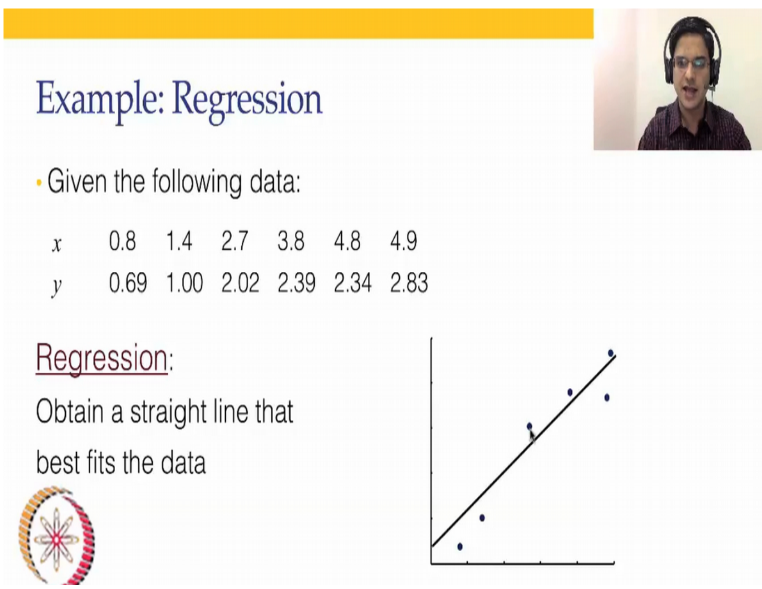
MATLAB Programming for Numerical Computation
Dr Niket Kaisare
Department of Chemical Engineering
Indian Institute of Technology, Madras

Module No. #06
Lecture No. 6.1
Regression and Interpolation – Introduction

Hello and welcome to MATLAB programming for numerical computations. We are in week number 6, and this week we are going to cover, regression and interpolation. Regression is something that is familiar to you, probably as curve fitting. So, for example, fitting, straight line to several data points, interpolation was probably familiar to you as, simply joining the dots. So, we are going to cover regression and interpolation in module 6 okay.

(Refer Slide Time: 00:47)

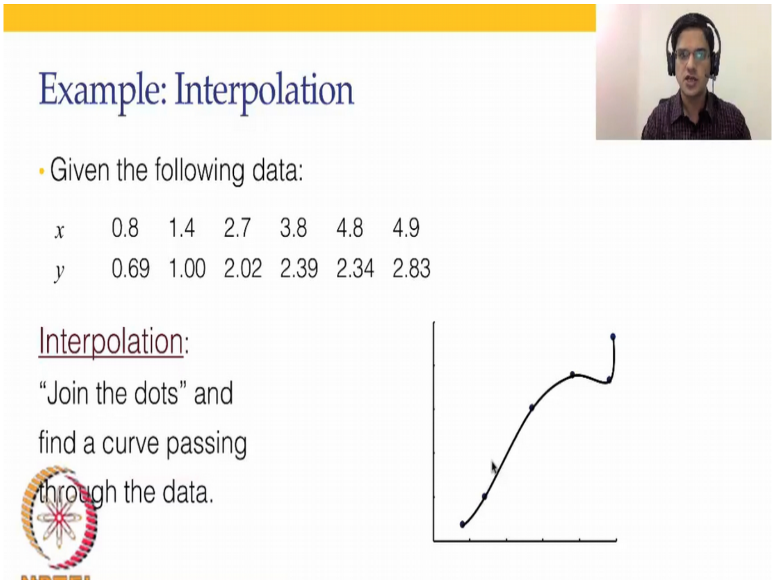


Let us first take an example. Let us say we have been given the following data. The x is given over here and y is given alongside with x. And this, when we plot in x versus y curve, we are going to get this as these data points on the 2-dimensional space okay.

What regression, tells us is, how do we obtain a straight line. For example, that best fits the data. So, this is, the straight line of the form a0 + a1x = y. And we want to find the values of a0 and a1.

So, that this particular line is, the best fit line for all these data points. So, what we are going to do in regression is, to minimize the error between the data points on this line and the corresponding data point that is shown over here.

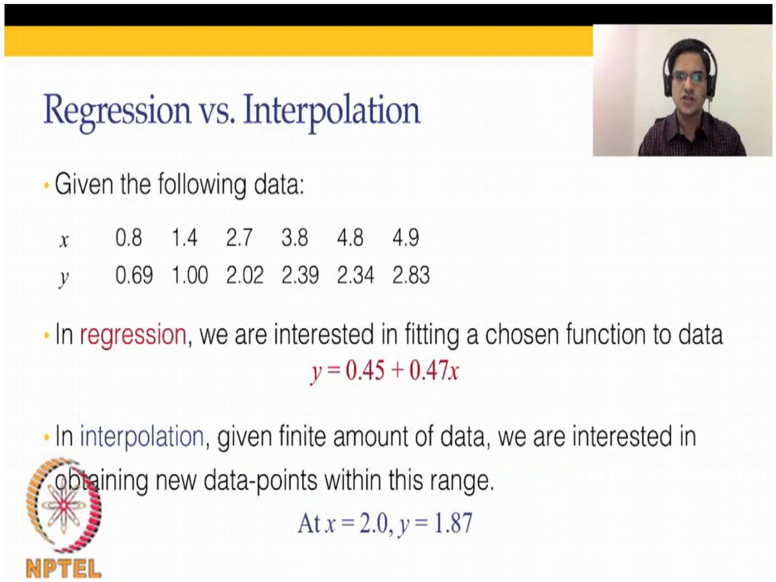
(Refer Slide Time: 01:45)



Now what is interpolation? Interpolation, is kind of joining the dots which basically mean that we want to find a curve that passes through all the data points. The reason why we would use interpolation is for example fill in the missing data. So, let us say this is the particular set of data, and we want something at 2.0. See there is no data available at x = 2.0. So, in that particular case, we are going to use interpolation.

As again interpolation, when we are using to use regression is, when we need some kind of a functional form as y = f (x) and we want to find that particular functional form or specifically we want to find the parameters, that best satisfy that functional form that is when we use regression okay.

(Refer Slide Time: 02:35)



So, in regression, we are interested in fitting a chosen function to the data. The function for example, in this case, might end of being y = 0.45+ 0.47 x. In case of interpolation, given a finite amount of data, we are interested in obtaining new data points, within this range, that is new data points within the range of 0.8 to 4.9 okay. So, that is the intend of interpolation and regression.

So, for example if you want to query at x=2.0, what is the value of y? What is the probable value of y? If this is the interpolant that we are going to use, than the value of y, is going to be 1.87 okay. So, what are we going to do in this module?

(Refer Slide Time: 03:21)

In module 6, the first lecture, that is this lecture, is primarily a demonstration lecture. I am going to, cover linear regression and Newton's interpolation, to just give ourselves some idea of how the regression is going to work okay. We are not going to use, this particular examples to see how interpolation and regression works.

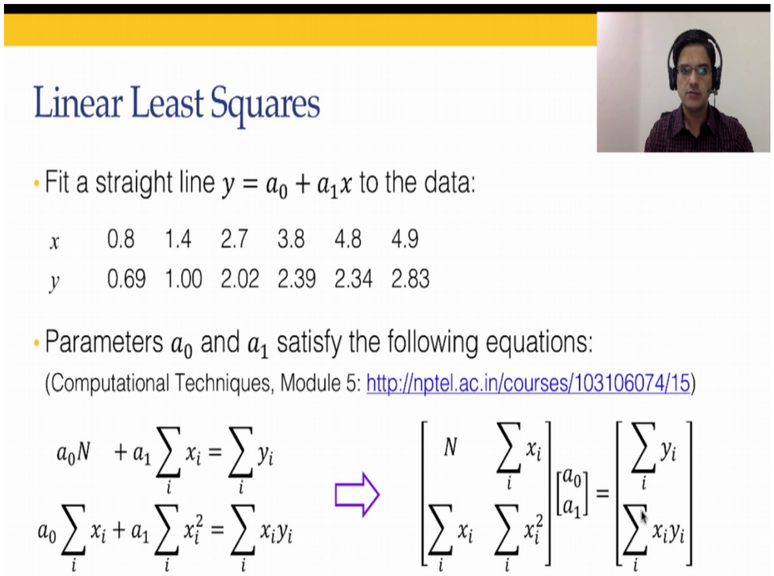
We are not going to use these examples in order to code something in MATLAB. I will quickly demonstrate a MATLAB code for you to show you, what regression means, what interpolation means and the difference between the 2 and primarily for the sake of completeness okay.

In the next lecture, we are going to cover, curve fitting with multiple parameters and we will take up one MATLAB function called lsq curve fit in order to solve this problem as well. So we will take a problem, and we will solve it in 2 ways. One is using a method that we would learn in the next lecture, as well as using a MATLAB based algorithm.

In lecture 6.3, we are going to use these concepts for parameter estimation. We will take the example of a reaction rate, something that you would have done in your chemistry course in eleventh and twelfth grade. We are going to take that forward in lecture 6.3 using MATLAB. So 6.2 and 6.3 will cover examples in linear regression.

In lecture 6.4, we are going to cover interpolation. In case of interpolation, we are not going to build our own codes to do interpolation. Instead we will use two powerful functions in MATLAB spline and pchip in order to do interpolation. We will cover what spline and pchip does in lecture 6.4 as well as methods to use them okay.

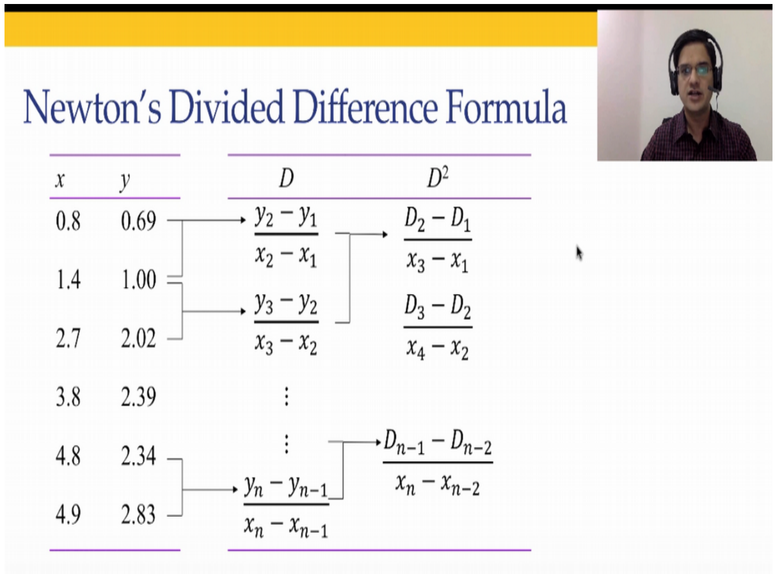
(Refer Slide Time: 05:43)



So, let us go over to the demo part, of for linear regression as well as Newton's interpolation and we will take this particular set of data, in order to demonstrate regression and interpolation. So in case, of linear least squares regression, the question that we want to ask is, how to fit a straight line y = a0+ a1x to the given data. In the computational techniques lectures, in module 5, lecture number 1 and lecture number 2, we have covered methods, to do linear regression and the link for which is given over here.

So, if you want to understand the derivation, for how to get this particular equation. You can go over to this link and you can look at the derivation for getting this equation. We are going to get 2 equations and 2 unknowns. This is the first equation, this is the second equation. We put these equations together. We are going to get this in the form a phi = b. So, our phi vector, which is a0 and a1, the two parameters is a0 and a1 is just going to be a inverse multiple by b okay.

(Refer Slide Time: 07:01)



In case of, Newton's divided difference formula which was covered in module 5, lecture 4. In that case, we built, what is known as divided difference tables. So, divided difference table, in the first divided difference, is based on difference between y2-y1 / x2-x1, y3-y2 / x3-x2, y4-y3 / x4-x3 so on and so forth. The next set of divided difference is based on the d's. So, that means d square is nothing but d2-d1 / x3-x1, d3-d2 / x4-x2 so on and so forth and so on we go until d ^ n. This is how we built the overall divided difference table.

And the top row of the divided difference table forms the divided difference coefficients. Again, and this is for demonstration purposes only, to show the things that were covered in the computational techniques course, we are kind of go showing parallels using MATLAB. And that is the intent of this demonstration lecture. So let us go to MATLAB okay.

(Video Starts: 08:20) And let us look at the example. Okay, unlike the previous lectures, today what I am going to do is, I will just show you, the final code and I will just go over some of the interesting aspects of the code, I am not going to discuss the code in detail, since this is a demo lecture. So, let us look at look at the various lines of code.

This is the x and y data points n is nothing but the length of the vector x. This is the part, where we compute linear regression and this is where we compute interpolation, using divided difference. Our matrix a was formed as shown over here. n we have, n summation xi, summation

xi. Summation xi squared as the 4 guys. n sum(x) sum (x) and sum(x).* x which is summation of xi squared.

Likewise, we form the b matrix that is shown over here okay. And then phi, is going to be nothing but inverse of a * b. We are plotting the original data points and we are plotting the value a0 + a1x. So a0, that is like phi 1+ phi 2 * x for this value of x which will be a red straight line okay.

So, that is what we will do in a linear regression part. And the divided difference part, we have created another function called ndd, to calculate the interpolant using Newton's divided difference. So, we will calculate the interpolated values, in the entire range 0. 8: 4 .9 in steps of 0 .01 so, that we can see the entire curve how the interpolated curve looks like. So let me now run the code okay.

So, the blue dots that you see over here, were the original data points, the red line is the interpolated line, that we have plotted. So this is the interpolated line, that we have plotted in red and sorry, that is the regression line. The best curve fit line that we have plotted in red. And in black, we are getting the interpolant okay. And the interpolant that we obtained is using Newton's divided difference formula.

So this is the straight line, these are the data points and this is the interpolated curve okay. The thing that you see over here, because these 3 points are located fairly close to each other in the y direction. That is why we see a very funny way, in which this interpolated curve looks like, instead of the interpolated curve going smoothly in this way okay.

That is one of the things that happens in interpolation, we have taken an extreme example over here, you will not get this type of behavior for most of the interpolation cases okay. Anyway, now to finish up this lecture, let us look at the Newton's divided difference formula okay.

Newton's divided difference, we create a divided difference table, first by initializing it as a 0, a vector of n by n array of zeroes okay. And at each step, we are going to calculate the d column. And the d column, is going to be nothing but, the difference of the previous column, divided by xi-x1, xi + 1-x2 so on and so forth okay.

If we go over here and we look at the divided difference table each of these guy is computed from the previous column of the divided difference table. And that is what the MATLAB code does. And finally, the coefficient is going to be, nothing but the first row of the divided difference table and we will use those coefficients, to do the interpolation okay. So, this is the, part where we are doing regression, this is the part, where we are using Newton's divided difference formula in order to do interpolation okay. (Video Ends: 12:47)

To recap, what we talked about in this lecture was, was this. We talked about linear regression which is to fit a straight line or a curve to the given set of data, the example that we took was of fitting a straight line. But regression, does not necessarily mean, only to fit a straight line. That is something that we are going to see in the next lecture. Newton's interpolation, we use in order to find the value of y, at intermediate points of x okay.

So, these are the things that we demonstrated in today's lecture. And in the next 3 lectures we are going to spend 2 lectures covering regression and parameter estimation. And the last lecture covering interpolation okay. So, with that I come to the end of this lecture and I will see you in the next lecture. Thank you and bye.