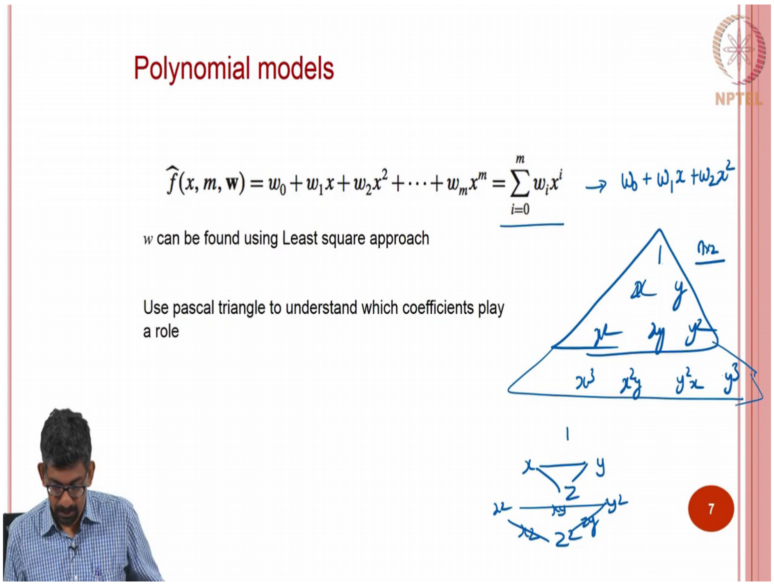
Surrogates and Approximations in Engineering Design
Prof. Palaniappan Ramu
Department of Engineering Design
Indian Institute of Technology, Madras

Lecture – 11
Types of Surrogate – Polynomial Models

(Refer Slide Time: 00:17)



Now, we are entering into the different types of meta models. The first one is very simple what we have discuss already y equals mx plus c, but if it is quadratic I will write it as beta naught plus beta 1 x plus beta 2 x square. So, I am just giving a generic form to that I am saying I am going to call it w i x i, ok.

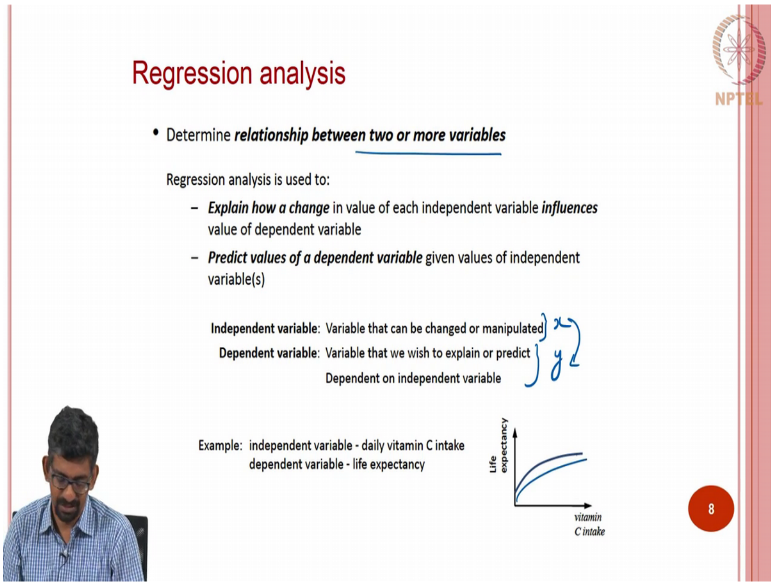
So, let us say that i runs from 0 to m ok. m is 2, let us say, so, your expression will be w 0 x 0 is what 1 plus w 1 x 1 plus w 2 x square that is all. So, now, you can use this guy for anything for any n dimensions all that, but this is assuming that this is only 1-dimension. If you have 2-dimensions then you have to have one of this plus interactions also need to have an a i a j and then you will have to have x i x j also ok.

So, that you and how do you decide upon that is you can use a Pascal's triangle. You know what a Pascal's triangle is, 1, x, y, xy, x square, y squared x squared y y squared x oh correct one no, no sorry x cube sorry and then you can keep building this. So, it is called a Pascal's triangle, right and for 3 dimensions beyond that which you cannot do for three triangles it will be a pyramid kind of a structure.

So, 1 so, it will go like this ok. So, you can build it like this x square y square z square and then it will be. So, you can build that and of course, beyond the second or draw 4 dimensions, but the deal is the question is if you are going to approximate your data points in 2-dimensions with a quadratic then I need to take 1 2 3 4 5 6 coefficients I need to find. So, minimum n times 2 number of coefficients was 6, 12 points I require or if you said you know I want to have a cubic then 6 plus 4 is 10. So, you can keep doing like that.

But, there are also other ways in which you can find out the number of coefficients because Pascal triangle is for our basic understanding, but there are expressions n times n minus 2 divided by d minus 3 d plus 3 or something you can find the number of coefficients, that is not a problem ok.

(Refer Slide Time: 03:12)



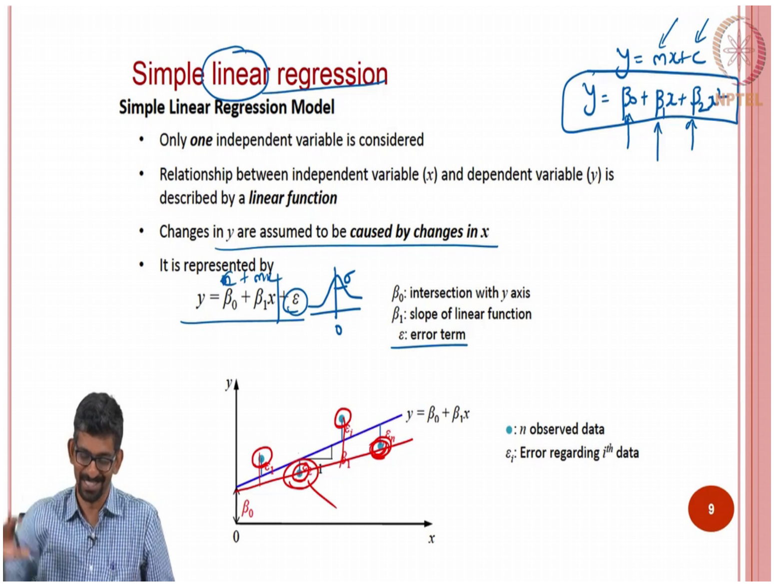
So, we will introduce the generic regression analysis now which is eventually which will also let you find your coefficients which is the w that we saw in the previous slide, what is the idea behind the regression analysis it is a relationship between two or more variables. Obviously, there is a dependent variable there is an independent variable, but what you should also understand this there is no causal relationship because this is just data mathematically it does that.

One example that people usually give us there was a people checked in the western they checked the blood pressure of people in different strata and then they figure out that people who have higher degrees had lesser blood pressures. They had a PhD or maybe double PhD or something like that they had lesser blood pressures then some ones who had a high school dropout or something like that ok. So, they just said ok, what is your education they took different attributes, right. So, one of the columns was your education. So, they looked at those two data, blood pressure and education. Then they saw that there was a direct relationship. So, whoever had higher degree they had lesser.

So, that is just a data analysis they also have an explanation for that what they are saying is remotely they are connecting not remotely they are saying that with the higher education probably you are exposed more you are in a better position. So, you know you have more exposure to staying healthy and all that. So, your blood pressure is under control whereas, the other ones to me get jobs and they do not have time to bother about their health and stuff like that, but the important point is it is not a causal relationship, ok. So, when you do a regression analysis you cannot say this is the cause and that is the effect.

So, the basic idea is your independent variable is your x and your dependent variable is your y, that is what we are talking about here. When I am changing my x I want to find out what my y is. So, this is this is a this is causal relationship vitamin C intake and life expectancy I want to get this curve, that is all.

(Refer Slide Time: 05:28)



Ok, this is something that I want you to pay attention. Linear regression, do not take your perspective of linear. Your perspective of linear is this is a linear equation. This is a non-linear equation and the quadratic equation, correct do you what your perspective is now just, but I did not say linear equation. What did I have said here? I have said linear regression. What are you regressing? You are regressing the coefficients regressing means taking errors and then finding.

So, you are regressing the coefficients, you are not regressing the equation. So, this non-linear equation is still linear regression because my coefficients are still linear, ok. So, do not say I do non-linear regression because I use a cubic response surface, no. You might still do you might do a non-linear regression, that is a different story. But, what I am saying is just because your equation is non-linear does not mean that you are doing a non-linear regression, it is still only linear regression oftentimes just figure that or not.

Anyway, in this particular case we are considering only one independent variable. The relationship between the dependent and the independent variable is linear in nature. The changes in y are I mean in this particular case not in linear regression in this particular case. The changes in y are assumed to be cause by changes in x that is the assumption basic assumption, right.

So, now you represent this as beta naught plus beta 1 x or if you want for your easy this thing I am sorry c plus m x that is what we are looking and often we stop here, but the general description is there is always an error that is associated with a model. And, people usually say that this error can be approximated using a normal distribution with 0 mean and some standard deviation.

And, we know that the beta naught is the intersection with the y axis and beta 1 is the slope, that you already know y equals m x plus c, right and the epsilon is the error term. Now, what we are saying is these are your points data points and you are fitting this blue line here, this is the equation of this line. What I want to find? I want to find my beta naught and beta 1. Why is that a finding problem because if I change my beta 1 and beta naught this equation is going to change, beta naught has changed. The slope has changed that you will agree.

Now, I will reduce it this way beta naught has changed this slope has changed. So, like this I can do I do not know, infinite number of lines. The question is which of this line is good. Finding which of this line is good is not a problem, provided you have given the criteria someone come ask into this room and asks me which of these students are the best student. It depends on the criteria that you are asking. Based on the exam that you gave, based on the interaction in the class, based on the knowledge that they have developed because these are all not related based on the knowledge that they have gained out of this, based on the amount of implementation that they have done after taking your class, this is the criteria this is what we discussed yesterday, right the criteria is important that let us you make the decision ok.

So, once you finalize the criteria then it will either be maxima or minima of the criteria. Let us say it is a grade is my criteria then the best student is the one who gets the maximum grade. Again, interaction it is maximum number of questions, interesting questions, who have kept me in conversation for long maximum of that, ok. So, now, what do you think will be the criteria? Simple, come on, what is a criteria we have already seen, it is the error.

So, it is a very interesting regulation analysis teaches us life, because it is a compromise estimate. You cannot find a line that will pass through all the points. If you make real life connections to each of these points, ok. This is what your friends want, this is what your parents want, if you have a girlfriend this is what she wants and out of everything this is what your professor wants ok.

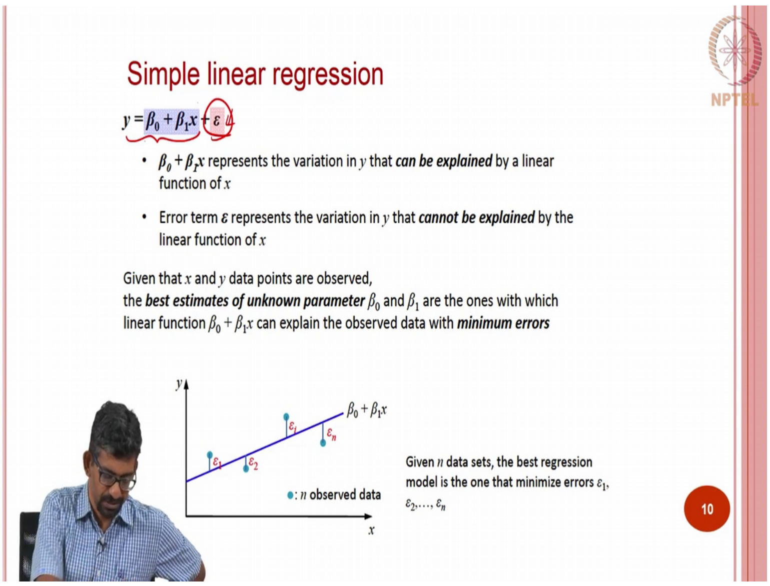
And, out of everything you have only 24 hours and limited capabilities to do each one of these. So, whoever you are, you cannot fit everything. For today you will do this. Since I have a class today, I will have a line that passes through this since I have a class today I will go and attend that class for sure. So, I will spend more time with the professor staff, ok. I cannot spend time with my parents or friends whatever I cannot spend with my, whatever I cannot go play today; I cannot work out today, ok.

So, life what it say this it is always a trade off and it is always a compromise. At that point in time you need to know what should be weighted more meaning weighted means not the wait weight, that that way that I am talking about. So, actually regression analysis teaches you life. So, that is an interesting part and you will never be take it from me and I am few years elder than you so that you can take it from me, you will never be able to satisfy everything, ok. So, it is a good idea to satisfy one at a time that is all, then you will have a better chance.

Sometimes you might have to leave out something in total that is an outlier in your data that will make the fit much better ok. Your friend who keeps on comes in the evening and then you know buys you something, but then keeps on negative influence, and this is happening that is my member of that is telling this. So, you go back and then you do not feel like coming back to the lab in the night, ok. So, that is really negative effect, ok. You should really get rid of that friend here that is a good idea actually.

So, it is an outlier sometimes you will have to have an outlier. So, that is why when you do a correlation plot your friend will be in the negative correlation side, you know what all influences me to go to the lab. Your friend will be in the negative correlation. So, you will have to eliminate that and then immediately you see their earlier performance increase.

(Refer Slide Time: 13:19)

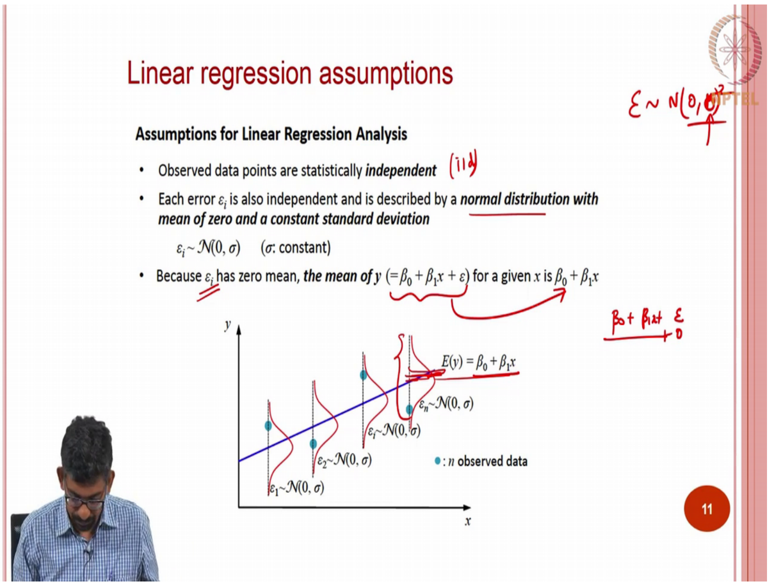


Anyway, going back to this, so, our simple linear regression as I pointed out we just use only this equation, but there is also an error. We will see how this error comes and how this error is usually built up. This particular term that beta naught plus beta 1 x it represents the variation in y that can be explained, that is all. There is also an unexplained term. That is what in principle your r squared captures. Your r squared says this is the information that is explained by the model.

If 98 means your model there are two – three different ways people look at it, but from a statistical perspective what it says is no model can be explained 100 percent, because it is random. Because tomorrow you generate one more point your model itself will have to be adjusted a little bit. So, what you need to knows is it is not enough if you predict you need to be able to tell: what is the confidence that I have in my prediction, ok. So, that is why I mean I want to make that comment.

So, it is not enough if you predict in a probabilistic sense you will also have to say: what is the confidence that you have, ok, that is when that r squared kind of a thing comes into picture. Given that x and y are observed you want to find the best parameters beta naught and beta 1. So, if this error is 0 then that is the best beta naught and beta one that you can have, that is what it says.

(Refer Slide Time: 14:55)



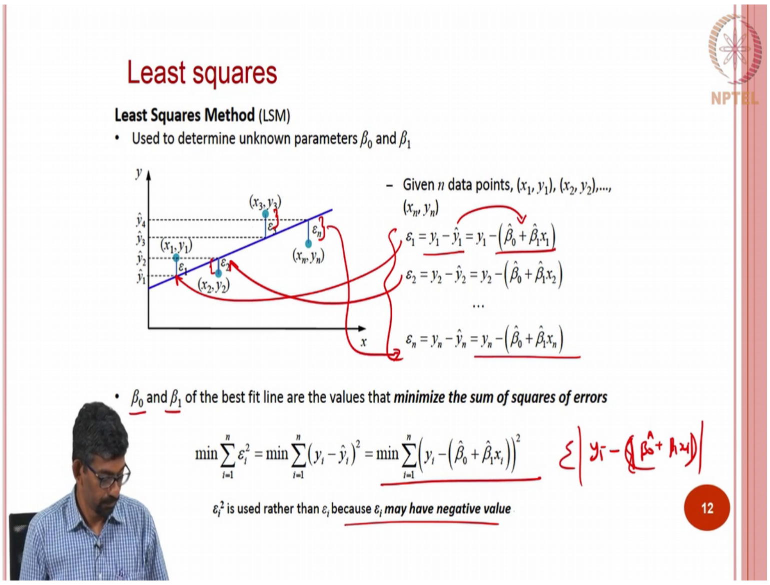
Ok. So, what are the assumption that goes into linear regression? If you remember, I told you that epsilon follows a normal distribution with 0 and some standard sorry some standard deviation. The observed data are statistically independent it is usually called iid ok. What is iid? Identical independent distributor they are identical samples independent and independently distributed.

So, each error is also independent this error is not dependent on this error. This error is not dependent on that error they depend on the fit, but not on each other. Because, if I change this error will this error also get better not necessarily this error might actually get worse. If you take a non-linear equation you can understand that thing. Each error is also independent and is described by a normal distribution that is what we told with a mean of zero and a constant standard deviation, ok. These are for theoretical purposes these discussion come into picture. There are also people who try to go and approximate this standard deviation and reuse this, ok.

So, what is it that I want to tell them this I forgot. So, each error is follows a normal distribution and then it follows, since this epsilon which is the error that we are talking about has a zero mean, the mean of this guy the line the mean of this guy for this particular stuff since I mean take an equation y equals whatever it is a linear equation. So, beta naught plus beta 1 x plus error, right; so, we are saying that the x is random. So, the mean of that ok, will be the mean of that mu also which we are saying that the error of that the mean of that error is also 0. So, it will be you are only adding this one. So, this is equivalent to this that is what we are saying.

So, that line will be just beta naught plus beta 1 x and then anywhere else you need to have this error estimate to understand. This r squared kind of gives you this error one that is what it says how much of the variation is explained. If it says r is 1, it says the variation is fully explained, meaning it is like this it will be here though; it will not be because the variation is fully explained. The line has to go through here and it will be a meaning it is a deterministic line it is not a distribution there the it is very clearly observed that is all.

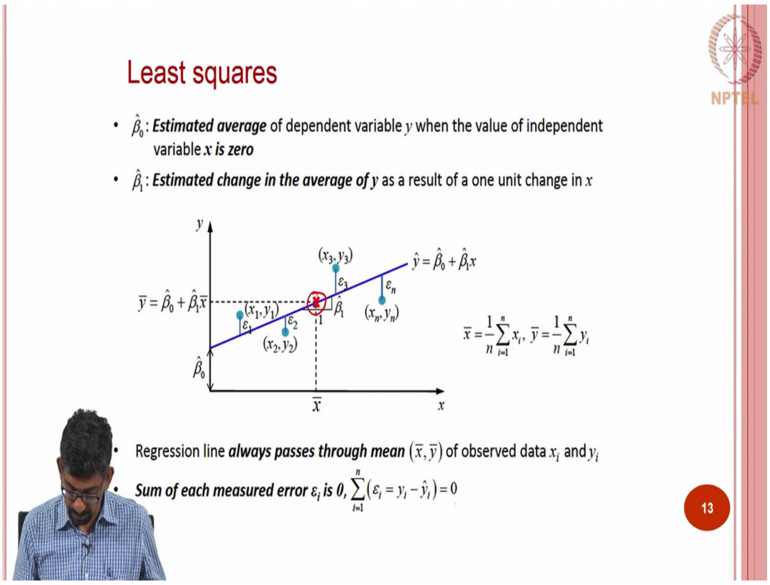
(Refer Slide Time: 17:40)



Now, the least square method this we have already discussed, but for the sake of completion we will do it here again. So, there are four points the distance in each of them x 1, y 1, x 2, y 2, x 3, y 3. So, you have error 1, error 2, error 3, error n, ok. Now, what you do this is important this arrangement of the data is important, ok. Error 1 is y 1 minus y hat error 2 is this error, this error and then yeah error 3 would have been this error and error n would have been this error.

Now, y 1 minus this guy is this ok, y 1 hat how would I get because I knew the form beta naught hat plus beta 1 hat x 1. This beta 1 and beta naught hat are the same for all of them I am just changing only my x 1, x 2, x n. Now, what I am going to do is this I need to beta naught and beta 1 of the best fit line are the values that minimize these sum of the square. So, I am just writing them as error squared sum them like this, ok. Epsilon i squared is used rather than epsilon i because you might have negative as well as positive values. As I told you can also use y i minus absolute value of beta naught hat plus beta beta i x i sorry, the absolute of the whole thing you can also used.

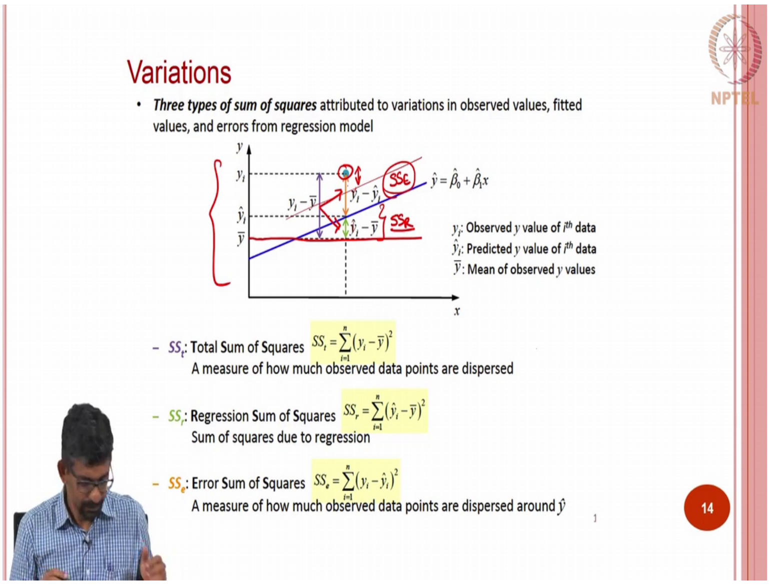
(Refer Slide Time: 19:17)



Ok, but there is a reason for using the squared it has it brings in very nice statistical properties and then you have this you know your normal distribution varying another that is the reason that it is used. People have shown that using the absolute deviation also will not make a lot of difference.

So, there are some interesting properties. For instance your x bar y bar they will intersect on the line for sure, ok. It is a mean of x bar sorry x and y, ok. They will pass through the line there are no other it might not pass through any other points, but x bar and y bar intersection it will pass through, some interesting properties are there. Passes through the mean sum of each measured error is your you are trying to force them to be equal to 0.

(Refer Slide Time: 20:09)

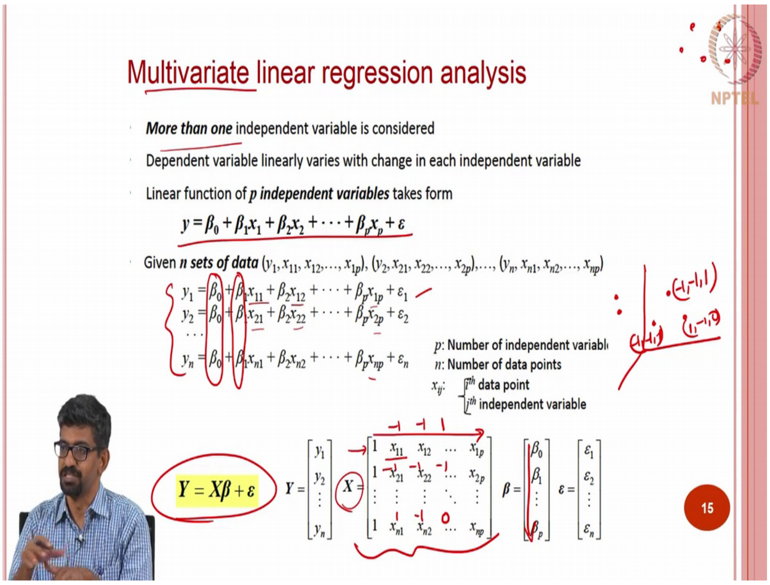


Now, this is what I we spoke about right the sum of the regression and the sum of the error and the total sum of the squares. So, this is y bar it is the mean of the y data. So, this is the y data whether we are talking about and this is my y bar. What I am going to do is, I am going to talk about my line from meaning the fit from this red line, how good it is. So, what I am saying y i minus y bar. So, that is my total sum of the squares, ok.

Now, I am decomposing in this into two. One is the actual point and the predicted value, ok. Similarly, the actual point minus sorry the predicted value minus the y bar the average that you can have. So, I am decomposing this into two and this guy is the orange one is the error sum of the squares and the green one is your regression. Because, this is comes from your regression in case if my line was here instead of this blue line I had this line then this is the regression difference with respect to my y, and this is a this still it is a different thing this error is still the error that comes as a residual sometimes. This is also called as a residual sum of the squares here r is not used because we do not want to confuse regression errors and residual errors, ok.

This error the between the point and your prediction the actual point and your prediction is the residual or the error sum of the squares. So, you can also write r square anova my expressions from this and they say I do not exactly remember SS e by SS r or SS r by SS e, if you are comfortable with that you can use that, but this is the idea, ok.

(Refer Slide Time: 22:17)

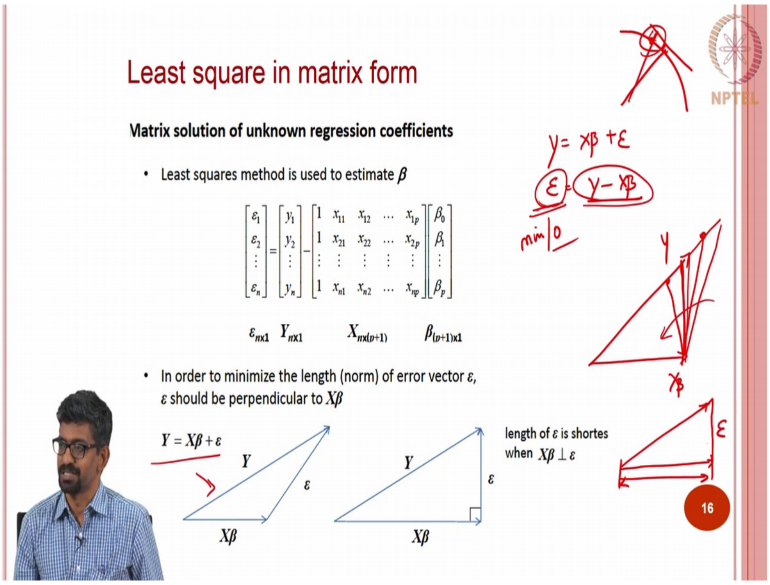


So, does that mean that I can only do one variable or I cannot fit complex curves or surfaces? No, that is not true. You can also do multivariate linear regression and we will also see how you can adopt this for non-linear regressions also. So, what does multivariate mean? It means that you have more than one independent variable, right now we just saw x, right. So, you can have x 1, x 2, x 3, x n also and this is what the equation looks like it is just beta a naught plus beta 1 x 1, beta 2 x 2 beta p x p plus epsilon. We are not taken the interactions.

So, we are just going to directly go here and we are we are going to take a geometry idea to see how this linear sorry, the minimum least square is achieved, ok. Given n sets of data y 1 to y n you have beta naught plus beta 1 you are able to predict this guy that is the whole point, ok. This beta naught beta 1, beta 2, beta p, beta n everything is the same for each of these data points you understand, right.

For each of this y 1, y 2, y n my beta naught, beta 1 will remain the same. What will vary? Only my x's are going to vary, that is all. So, this immediately tells you that you can write this in a matrix form. So, what we are going to do is we are going to write it in this form, ok. So, Y I know is a vector beta of course, you know it is a beta beta naught beta. So, now, I just need to write it in this form, that is all. So, I will say beta naught plus beta 1 x 1 1, beta 2 x 1 2, beta p x 1 p, that is all you need to write, ok. There is a reason to write it this way, we will see what it is.

(Refer Slide Time: 24:05)



So, I am just writing because you remember this, right. We write it like this. So, I am just rewriting this as error equal to y minus x beta we wanted to minimize this error that is all you wanted to do there is a reason to write it this way. So, error equal to y 1 minus y 2 minus this matrix x matrix times beta naught to beta p.

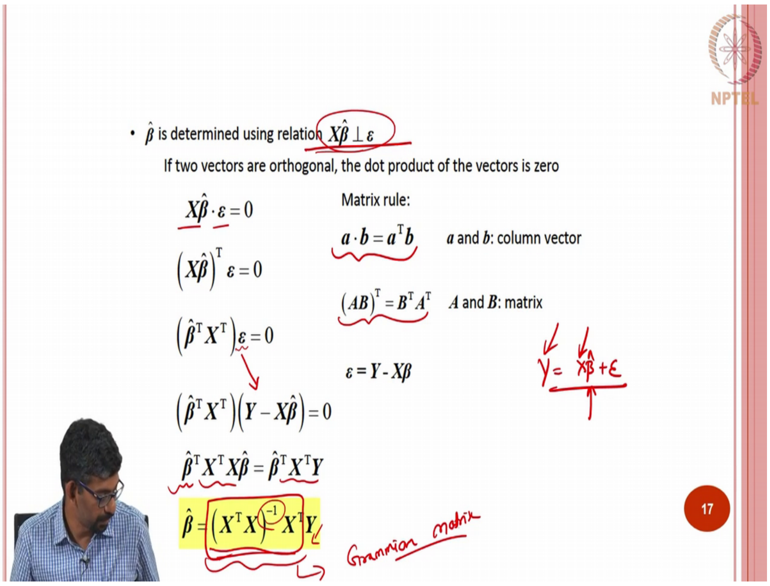
Now, what is it that you want? You wanted your error to be minimum or minimum or 0. This is what you want this guy to be. So, this y and x beta in a geometric sense can be written like or can be represented like this in a vector form, ok. So, I can have x beta do not look at the bottom part at y, and the error is the line connecting this guy and this guy.

So, the question is when will that error if this is a line that is going to connect this guy and this guy, when will that connecting line have a minimum value it is here, it will be like this, if you do is here it will be like this, if it was here it will be like this. So, which of this will be minimum? It is not 0, please understand. The questions the question is different I am not asking when that guy when that guy will be 0 when y is on the on x beta, but the minimum this is a turned out to be a good example. The minimum one way of looking at that is also projection correct when will that guy lie on top of this guy, when you project he should fall exactly on top of it ok, right now with an extender version also you project it, ok.

So, when you project it should fall on that guy exactly ok. So, when will that happen? When will your when you project this guy you should fall exactly on this, when will that happen, when? So, this eta will take minimum value when it is perpendicular your normal variant everything will be like that only.

So, sometimes when people make pictures they make no mistake they do this and then they will not draw it perpendicular they will just draw it something like this they try trying they are just saying conceptually they are explaining right that I have the explanation was not this. So, they will leave out that and people will say that is not optimum because it is not perpendicular. So, that is that is why the idea is ok. So, when you when your error is perpendicular to your x beta between your x beta and y that is when your error is going to be minimum, ok. So, this is the geometric sense, right. How are we going to use it to in a mathematical sense? We will see how to do that.

(Refer Slide Time: 27:47)



We are going to just use this relationship X beta should be perpendicular to your or the other way around the error should be perpendicular to your X beta for minimization. You need to know little bit of matrix you need to be able to appreciate, but I am sure you can do that, ok. If two vectors are orthogonal what do you know the dot product of the vectors is 0.

So, the two products that we are talking about is X beta and epsilon, ok. They are equal to 0, but there is also a matrix rule that says a dot b is equal to a transpose b. So, what I am doing is, I am making this guy transpose time error equal to 0. There is another rule that says AB transpose is equal to B transpose A. So, I am just writing this in a slightly different manner this is AB transpose. So, I am going to write it as A transpose B transpose. So, beta transpose time X transpose epsilon equal to 0 then what I can do is I just keep this guy and I replace this epsilon as Y minus X beta because we know that Y equal to X beta hat plus epsilon.

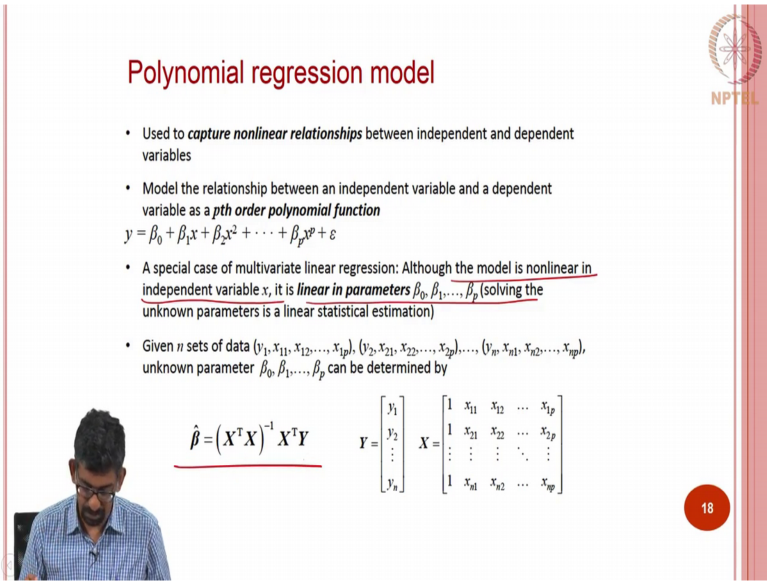
So, I am just replacing this epsilon by Y minus X beta. Then, I can write this expression as beta transpose X transpose times Y, right that I bring it to the right side and then I say beta transpose X transpose X times beta here. It is just I am just taking a dot product here, that is all ok. There are two quantity and we are just equating each one of them ok. Now, what happens says what is that we wanted to find we always want to find only beta, that is what you want to find in the entire equation you know Y you know X you want to find your beta that is what you want to find. So, I want to find beta and I am isolating the other things. What is the other things X transpose X inverse X transpose Y. What is this quantity got to do? It has only X values, that is all.

This capital X is nothing, but this guy you already know this. This is nothing, but your coordinate points, ok. You take point number 1 this is x 1 1, 1 2 meaning the second dimension the n-th dimension, I take this point nth point in dimension 1, in dimension 2, in dimension n what is it is coordinate you understand right. So, let us take just three coordinates ok, I give you three points, but imagine the these are in space. So, I am saying minus 1, minus 1, minus 1 and then I am saying minus 1, minus 1, 1 and then I am saying 1, correct yeah it can be 1, 1 minus minus 1, 0. So, it can be anything that is not 1.

So, for this point this guy will be minus 1, minus 1, 1 for this guy it will be minus 1, minus 1, minus 1 and then 1, 1, 0 minus 1, 0 that is all. So, it is just the spatial coordinates that is all your X matrix is nothing, but the spatial coordinates that is all, ok. Of course, you will have issues if you have two points here and then you had ten points here in a design space, because you are going to invert this you will transpose this and all that they you will get some because you are going to invert this you will have a transpose of that and all that. So, there will be numerical similarities.

So, this you know it is only matrix inversion and transpose that is all from your coordinate matrix and Y you know the values at those points. So, by just doing this you can get your beta by using this simple principle, that is all. You are not doing any optimization no minimization nothing just by using a simple projection algorithm not even algorithm approach we are able to do the simple. So, this is called the Grammian matrix, this is for your information not lot of people use the terminology it is called the Grammian matrix.

(Refer Slide Time: 32:15)



So, yeah so, you can also use a polynomial regression, it is used to capture the non-linear relationship between dependent and independent variables. A special case of multivariate linear all though the model is non-linear see this is what I told although the model is non-linear independent variable X, it is a linear in parameters beta naught, beta 1, beta p solving the unknown parameters is still a linear statistical problem estimation. This is the same idea, oops I am sorry this is all you need to solve for, that is all.

So, whatever we discussed right now it covers polynomial response surface, ok. So, usually when people say response surface modelling they talk about polynomial response surface, but off let then people move to surrogates, they call it meta models. Now people are also calling emulators, they are using data mining techniques are data learning machine learning techniques to do stuff.

So, in that there are going to be three techniques that we will talk about. One is polynomial response surface which we have already talked about polynomial models, the second 1 that we will talk about this radial basis function. So, if you talk in if you think about all these things the type of functions are going to change that is all the form, but there is little bit more to that, meaning the moment the type of the function changes you have more control there is a little bit more parameters it is intelligent parameters are built in by changing those parameters you can get some shapes that are not that you might not be able to get with polynomial as simple as polynomial, but you need to know it should not be an over (Refer Time: 34:00).