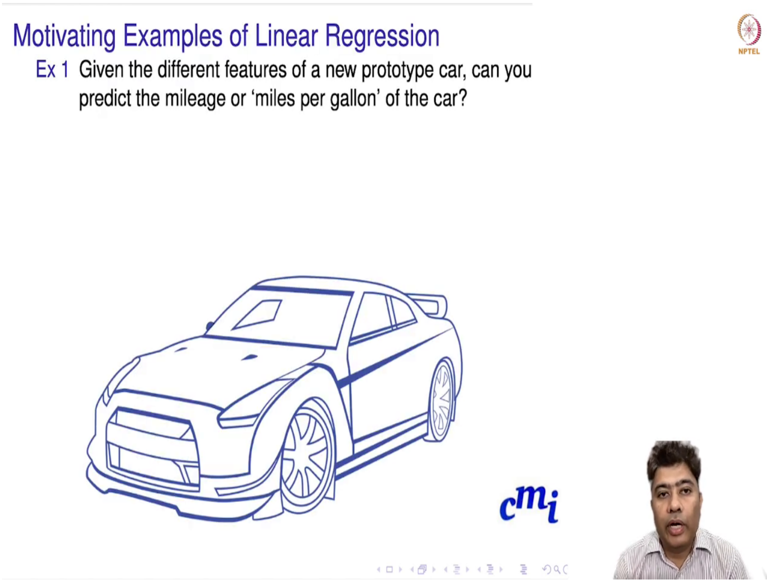
Predictive Analytics - Regression and Classification
Prof. Sourish Das
Department of Mathematics
Chennai Mathematical Institute

Lecture - 02
Least Squares method

(Refer Slide Time: 00:16)

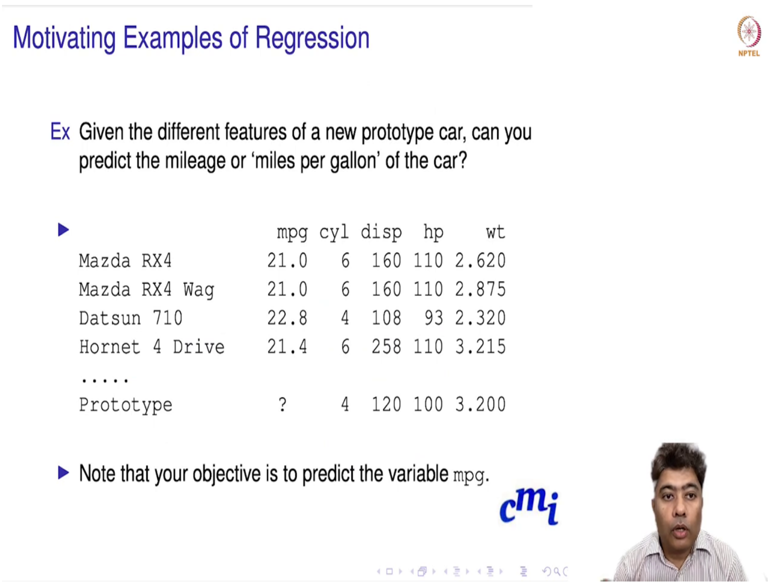
Welcome back.

(Refer Slide Time: 00:17)



So, we were continuing on regression examples on empty cars data sets. So, as we know that given a different features of new prototype car, can you predict the mileage or the miles per gallon of the car?

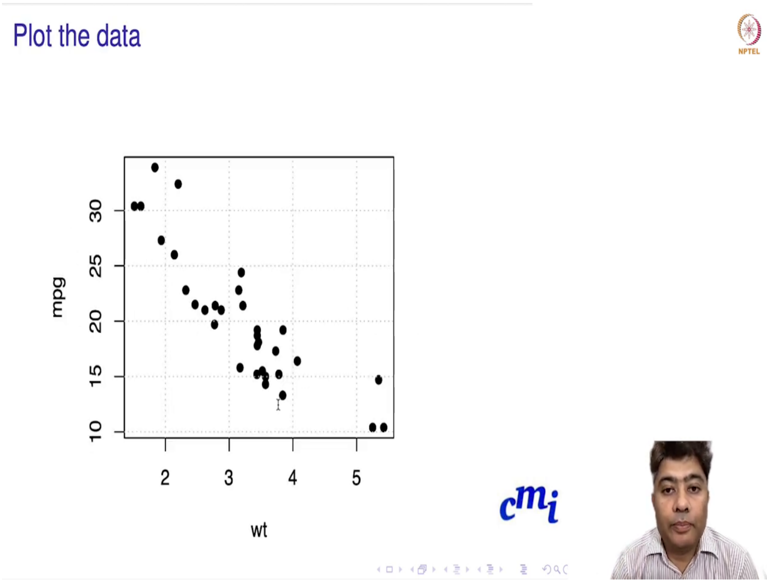
(Refer Slide Time: 00:31)



Now, the data that you have in your hand is some sort of you know like this kind of typical table format that you have and say maybe Mazda RX4 which gives you 20 miles per gallon, 21 miles per gallon. It is a 6 cylinder car, it has a displacement of 160, it gives a horsepower of 110 and weight is 2.62. So, there could be another car Hornet 4 Drive which is has a miles per gallon 21.4 miles per gallon and 6 cylinder car, displacement is about 258 and horsepower 110, weight is 3.215.

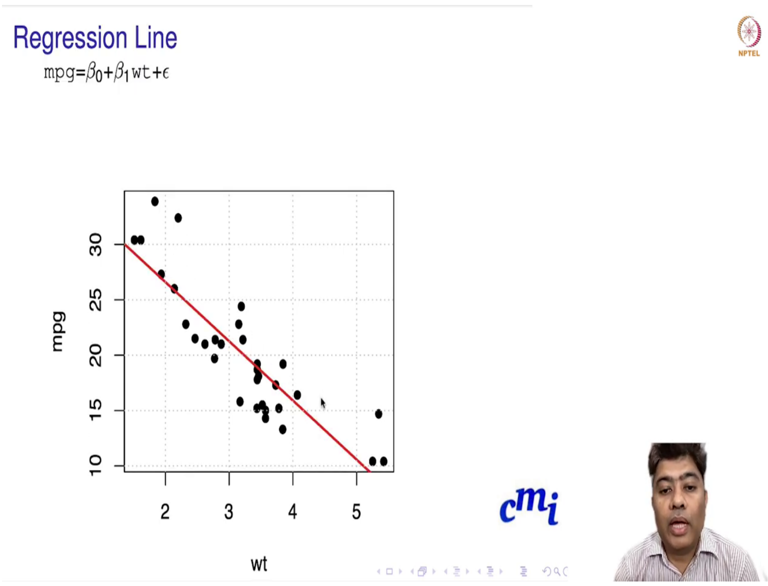
Now, the prototype car that you have built maybe in it is in only in available in your you know computer models has a 4 cylinder car, displacement is 120, horsepower is 100 and weight is 3.2. Now, you want to estimate what is the miles per gallon. So, you want to predict or estimate the variable miles per gallon.

(Refer Slide Time: 01:50)



So, like any data science and project what we will do, we will start with some visualization. So, we take this weight; suppose you take this weight variable and the miles per gallon and we plotted them. So, what we are seeing that as the weight of the car increases, the miles per gallon decreases. So, we see a negative relationship between the weight and the miles per gallon.

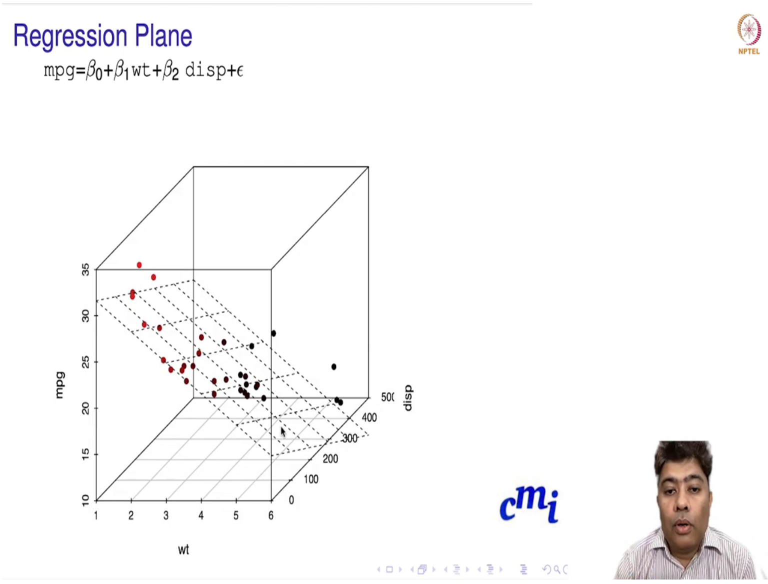
(Refer Slide Time: 02:23)



So, naturally we would like to fit a straight line to begin with miles per gallon as a function of beta naught plus beta 1 times weight plus some error. So that means, there will be some error because you can see that this there is a some sort of randomness in the system and you know. So, like this point we would like to have it on the straight line, but that is this not exactly there. So, there will be this difference is the error.

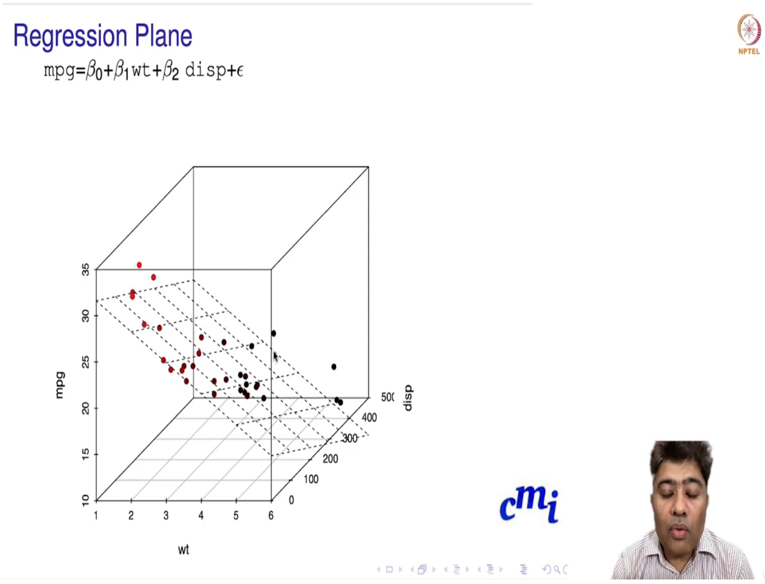
So, for this point, it is the error. Interestingly, this point is almost on the line so, error for this point will be almost 0. So, there will be some point which will be on the line or very close to the line, for them the error will be 0. And, then there will be some point here, some point here, some point here for them error will be some error will be there.

(Refer Slide Time: 03:25)



Now, if we consider a third variable say displacement in our analysis. So, the model will be beta naught plus beta 1 times weight plus beta 2 coefficient time displacement plus error. Now, two things you must notice. First, in the previous case we had only two variables weight and miles per gallon. So, the all the points were on the two-dimension.

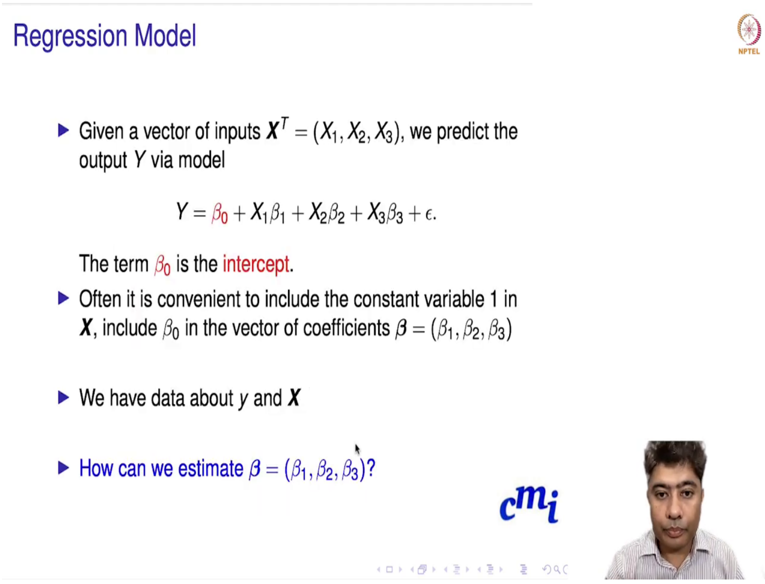
(Refer Slide Time: 03:59)



Now, we are bringing third variable displacement in this axis displacement, this is a weight and this axis is miles per gallon. In this so, naturally all the points that we are seeing there in the three-dimension. And so, its likely you can if you can imagine yourself you are in a room in one length is weight you know x axis is weight, y axis is displacement and the z axis is miles per gallon.

Then all these points are somewhere kind of you know hanging in the three-dimension space. So, this is what you are seeing in the graph. Now, this model is trying to fit a plane through this points hanging in the 3D space. So, previously it was a line, in two-dimensions prop space it was a the model was a line; in three-dimension space, it is a plane ok.

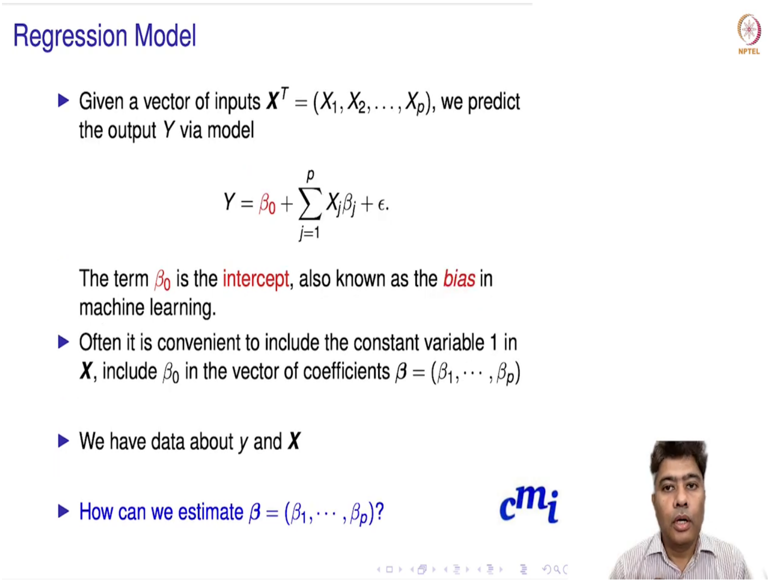
(Refer Slide Time: 05:14)



So, given a vector of inputs X 1, X 2, X 3; now we are kind of putting some kind of abstraction, we predict the output Y. So, Y equal to beta naught plus X 1 beta 1 plus X 2 beta 2 plus X 3 beta 3 plus epsilon. The term beta naught is the intercept. Often, it is convenient to include a constant variable in the X matrices, that includes the beta naught in the vector of coefficient beta which is only the beta 1, beta 2, beta 3. So, in these things.

So, we have data about now y and x and we have a model. What we do not know? The value of these coefficients. So, we want to estimate this parameters beta 1, beta 2, beta 3.

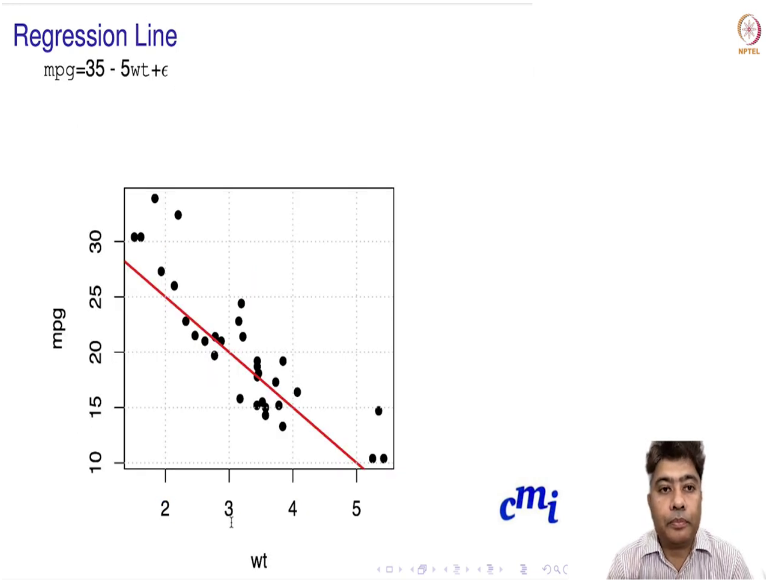
(Refer Slide Time: 06:09)



Now, we are going to expand this model to the p many features. It is just you had a model and three-dimension model. Now, you have a p dimension model. So, in three-dimension in two-dimension you are fitting a straight line, in three-dimension you were trying to fit a sort of a plane, where in the p dimension more than three-dimension we cannot visualize. So, what we have to do?

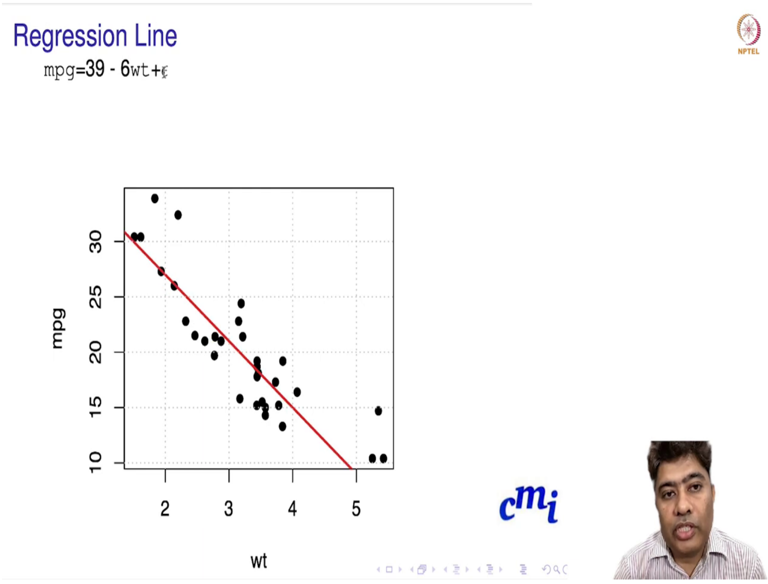
We can just you know imagine that in a p dimension, it will fit a p minus 1 dimension hyper plane. So, that is what exactly we are going to do. So, why is this model is going to fit a p minus 1 dimensional hyper plane in a p dimensional data space. So, this is the exact same problem. We have bunch of X's, we have bunch of and a Y, the given X and Y's we want to estimate the betas. So, that is where the whole problem lies.

(Refer Slide Time: 07:20)



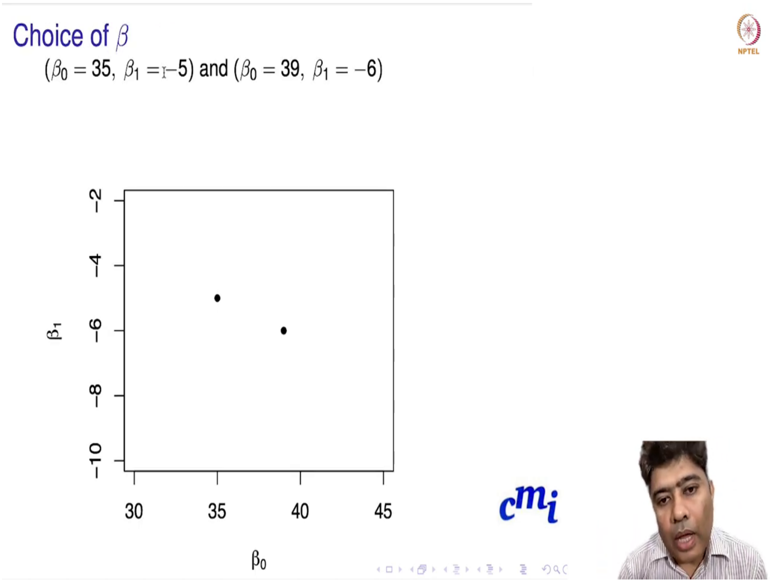
Now, in when it comes to estimation of beta, let us try to understand how the values of beta is going to affect my model. So, let us take this line beta m naught as 35 and beta 1 is minus 5. So, my model is mpg 35 minus 5 times weight plus epsilon.

(Refer Slide Time: 07:49)



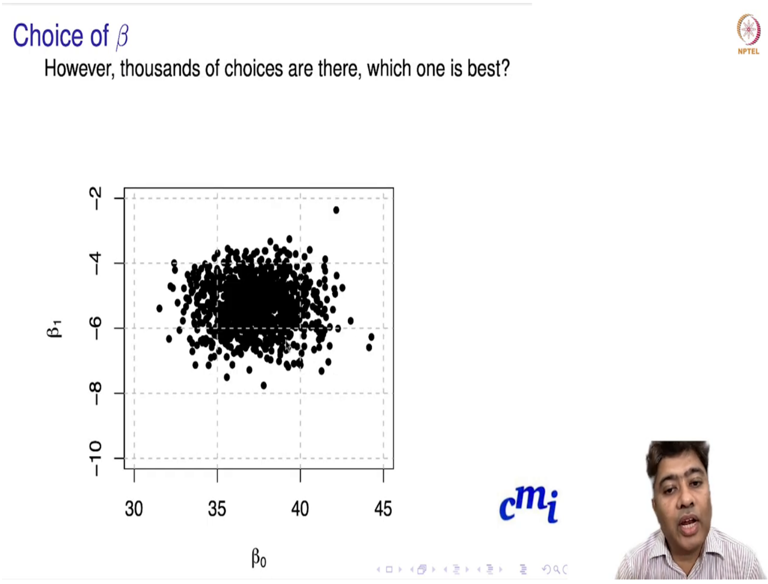
Now, if I change it to 39 and minus 6, now the model is also changing. You can see. So, there are only two data. One is two values that I am considering for beta naught and beta 1; 35 and minus 5 and 39 and minus 6. And, you can see the model is kind of you know changing accordingly.

(Refer Slide Time: 08:13)



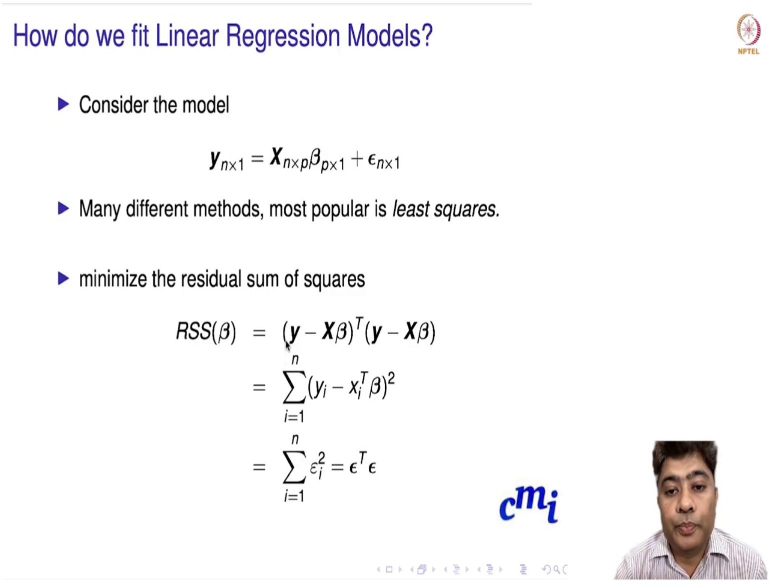
Now, what are the choices of beta that can I have? So, beta can beta naught can take values around the say x axis and suppose beta 1 takes the values around this axis. So, one possibility is beta naught is taking 35 and minus 5. So, which is this value, another possibility is beta naught is taking 39 and minus 6. So, this is varies this value and this gives us two possible line. So, the question is which one to choose? But, one other question is why should I choose between these two only?

(Refer Slide Time: 08:54)



There could be infinitely many possible choice of line and for each line I will get a different different each choices of beta, I will get different choices of line. So, which beta to choose? Each choice of beta will give me a line.

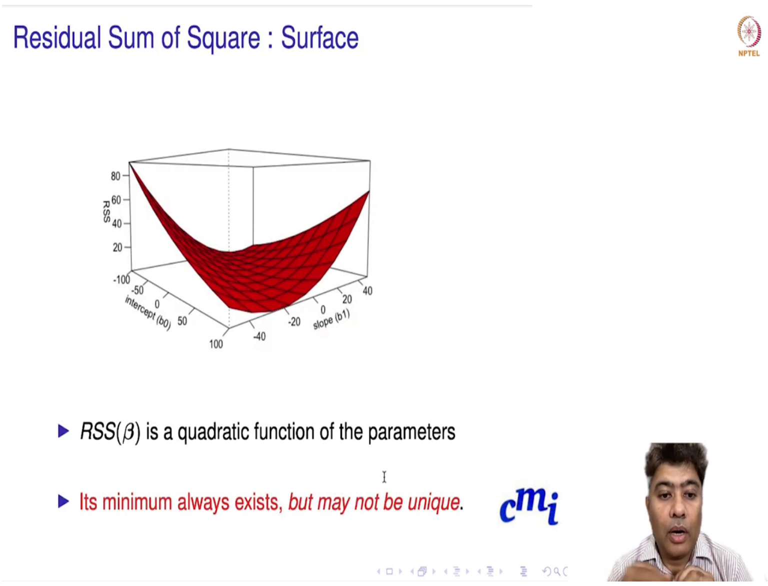
(Refer Slide Time: 09:12)



Now, question is which one to choose? So, there comes an interesting concept called residual sum of squares of error and from where the concept of least square comes. So, we have the model y equal to X beta plus epsilon, y is a n cross 1 vector, X is a n cross p matrix, beta is a p cross 1 vector and epsilon is a n cross 1 vector. Now, what we are going to do? We are going to understand the residual sum of squares of error.

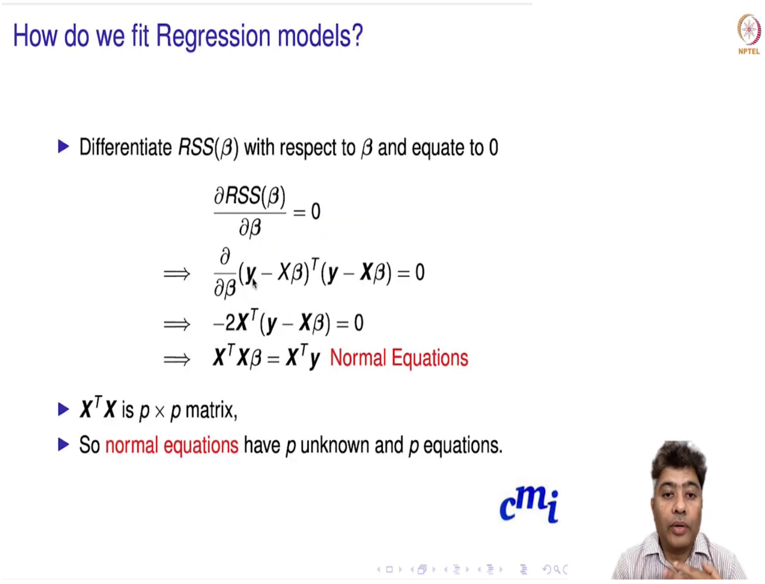
So, y minus X beta. So, we can just take the X beta on the left side, that gives us a y minus X beta which is epsilon effectively, transpose y minus X beta effectively what is it? This is y minus x i transpose beta whole square and some of that. Now this is nothing, but epsilon i square or epsilon transpose epsilon. So, this is called residual sum of squares of error.

(Refer Slide Time: 10:20)



Now, if I plot this residual sum of squares against different choices of beta naught and beta 1, we get a valley like this kind of a valley. And of course, we would like to choose a beta naught and beta 1 somewhere in the valley which is the; which will give me the minimum sum of squares of error. And, turns out it is the it its minimum always exist and it may not be unique, but minimum always exist. And we will talk about it, but this we are going to talk about it; but this is let us talk about more about it.

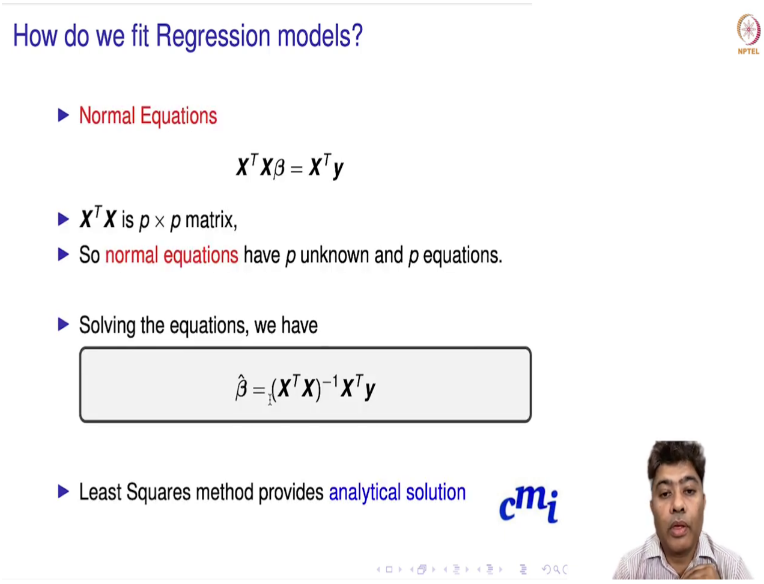
(Refer Slide Time: 11:07)



So, how can we get this minimum? So, easiest thing is you differentiate this residual sum of square with respect to beta and set it equal to 0. So, what is my residual sum of squares of beta? y minus X beta transpose times y minus X beta and differentiate it with respect to beta and set it equal to 0.

So, after differentiating what I have is minus 2X transpose times y minus X beta equals to 0. Now, this we can write it as X transpose X beta equal to X transpose y. This equation, set of equation is called normal equations and X transpose X is a p cross p matrix. So, the normal equations have p unknown and p equations.

(Refer Slide Time: 12:02)



Now, solving this equation, we can have beta hat equal to X transpose X inverse X transpose y. And, this solution is called Ordinary Least Square solution on OLS or least square method provides an analytical solution. We have a exact analytical solution. Because, you see that all we have to do we have the data X, in the data all we have a X and y which is plug in X and y in this formula. This will give you beta hat and in that is it your model is ready.

You can plug it in this model, you can deploy it in a production setup. So, we will stop here now and we will do some hands on to see how these things work.

Thank you.