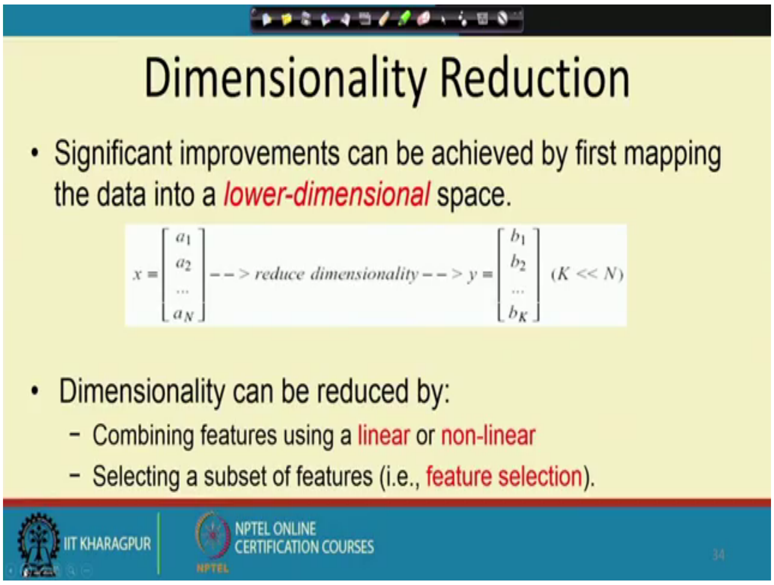
Data Mining
Prof. Pabitra Mitra
Department of Computer Science & Engineering
Indian Institute of Technology, Kharagpur

Lecture - 42
Dimensionality Reduction- II

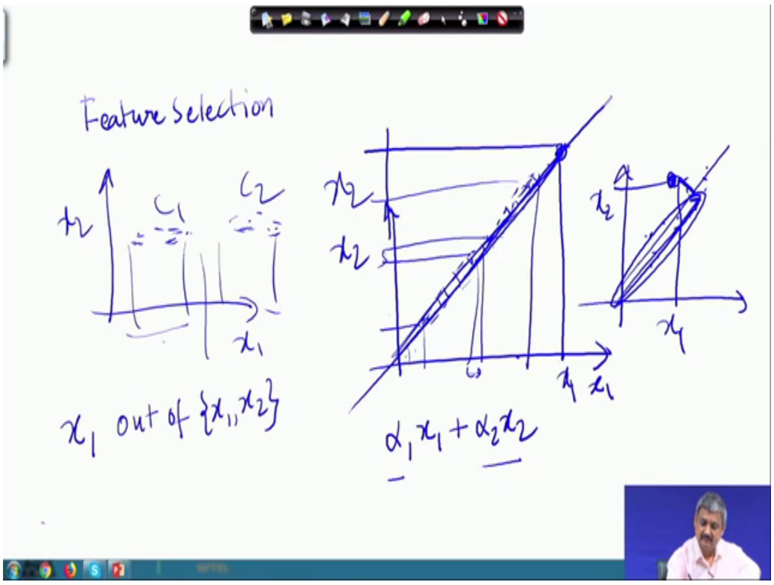
We discussed in the previous lecture how you select a subset of features for the purpose of dimensionality reduction.

(Refer Slide Time: 00:28)



So, let me give you a particular view of this approach.

(Refer Slide Time: 00:32)

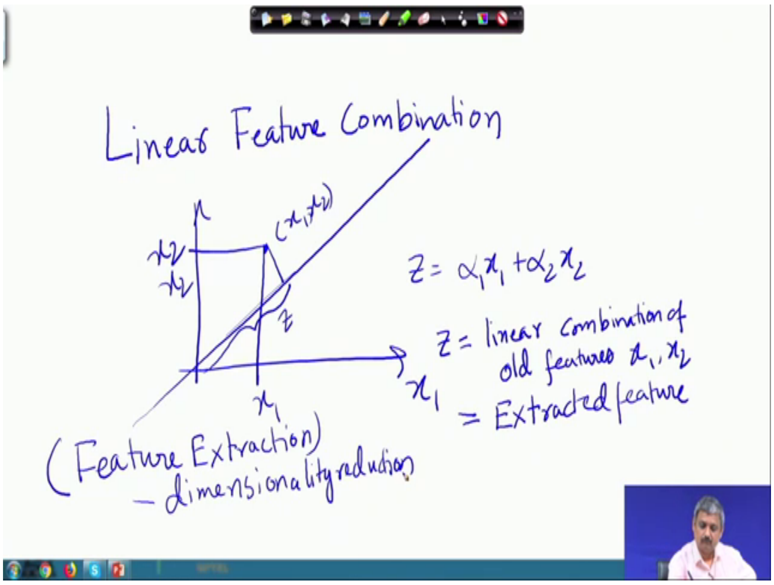


So, what we did in future selection is the following. So, if we had two features for example, and if the features were like this I project it to x 1 and take only x, because only x 1 is sufficient to say classify these two classes if this is class 1 if this is class 2 which are x 1 only is enough to separate between class 1 and class 2.

But suppose my data is distributed like this distributed like this the same thing, but tilted. So, now, if I project to x 1 the class separability is much reduced, if I project to x 2 also the class separability is much reduced. So, what do I do? So, earlier I was choosing x 1 out of x 1 x 2, here what I do to do that thing. So, can I do one thing here is that instead of projecting to x 1 or projecting to x 2 if I project in a direction like this if I project the data in a direction like this now there is much more classifier ability ok. So, here what I do I look at values of x 1, and I decide class 1 or class two here I do not look at value of x 1 I do not look at value of x 2, but I look at some value which is alpha 1 x 1 plus alpha 2 x 2 which is this direction because any point here will be depending on the slope of the line there will be alpha 1 x 1 plus alpha 2 x 2.

If you if you take the curve this distance that is the projection from the origin distance from the origin of this point it is this value x 1 alpha 1 plus this value x 2 alpha 2 ok. I can express this. In fact, to be particular alpha 1 is x 1 by sin theta this is theta plus x 2 by sin theta that will give me this sun this line ok. So, so instead of projecting to axis I project to a axis which is not this, but some other tilted axis. And what happens if this is this tilted axis? The projection of this point to this tilted axis will be some function of projection of; so this is my point projection to this x 1 projection to this x 2 some combination of x 1's into alpha 1 x 2 plus alpha 2 will give me this value projection to this ok, maybe I will properly write it down.

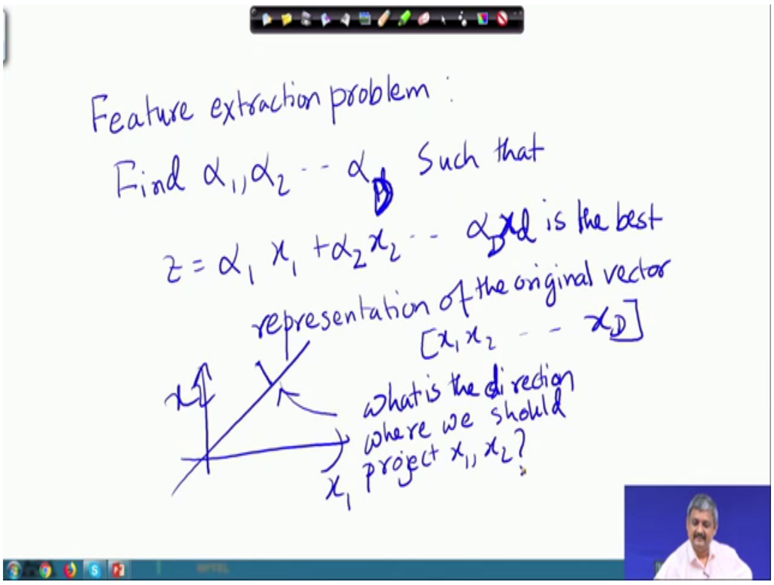
(Refer Slide Time: 04:30)



If this is a point x 1 x 2 this is x 1 this is x 2 the projection of that point to another line this value can be written like this ok.

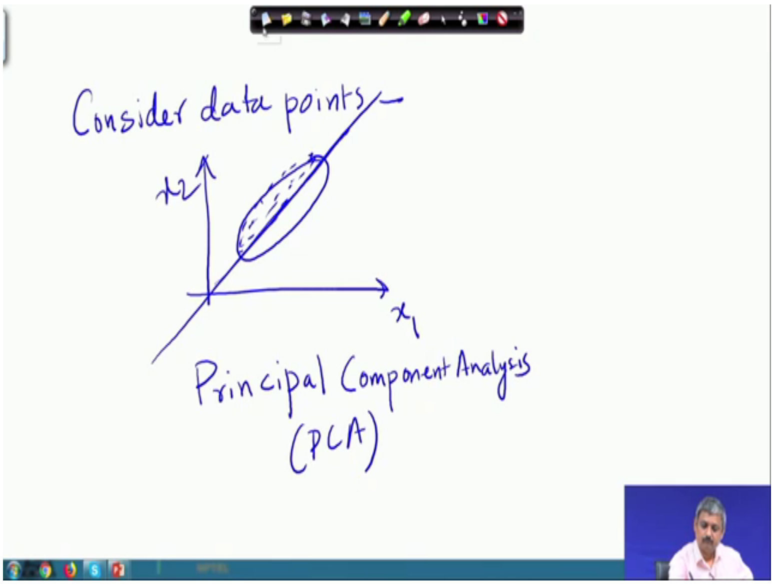
So, this new feature Z also known as the extracted feature, this entire approach is known as feature extraction. Note that this is also dimensionality reduction because it reduces dimension Z is only one value x 1 x 2 are two values.

(Refer Slide Time: 06:43)



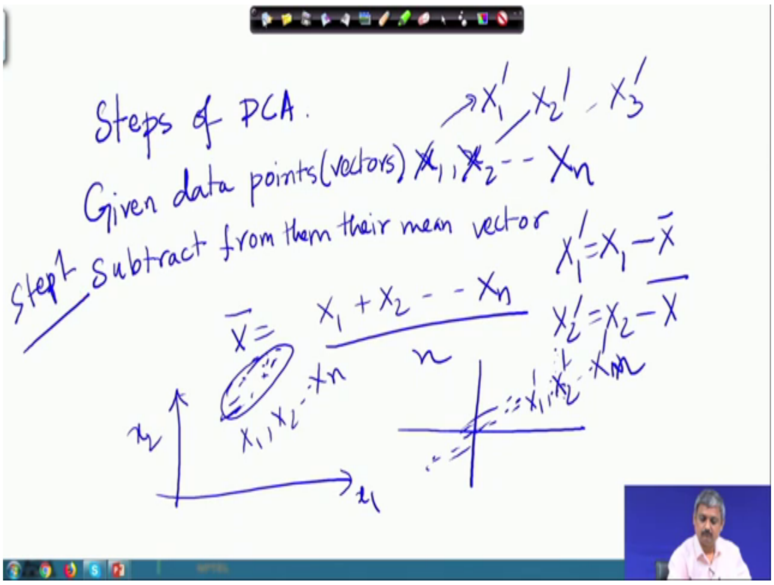
X D, what is the combination which is best? In other words x 2 (Refer Time: 08:28).

(Refer Slide Time: 08:30)



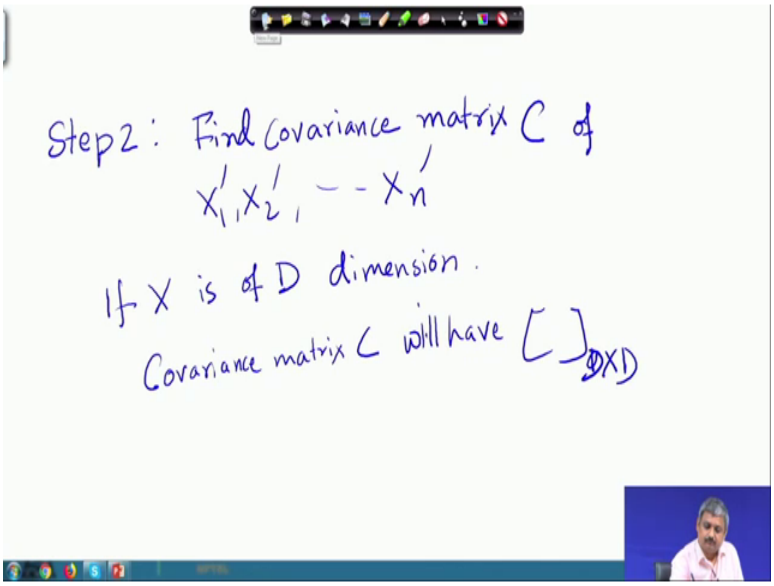
Suppose my data is distributed like this training data. So, which direction I should project? Naturally your answer will be this. Why this? Because suppose if I fit ellipse this is the principal axis. So, that is the idea of something called a analysis.

(Refer Slide Time: 09:40)



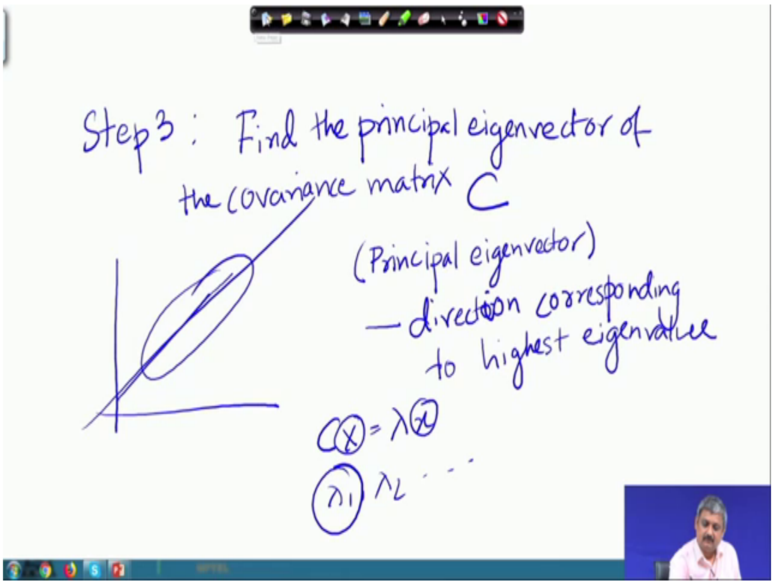
In other words if it is your data point you subtract the mean and get your data points like, where you get. So, these points, points you centre them in step one.

(Refer Slide Time: 11:51)



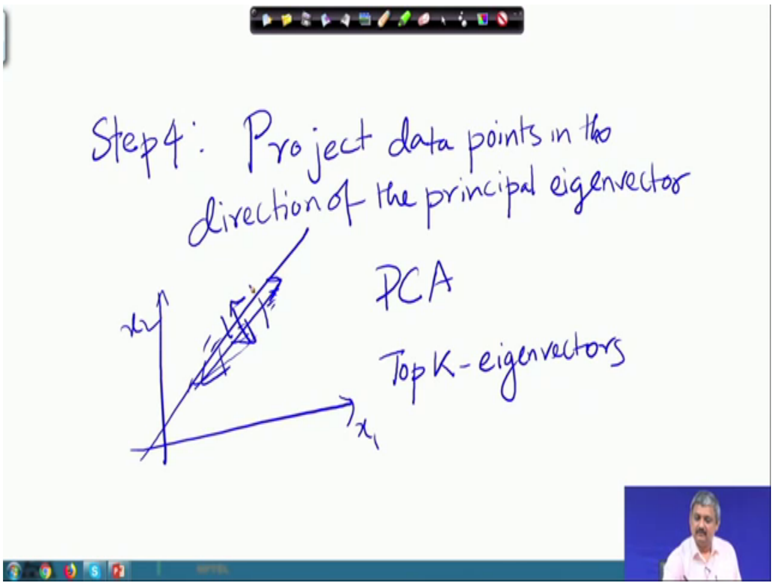
So, if X of; D by D matrix ok; so you know how to get the covariance matrix, you basically you take the cross terms cross terms covariance matrix.

(Refer Slide Time: 13:13)



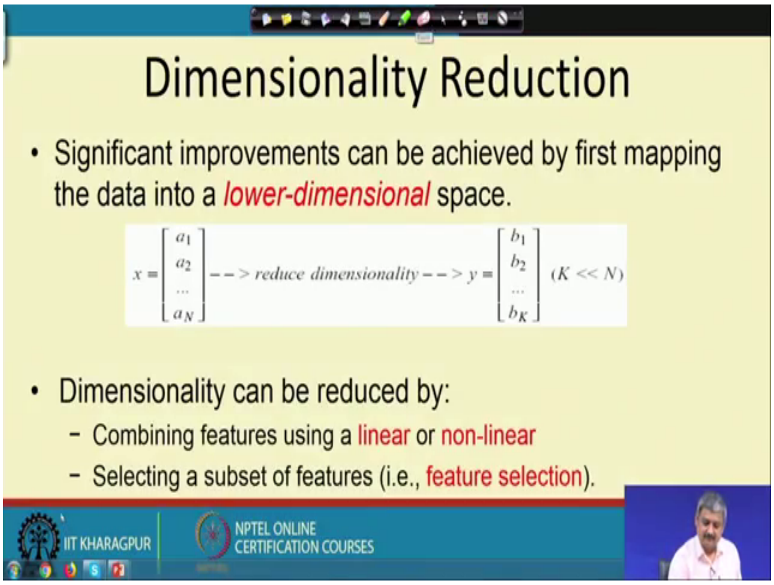
What is principal eigenvector? Value, so we will basically solve C x equal to lambda x you will get all the eigenvalues lambda 1, lambda 2. So, take the highest eigenvalue and any vector the unit vector satisfying that direction will is that principal direction ok. So, if your data is like this. This will be the principal direction, eigen means natural direction ok. This is clear, this will give you the principal direction.

(Refer Slide Time: 14:55)



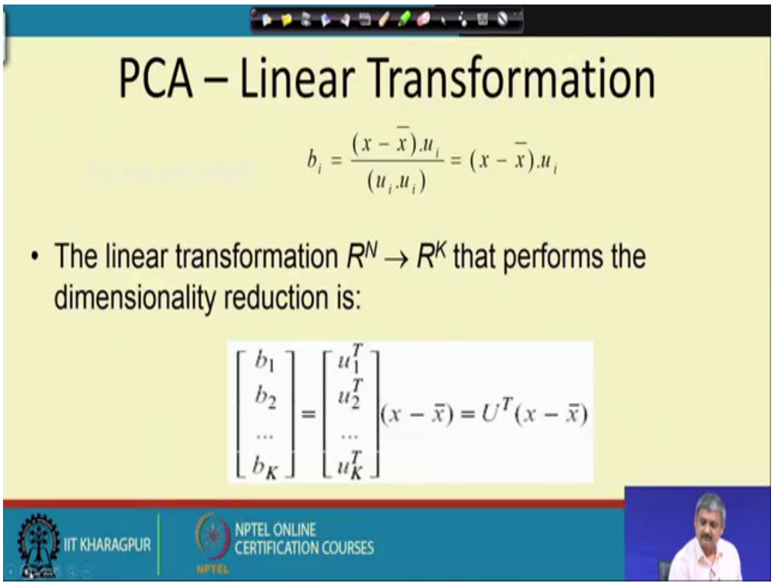
You project your data points to this direction that is your reduce dimension PCA. In general if you want to proda produce on a project to two dimension you take the top two eigenvalues. If you want to project to K direction take eigenvectors ok, this is the maximum spread naturally. So, this direction is the highest variance of the data, this direction has the least orthogonal direction, this has the highest. So, that is the idea. So, this is the basic idea behind principal component.

(Refer Slide Time: 16:17)



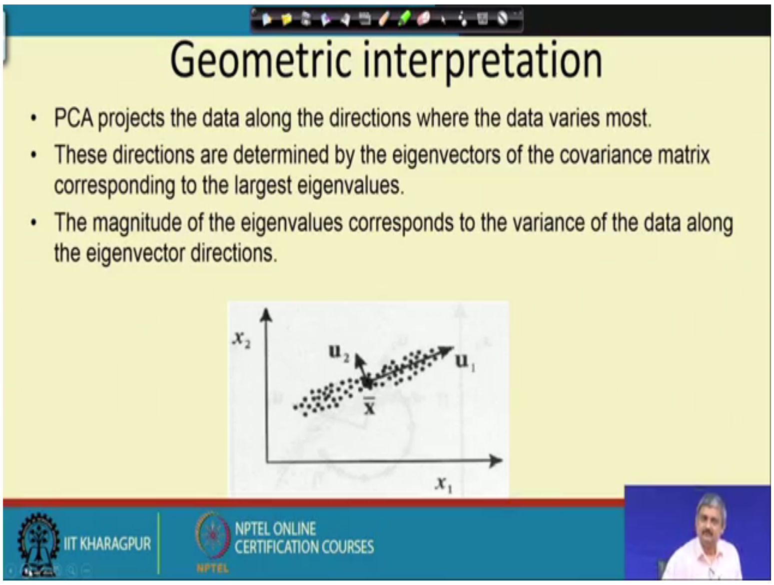
I will I have all these mathematics you should go through them. So, these are the steps.

(Refer Slide Time: 16:26)



So, basically it is a linear what you (Refer Time: 16:29) direction.

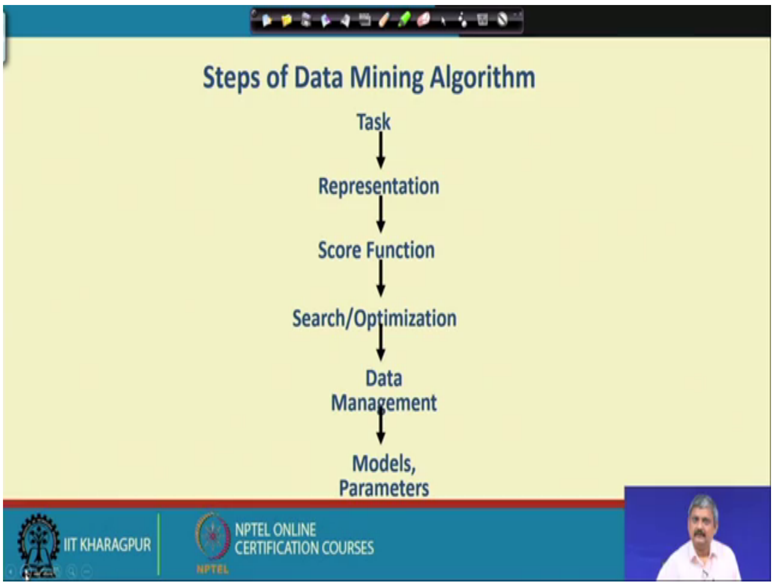
(Refer Slide Time: 16:28)



There are many extension to this for example, non-linear principal component analysis, it projects not to a light but to a curve, there are neural networks like self-organizing map which does this many of these are used. So, this is the step.

In summary I would like to tell that if you want to apply data mining algorithm you have to do this go through these tasks.

(Refer Slide Time: 16:56)



You have to first see what is the task, then do a representation, so either to dimension attraction or something find a good set of vectors attributes. Then see how to define a goodness of this function of certain model mathematical model to fit this data which is measured by the score function of the right classification accuracy or clustering. Then find for the best model which gives this do the model parameters and return that. There is one more step after this which is interpretation, visualization. And actual finally, deployment in your business process; then you are done with your data mining algorithm. So, this is a iterative you go back and do it and finally, you solve it ok.

So, with this I close the course I will give tutorials and on case studies on some programming and other problems which you can follow, but this is the end of the theory.

Thank you.