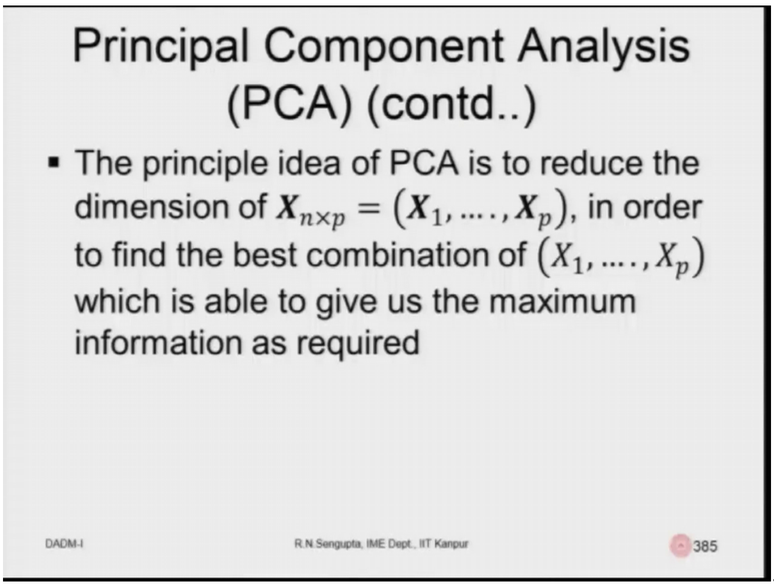
The method of PCA then goes on to look for a second combination those who just or what I mentioned. So, first of all once you find out the dependent structures on the combinations for the first stage we glean out take the information, in the next stage after the left amount of dependence, we find out the second combinations we give the second ranking of the information structure. So, that is what is mentioned it then, goes on to look for a second combination uncorrelated with the first.

So, once I am mentioning uncorrelated basically, means it is something to do with orthogonality in the in that graphical framework. So, that is why I mentioned that we would definitely would have done such small things in class 10, 11, 12 such that, pictorially it will be really easy for us to basically have a look at how the principal component analysis looks at. So, it go then goes on to look for a second combination uncorrelated with the first, which accounts for as much as the remaining variation as possible and so on and so forth. If the greater part of the variation is accounted for by as small number of combinations of the random variables the independent random variables are then, they will may be used in place of the original variables and we are able to give the dependence structure with the least number of variables such that our overall task is reduce.

So, what we are trying to do is that, have a set of variables which are already given, combine each of them at stage by stage such that, we give the minimum number of combinations which are each orthogonal to each other, using that minimum number of combinations we are giving a we are able to give the maximum amount of information about the dependent structure of those x 1 to X p and the dependent variable. That is all what we mean by principal component analysis.

And again I am repeating, we will try to do the problems, you the steps of the calculation in the way as I mentioned verbally we will see that. But pictorially it will be easy for us to see that, as we are doing the calculation pictorially it will also be give us a lot of confidence, when we try to understand that how the combinations of the random variables x 1 to X p are being done; such that, they are independent of each other means they are orthogonal on the graphical plane such this 1 to 1 correspondence between orthogonality in the graphical plane and the independence and uncorrelated structure in the calculations will come out very easily.

(Refer Slide Time: 20:01)



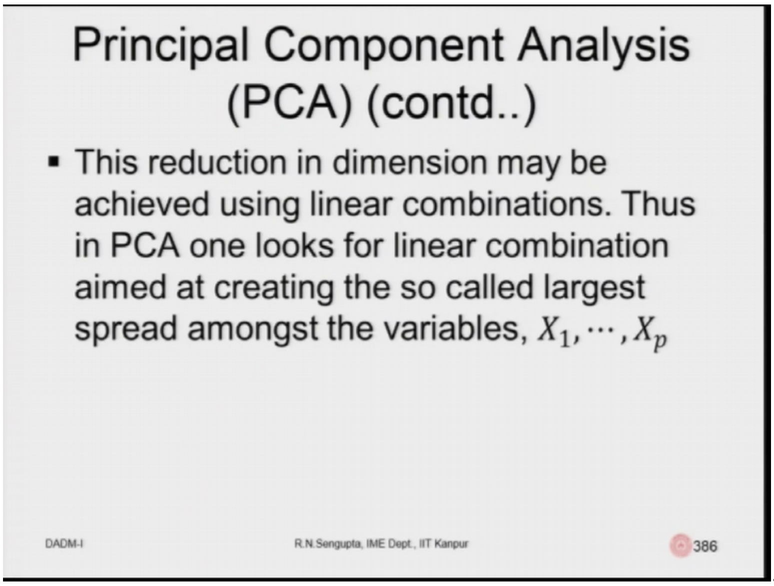
.

So, the principle idea how the principle component analysis is basically, to deduce the dimension of the actual matrix which is given, which is X the matrix X n comma cross p where, n is the number of the readings, the readings will not decrease remember that. What we are trying to reduce is p and p are the number of variables which we have. So, here we have this matrix is broken down into vectors.

So, these are all bold X 1 till X p where, X 1 bold is basically the first vector or the column whichever, you denote depending on what does the nomenclature of the of trying to basically predict the matrix or portray the matrix it is. So, X suffix 1 is basically the first set of values for the first variables and they are of size n cross 1; that means, is a column vector. Similarly the last one would be X suffix p which is also n cross 1, but here the random variables are for the pth 1 so, obviously, we will have x 1 x 2 x 3.

So, I am talking about the vectors, capital X 1, capital X 2, capital X till capital X p. Now we want to find out the best combinations of them, which is basically combinations that means, we want to combine in such a way, so, that they will give us the maximum informations are required depending on the dependent structure.

(Refer Slide Time: 21:29)

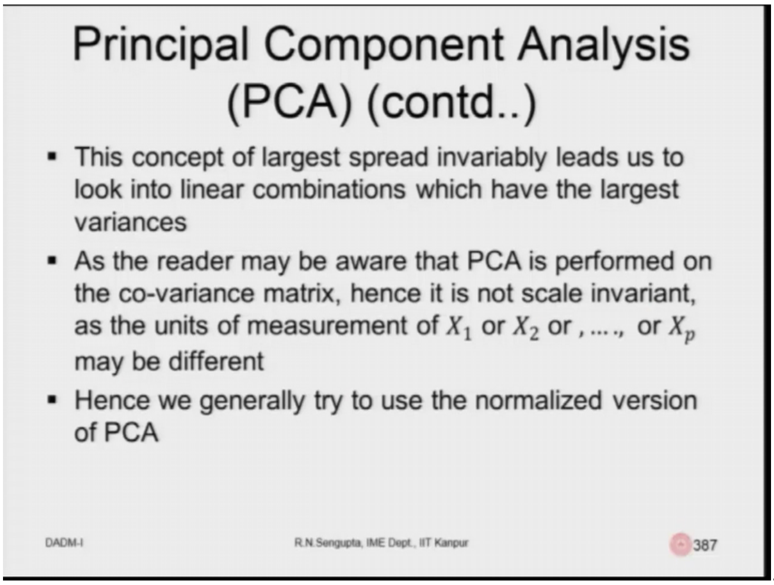


This reduction in dimension or dimensionality may be achieved using linear combinations. We are considering linear combinations of this random variables only no non-linear combinations. Thus in PCA one looks for linear combination aimed at creating the so called largest spread among the variables X 1 to X p stage by stage such that, we are able to find them in such a way that each combinations are orthogonal to each other just stage by stage such that, the overall first set of information which I glean from the first combination can be kept aside such that, the effect of the first set onto the second or second to the first will always be of no significance because they will be the orthogonal.

Similarly, when we take the second one, considering the first is already in a move, we will place the second place, means place it in the in the dimensionality in the graph in such a way, that it will be independent for the third also, the third combination of the set of X 1 to X p. So, if the first and second are orthogonal and second or third to are orthogonal, we will ensure in the calculations that, it will also be orthogonality would also be maintained between the first and the third.

So, once we take out the third one, we basically go for the fourth combinations. So, if the set of combinations for X 1 to X p which is the fourth combinations which we have, if it is orthogonal with the third; obviously, it will be orthogonal to the second it will be orthogonal to the first. So, we go step by step and the variability would be maximum in the first case, would be a little bit reduced in the second combination, would be a little bit further reduced in the third combination and go on in such a way, that the cumulative effect which we want to find out for the combinations of this X 1 to X p can be found out in the least combinations such that the dimension dimensionality is reduced and the maximum set of information is gleaned or gathered as fast as possible.

(Refer Slide Time: 23:35)

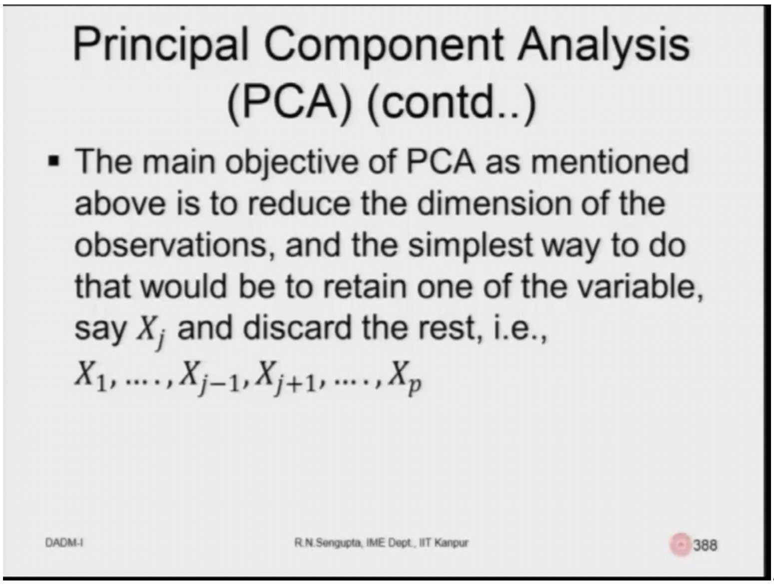


The concept of largest spread invariably leads us to look into linear combination which have the largest variance. So, the spreads being large means variability is large, spread being low means variability is low. As the reader may be aware the concept of principal component analysis is performed on the covariance matrix. So, obviously when we have X 1 to X p remember that, these X 1 to X p the matrix which we have may not be independent on each other.

So, the combinations we are have formulating is made in such a way they are being forced to be orthogonal to each other. Hence it is not scaling invariant, so the scaling would basically affect them because, if you reduce or increase one of them by unit of 2.5 and decrease one by unit of 0.5, obviously this concept of a scale invariance will not be true. So, as the units of measurements of X 1 or X 2 or X 3 or X 4 or X p may be different such that, their scale concept would be important when you are trying to basically do the PCA method.

Hence we generally we try to use the normalized PCA method; that means, we try to find out the normalize concept of PCA such that, the units do not make any sense or units do not basically effect when we are trying to basically combined in the first set second sets so on and so forth. Because in that case trying to find out the variability and then find trying to find out the orthogonality, would be much easier than in the case when they are non normalized or not normalized.

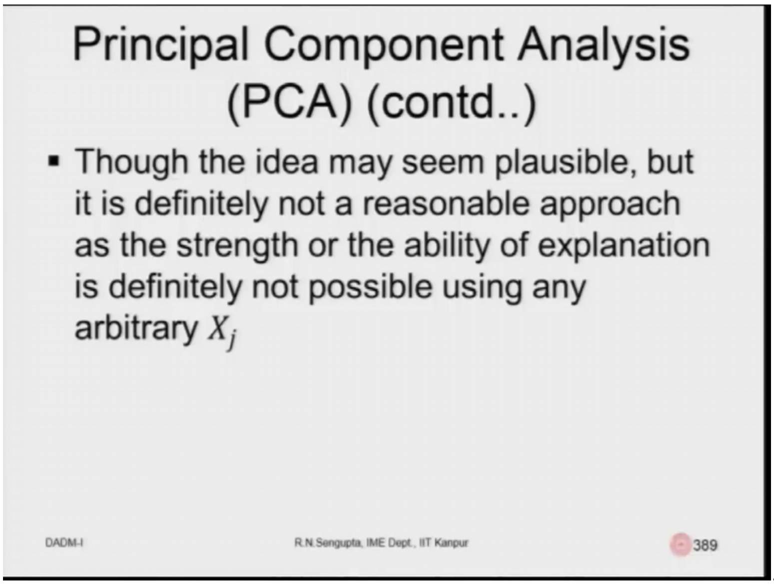
(Refer Slide Time: 25:13)



The main objective of PCA as mentioned above is to reduce the dimensionality of the dimension on the observations, and the simplest way to do that is to retain one of the variable say X j and discard the rest. So, we will basically take randomly X jth 1 and discard X 1 to X j minus 1 and also X j plus 1 to X p such that, the maximum the randomly the word randomly I am saying is that it would not be do not done randomly it will have some logic.

So, once we take out that random variable, we are able to basically take out at one go the first step the maximum set of informations which is available. And we continue doing it in such a way, that we glean out step by step thus maximum amount of information of dependence structure by the least number of variables.

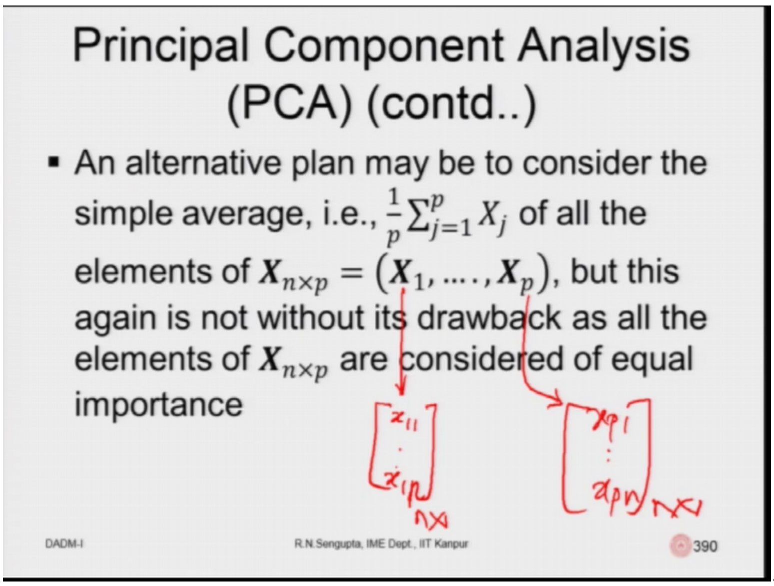
(Refer Slide Time: 26:05)



Though the idea may seem plausible, but it is definitely not a reasonable approach as a strength of their or the ability of the explanation is definitely not possible using any arbitrarily X j, that is what I mentioned.

So, arbitrary picking up any X j may not give us the best starting point based on which you want to find out the maximum variability and then go step by step, as we try to basically find out the maximum variability coming out in the minimum number of such random variables. So, that may not be possible.

(Refer Slide Time: 26:42)



So, there is an alternative plan. So, what is the alternative plan and an alternative plan may be to consider the simple average; that means, we find out these combinations of the variables X 1 X 2 X 3 till X p in such a way that, we find out some convex combinations of them.

So, that is what, we will come to that. So, an alternative plan may be to consider the simple average; that means, we are giving weightages and consider there are p number of variables; obviously, the weightages for each of them would be one by p 1 by p; obviously, it can change and that will depend on what the variables you think are have maximum amount of importance. So, such that simple average of all the elements in X 1 to X p, So, X 1 to X p are the vectors which I have just mentioned.

So, I just and know that you have understood it, but I still I will I will draw. So, this X 1 this means so this is n cross 1 so, but when we find out equal weightages, but this again is not without its drawbacks as all the elements on X p n are considered to be of equal importance will which not be true. So, the first is randomly pick up giving up any X j s would not give us the best solution, given us equal weightages for to x one to X p may not give us the equal solution.

So, we will basically see it in the next class that how we can do it this principal component analysis considering, the fact that on two main tasks are there, which I am repeating again. Number 1 take the combinations in such a way that, we are able to glean up maximum variability in the first case, then reduce variability, the next step reduce variability the next second step then, the third step so on and so forth point 1.

Dimensionality is reduced, maximum amount of information is possible and when we basically find out the combinations such a way, that the first set would be independent of the second set that is a orthogonal, second set would be independent on the third that is a orthogonal, but; obviously, we will ensure that first and third is also orthogonal. When we find out the fourth stage or combinations it will be orthogonal to third second first and we will continue doing this such that we are able to clean the maximum set of information. So, with this I will end this 37th lecture and continue discussion more about in 30th lecture and so on and so forth have a nice day and.

Thank you very much.