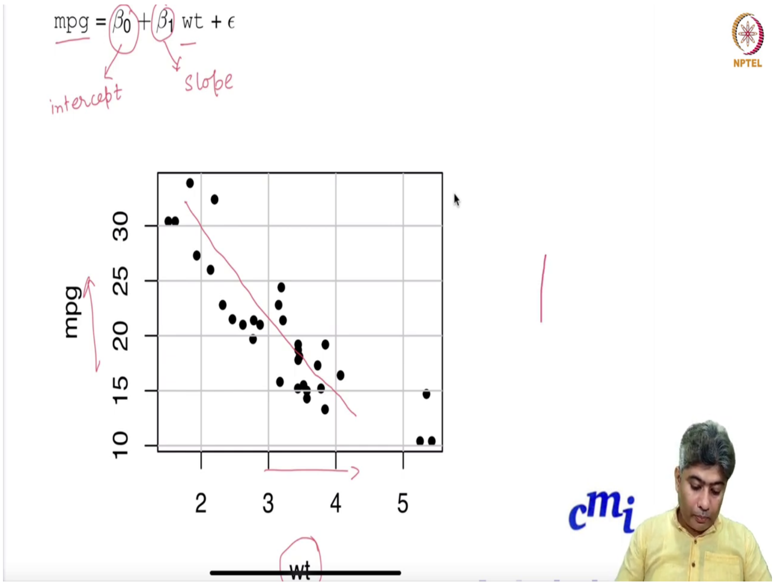
Predictive Analytics - Regression and Classification
Prof. Sourish Das
Department of Mathematics
Chennai Mathematical Institute

Lecture - 14
Geometry of Regression Model and Feature Engineering

Hello all. Welcome back to Regression and Classification course. This is lecture 4.

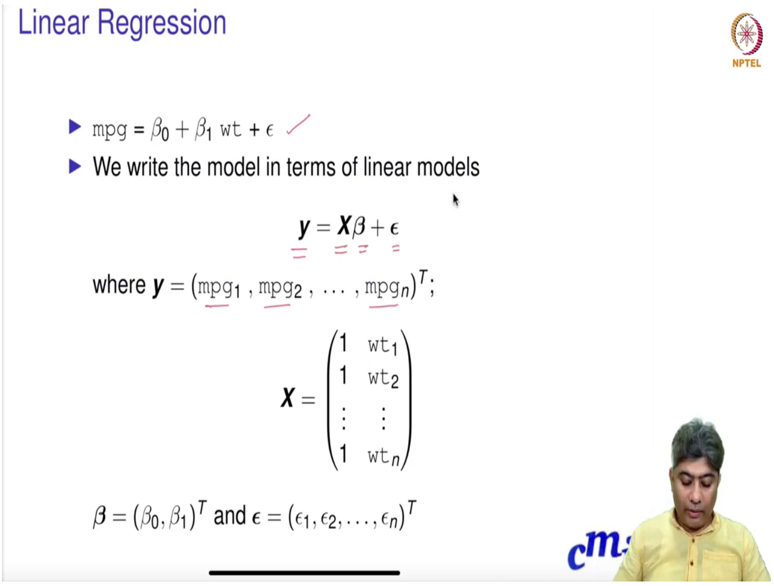
(Refer Slide Time: 00:23)



We will start with a basic simple models with mtcars data set. So, in this mtcars data set, our target variable is miles per gallon and we are trying to fit essentially the weight on the x axis and miles per gallon on the y axis. So, the model that we are fitting is weight m miles per gallon as a function of weight.

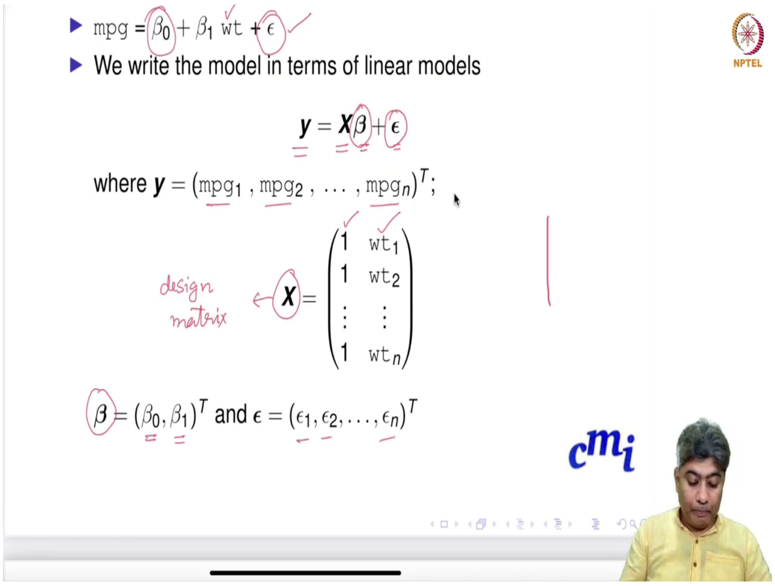
So, miles per gallon equal to beta naught plus beta 1 times weight. So, beta naught will be the interceptor, it will be the intercept term and beta 1 is the slope. So, we have some kind of line that is what we will try to fit in this model with this model.

(Refer Slide Time: 01:34)



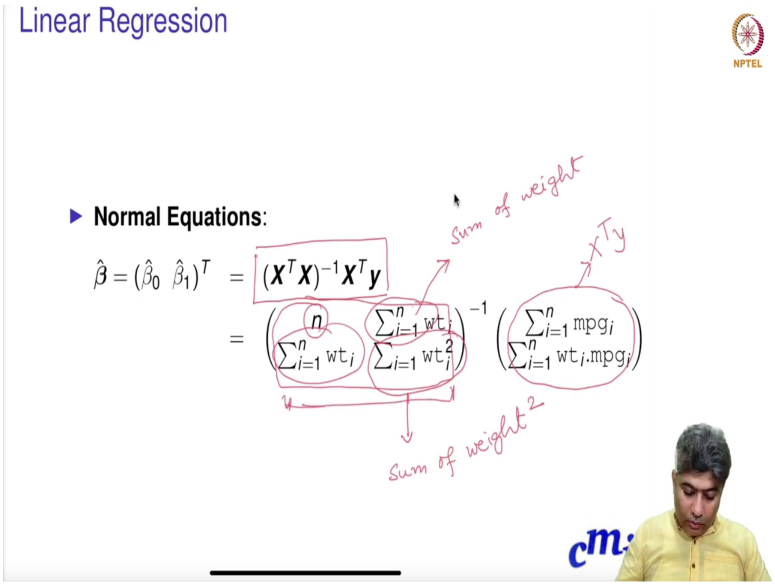
Now, this model as we have seen in the previous example, previous lectures that can be define these models can be defined using the matrix notation y equal to X beta plus epsilon, where my y's is a vector of you know n dimension; miles per gallon 1, miles per gallon 2, miles per gallon n. So, we have n samples. In each samples, we have about you know like miles per gallon. Now, what we have is X, X is has 2 column; the design matrix. X is the typically called design matrix design matrix.

(Refer Slide Time: 02:30)



And, the first column all 1s, this first column with all 1s represents the beta naught and the second column is weight, that is the weight of the independent weight. And, then the beta here have two parameter. We want to estimate beta naught and beta 1 and the epsilon here the my epsilon is has n elements; epsilon 1, epsilon 2, epsilon n ok.

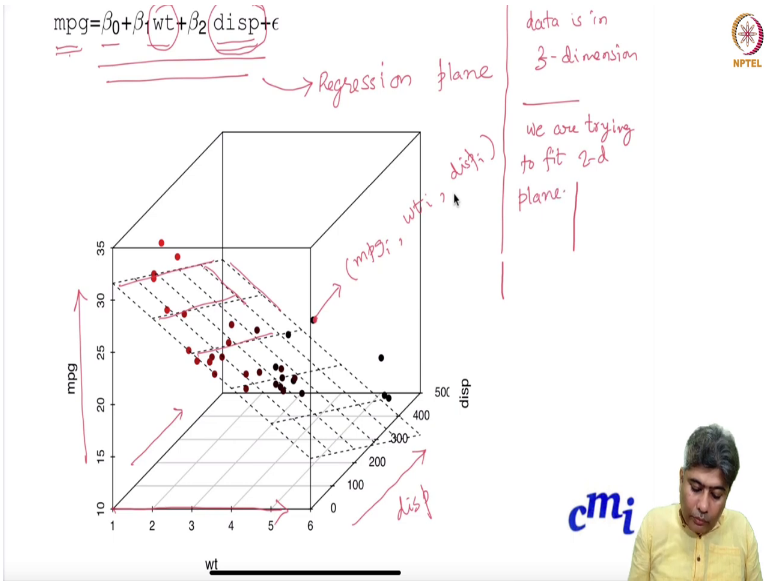
(Refer Slide Time: 03:19)



Next, what we will we can do? We can we have seen that in normal equations will give you this is the analytical solution X transpose X inverse X transpose y. Now, in the X is we have defined this is our X, this is our X. So, all you have to do just write X transpose X and if you do X transpose X, that will give you this may break the first.

First element is n, second element is sum of the weights and the this is sum of weights and this is sum of weight square. And, this is same as this one because it is a because this is essentially symmetric matrix; X transpose X is going to be your symmetric. Now, this is your X, this is our X transpose y, this is our X transpose y and this is our X transpose X inverse ok.

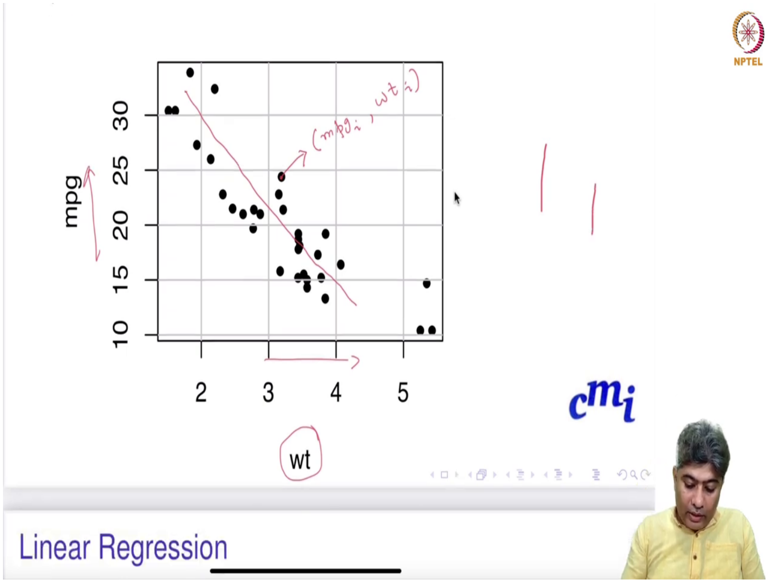
(Refer Slide Time: 04:49)



Now, next what we are going to do? We already had beta naught plus beta 1 times weight, now we are going to add a third a second independent variable; not third, second independent variable. So, this is miles per gallon is our target variable or dependent variable. First independent variable that we considered in the model is weight and the second independent variable we are going to consider is the displacement.

Now, if you do that, essentially the geometry will look somewhat like this. So, previously we were essential considering weight on the x axis and miles per gallon on the y axis, now we have added a z axis here. We have added a z axis. The z axis is displacement. A new variable is being brought. So, the points which were previously on the 2-dimension, you see these are the points.

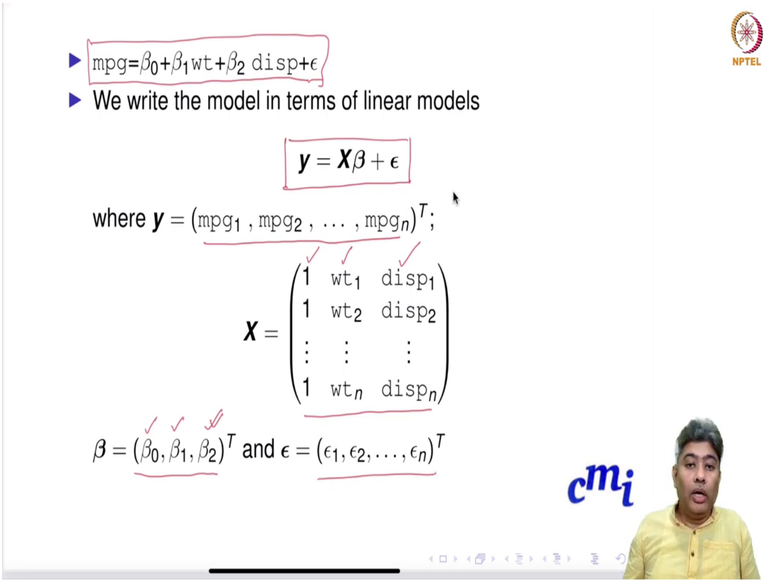
(Refer Slide Time: 06:10)



Each of the say, if you take these points, these points will be typically suppose this is the ith points; it will be like mpg i comma wt i. It is a tuple of you know 2-dimension. Now, what we have is essentially a third variable and the if this is the point that we were talking about, this point is mpg ith value wt i and displacement is the third value; displacement of the ith sample.

Now that means, this is on the third dimension and hence the entire geometry is in the third dimension. In that third dimension, we are trying to fit this model, we are trying to fit this model. This means this model is actually fitting a regression plane, regression plane or simply 2d plane. So, in a three the data set is in a 3-dimension and we are trying to fit a 2d plane. So, data is in two 3-dimension and we are trying to fit we are trying to fit 2d plane ok.

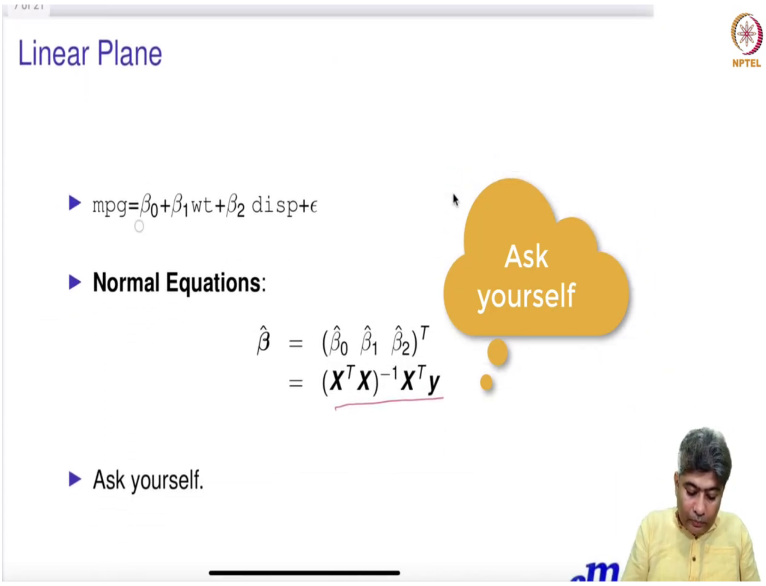
(Refer Slide Time: 08:10)



Next, we how can we write it in matrix notation? So, we can take this model and we can write it in matrix notation, that if we can write it in matrix notation that will clarifies our concept more clearly. So, we can write this one using y equal to X beta plus epsilon, y is as usual all the miles per gallon vector of dimension n.

And, now this is what we have the design matrix, previously we had the intercept column and width and the third column in my design matrix is displacement. In the intercept the regression coefficient has now a third parameter beta 2. So, previously I just had beta naught and beta 1, now we have beta 2. The residual factor is as usual epsilon 1, epsilon 2, epsilon n.

(Refer Slide Time: 09:17)

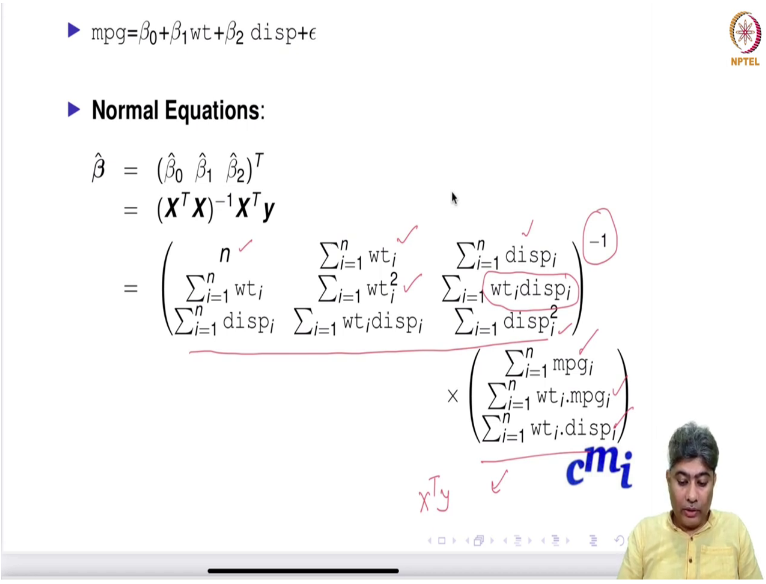


Now, if we solve this as a normal equation so, it will be X transpose X inverse X transpose y. So, here is ask yourself question; how the analytical solution of X transpose X inverse X transpose y would like? So, if you remember we for 2-dimension for the previous problem, we solve this, we solve this and this was our solution right, this was our solution. This was our solution. Can you develop the solution for the 2-dimension case? Can you develop the solution for this case, this model? So, this is my question, this is my question.

(Refer Slide Time: 10:24)

So, what I am going to do? I am going to request you to stop yourself and for about 2, 5 minutes pause your video, think about it how you can solve this ok. Now, let us come back, I hope you have tried and you have the solution.

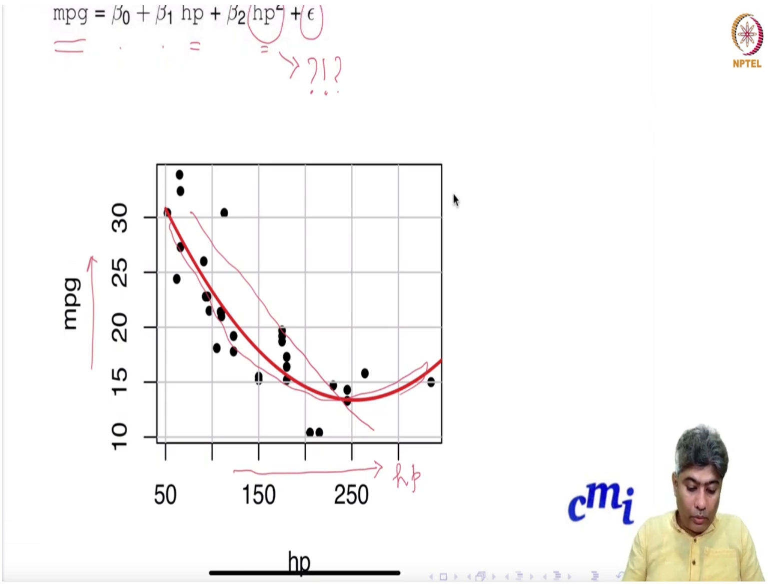
(Refer Slide Time: 10:56)



So, let us see how the solution would like look like. It is slightly cumbersome, but we can we can figure it out. So, we can see this first will be n, second will be weight, summation of weight, third will be the summation of displacement. Then, this will be again the weight square and displacement square and this will be cross product of weight and displacement.

So, this matrix is the X transpose, this matrix is X transpose X. You have to take the inverse of that and this matrix is your X transpose y. So, the first element is simply sum of weights, sum of miles per gallon. The next is weight times miles per gallon and the third one is weight times displacement. I hope your answer was correct.

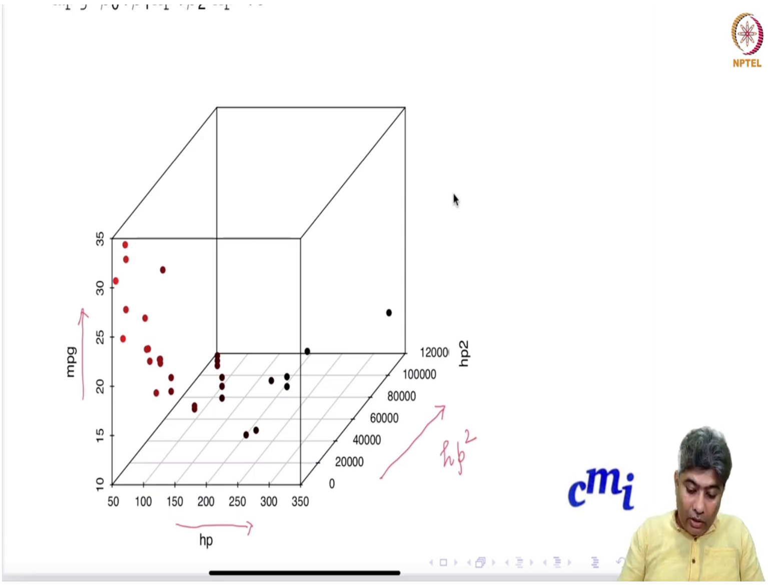
(Refer Slide Time: 11:58)



Next, we will move on to a new topic called quadratic regression. In quadratic regression what we will consider now in the same data set miles per mtcars data set, we have miles per gallon and horsepower. In the miles per gallon and horsepower what we are doing? We are plotting the horsepower on the x axis and miles per gallon on the y axis.

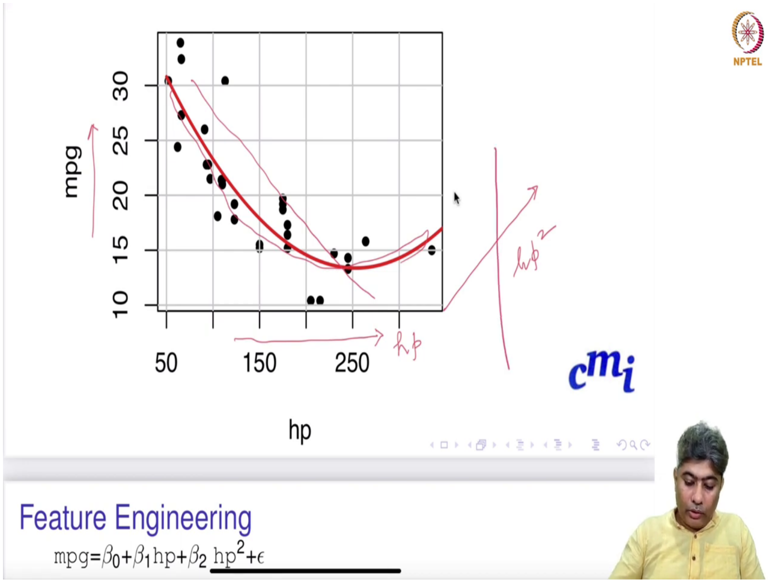
We could have easily fit a straight line here, but looks like one can try also the quadratic regression. Now, how this quadratic regression can be fit? The simple way of doing it miles per gallon as a function of beta naught plus beta 1 horsepower plus beta 2 horsepower square plus epsilon. Now, this horsepower square is an very interesting term. We will let us try to understand what are we doing adding horsepower square.

(Refer Slide Time: 13:20)



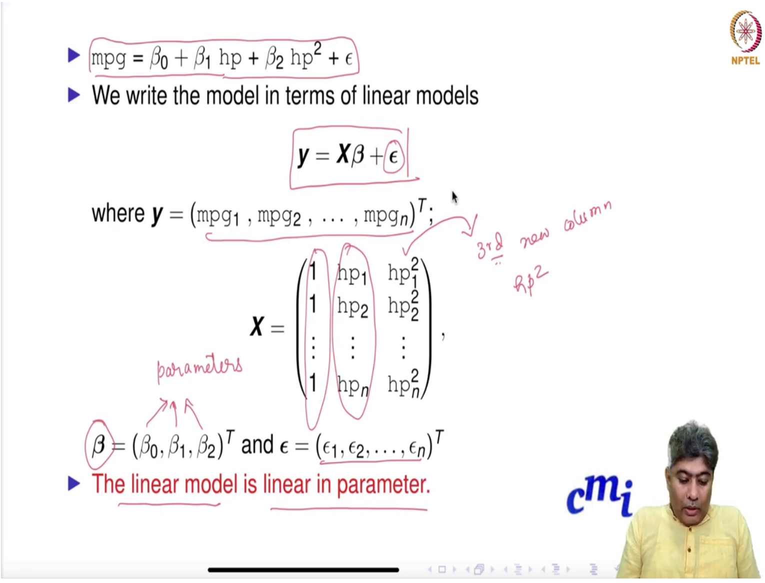
So, essentially what we are doing? This was our original 2-dimension plot horsepower on the x axis, miles per gallon on the y axis. Then, we are putting horsepower square on the z axis, this is our horsepower square. So, we are creating a third dimension. So, this is our third dimension. So, our original data was on the 2-dimension, if this is the original data is in the 2-dimension right. Now, what we are doing? We are putting, we are creating a third-dimension horsepower square and by doing that we are putting the 2-dimensional data into third dimension.

(Refer Slide Time: 14:04)



We are putting it into third dimension.

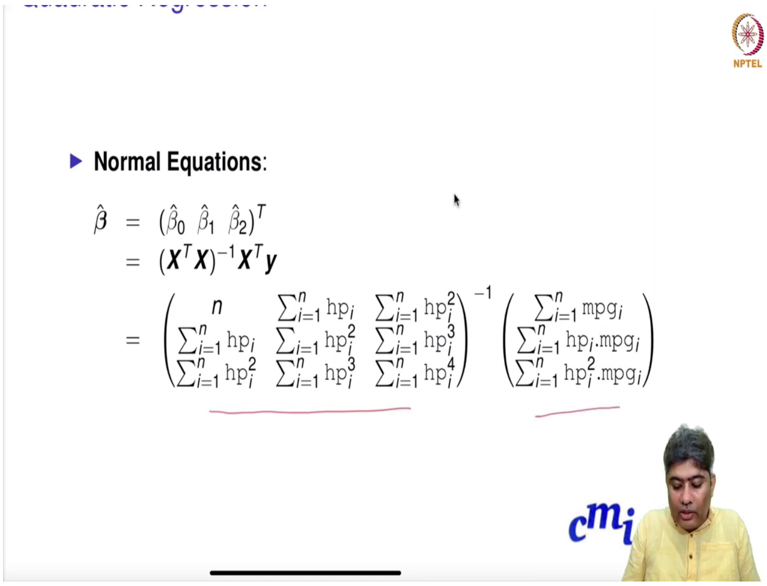
(Refer Slide Time: 15:22)



Now, how can we model it? Simply, what we are doing here so, this is my model, this is our model right; what we and we can write it in y equal to X beta plus epsilon or my y is exactly same. So, the first is intercept column, second is the horsepower and the third column we are creating a new column, third new column called horsepower square. We are engineering this new column ok and rest of the thing is as usual previously.

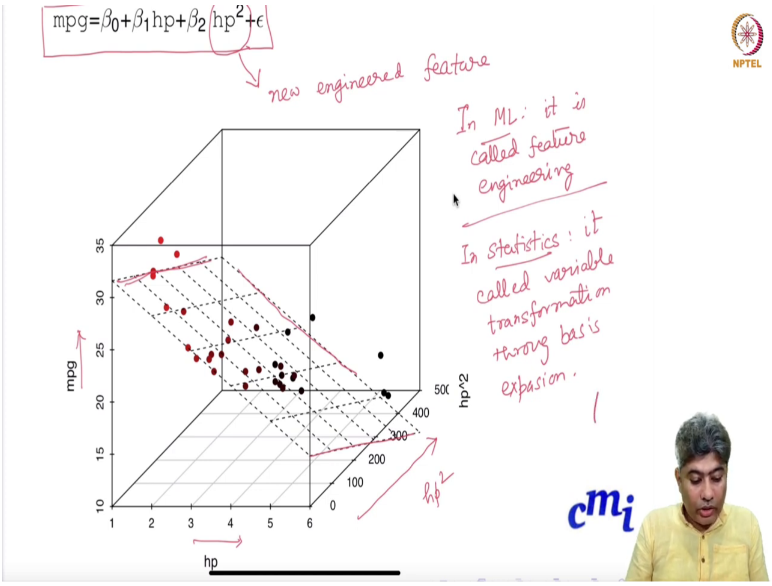
So, beta naught, beta is my regression coefficient which has three parameters, three parameters and epsilon, bunch of epsilons that we have here. These are n dimensional epsilon. So, the linear model is actually linear in parameter. When we call a linear model, it is not necessarily it is linear in predictor or linear in variable. When we call it as linear model by saying linear model, we mean it is linear in parameter.

(Refer Slide Time: 15:56)



So, how can we solve this quadratic regression? Same, if way X transpose X inverse X transpose y and you can solve this by yourself. This is your X transpose X, you have to take the inverse of that and this is your X transpose y. And, then if you solve these two multiply these two matrix, what you will get? Your beta hat.

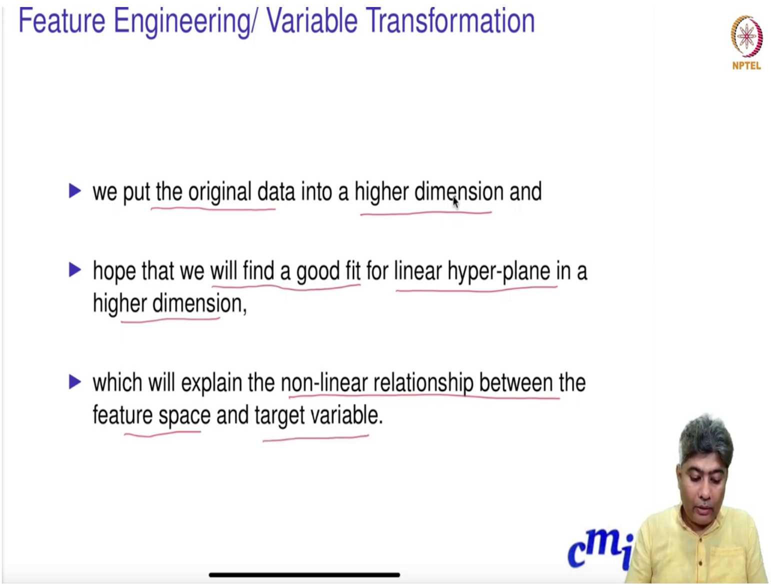
(Refer Slide Time: 16:23)



So, and if you do that you will fit a plane. So, what are what is it actually what we are doing? So, essentially my data is in the 2-dimension, in the 2-dimension in the x axis and y axis. We are engineering a new axis called horsepower square and then what we are doing is essentially fitting a hyper plane in the 3-dimension. This model is a 2d plane in 3-dimension, but quadratic in 2-dimension and this is. How are we achieving this? By engineering this new feature.

This is called this is new engineered feature engineered feature. In machine learning, it is called feature engineering, in machine learning it is called feature engineering feature engineering. In statistics, in statistics it is same thing it is called variable transformation, variable transformation through basis expansion basis expansion. And, we will talk about it the both thing, but essentially we are creating a new variable and that is why it is called variable transformation or feature engineering.

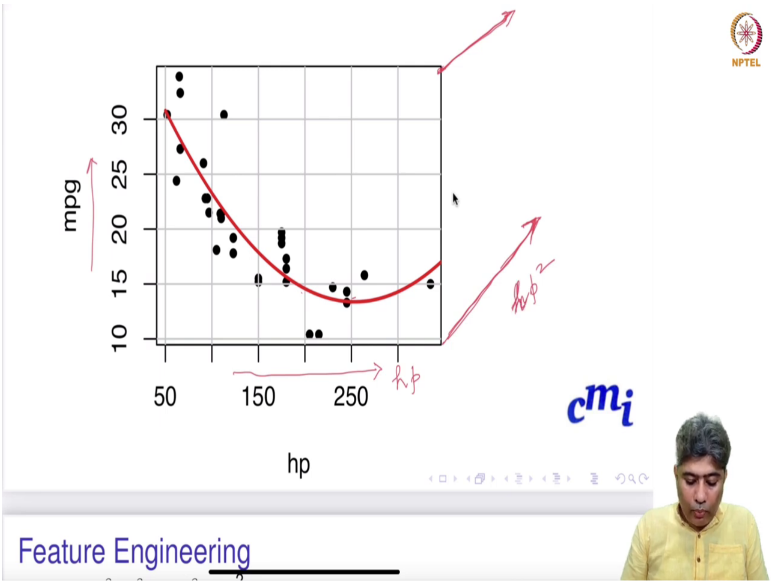
(Refer Slide Time: 18:34)



So, weight feature engineering or variable transformation what we are doing? We put the original data into higher dimension. So, this is the first trick. The trick is we put the original data into higher dimension and you hope that we will find that a good fit for a linear hyper plane in a higher dimension and which will explain the non-linear relationship between the feature space and the target variable.

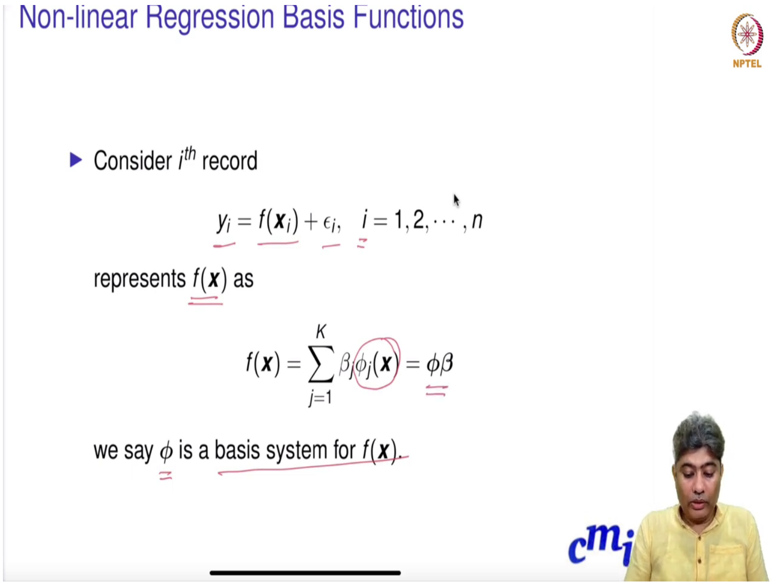
So, if I go back, if I go back in this example. So, clearly in this example the relationship between horsepower and the miles per gallon is not really a not really a you know straight line, it is not linear. It looks like they have some kind of a non-linear relationship. And, if they have a non-linear relationship, then it is better we should use a we should use the feature engineering technique and what we are doing we are creating a new dimension.

(Refer Slide Time: 20:06)



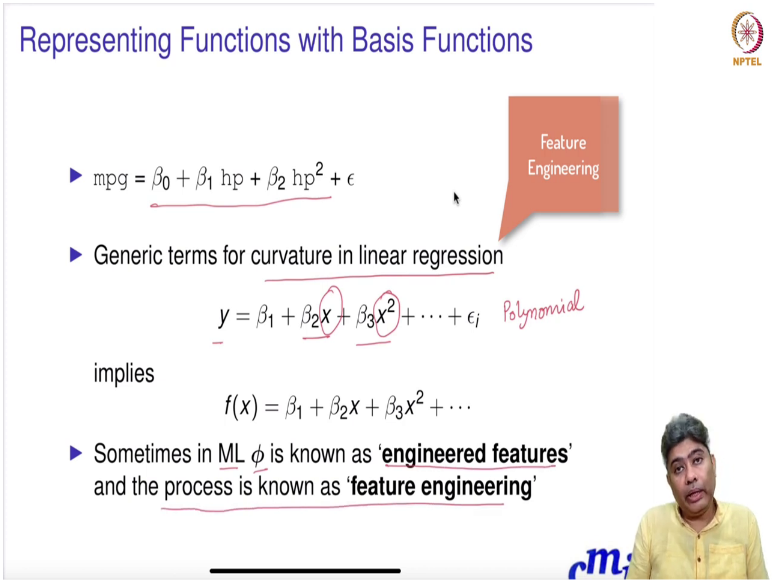
A new dimension by creating featuring with the feature engineering, we put we are putting the 2-dimensional data into high dimension into 3-dimension. And in that 3-dimension we are trying to fit a hyper plane in this case 2d plane. But, fitted 2d plane, this fitted 2d plane is essentially a quadratic regression for this is the; this is the model. Remember, that this 2d plane, this quadratic regression is actually a 2d plane in in 3-dimension space.

(Refer Slide Time: 20:51)



Now, we are generalizing this concept, what we I just spoke to you regarding the quadratic regression that y i f x i plus epsilon i. So, there is a some relationship between y i and x i, i is the ith record of the data set and we suppose we can write this f of x as summation beta j phi of x j. So, phi is basis system of the f of x so; that means, we can write it as phi of beta. So, this is how we can typically write.

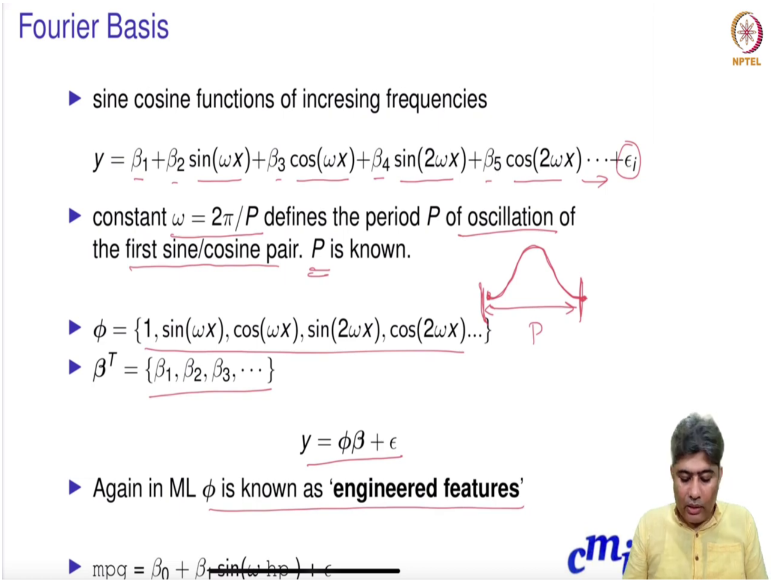
(Refer Slide Time: 21:32)



So, what is the example? The example is if we take this simple quadratic regression type thing. So, we can generate a generic terms of curvature in linear regression. So, we can just say y equal to beta naught plus beta 2 x plus say beta 3 x square plus beta 4 x cube. So, basically higher order polynomial. So, this is a polynomial right.

So, you can fit a polynomial and this x, x square, x cube these are the basis of the polynomial. In machine learning, this phi or the basis is known as engineered feature and the process is known as feature engineering.

(Refer Slide Time: 22:27)

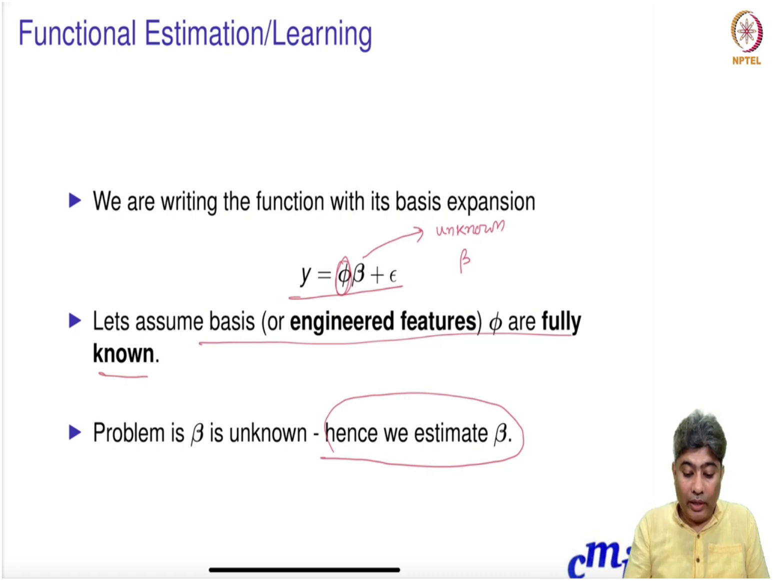


Now, if you have a function which is very sinusoidal in nature, you can define a function like beta naught or beta 1 plus beta 2 sin omega x plus beta 3 times cos omega x plus beta 4 times sin 2 omega x plus beta 5 times cos 2 omega x. And, you can go on as you want as many Fourier as you want, you can consider plus a epsilon.

Now, consider omega 2 pi by P, its defines the period, its defines the period. Suppose, this is the; this is where a period starts and this is where a period stops. So, this period is this is this period is your our P. So, defines the period P of oscillation of the first sine cosine pair and P is known. We are assuming P is known, otherwise this unknown P will create some problem.

For the time being, we are just assuming the P is known or for many examples from any real-life examples P is indeed known. So, our basis system is phi and beta is beta 1, beta 2 beta 3 and you can effectively write y equal to phi beta plus epsilon and this phi is as discussed is known as the engineered feature.

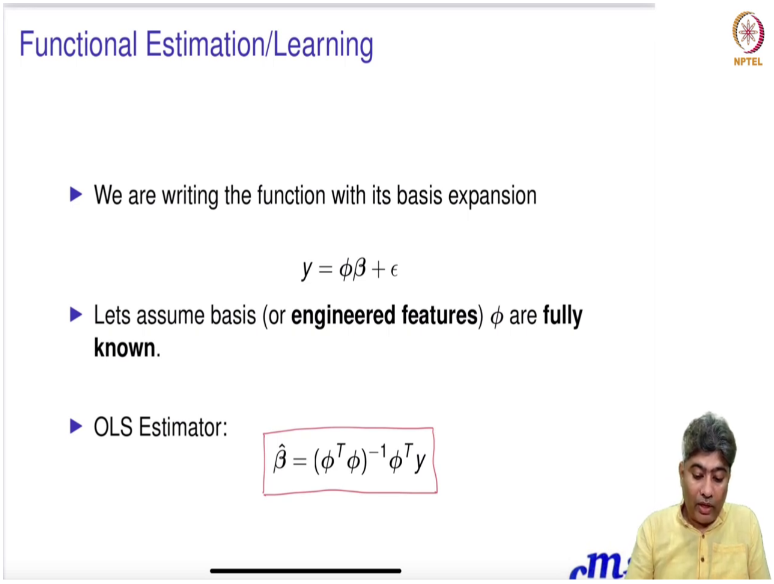
(Refer Slide Time: 23:59)



Now, you can define a model, no harm; mpg equal to beta naught plus beta 1 sin omega hp plus epsilon. So, you can always define a model like that and you can try out whether its good fit or not. So, we are writing the function with its basis expansion, we call it phi y equal to phi beta.

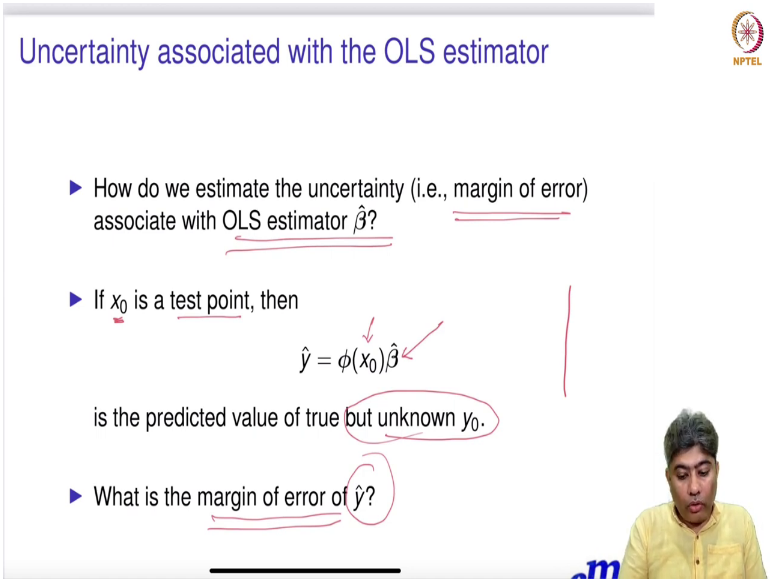
Let us assume these basis or the engineered functions are fully known to us ok. So, this is this is something we are making this assumption, this phi system is completely known to us and the beta is unknown unknown. So, beta is completely unknown and our what we want is we want to estimate beta. So, we want to estimate beta.

(Refer Slide Time: 25:01)



So, this comes into the something called functional estimation or functional learning. We will talk about it later more into the later. So, we can we can apply here also the OLS estimator. There are different method of estimation for beta is there and we can simply apply the OLS estimation of phi transpose phi inverse phi transpose y. We have done that same for the quadratic regression.

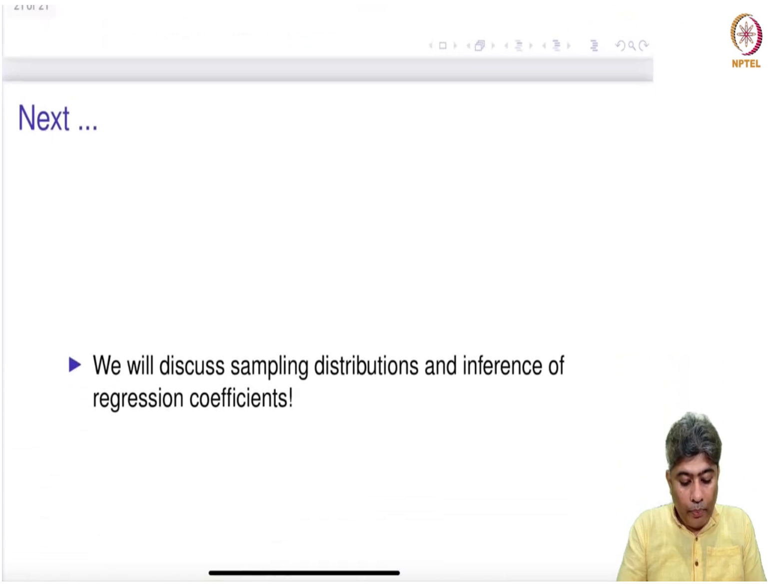
(Refer Slide Time: 25:30)



Now, question is uncertainty associated with these OLS estimate. How do we estimate the uncertainty or that is margin of error associated with the OLS estimator beta hat? So, if x naught is a test point, you plug the x naught here ok, you take the beta hat and you got the y hat. So, y hat is the predicted value, but the true y naught is unknown, the true y naught is unknown.

So, what is the margin of error of y hat? Can we say anything about the margin of error of y hat? And, if we can say something about the margin of error of y hat, if we have some idea about the beta hat, if we can say something about the beta hat; then only we will be able to say something about the y hat.

(Refer Slide Time: 26:45)



So, with this we will stop and we will move on to something called the sampling distribution and the inference of the regression coefficient. So, the sampling the distribution will give us that how much margin of error you can expect along the beta hat, how do you predicted y hat. So, in the next video we will resume with the sampling distribution of beta hat.

Thank you very much. See you in the next part.