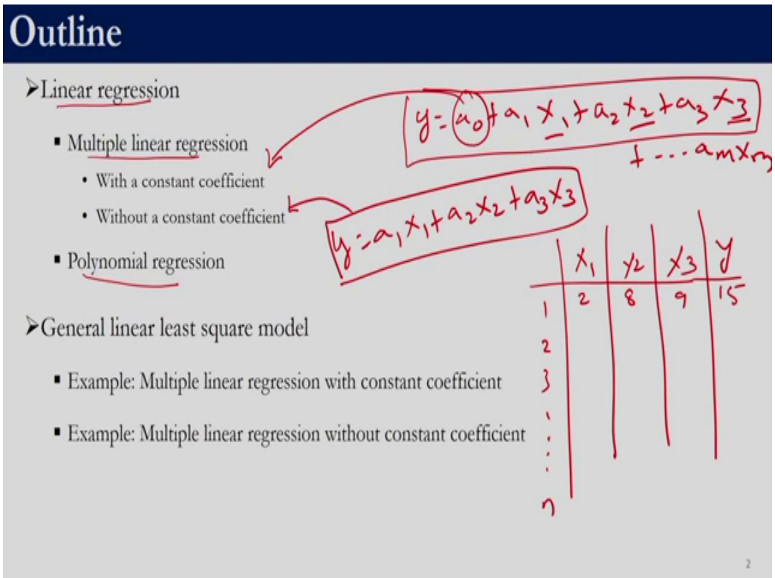
**Computer Aided Applied Single Objective Optimization**
**Dr. Prakash Kotecha**
**Department of Chemical Engineering**
**Indian Institute of Technology, Guwahati**

**Lecture - 03**
**Curve fitting: Regression**
**Multiple, Polynomial and General Linear Least Square Regression**

Welcome, in the previous session we had looked into simple linear regression where in we were given a data; wherein we were given a set of data points right, let us say x 1, y 1, x 2, y 2 all the way up to x and y 1 and we saw how to fit a; how to fit a model y is equal to a naught plus a 1 x. So, our objective was to find out the values of a naught and a 1; such that the model base represents the data point, that is what we had looked into it, that is going to form the basis of this session right.

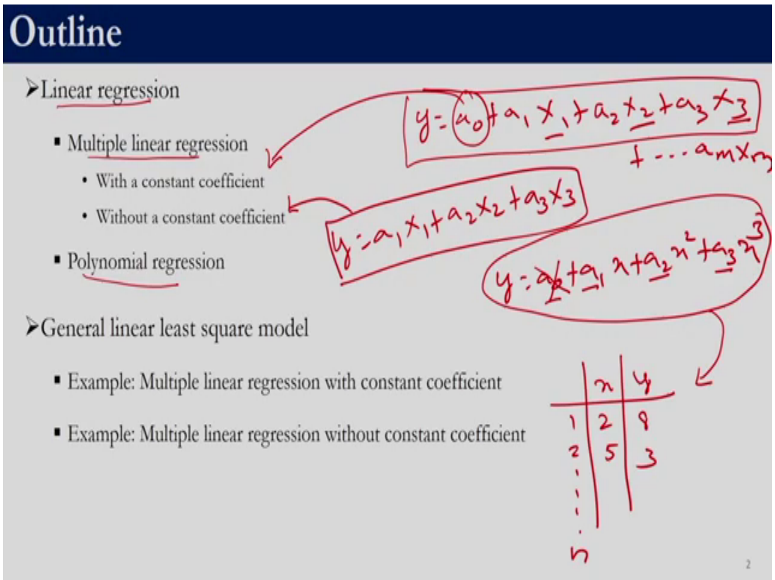
(Refer Slide Time: 01:01)



The outline of this session is going to be in under linear regression we are going to discuss multiple linear regression and polynomial regression. So, in multiple linear regression we are essentially trying to fit a model y is equal to a naught plus a 1 x 1 plus a 2 x 2 plus a 3 x 3 and so on. Let us say till a m x m right, where x 1, x 2, x 3 are the independent variables. So, the data points are given like this x 1 x 2 x 3, and for this. So, let us say this is 2 8 9 the y value is given, the dependent variable value let us say 15. So, like this we are given n points right. So, our task is to fit a model; y is equal to a naught plus a 1 x 1 plus a 2 x 2 plus a 3 x 3. We can either fit this type of model right wherein we have a constant coefficient.

So, we will discuss that first or our model can also be y is equal to a 1 x 1 plus a 2 x 2 plus a 3 x 3. So, here if we see, we do not have a constant coefficient, so we will also discuss that case right. So, that is what we are going to do in multiple linear regression. In coming to polynomial regression; in polynomial regression we are required to fit a model y is equal to a naught plus a 1 x plus a 2 x square plus a 3 x cube right.

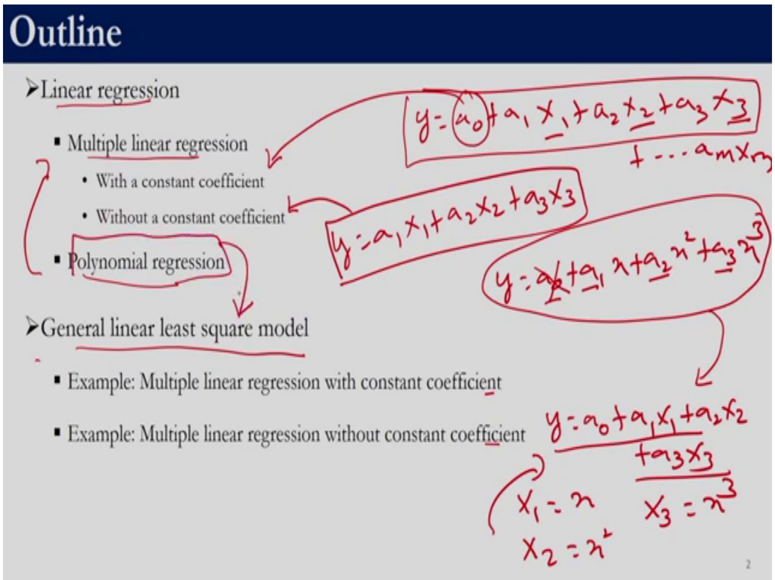
(Refer Slide Time: 02:39)




So, in this case the data that is given to us is the independent variable x and y, let us say 2, 8, 5, 3 like this we have been given n points. So, our task is to find out the values of a naught a 1, a 2, a 3 such that this model base represents this data points right. So, that is going to be polynomial regression.

Again in polynomial regression we can do it with constant coefficient or without constant coefficient, depending upon the need we can either have a constant coefficient or not or the constant coefficient may not be there in the model. So, that would be an extension of the discussion that we would have till that point of time, so that can be easily handled right. So, in this case what we will be essentially doing in polynomial regression is, we will be converting it into a multi linear regression right.

(Refer Slide Time: 03:43)



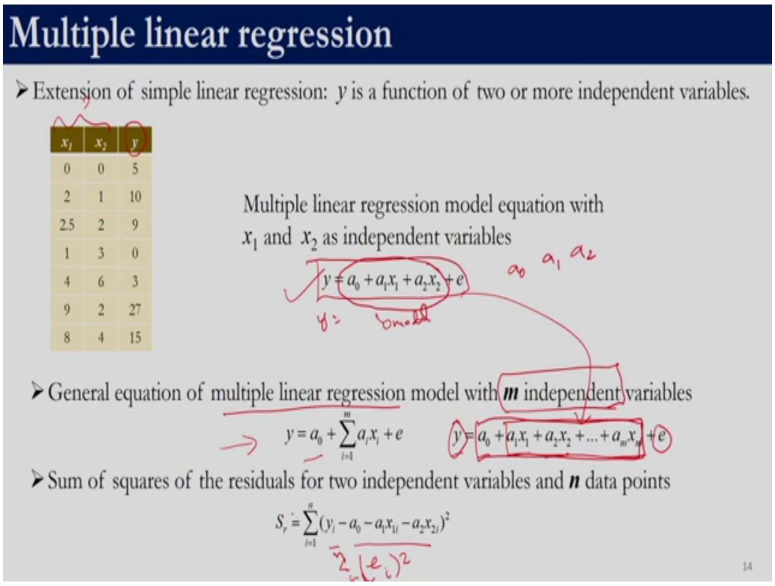

So, y is equal to, it is going to be let us say if we have the constant coefficient a naught, then we are going to say plus a 1 x 1 plus a 2 x 2 plus a 3 x 3. Where x 1 is nothing, but the data point x which is given to us, x 2 is the data point each data point squared, so x square and x 3 is going to be x cube right. So, once we do this transformation right, this is nothing, but the multiple linear regression which we will be do discussing in the first half of the session right.

So, polynomial regression is essentially being converted into multiple linear regression. After that we will be looking into general linear least square model. So, this is this was for simple linear regression. When we do multiple linear regression and for polynomial regression, there will be. When we do this multiple linear regression or polynomial regression this coefficient matrix is going to contain lot of terms which need to be evaluated right.

So, either we can stick to that process or we can adopt this general linear least square method. Wherein from the data points we will be able to directly get this a matrix right, the coefficient matrix and the right hand side can be easily obtained if we know general linear least square method.

So, under general linear least square method we will be discussing multiple linear regression with constant coefficient and also multiple linear regression without constant coefficient. So, that is going to be the focus of this session. So, in this case whatever we had considered so far we had only one independent variable right; one dependent variable and one independent variable. So, in multiple linear regression we will have more than one independent variable, the dependent variable is still 1 right. So, we are talking about models like y is equal to a naught plus a 1 x 1 plus a 2 x 2 plus a 3 x 3 and so on. So, now, let us look into multiple linear regression right.

(Refer Slide Time: 05:45)



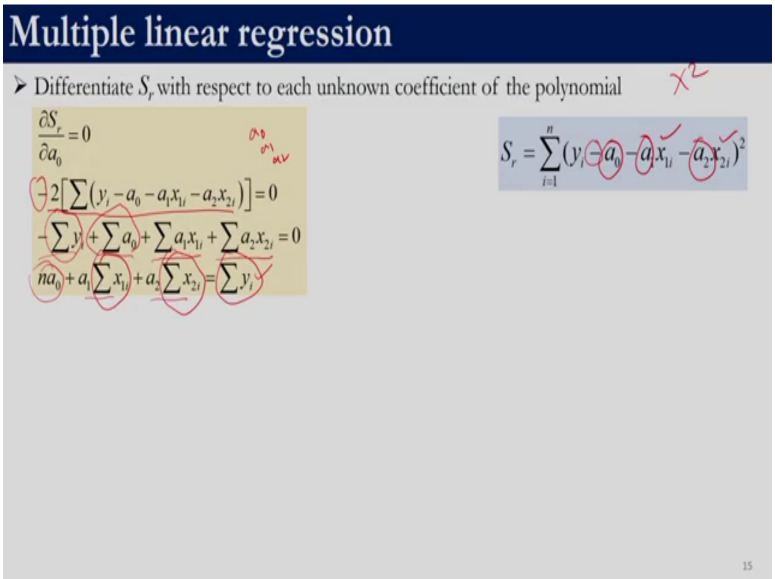

So, in multiple linear regression these two are independent variables x 1 and x 2 right. So, for example, that can be temperature and pressure. So, these x 1 and x 2 are independent variables right and y is a dependent variable. In this case the model that we want to fit is, the model is a naught plus a 1 x 1 plus a 2 x 2. So, this is y model right.

So, there depending upon how accurately this data was measured, there may or may not be an error associated with each data point right. So, our model is a naught plus a 1 x 1 plus a 2 x 2 and our task is to find out the values of a naught a 1 and a 2 right. So, this is equation involving two independent variables. We can have an equation involving m independent variables. So, for example, we can have y is equal to a naught plus a 1 x 1 plus a 2 x 2 all the way up to a m x m right and since this is the measured value and this is our model the error vector would be there right.

So, this can be compactly written as a naught plus summation i is equal to 1 to m ai x i. So, that will capture this part completely right, except for the a naught which we have separated out and then plus e. So, this is known as multiple linear regression and it has m independent variables. So, first we look into this case. once we are comfortable with this case this is merely an extension of, this can be extended to this one right. So, even in this case we need to define, we need to set a criteria right.

So, the criteria is again going to be the sum of square of errors right. So, this is the model value, this is the observed value. So, this is the error right, error have associated with the ith point. So, we are going to square it and then we are going to sum it up right. So, S r indicates the sum of square of the residuals or the sum of square of errors.

(Refer Slide Time: 07:51)

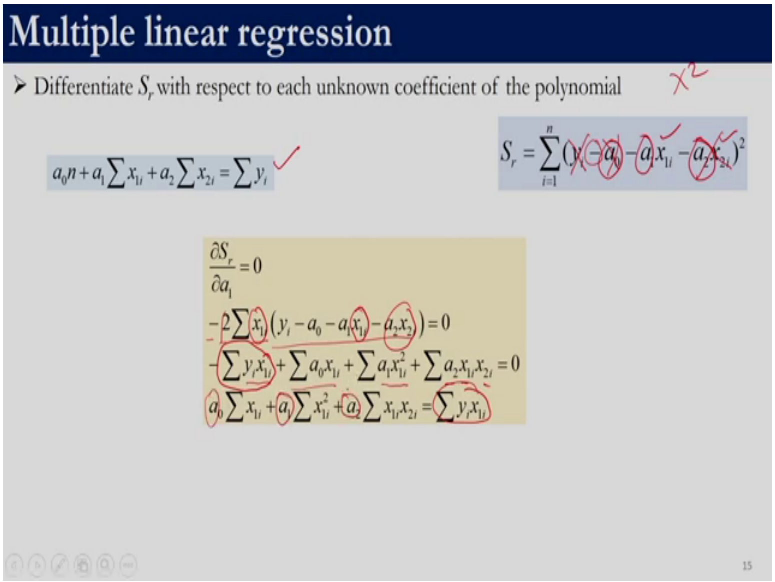


So, for this again we need to apply the stationary condition. So, here we have three unknowns; a naught, a 1, a 2. Remember x 1 i and x 2 i are known right, so those data points are known. So, we need to differentiate this with respect to a naught a 1 and a 2 right. So, when we differentiate this with respect to a naught right, so it is 2 times this expression right, 2 times the same thing x square differentiation of x square is 2 x d x right. And then the differentiation of a naught with respect to a naught will give us 1, but since we have this minus sign over here we will get minus 1 that is why this minus is over here.

Now, if we expand this is minus summation of y i plus a naught, because of this minus and this minus this will become plus. Similarly plus a 1 x 1 i summation plus summation of a 2 x 2 i right. So, since, so the summation of a naught is n a naught and then a 1 and a 2 can be moved out of the summation and since y i is known the data is known. So, summation y i can be calculated. So, we take it to the right hand side.

So, this will be a constant term, it does not have any coefficient associated with it, so we take it to the right hand side similar to what we did earlier right. So, now, this is equation linear equation right, this term is known, summation x 1 i is known, n is known and summation x 2 i is known right. So, this equation if we see it is a linear equation in a naught a 1 and a 2. Those values are not known.

(Refer Slide Time: 09:31)

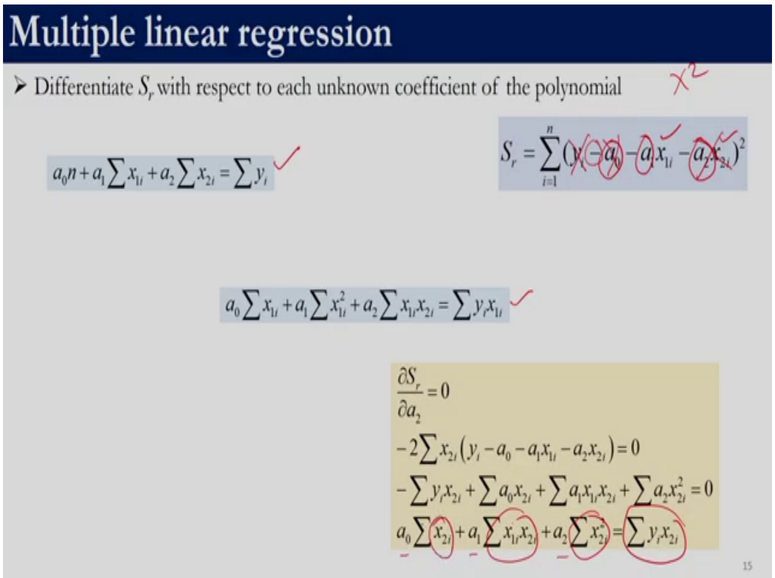



So, similarly we need to do dou S r by dou a one. So, whatever equation we have derived is given over here right. So, let us calculate dou S r by dou a 1 equal to 0 right. So, we apply this stationary point condition right. So, in this case again 2 times this value. So, this two and this should be evident and then we need to differentiate with respect to a naught right. So, differentiation of y i will be 0, differentiation with respect to a naught would be 0, differentiation with a 2 will be 0. So, differentiation with respect to a 1 will give us the coefficient minus x 1 i.

So, that is why we have this x 1 i and this minus over here right. Similarly 2 can be eliminated because the right hand side is 0 and then we expand this right. So, here we have minus summation of y y i x 1 i plus a naught x 1 i plus a 1 x 1 i square, because this is x 1 i and this is also x 1 i right, so square plus summation of a 2 x 2 i a 2 x 2 i and x x 1 i right. So, in this equation if we see, since x 1 is completely known and y is completely known this term can be calculated right and it does not involve any coefficient. So, this can be taken to the right hand side. So, that is why we, that is what we have here.

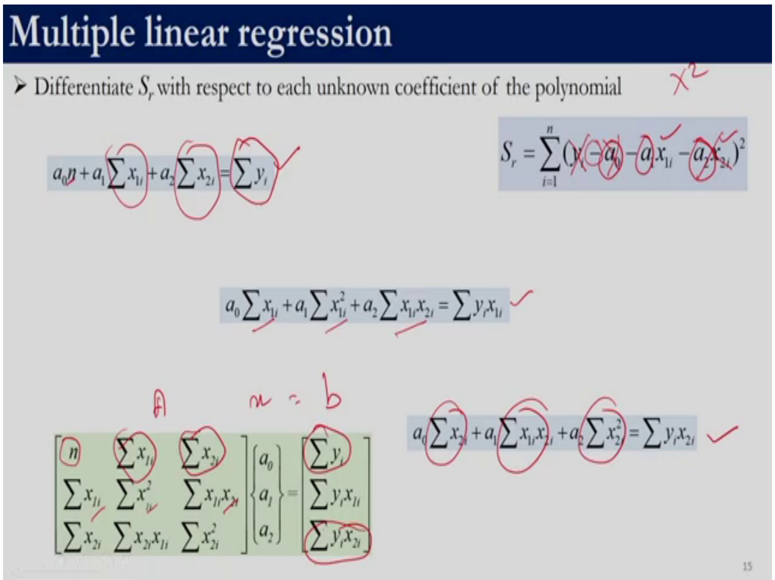
And similarly a naught can be taken outside over here summation of x 1 i plus a 1 which is which is unknown a 1 summation of x 1 i the whole square plus a 2 summation of x 1 i x 2 i. So, this equation is again a linear equation in a naught a 1 a 2 and a naught a 1 a 2 are unknown. So, that is our second equation.

(Refer Slide Time: 11:25)




Similarly, we can dou dou S r by dou a 2 equal to 0 right I leave the, I leave the differentiation to you right. So, here also we will get equation which involves a naught a 1 a 2 they are linear. The right hand side is summation of y i x 2 i right and all these coefficients can be calculated. Since x 1 x 2 is known this things can be calculated right.

(Refer Slide Time: 11:53)

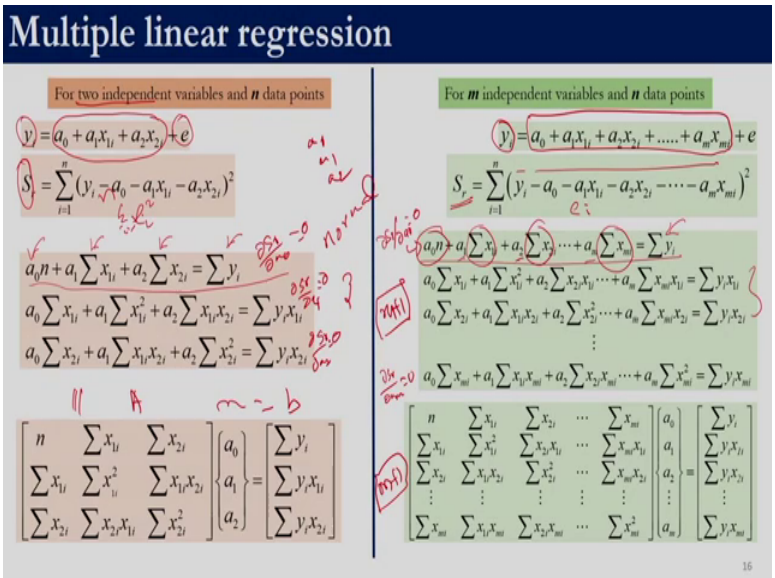



So, now we have three equations in three unknowns, the unknowns are a naught a 1 a 2. All the other terms can be determined from the data itself right. So, these three equations, since they are linear equations we can put them in the conventional form right. So, the first equation n is known, so it comes into the coefficient matrix right. This is our coefficient matrix, this is the x vector and this is going to be the right hand side vector.

So, A x equal to b. So, the coefficient of x 1 is sigma x 1 this one the coefficient of a 2 is sigma x 2 i, so, that is over here right. So, a naught a 1 a 2 and the right hand side is summation of y i. So, that is over here. So, similarly the other two equations we can write the coefficient of a naught is this, coefficient of a 1 is this coefficient of a 2 is this. So, those things will go into the coefficient matrix and similarly the third equation. These three values can go into the coefficient matrix and this right hand side is given over here right.

So, just like in the last linear regression we had a non we had a non-linear optimization problem right, because it involved the squire term in the objective function that when we applied stationary condition, last time we got two equations in two unknowns. In this case multi linear regression we did the same thing, we applied the stationary condition. When we apply the stationary condition since there are three unknowns a naught, a 1, a 2 we get three equations in three unknowns. All the equations are linear which can again be put in the standard format A x equal to b.

(Refer Slide Time: 13:37)




So, this part shows what we have discussed so far right. So, we had, this was our model, this is the measured value, this is the error. Our objective function was S r is equal to sum of square of error, so this was e i, so square and then summation right. And then we applied the three stationary conditions dou S r by do a naught equal to 0, dou S r by do a 1 is equal to 0 and dou S r by dou a 2 equal to 0 and we got these three equations, three linear equations in three unknowns which can be put in this conventional form A x equal to b.

Let us assume that if there were more independent variables, instead of two independent variables if there had been m independent variables right. Let us say this was our model a naught plus a 1 x 1 i plus a 2 x 2 i plus all the other terms till a m x m, a m x m i plus an error. Remember this is the model right, this is the observed data point or the measured data point. So, there may be some mismatch between what we have observed and what our model is capturing. So, we have this error.

So, again this is the same as sum of square of errors. This is the model the model part, this is the measured value y, so the error with respect to each data point squaring it up and summing it up, so this is our S r. So, over here if we are to apply stationary condition we need to do dou S r by dou a naught equal to 0, dou S r by dou a 1 equal to 0, dou S r by dou a 2 is equal to 0 all the way up to dou S r by dou a m equal to 0 right.

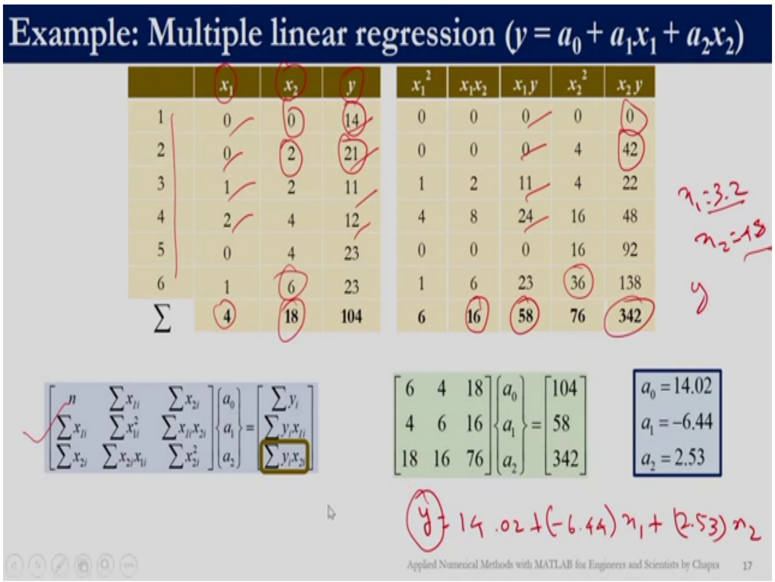
So, if we do that. So, for example, if we do with respect to a naught, so this is dou S r by dou a naught equal to 0 this equation corresponds to that. So, it will be a naught into n plus a 1 summation of x 1 i plus a 2 summation of x 2 i all the way up to a m summation of x m i equal to sigma y i. So, you can see the analogy between this equation and this equation right. So, this equation it said like the first term is n into a naught, the second term is summation of first independent variable, the second one is summation of the second independent variable and the right hand side is summation of the dependent variable.

So, over here also the right minus hand side is summation of the dependent variable n into a naught is the same right. And this summation of the, all these summations are the sum of individual independent variable and multiplied by their corresponding coefficient right. So, similarly you can do dou S r by dou a 1 equal to 0, dou S r by dou a 2 is equal to 0 and you will be able to get these two expressions and you can do all the way up to dou S r by dou a m equal to 0 right.

So, in this case you will get a naught summation of x m i plus a 1 summation of x 1 i x m i plus a 2 summation of x 2 i x m i plus, I mean the other terms plus a m summation of x m i the whole square is equal to summation of y i x m i right. So, this equation again these set of equations, so here we will have m equations right. So, here we had three terms, three unknowns a naught a 1 a 2 right, so here we will get three coefficients. Here we have m plus 1 coefficients right m plus 1, because we are starting with a naught right a 1 to a m are m coefficients and then we have this a naught. So, we will have m plus 1 coefficients. So, here also we will have m plus 1 equations.

So, just likely form this, just likely put this normal equations right into this matrix form. These normal equations can also be put into this matrix form. So, now, if you see there is actually a symmetry over here right. So, if we know how to do for two independent variables we can do it for m independent variables right.

(Refer Slide Time: 17:45)




So, now let us look at an example. So, in this case we have been given 6 data points right, we have been given x 1 and x two. So, x 1 is one dependent variable, x 1 is one independent variable, x 2 is the other independent variable and y is the dependent variable. So, these values are given and we have established this.

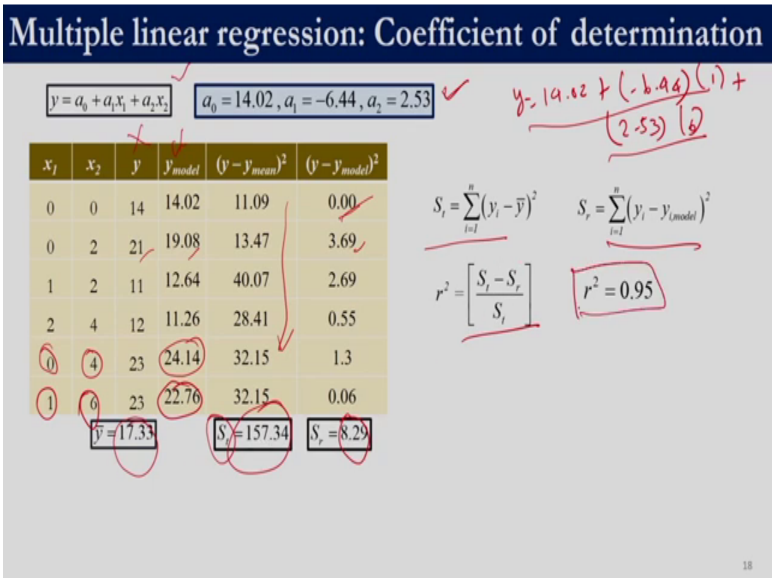
Now, our task is to find out these individual values and then we need to solve it right. So, in this case n is 6. So, summation of x 1 is 4, summation of x 2 is 18, summation of y is 1 naught 4, summation of x 1 square. Remember again it is not 4 square right, it is each element has to be squared and then their sum has to be taken right. So, it is 6 summation of x 1 i x 2 i is x 1 into x 2 0 into 0 0 into 2 1 into 2 2 into 4 0 into 4 1 into 6 right and then we need to sum it up right.

So, the 16 will go in these two places and then we have this y i x 1 i, so that is 14 into 0 is 0, 21 into 0 is 0, 11 into 1 11, 12 into 24 12 into 2 24 and similarly you can calculate. So, that summation happens to be 58 and then we will require x 2 square over here. So, each element of x 2 has to be squared. So, we have 6 over here, so 6 square is 36 and then we need to sum this. Remember again it is not 18 square we need to sum this vector right, so that is 76. Similarly we will have to calculate x 2 into y. So, in that case it is 14 into 0, so 0, 21 into 2 42 and similarly we can calculate the other values and that summation would be 342.

And so now, we have all the values, if we plug them this is these are our three equations in three unknowns; three linear equations in three unknowns. So, if we solve this we get these coefficients; a naught is equal to 14.02, a 1 is equal to minus 6.44 and a 2 is equal to 2.53. So, our model is y is equal to a naught. So, 14.02 plus a 1 which is minus 6.44 x 1 plus a 2 x 2 is 2.53 into x 2 right.

So, given any other value, so for example, if you say at x 1 is equal to 3.2 and x 2 is equal to let us say 4.8 right. We do not know the value of y from this data set right, but if we can plug these values 3.2 and 4.8 into x 1 x 2 and we will be able to calculate y. So, that is the benefit of having a model. Now, that we have calculated the model right we can also use the same concept coefficient of determination which we discussed earlier for multiple linear regression right, its valid over here also.

(Refer Slide Time: 20:57)




| $x_1$ | $x_2$ | $y$ | $y_{model}$ | $(y - y_{mean})^2$ | $(y - y_{model})^2$ |
|---|---|---|---|---|---|
| 0 | 0 | 14 | 14.02 | 11.09 | 0.00 |
| 0 | 2 | 21 | 19.08 | 13.47 | 3.69 |
| 1 | 2 | 11 | 12.64 | 40.07 | 2.69 |
| 2 | 4 | 12 | 11.26 | 28.41 | 0.55 |
| 0 | 4 | 23 | 24.14 | 32.15 | 1.3 |
| 1 | 6 | 23 | 22.76 | 32.15 | 0.06 |

So, these three definitions we already know right. We know the model over here ah, the model coefficients over here and the model is given over here. So, now, from the model from these values we can calculate these values. So, for example, this 24.14 is y is equal to a naught is 14.02 plus a 1 is minus 6.44 into x one; x 1 in this case happens to be 1 plus a 2 that is 2.53 into x two. So, x 2 is 6.

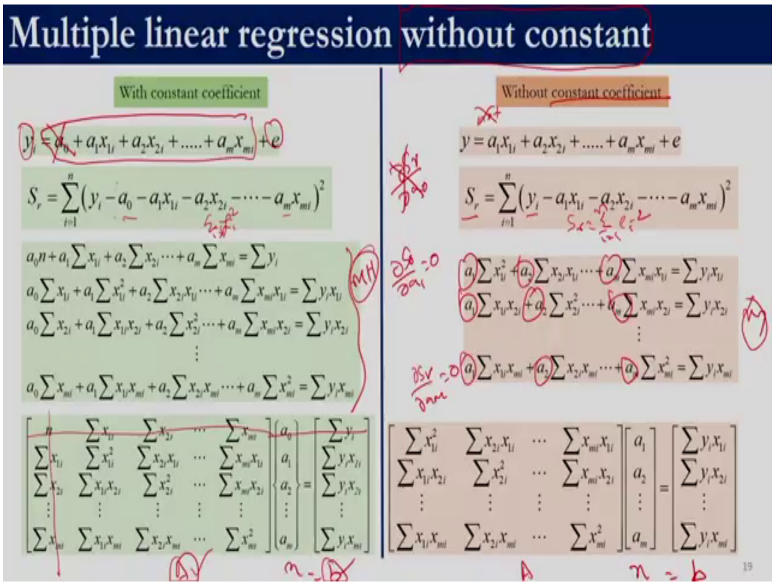
So, if we calculate these values we will get 22.76. So, for 24.14 we should have used x 1 is equal to 0 and x 2 is equal to 4 in this equation and we would have obtained y model right. So, for determining this y model we will not require this y right. So, this is the value predicted by the model right. So, the error we can find out right. So, the mean of y in this case is 17.33. So, y minus y mean.

So, again 14 minus 17.33 the whole square will be 11.09, 21 minus 17.33 the whole square is 13.47 and this can be calculated, and then if we sum this is nothing, but the definition of S t. Similarly we can calculate model right. So, here since we are showing you only two decimal you do not see a value over here, but there is a difference over here right; 14 minus 14.02 the whole square, 21 minus 19.08 the whole square and similarly right. So, this happens to be 8.29.

So, here itself we can say if we had considered the model to be nothing, but the mean this is the error and over here if we consider the model with these coefficients then this is the residual error. So, we can calculate r square over here. So, if we plug in these values, S t value S r values in this we get our square of 0.95. So, far what we consider is like, so multiple linear multiple linear regression if we are to summarize, we started with two independent variables, then we extended it for m independent variables right.

In both the cases we had constant a naught right. The first time we had a naught plus a 1 x 1 plus a 2 x two, the second time we had a naught plus a 1 x 1 plus a 2 x 2 plus a 3 x 3 all the way up to a m x m. And then we looked into an example as to how to exactly calculate the coefficient values and the coefficient of determination right. Now, we will see like what if the model did not have a constant coefficient. So, model is not y is equal to a naught plus a 1 x 1 plus a 2 x two, but my model is y is equal to a 1 x 1 plus a 2 x 2 right.

(Refer Slide Time: 23:53)




So, let us look into that. So, this is what we have previously derived right. So, this was our model, this is our model right, this is what is the measured value. So, the error is e and then we said y minus the model right that is error right with respect to the i th points square it up and sum it up.

So, that is nothing, but sum of square of residuals and we had this n plus 1 equations due to the stationary condition. There were there are n plus 1 unknowns, we applied the stationary conditions and determined this n plus 1 equations right. And then we just plug and then just we put it in the conventional format A x equal to b, because all these equations are linear right.

So, this is the coefficient matrix given a data set this can be completely determined, this can be completely determined. So, the only unknown as this x or this coefficients of the model.

So, that can be calculated if we know how to solve linear equations right. So, now, the cases what happens if there is a no constant? So, if there is no constant, let us say if the model does not have this a naught right, everything remains the same right. So, this a naught is not there; so now, we have. So, the definition of S r is still the same right y i minus the model without a naught right.

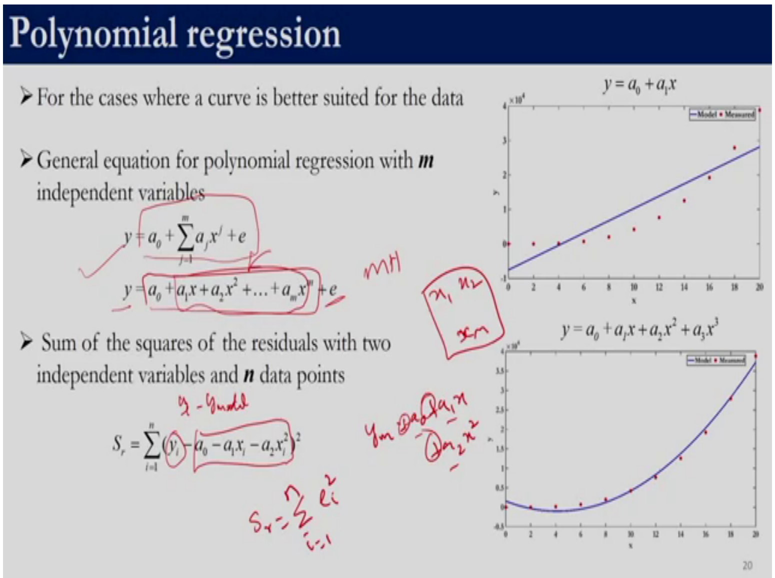
So, this a naught is what is naught here, so without constant coefficient right. So, y minus, this model y i minus the value obtained from the model e i square it up and sum it for all the n data points right. So, that is the same thing right. In this case now we need to differentiate only m times; dou S r dou a 1 dou S r by dou a 2 dou S r by dou a 3 all the way up to dou S r by dou a m and equate it to 0. There is no dou S r by dou a naught right, this cannot be determined because the model does not have a naught right. So, now, we have m equations right.

So, similarly if we put them in the conventional format, all these equations would be linear nothing changes right. All this equation are linearly with respect to a 1 a 2 and the coefficient model coefficients all the way up to a m. So, this is again in that format A x equal to b right. So, here we had m plus 1 equation, here we will have m linear equation, here it was m plus 1 linear equation, here it is m linear equation right.

So, here if we see it is nothing, but whatever we derived for the constant coefficient except for this row and for this column right. So, all the other terms would be the same right. So, that is how we can still work with without constant coefficient right. If the model does not have a constant coefficient still we can apply the same concepts which we have discussed so far to fit those kinds of model. So, that was multiple linear regression right.

Now, we will move on to polynomial regression, where we still have only one, wherein we have one independent variable and one dependent variable, but the model is polynomial with respect to the independent variable right.

(Refer Slide Time: 27:15)




So, for example, if the data set is like this, we might choose to fit a polynomial model let us say a cubic polynomial model rather than a straight line. So, remember in regression you need to know the data points right and you need to know the model right. If you do not know the model, there is you cannot even apply regression. So, the nature of the model should be known right only the coefficients are unknown.

So, for example, if you give them some data points and ask a model to be fit to it. The first question is what is the type of model right? Whether you want a linear model or a non-linear model. So, that has to be first fixed, only when you fix the model can we attempt to determine the coefficients of the model. So, regression requires the data points as well as a model.

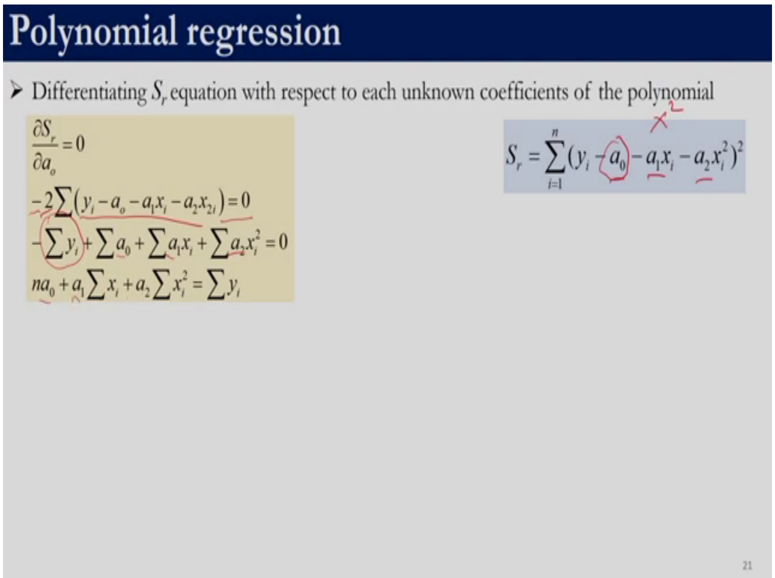
Now, if you do not know the model you might choose to fit 2 or 3 different models and see whichever model best represents your data you might choose to take that as the final model, but to apply regression you require a model right. So, this is the model right. So, a naught plus a 1 x plus a 2 x square all the way it can go up to a m a m x power m. So, here also there are n plus 1 unknowns right. This is the measured value or the observed value, this is what is from the model right. So, there could be a difference between the measured value and the model, so that is the error right.

So, this is the compactly it can be written like this right; a naught plus sigma j is equal to 1 to m because there are this m terms a j x power j plus the error right. So, this is the generic polynomial regression right, but first we will work with y model is equal to a naught plus a 1 x plus a 2 x square, similar to multiple regression wherein we started with just x 1 and x 2 and then extended to x m. We will do the same thing over here that we will start with just a naught a 1 and a 2 and then we extend it all the way up to a m right.

So, by now you should be familiar with this right. So, this is y is the observed value, this is what we get from the model right. The minus sign is because y i minus y model right. So, all this positive over here, all these positives over here would become negative right. So, the error is, the observed data point minus model the whole square right, so the error square. So, what we are going to do is the same thing that we are going to minimize the sum of square of errors i is equal to 1 to n e i square right. So, error in this case is y i minus this model.

So, for the sake of completeness we will do it, but otherwise you should be able to do it by yourself.

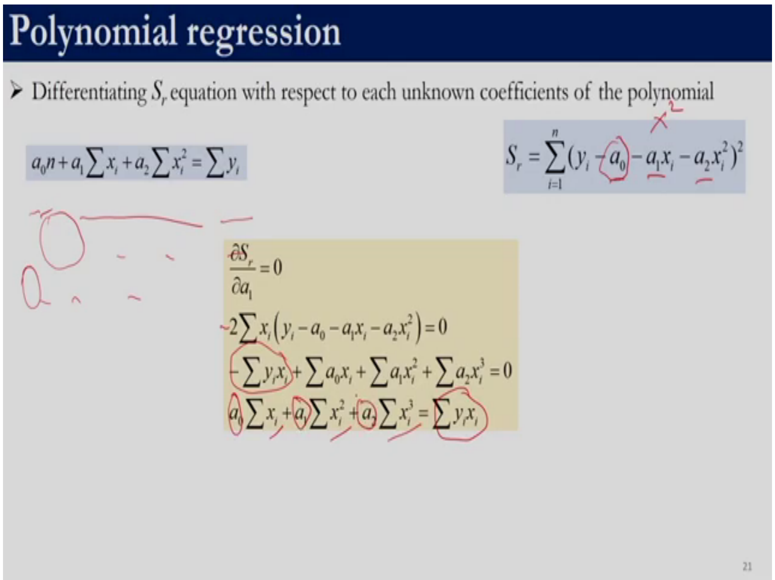
(Refer Slide Time: 30:19)




So, its just that now we need to again apply the stationary points right dou S r by dou a naught equal to 0, dou S r by dou a 1 is equal to 0 and dou S r by dou a 2 is equal to 0 right. So, dou S r by dou a 1 equal to 0 the same thing 2 times this x, so x square 2 times this expression right and when we differentiate a naught minus a naught with respect to a naught we get a minus sign. So, that is why this minus sign is over here, so that has to be equated to 0.

And then again do the usual rearrangement, here we again have the term sigma y i right which is completely known that can be taken to the right hand side; otherwise these a naught a 1 a 2 are unknown right all the other terms are known, n is the total number of data points.

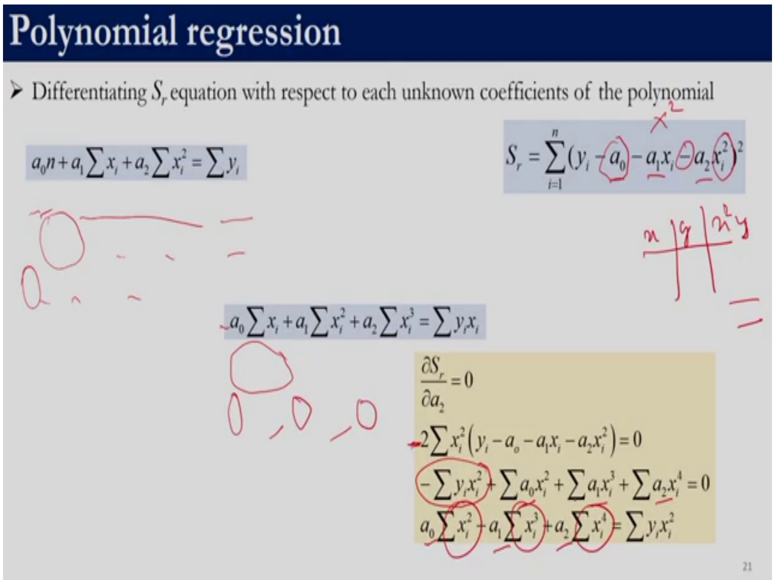
(Refer Slide Time: 31:07)




So, similarly we need to calculate dou S r by dou a 1 is equal to 0. So, 2 times this entire expression. So, differentiation of y i with respect to a 1 would be 0, a naught would be 0 over here we will get a minus x i right. So, that minus can be written over here.

So, minus 2 times x i this expression and then x i can be multiplied and this term if you see it can be taken to the right hand side, because it is completely known. Again this term, this term and this term are known because the data points are known, so we can calculate it. The only three unknowns are a naught a 1 and a 2 right. So, that this is the second equation.

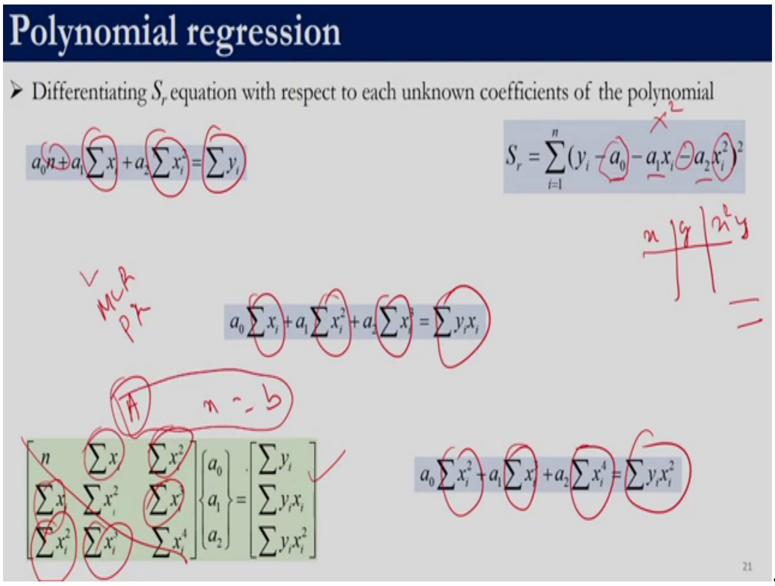
(Refer Slide Time: 31:45)




And the third equation is dou S r by dou a 2 is equal to 0, again the same concept we need to apply right. So, x i square, since it is differentiation with respect to a 2. This i square is a constant and because of this minus sign this minus sign appears over here right. And then over here this term would be completely known, because you have x and y right. So, you can calculate x square y.

So, square each element of x multiply it with y and then sum it up over here right. So, this term is known, so that can be taken to the right hand side otherwise again a naught a 1 a 2 are the unknowns, a naught a 1 a 2 are the unknowns and these coefficients can be determined.

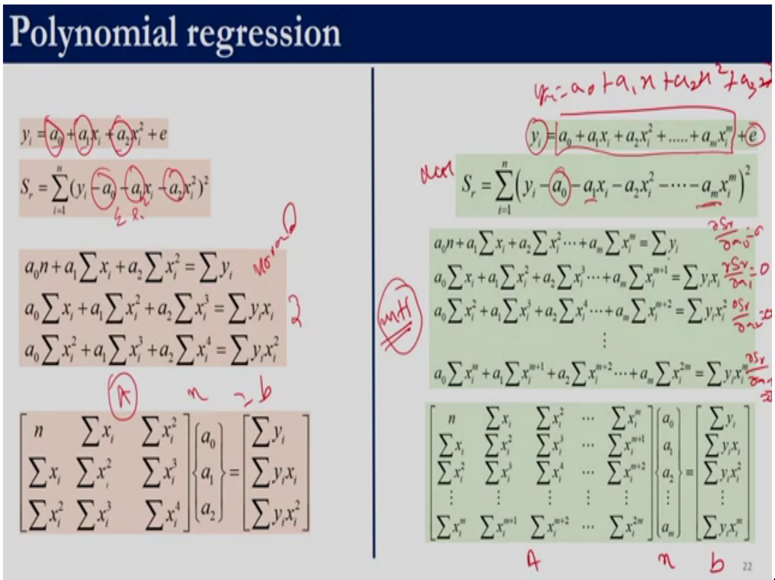
(Refer Slide Time: 32:35)




So, these are the three normal equations in this case right. Again they can be put into conventional A x equal to b form right. So, these three things are completely known. So, that will form our b vector and this n sigma x i sigma x i square are known. All these terms are known, so they would come in the coefficient matrix right. So, if we know how to solve this then we can know given data points if we know how to solve this we can find out a naught a 1 a 2 right. So, in all the three cases right; in linear regression, multiple linear regression and polynomial regression.

If you see the coefficient matrix it would be symmetric right, so sigma x i sigma x i sigma x i square sigma x i square sigma x i cube sigma x i cube right. So, this is the diagonal, so it will be symmetric and this matrix will also be positive definite and there are better methods to solve this ax equal to b efficient methods to solve this A x equal to b.

(Refer Slide Time: 33:41)



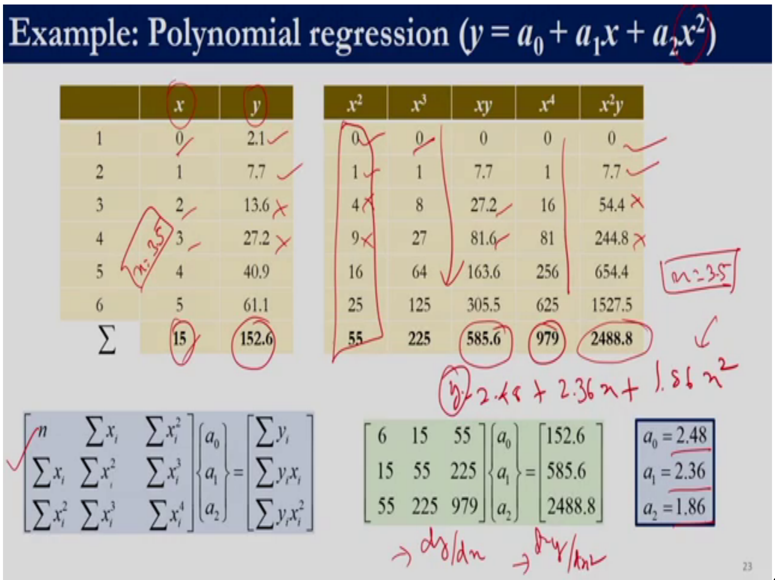

So, what we currently saw was a naught plus a 1 x i plus a 2 x i square. So, we restricted to only the quadratic term right. So, this is nothing, but the error square sum of square of errors, and these were the three equations we obtained because there are three coefficients which we do not know a naught a 1 a 2.

So, these were our normal equations right and that can be put in this A x equal to b form, and this a is again positive symmetric positive definite. So, what if we did not, what if we had more terms right. So, let us say a 1 x 1 plus a 2. For example, if we had y is equal to a naught plus a 1 x plus a 2 x square plus a 3 x cube all the way up to a m x m right, so this is model right. So, if this is the observed value then we also have the error.

So, what if our model is this one. So, conceptually everything remains the same its just the math that we will have to do right. So, in this case it is y minus y model right, so that is the error error square. So, again we need to minimize this, to minimize this we need to apply the stationary conditions right. So, the stationary conditions, here again it would be m plus 1 equations right, because remember we have m constant coefficients from 1 to m and then we also have this a naught right.

So, right now we are not talking about without constant coefficient we are only talking about higher order polynomial terms being present in the model. So, we will have we can derive this n plus 1 equations right. So, its the same thing dou S r by dou a naught equal to 0, dou S r by dou a 1 equal to 0, dou S r by dou a 2 equal to 0 all the way up to dou S r by dou m equal to 0. So, these are our, this will be our m plus 1 normal equation again they can be put in the standard A x equal to b form, because all this m plus 1 equations are linear.

(Refer Slide Time: 35:53)

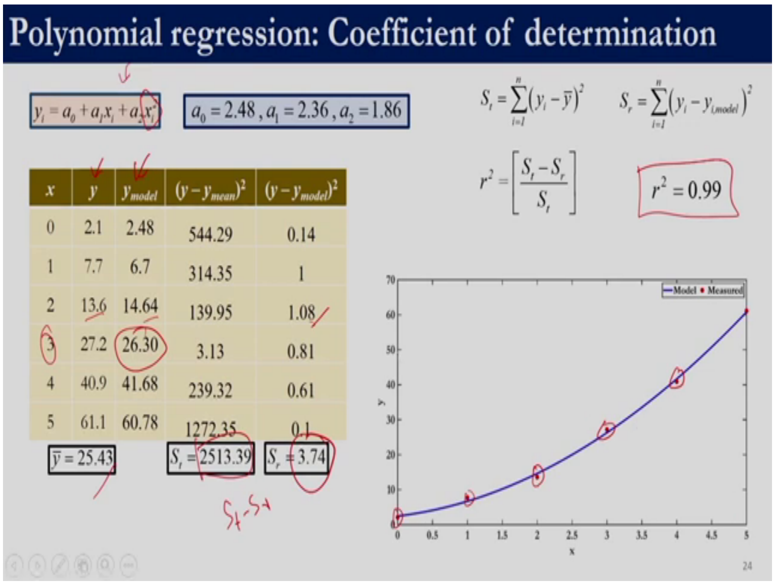


So, now let us look at an. Now, let us look at an example for polynomial regression right. So, these are our data points, we have 6 data points right and this is our these are our three normal equations. So, the model involves only till the square term y is equal to a naught plus a 1 x plus a 2 x square. So, the number of data points is 6, summation of x is required which is 15 summation of x square is required, again it is not 15 into 15 right, it is each element has to be squared and then it has to be summed up. So, that will be required in these three places.

We require summation of y which is 152.6, then we will require summation of x cube. We need to calculate this and the summation of it is 225 and then we will require y into x. So, y into x is again not 15 into 152.6 right, it is 0 into 2.1, 1 into 7.7, 2 into 13.6, 3 into 27.2.2 and then this has to be summed up right. So, that comes out to be 585.6 6 x power 4 again this to the power 4 and this summation happens to be 979 which will be required over here. Then we require x square, so x square is over here y is over here. So, 0 into 2.10, 7.7 into 1, 7.7 136.0 into 4, 54.4, 27.2 into 9, 244.8 and then we need to sum this vector. So, that will be 2488.8.

So, if we plug in those values and solve for these three equations in three unknowns we will get 2.48, 2.36 and 1.86. So, our model is y is equal to 2.48 plus 2.36 x plus 1.86 x square right 4 x is equal to 3.5 we do not have the value of y so, but that can be plugged over here x is equal to 3.5 can be plugged into this expression and we can find out what is the value of y right and if we are interested we can also find out d y by d x. So, that will change that will tell us the behavior of y with respect to change in x and we can also calculate d square y by d x square and if there are any influential. So, for example, this x plus y is going to be some physical variable right. So, these can give additional insights. So, once we have this model we can calculate all, we can calculate these values d y by d x and d square y by d x square.

(Refer Slide Time: 38:37)




So, the coefficient of determination is the same concept right. So, we have this observed value, we know the model value right ah, we can calculate the model value because we know the constant coefficient right.

So, for example, this 26.30 is going to be 2.48 plus 2.36 into 3 plus 1.86 into 3 square, because of this model right. So, that would be 26.30. So, that is how we calculate y model and then we know the mean of y can be calculated which is 25.43. So, this terms can be calculated y minus y mean the whole square can be calculated. So, that works out to be 2 5 1 3.39, similarly y minus y model right.

Y minus y modulus. So, for example, 13.6 minus 14.64 the whole square would be 1.08 and the summation would be 3.74 from here itself we can see that S t minus S r is significant right. So, we have, so, the model coefficients which we determined r actually better than considering the mean itself as model. So, r square in this case should turn out to be very good right.

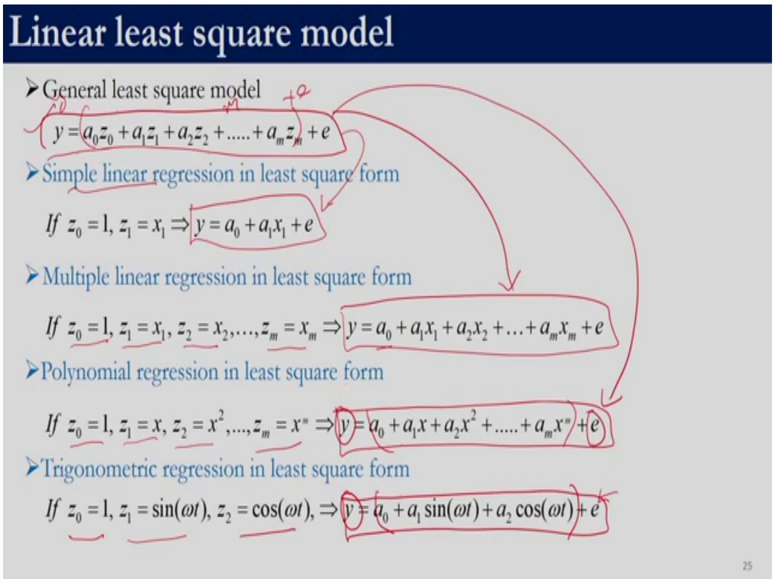
So, if we plug in these values in this expression S t minus S r by S t we get an r square as 0.99 right. So, we have these data points right and now we have this coefficient. So, between 0 and 5 we can generate let us say1000 2000 points and we can plug in the model, because the model is fully known once a naught a 1 a 2 is known ah, if you plug various values of x we get various values of y with respect to that we can actually plot the entire model in this range right. So, given any value of x between 0 and 5 we will be able to predict the value of y. So, that is the use of regression.

So, to consolidate whatever we have seen so far right what we have, what we need to do is define the objective function right. In this case the objective function was sum of square of errors right or sum of square of residuals which is nothing, but the error or residual is nothing, but the difference between the observed value and what the model would predict right, so that has to be minimized. So, in order to minimize that we apply the stationary condition.

So, in linear regression in simple linear regression wherein there were only two coefficients; a naught and a 1 because the model was a naught plus a 1 x. We found out dou S r by dou a naught, we equated dou S r by dou a naught and do a S r by do a 1 to be 0 equated it to 0. So, we got the normal equations and the normal equations were linear. We had to two unknown coefficients a naught and a 1 and two linear equations, so we were able to find them out. The same concept we applied in multiple linear regression and polynomial regression.

We frame the objective function which is nothing, but minimization of sum of square of error and then we applied the condition for stationary we applied the, we determined the stationary points and that led us to the number of that led us to the normal equations right. So, in all the three cases whether it is with constant coefficient or without constant coefficient we ended up with set of simultaneous linear equations right. So, that was the whole concept right. We kept applying it multiple times to make sure you get the concept. All these three cases linear simple linear regression, multiple linear regression and polynomial regression can be fit into something called as general linear general least square regression. So, now, we will look into that right.

(Refer Slide Time: 42:27)



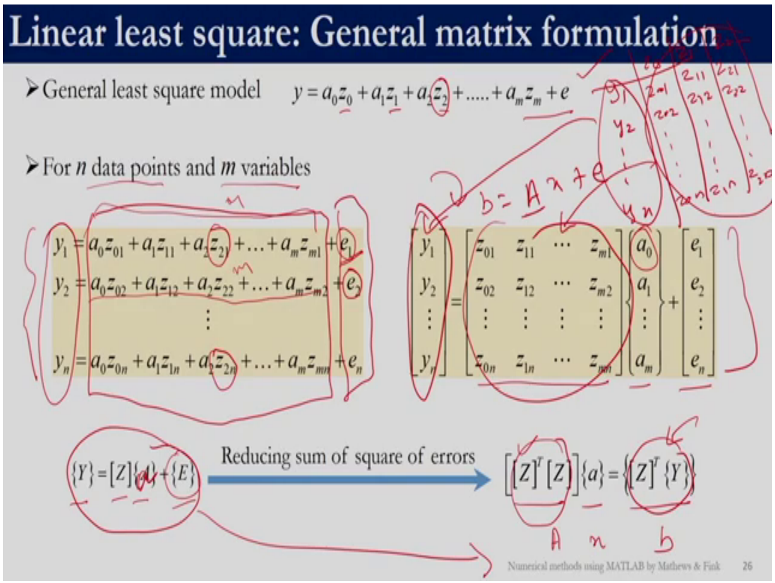

So, for example, all those three models are a subset of this particular model y is equal to a naught z naught plus a 1 z 1 plus a 2 z 2 all the way up to a m z m plus e right. So, this part constitutes the model right. So, the value predicted by the model plus error would be the measured value. So, if this for this to fit into linear simple linear regression, this z naught is 1 z 1 is x 1 right. So, this z 2 onwards z 2 to z m 0. So, in that case this reduces to simple linear regression right. For multiple linear regression z naught is again 1, because we have this we want this constant coefficient; z 1 is x 1 z 2 is x 2 and similarly z m is x m right. So, this is what we solved.

So, if we can solve this one, if we have a generic expression for this one that would be valid for this one also right. So, under the condition z naught is 1 z 1 is x 1 z 2 is x 2 and z m is x m all the way up to z m is equal to x m, and similarly polynomial regression can also be represented by this one with z naught equal to 1 z 1 is equal to x right and then we require x square over here, so z 2 is equal to x square all the way up to z m is equal to x m. So, this is the model which we work with right. This was the model part and this was the error right and this is the observed value.

So, all these three cases which we saw is actually a subset of this general linear least square. So, even it can capture even we can fit models like this right y is equal to a naught plus a 1 sine omega t plus a 2 cos omega t plus error, again this is the model part, this is the error part, this is the observed value right. So, here for this model to be represented by this z naught has to be 1 right, because here we do not have any coefficient, so it has to be 1, z 1 is sine omega t right and z 2 is cos omega t right. If I take z 2 as cos omega t this model this general model boils down to this model.

So, what I am trying to tell you is that all these four case all these three cases which we which we independently discussed right or cases like this can be solved if we are able to solve this one right. So, this is the. Once we are able to solve this one the all the cases can be deduced from that. So, let us see.

(Refer Slide Time: 45:17)




So, this is that general model which we discussed in the previous slide right. So, if n data points are given and if there are m variables right. Let us say we will have, let us say for this the data point is going to be y 1 y 2 all the way up to y n and then we are going to have. Here if you see z naught z 1 z 2 all the way up to z m are the m plus m independent variables right. So, here we will have z naught. So, this is, so this is going to be z naught 1 z naught 2 all the way up to z naught n, then we are going to have z 1 right.

So, this is that 1 this is z 1 2 all the way up to z 1 n and then let us say z two. So, z 2 is the variable, the first point of the z 2 variable the second point of the z 2 variable and the n point of the z 2 variable, similarly we can extend it to m variables. So, this is given either this is given or this can be obtained from whatever data is given right. So, these are our equation, so these are our data points right. So, if we are to apply this model to each of this point.

So, these are the n equations which we are actually specifying, so y 1 is equal to a naught z naught 1 plus a 1 z 1 1 plus a 2 z 2 1 all the way up to a m z m 1 right, so this is again the model part and there may be some error associated with the model for especially for the first point, so that we indicate e 1 right. And for the second point if we is applied the model right this is the value obtained from the model, so that is, and the error need not be the same as the error for 0.1.

So, remember the first slide we are in the error associated with each point was different right. So, this that is why this error is being separately written for each of the n points right. So, these are the observed values. This is what is we are getting from the model for each of the value by substituting the point right.

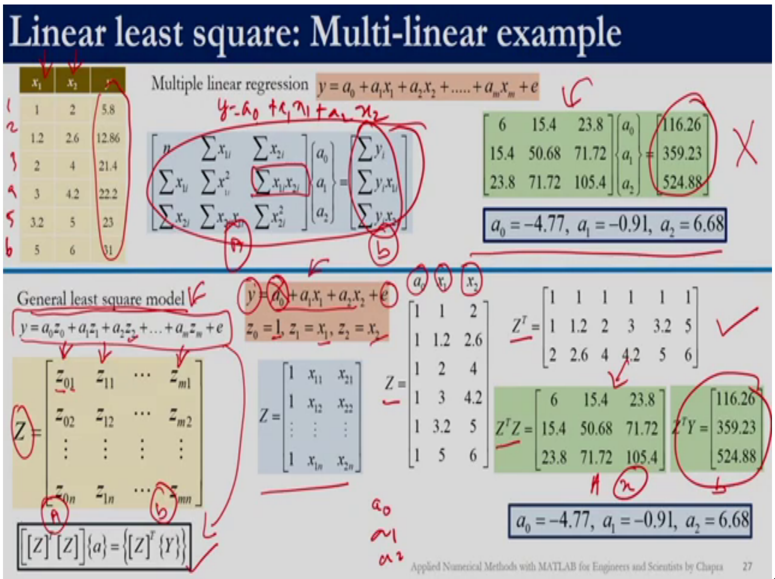
So, z 2 is the variable when we set z 2 1; that means, we have substitute that the first point, when we set z 2 n; that means, we have substituted the nth point in the model right. So, this is the model part and this is the error vector right. So, these n equations can be written in this matrix form b is equal to A x plus e right. So, a from this one if you see it is nothing, but this coefficient matrix right which is nothing, but this data points, so that is what is a right. And this y vector is nothing, but the observed points right these are the unknown coefficients m plus 1 coefficients because we are starting with a naught plus we have this error vector right.

So, this can be compactly written as the vector y, the Z matrix, the vector a plus the error vector right. So, this a is lower case a right. So, this set of equations can be represented by this right. So, this is what is our; this is what is our regression; this is what is our problem right. So, now, our job is to find out this a from this right and we do not know the error. So, this is a analytical solution right, this can be solved and the solution to this is Z transpose Z into a is equal to Z transpose Y. So, if you are interested in how did this derivation happen, how did we get this analytical expression you can look into this book by Mathews and Finks right this derivation is given there.

So, for this problem, instead of finding out all the derivatives and stuff if we solve this set of equations. So, this if you see it is linear equations in which Z is known right, this is the Z that is known. So, Z transpose Z is also known, a is what we are trying to find out right and Y is known, so this is Y and Z is again known. So, Z transpose Y can be calculated right and Z transpose Z can be calculated right. So, we directly get this as a this as b x right. So, if we use this analytical expression we do not need to find out the summation of e.

If you remember we solved lot of calculations to actually find the coefficient matrix and the right hand side vector. So, all that can be avoided right if we make use of this analytical expression right. In view of the nature of this course we are not showing you the derivation, but you can have it have a look at the derivation. Now, that we have looked into general linear least squares right. We will let us use the concept which we learnt just now general linear least square to solve the multiple linear regression problem which we have solved earlier right.

(Refer Slide Time: 50:25)



So, this is the problem which we have solved earlier right. So, x 1 is an independent variable x 2 is an independent variable. We have 1 2 3 4 5 6 data points right and this is the model which we were which we had to fit; y is equal to a naught plus a 1 x 1 plus a 2 x 2 right. So, for that if you go back and look at all the derivation and the matrix which we obtained, this was the coefficient matrix, this is the right hand side vector when we substitute these values right.

So, all that we have done previously, so we are not repeating it over here right so, but if you calculate all of this we will get this coefficient matrix and this right hand side vector which can be solved to obtain this coefficient these coefficients right. So, that is what we have done previously right. So, now, that we know general linear least square. For general linear least square this is our model right.

So, in the previous few slides we have seen this is the model and the solution to this is this Z transpose Z into a is equal to Z transpose Y where Z matrix is given by Y, this y vector is nothing, but the dependent variable right. So, that is that is something that we are always going to get from the data points right over here and the Z matrix is z naught, so the variable z naught the first point right.

So, this is this entire column is the variable z naught and the nth points, this entire column is the variable z 1 and the n points similarly this z m is the mth variable all the n points. So, that has to be if we stack that in this form that will be our Z matrix. So, once we have the Z matrix we can just we can merely compute Z transpose Z that will be our coefficient matrix ah. This coefficient matrix will be equal to whatever we obtain over here right and the right hand side vector also same if we do Z transpose Y; Y is a vector column vector which is known to us and Z is this matrix.

So, if we do a transpose Y we will get a column vector that column vector will be the same as this right hand side b vector. So, rather than computing this all this terms. So, for example, rather than computing this will be using directly using this analytical solution to solve the problem which you are currently seeing right. So, the current model that we have is. So, this is our data point right. So, whenever we have data point we write the model and then we write the error term right, but if you are merely writing the model then we do not include the error term right. So, right now it is the data point which is observed equal to model plus some error that might be there right.

So, for this to be equal to this expression, the current problem to be equal to this. We need to take z naught as 1 z 1 as x 1 and z 2 as x 2 right. So, if we take that our current problem is a general linear least square model then our Z matrices a column of once right and the x 1 variable stacked all the all the n points stacked as the second column. The second variable stacked the second variable all the n point stack as the third column right. So, that would that that will be our Z matrix.

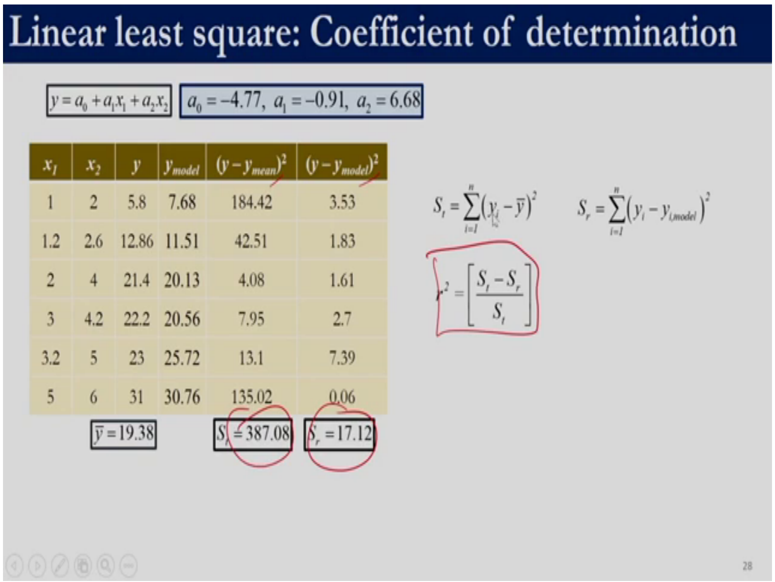
So, here if we apply the same thing to our current problem, so Z matrix in this cases is a column of 1 because we have this constant coefficient, had we not had this constant coefficient we do not need to include this column of ones right and then we need just need to stack these two columns x 1 and x 2 right. So, the way we have stacked this the constant, the first variable and the second variable right.

So, the solution which we will get will also be of the same form that the first value that we get will be a naught, the second value will be a 1 and then we will get the final coefficient a 2. So, if this is Z; obviously, this will be Z transpose. So, if you multiply Z transpose Z we will get a 3 cross 3 matrix and that 3 cross 3 matrix will be exactly similar to what we have got here right. So, these two matrix are similar.

So, without computing all these terms over here right we can directly get by doing Z transpose Z right. Similarly the right hand side vector instead of computing this right we can merely do Z transpose Y right. So, Y is the set of values which we have got for the dependent variable. So, if we do Z transpose Y again we will get a 3 rows one column a column vector and this value will be similar to what we have over here. So, since the coefficient matrix a and the right hand side vector b are identical our solution x will also be identical right.

So, instead of employing this approach right we can directly employ this approach to solve linear least square problem. Again it is one and the same thing, its just that this might be a little bit more convenient.

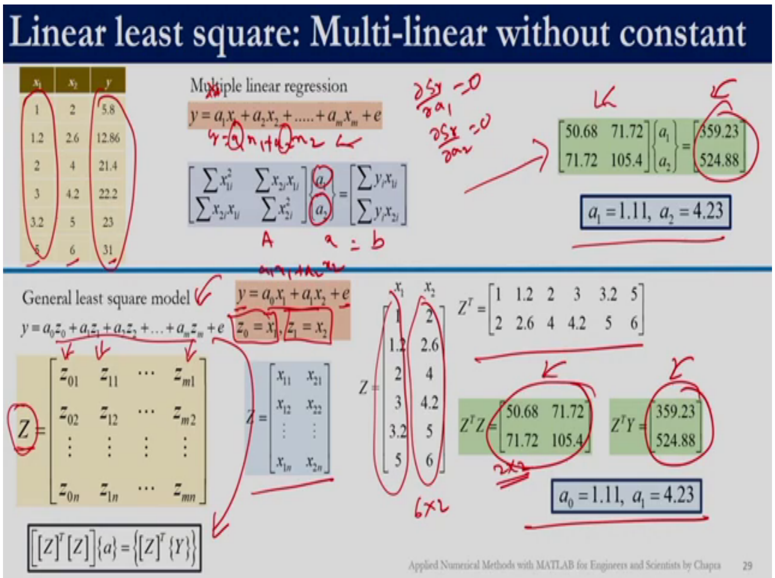
(Refer Slide Time: 55:31)




So, once we do that we can calculate the r square right. Once modulus known then its the same thing y minus y mean the whole square y minus y model the whole square and then sum it up right and then plug it into this expression that will give us the r square value. So, now, let us look into a problem right which we have previously solved without constant coefficient. Remember previously we had fit a model where there was constant coefficient and we also had fit a model where there was no constant coefficient right.

So, with constant coefficient we have seen how we can apply general linear least squares right. So, without constant coefficient it is going to be similar right. For the sake of completeness we will just show you the calculation.

(Refer Slide Time: 56:15)




This is the problem which we had previously right, so x 1 x 2 y. This time we would know we have a model variant there is no constant coefficient, there is no a naught right, there is no constant coefficient right. So, in this case our model was y is equal to a 1 x 1 plus a 2 x 2 right. Our task was to find out a 1 and a 2, the values of a and a 2 such that this model best fits this data point right.

So, if you do dou S r by dou a 1 dou S r by dou a 2 and equate to them to 0 you will get two linear equations into 2 2 1 ones a 1 and a 2 and if you write them in the conventional A x equal to b form this is what we will be getting. So, the matrix which we got is this one, this is the a matrix, this is the right hand side vector right and this was the solution right.

So, now let us do the same problem using the general linear least square model. So, again as we discussed in the previous slide this is our model, this is the analytical solution of that model and in this analytical solution the Z matrix is the z naught variable the z z 1 variable all the way up to z m variable stacked one after the other and all the n points. So, that is our Z matrix.

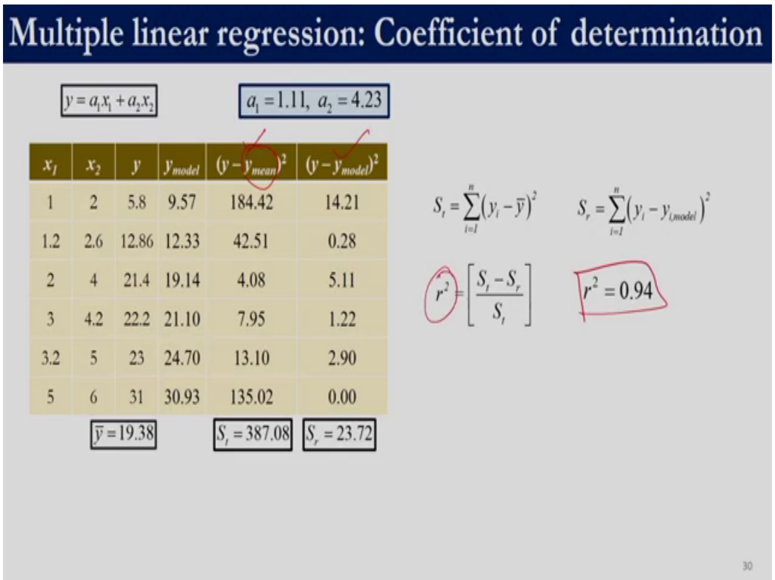
So, in this case our model is a naught x 1 plus a 1 x 2, you can even say it is a 1 x 1 plus a 2 x 2 its just the notations that are different. This is our model, this is the data point and, so this is the error term has to be over here right. So, for this problem to be equivalent to this problem z naught has to be x 1 right, because there is no constant coefficient term right. If constant coefficient term had been there then z naught is equal to 1, here there is no constant coefficient and z 1 is nothing, but the second variable x 2 right. So, in this case this is going to be our Z matrix right.

So, for the current problem the x 1 column all the all the 6 points have to be stacked over here. Similarly the x 2 variable has to be stacked over here, so we get a 6 by 2 matrix right, so this is the Z transpose. So, once we do Z transpose Z we will get 2 cross 2 matrix right and again we can calculate Z transpose Y. So, if we calculate Z transpose Y we will get this right hand side vector right.

So, now if you see this is exactly identical, this a and b is identical to what we would we obtained over here right. So, solution would be the same a naught is equal to 1.1 1 plus a 1 is equal to 4.23 right. So, a 1 is equal to 4.23. So, that completes multi linear regression with constant coefficient and multi linear regression without constant coefficient without constant coefficient in terms of general linear least squares. Without general linear least squares also we have seen how to get that.

Now, we have seen if we know the analytical solution for general linear least squares we can directly employ that for multiple linear regression problem right. And similarly this can be extended for polynomial regression also, because polynomial regression at the end of the day we are converting into a multiple linear regression problem right. So, whatever we have discussed for multiple linear regression with the analytical solution of general linear least square is also valid for polynomial regression.

(Refer Slide Time: 59:51)




So, again this is just for the sake of completion, you can calculate the r square value right. So, model is known you can calculate how the model ah. Had we taken the mean as the model what would be the error and now that we have a model what is the error right

So, the difference S t minus S r shows the improvement, S t minus S r by S t is the coefficient of determination right, so that is 0.94 in this case, that concludes the linear regression part right. So, now, we look into non-linear regulations. So, linear regressions it boils down to set solving a set of simultaneous linear equations, either we can do it as we did in the first part of this session or we can use the analytical expression Z transpose Z into a is equal to Z transpose y.