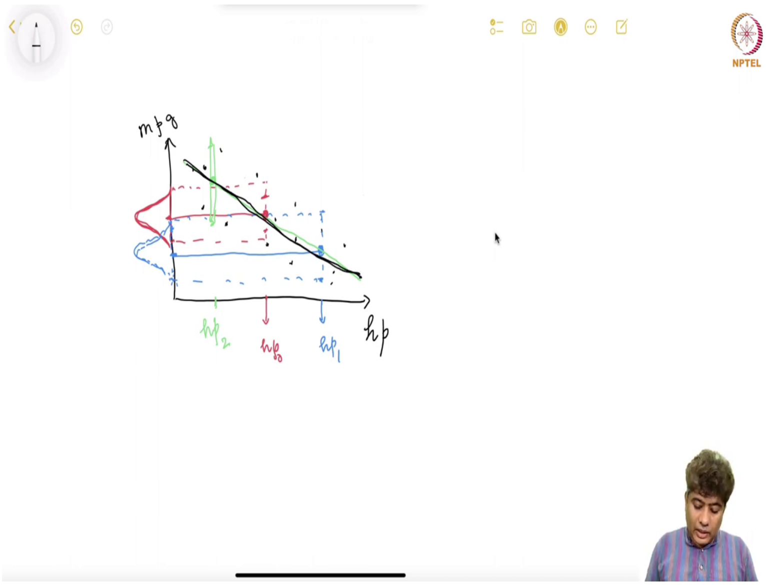
Predictive Analytics - Regression and Classification
Prof. Sourish Das
Department of Mathematics
Chennai Mathematical Institute

Lecture - 10
Regression Line as Conditional Expectation

Welcome back. Now, I am going to discuss the concept of Regression Line from Joint Distribution.

(Refer Slide Time: 00:26)



Let us consider this data set that we are considering. Suppose, this is on the x axis, we have horsepower, and y axis we have miles per gallon. Now, we have sort of data something like this. Now, what we have is suppose we have a prototype card for which we know the horsepower is somewhere here. This is the horsepower.

Now, the question is if this is the horsepower, then which could be the miles per gallon in this based on this data set. So, if this is the horsepower, most likely the value of miles per gallon will be somewhere in this range, in this range.

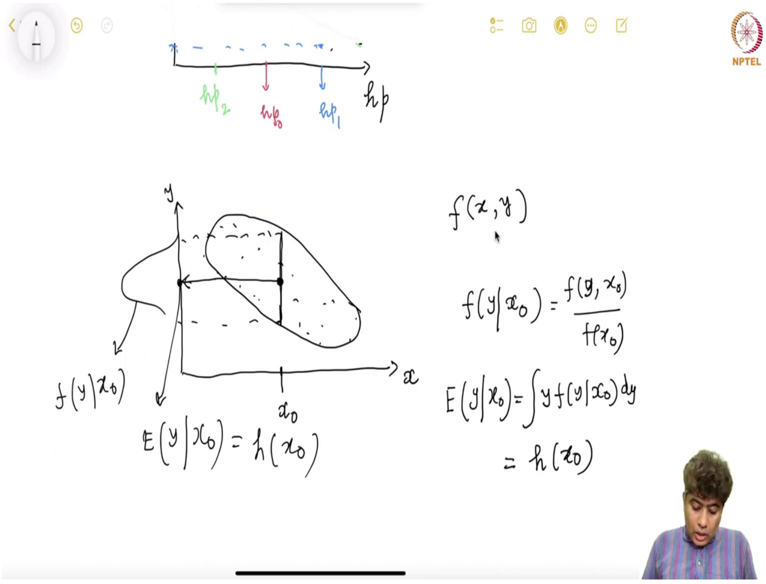
So, we can think of that most likely that these value of miles per gallon will be in this range, with mean could be here. And then, somewhere you can think of a distribution because the miles per gallon will not be exactly on this point if you have a horsepower here, your observed miles per gallon could be somewhere here or somewhere here or somewhere here. So, we can think of it is a distribution like this.

Now, so, suppose this is horse, one particular test point. Now, if you have another test point say hp 1, this is the another test point, now what would be the miles per gallon you would expect? So, most likely this would be somewhere here, in this range, right, somewhere in this range, with a expected value to be somewhere here.

Now, so, that means, essentially it is like this and you expect a somewhat a conditional distribution, a distribution some sort of a distribution of probable values of miles per gallon in this range with the expected value to be somewhere here. So, now, essentially, we are getting for different values of horsepower, possible horsepower values like here maybe this is hp 2, we will get a expected value to be here with a sum range of values probably in this range.

So, now if we join this expected values, then what we get is the expected line or what we call regression line. So, generally, we can think of this is the line that is passing through the middle of the point. This is the expected line that we had explained, we can think of. So, we have this distribution essentially.

(Refer Slide Time: 04:22)

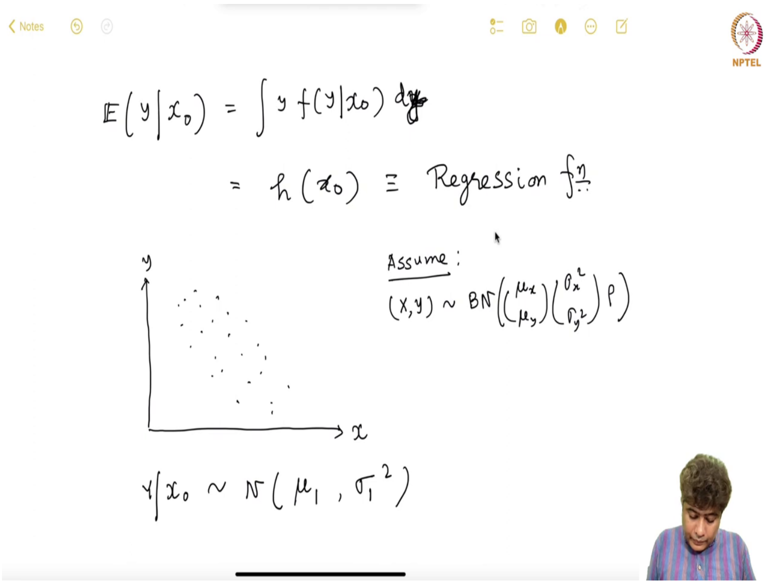


Now, if we generalize this, if we have a distribution x and y here, the points are somewhat like this. So, we can think of a joint distribution of x and y somewhere here. Now, for a given value of x, somewhere here for x naught, what would be the possible values of y that we can think of? So, that will be sort of a, what would be the distribution of y given x equal to x naught? So, that would be; so, this is the conditional distribution that we are thinking of.

Now, we can think of this will be like f of y comma x naught divided by f of x naught. We can think of the conditional distribution to be like this. And what would be the expected value of this? So, the expected value of this will be, expected value of y given x naught would be integration of y f of y given x naught d of y. Now, this will be some function of x naught, ok.

Because; so, this is my; so, this will be definitely some function of x naught and you know we can think of this is my f of y given x naught this distribution, and this expected value is my expected y of given x naught. This will be some function of x naught after integrating out the y.

(Refer Slide Time: 06:44)

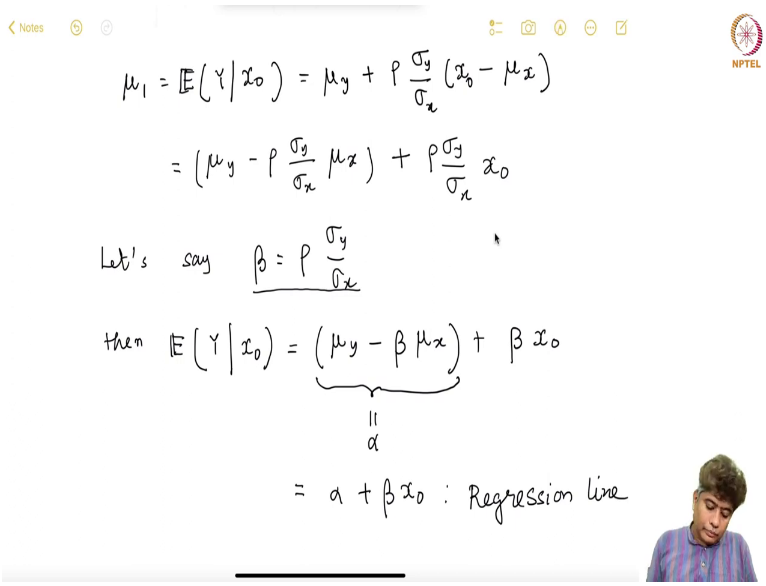


Now, so, this if you look at carefully this expected value of y given x naught which is integration of y f of conditional expectation y given x naught sorry d of y. So, this will be some function of x naught, and this is typically called regression function. This is not necessarily has to be regression line. This is typically called regression function. And our objective is to estimate this regression function in the regression line.

Now, interesting thing is, interesting thing is, if we assume that the relationship between x and y, if the relationship between x and y is some sort of bivariate normal, if we assume that relationship between x and y is some sort of you know have some distribution. And we believe looking at the data looks like this distribution x and y follow bivariate normal with mu mean mu x mu y variance sigma x square sigma y square and the correlation rho.

So, there are 5 parameters in the bivariate normal mu x, mu y, sigma x square, sigma y square and a rho. If you assume that x and y follow bivariate normal, then you can show that y, conditional distribution of y, conditional distribution of y given x naught will follow a normal distribution with some mu some sort of mu conditional y distribution and conditional variance.

(Refer Slide Time: 09:28)



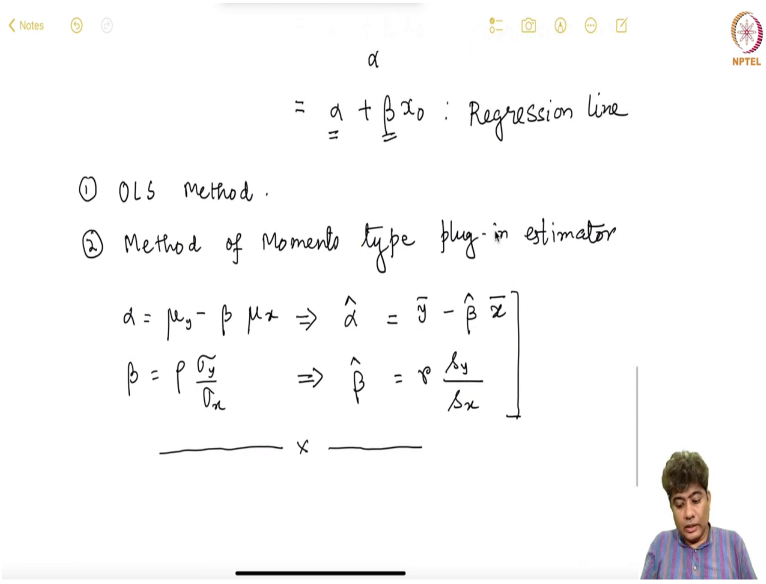
Now, what is this conditional mu 1? Mu 1 is expected value of y given x naught which is essentially turns out to mu y plus rho times sigma y by sigma x, x minus mu x, ok. So, this is the turns out that conditional expectation of y on x, y given x naught. So, we can call it x naught here, is a regression line. Why I am calling it regression line?

So, you can carefully look into it. I can rewrite it as mu y minus rho sigma y by sigma x times mu x plus rho sigma y by sigma x, x naught. Now, let us call, let us say beta equal to rho of sigma y by sigma x, then then my conditional expectation of y given x naught. I can write it as mu y minus beta mu x plus beta times x naught of we can.

Now, this if you look into carefully beta is rho sigma; beta is rho times sigma y by sigma x which is completely constant, mu x is constant, mu y is constant. So, it is actually a constant. We can call it alpha. So, essentially it turns out alpha plus beta x naught. So, if you assume underlying distribution between x and y is bivariate normal, then we can argue that the conditional expectation of y given x or the regression function is a line and this is typically called regression line.

Now, one way of, question is how can I fit alpha and beta. Obviously, you can fit alpha and beta using OLS method.

(Refer Slide Time: 12:43)

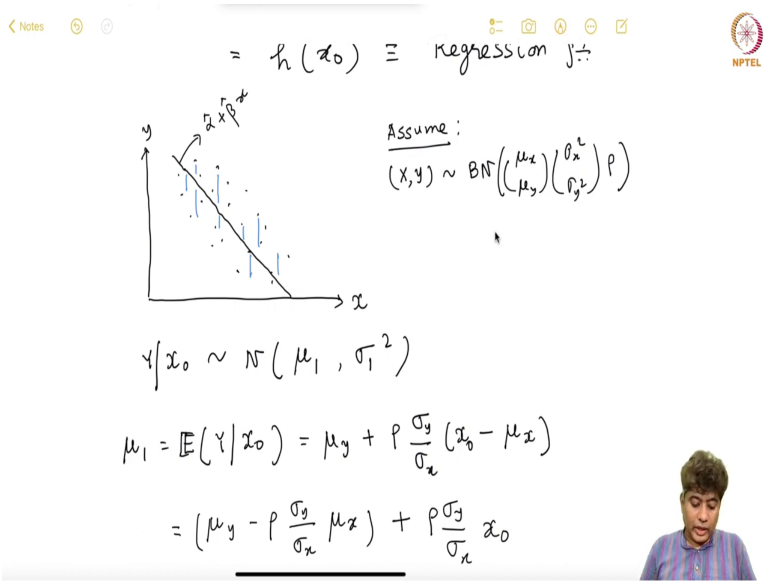


So, one method is ordinary least square method, that you can always use. We can also have some plug in estimator method of moments, method of moments type moments type plug-in estimator, ok. For example, what is my alpha? Alpha is mu y minus beta mu x. And what is my beta? Beta is rho sigma y by sigma x.

Now, I can estimate beta hat using sample correlation coefficient into standard deviation of y by standard deviation of x. Similarly, I can calculate alpha hat using this is mu y. So, y bar minus beta hat x bar. Using these estimates, these are the plug in estimates that we can use to estimate alpha and beta.

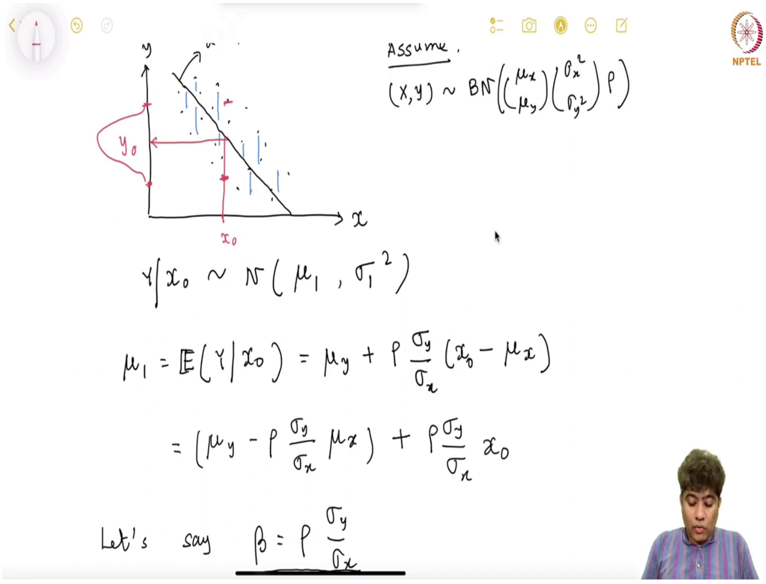
This works very nicely for at least the simple linear regression model with 1 x and 1 y. And you know you do not have to, I mean it could be like off the shelf estimator that you can estimates. Once you have the estimators, the next thing we want to discuss is still there is another issue of a regression line is still left with. Because if you see, the this is the regression line that we can estimate using alpha hat plus beta hat times x.

(Refer Slide Time: 14:54)



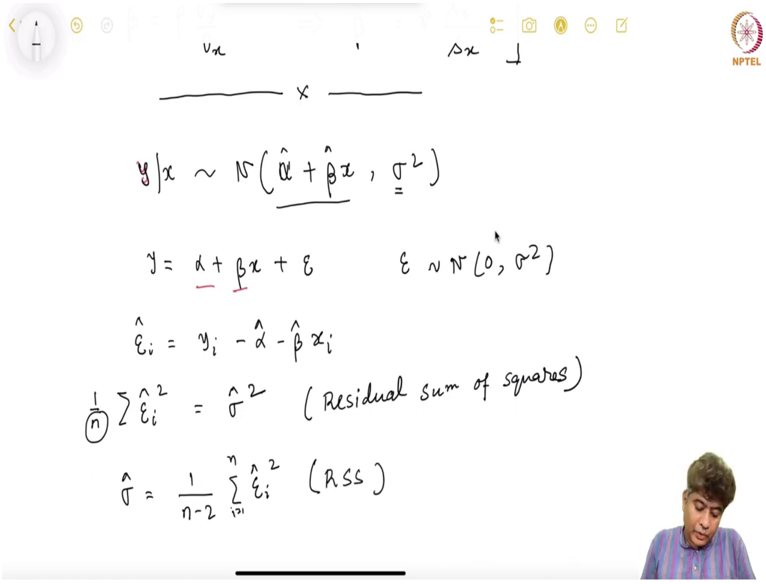
Now, you still have some of the one another issue still left with that is called the residuals. So, these are the, that these differences are called residuals, different points you are still left with. Now, how can we calculate these residuals will give us the idea that how much variability we can expect from our expected conditional expectation, because what happens carefully, you note that if I have; if I have say x naught here, if I have a x naught here, so this x naught the line will tell me that, ok this is the expected y naught.

(Refer Slide Time: 15:58)



But how much deviation I can expect? So, I have to get some kind of deviation, so that I can say that, ok this is the interval within which a 95 percent chance that my true y will up will lie. So, how can we do that?

(Refer Slide Time: 16:37)



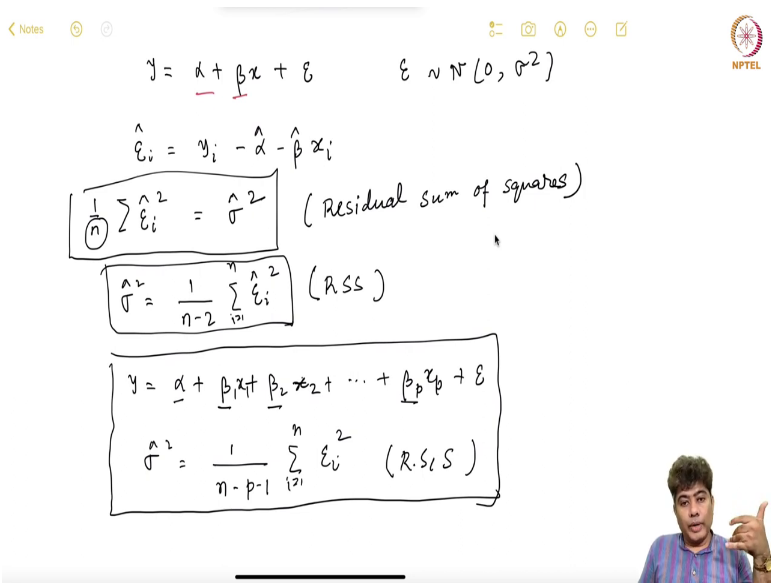
Now, you look into carefully that y, let me take a black color, y given x follow normal alpha plus beta x comma some sigma square. This is my conditional distribution. So, I have to estimate this sigma square.

So, I have estimated alpha and beta with alpha hat beta hat. So, what we can do? We can think of this model as y equal to alpha plus beta x plus epsilon, where epsilon follow Norman is 0 sigma square, and then automatically, y will give me follow normal alpha plus beta x and this.

So, what we can do? We can calculate epsilon i hat as y i minus alpha hat minus beta hat x i. And then, this will give me, we can use this square and we can just take the residual sum of square and this could be a sigma hat square. So, this will give me a idea about the residual sum of squares. This is this sometimes called residual sum of squares.

Now, I understand some of you may object here that I am using a 1 by n as a denominator. In some of the books the formula is given as, for residual sum of square is 1 over n minus 2 sigma summation i equal to 1 to n epsilon i hat square. Now, why n minus 2? Because you are estimating 2 parameters here you see, you are estimating 2 parameters here, alpha and beta, because of that you are losing 2 degrees of freedom. So, this is the typical you know formula for residual sum of square or RSS.

(Refer Slide Time: 19:27)



And if you have essentially a model which is beta alpha, we should consider alpha plus beta 1 plus beta 2, beta 1 x, beta 2, x 1, beta 2 x 2 plus dot dot dot beta p x p plus epsilon. Then, what will happen, your sigma hat square will be accordingly adjusted to n minus p minus 1 summation epsilon i hat square i equal to 1 to n. And this is the residual sum of square because you are estimating beta 1, beta 2 pa estimator plus 1, so n minus p plus 1. So, that is result to n minus p minus 1 as the divisor.

So, we will discuss about this more when we will generalize this. But for the time being we can focus on either this formula or this formula to estimate the residual sum of squares of error or the how much we can expect to deviate from the expected predicted value of y.

Thank you very much. Let us continue.