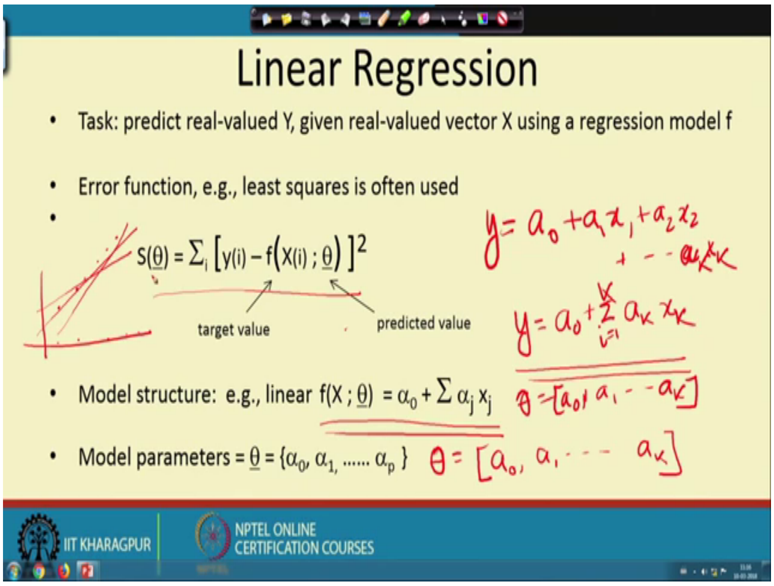
Data Mining
Prof. Pabitra Mitra
Department of Computer Science & Engineering
Indian Institute of Technology, Kharagpur

Lecture - 39
Regression- III

We will continue our discussion on the linear regression. As we discussed we had a model of the form, we had a model of the form of this type where y is a naught plus a 1 x 1 plus a 2 x 2 up to a k a k, x k ok.

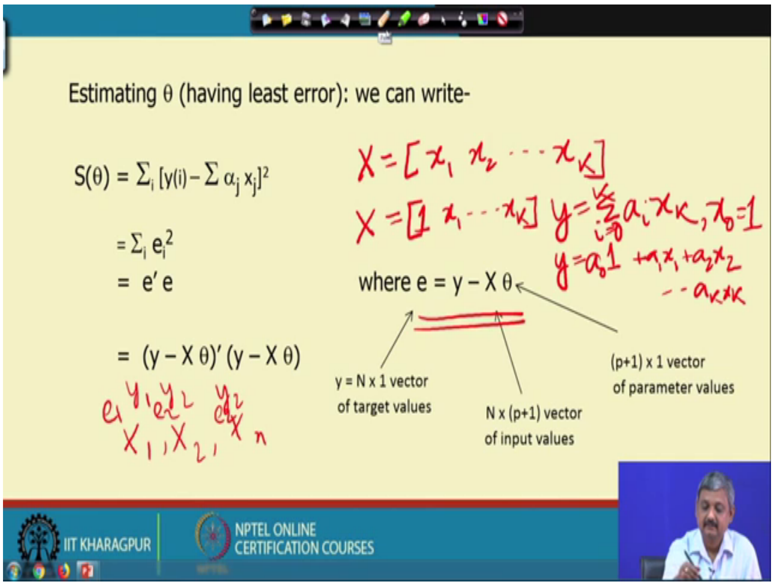
(Refer Slide Time: 00:33)



So, you can write it down as a particular form y equal to a naught plus sigma i equal to 1 to k a k x k ok. So, this set of value, so this is the regression model or the regression equation this one is the regression equation.

Well, I have written alpha here. So, you can as well put alpha here. So, the model parameters we call it as theta would be denoted by this a naught a 1 up to a k these are the these are the model parameters. So, theta is a naught a 1 up to a k. What we are supposed to do is to given a set of points x 1 y 1 x 1 x 2 x 3 and the corresponding y 1 y 1 y y 3 we have to find this parameters theta a naught a 1 a k. So, that the error that is the actual value minus the model predicted value which is this quantity model predicted value square of that and some of that of course, it will be a function of theta because if we change the slopes of the line here we will change this quantity sum squared error is minimum. We have to find that theta.

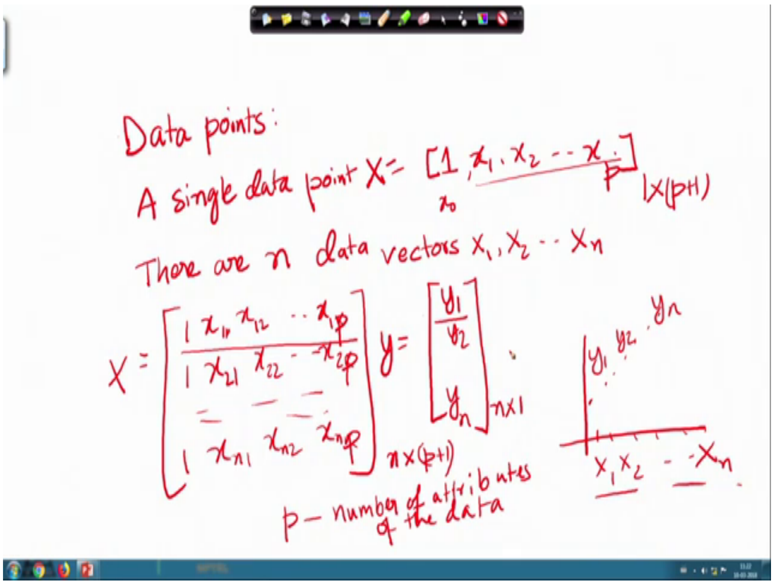
(Refer Slide Time: 02:34)



And what we did quickly was that and so you if you way of estimating the theta you can write down your error e as y minus summation alpha j, x j. So, what we will do actually is that, so each data point x each data point x even a multivariate class case is a collection of k or d components ok. And for the ease of the algebra what we do is that I sort of write x in a slightly different form for this a naught. So, you remember y is a naught plus a 1 x 1 I introduce a dummy component which is always 1 ok. So, then I can write y as summation i equal to 0 to k a i x k, where x 0 is always one when x 0 is always one, so just another way of writing this equation ok.

So, here x vector the multivariate x vector becomes 1 at x 0 and remaining values in the remaining positions slightly augmented. So, if we write in that form I can write my squared error e as this, squared error e as some sum this square which is summation e i square and if your e is a, for every point you have a vector. So, you have x 1 x 2 x n and you have y 1 y 2 y n and every point there is a error between these two that error vector I call that is a e vector which is nothing but y minus I will explain what is the e vector.

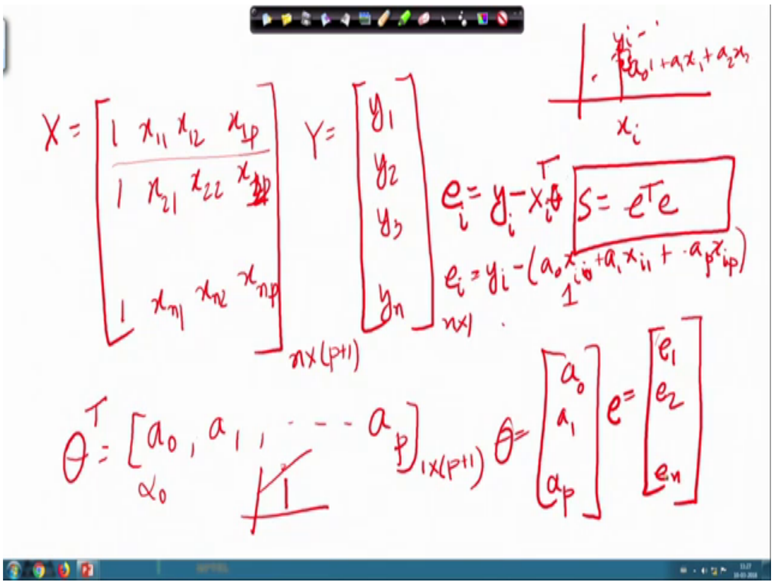
(Refer Slide Time: 05:10)



Let me write down the notations a single data point is a if we have or let me write it as p the notation. So, suppose there are p components. So, this is 1 cross p plus 1 p of this and 1 for x naught y n. So, I can write a big vector x like this. So, this is first vector first component first vector, second component, first vector k component ok. So, this is all these vectors I have write as a matrix. Let me call the capital l a a a x matrix.

So, what is the dimension of this matrix? It is it is n cross naught k p, let me write p is the number of components of the multivariate input number of attributes of the data. So, we have p plus 1 and n such vectors. So, this is x. What is y? So, each of this x have a desired y, so y is a n cross one vector ok. So, what is e? Let me clear (Refer Time: 08:32) x n p at each of this point what is the y.

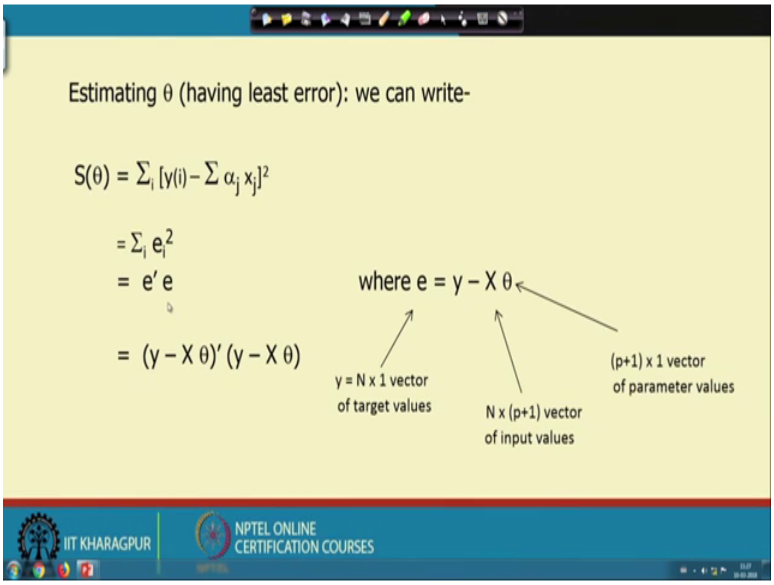
(Refer Slide Time: 08:30)



I have a model vector theta which is the coefficient of the regression. So, I have a naught or alpha naught whatever we want to write a p there will be p coefficients for a p dimensional p attributes plus 1 extra coefficient. So, it is 1 cross p plus 1. So, what is the error vector? Error vector is y minus if you take each of these small x's, x i let me call it. So, y i, e i is y i x i into theta because see they sorry x i transpose x i into theta transpose I can I can write because wait not x i into theta transpose x i transpose into theta transpose into ok, theta also let me make it the other way round. So, theta is actually let me write it as this vector.

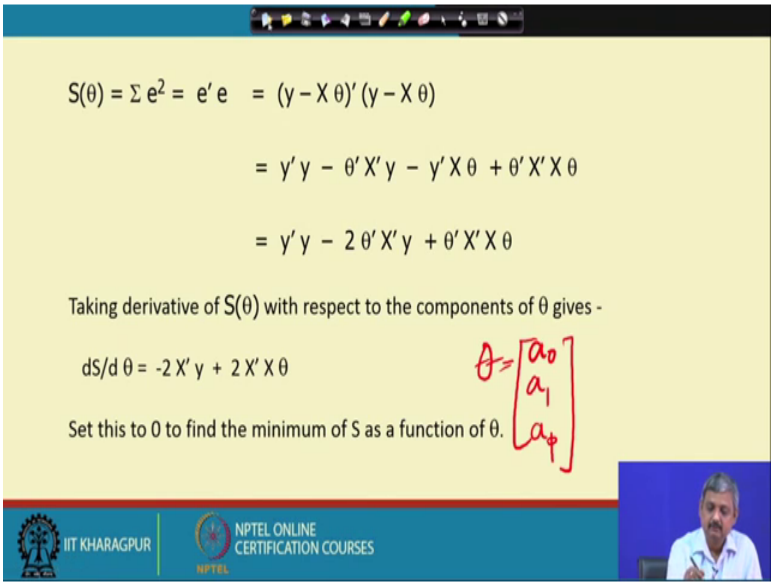
So, e i is y i minus sorry a 0 x i 0 which is one a 1 this is my e i error at point i, error at the I eth point if you remember the diagram error at the eth point is take x i find out a 0 1, a 1, x i 1, a 2, x i 2 and so on. Find out this and find out the actual y i the difference between that 1 ok. So, this e i, so I can also write a e vector as each of the point have an error ok. So, I take y i I take x i transpose multiplied by theta that is this quantity subtract that from y i, I will get my e i I will write it down e i as a vector. So, my sum squared error is you can imagine is sum of e i square e 1 square e 2 square e 3 square. So, I can write it as e transpose e this transpose e it will be this square plus this square if you do the vector multiplication ok. So, that is my error that is my error e transpose e if you match the dimensions you will get it.

(Refer Slide Time: 12:43)



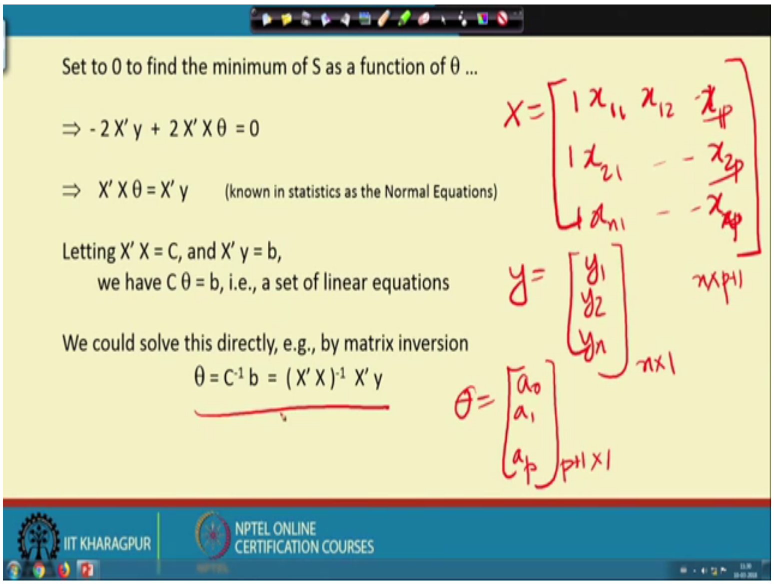
So, then we did all the minimizations; that means, take the derivative with respect to theta.

(Refer Slide Time: 12:55)



Note that, note, note that theta is this a p this vector ok. So, you take derivative partial derivative with respect to each of this theta you will get this and if we equate that to 0 you get if you equate that to 0 you get this thing you get theta equal to, you get theta equal to this quantity.

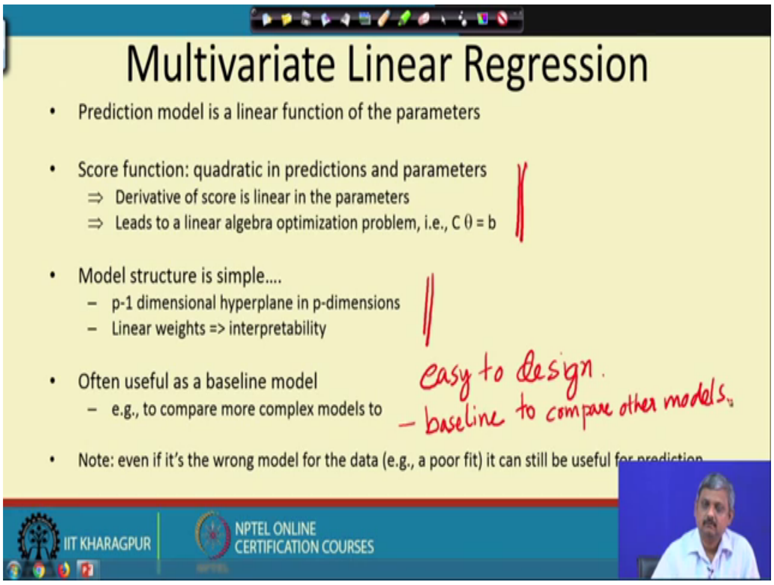
(Refer Slide Time: 13:42)



So, so this is defining this. So, the theta is equal to this quantity. So, basically you note the dimensions again. So, x as I told is 1 x 11, x 12, up to x 1p; 1 x 21, up to x 2p; 1 xn1 up to x2 xnp. So, x is this y is this and this quantity theta is this. So, thetas dimension is p plus 1 cross 1 this dimension is n cross 1 this dimension is n cross p plus 1 ok. So, if you just do this multiplication matrix do the matrix inversion multiply you can figure out that they will be the same if you multiply their dimensions will match you get your parameters of the equation theta ok.

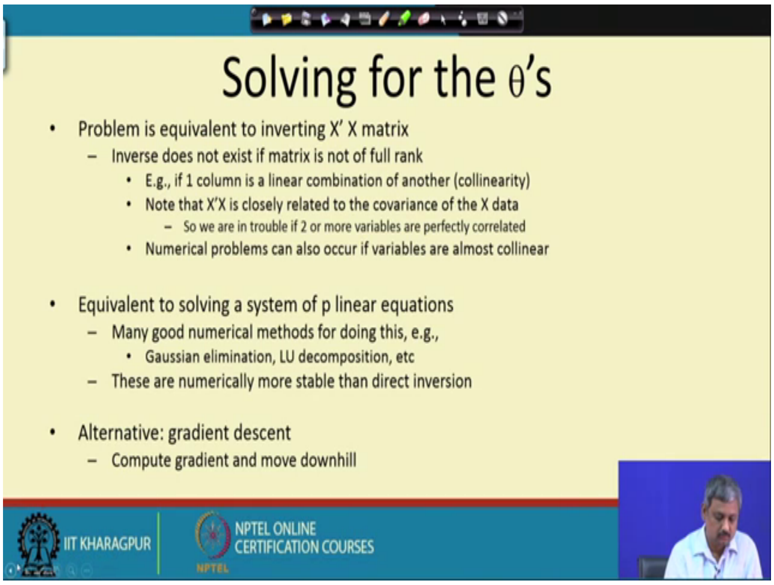
In the in the exercises I will in the assignments I will give you an problem where you have to actually work this out and find out the value of the theta.

(Refer Slide Time: 15:55)



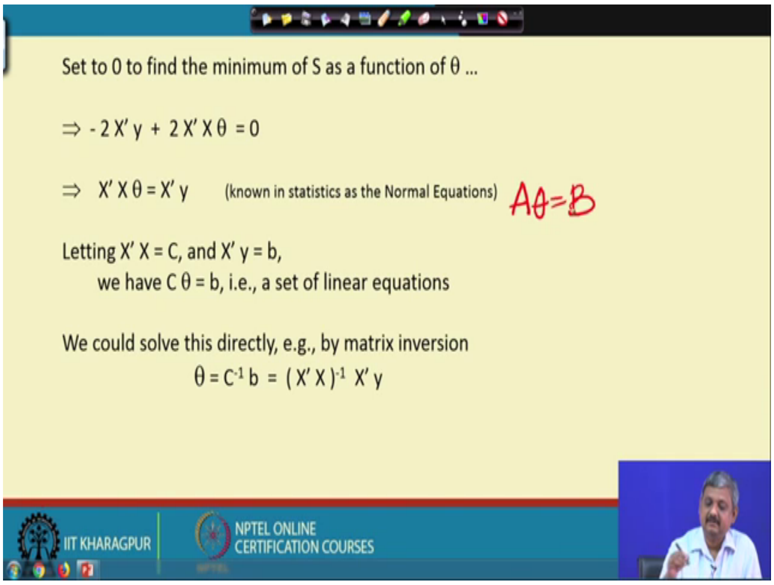
So, people solve it by matrix inversion or by some numerical methods.

(Refer Slide Time: 16:03)



Basically it is like solving a set of linear simultaneous equations of the form if you. So, this equation is like solving something of the form A theta equal to B ok. So, theta is a vector theta one theta (Refer Time: 16:42).

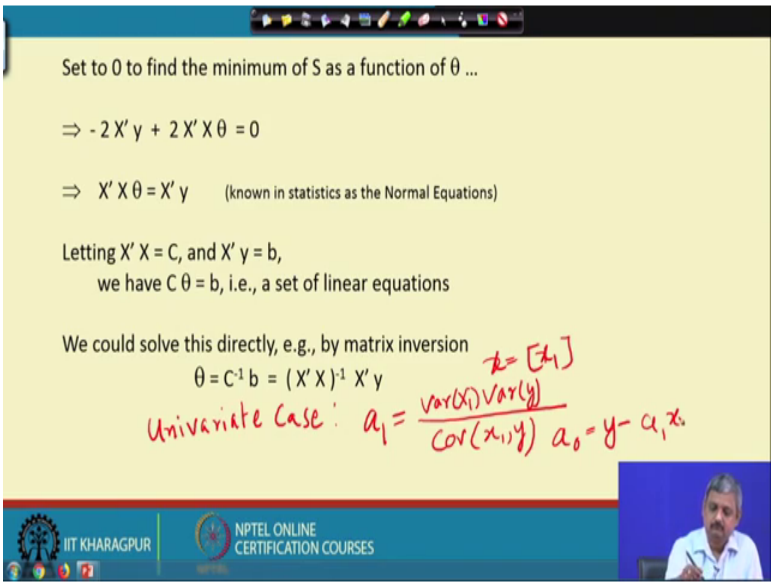
(Refer Slide Time: 16:25)



This is like a linear simultaneous equation. So, that people solve by matrix inversion or someone other method ok.

So, what are the, so you can usually people do it for multivariate there are more than one attributes by the way if you if you go back to this definition if it is for a univariate so x is just 1, then your slope of the line a 1 becomes variance x 1 variance y y slope of the line and a naught you can find by y minus a 1 x.

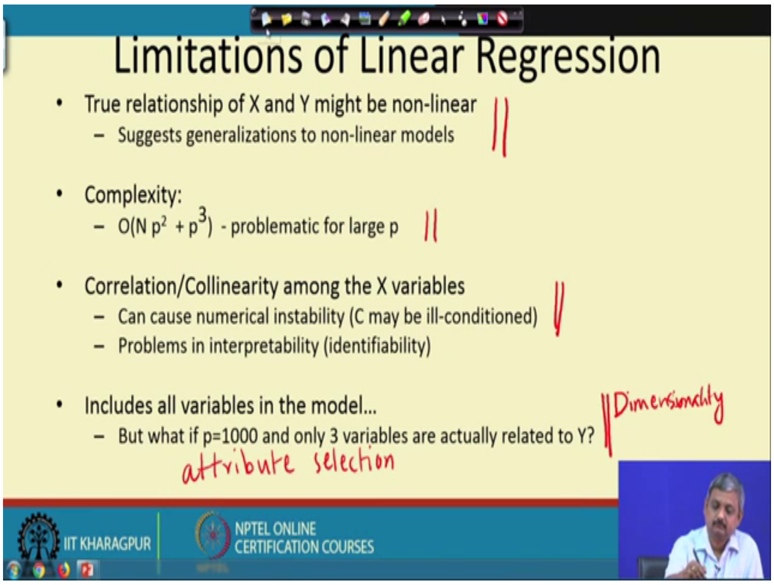
(Refer Slide Time: 17:18)



So, these are some of the advantages. It is linear, it is because it is linear it is simple and interpretable and it is it is easy to design ok. So, often it is used as a base line (Refer Time: 19:10).

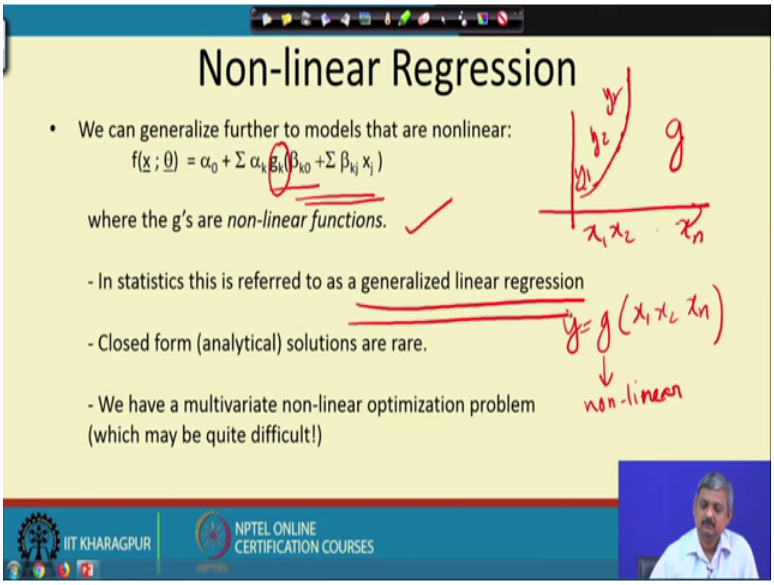
What are some of the limitations?

(Refer Slide Time: 19:30)



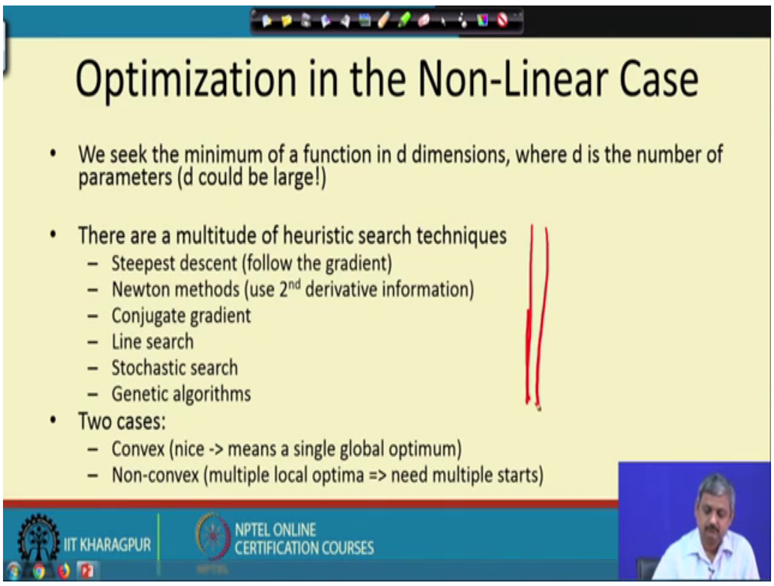
The actual laser might not be linear within y and x. For large dimensional problem linear regression is a problem, if x and if the x 1 x 2, x 2 they are correlated there can be numerical instability and the final problem which we will discuss in our future thing is the problem of dimensionality. And you have to do something called a which attribute to use for equation. For example, if you are detecting whether some person how much loan amount he will repay the name of the person is irrelevant attribute. So, you have to do an attribute selection.

(Refer Slide Time: 20:50)



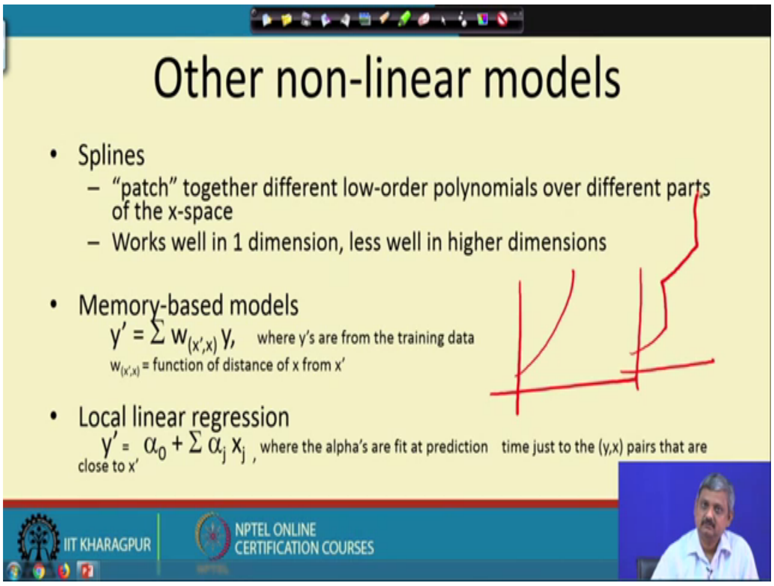
So, people extended this methodology to the to something called a non-linear regression ok. So, earlier we had only this much alpha k x j now you can have some function g on this quantity which is a non-linear function. So, you try to find a y as some g of, where g is in the linear case g was alpha 1 x 1 a 0 x 1 a 1 a 0 plus a 1 x 1 a 2 x 2 and so on; so even though that when the actual data is non-linear, you would like to go for this generalized linear regression. Problem is the solutions of finding the optimal parameters of g is not so simple as in linear equation ok.

(Refer Slide Time: 22:35)



So, if you want to find the best parameters of g which will fit the data you have to use more complex methods like some of these techniques, some of these techniques.

(Refer Slide Time: 23:03)



There are some other non-linear methods which instead of a single function fit piecewise linear functions ok. So, like local linear or splines they do that. The methods for finding them are even more complex, all right.

So, this is in general the methods of regression. There are some more extensions I will discuss in the next lecture, but it is a very powerful technique.

Thank you.