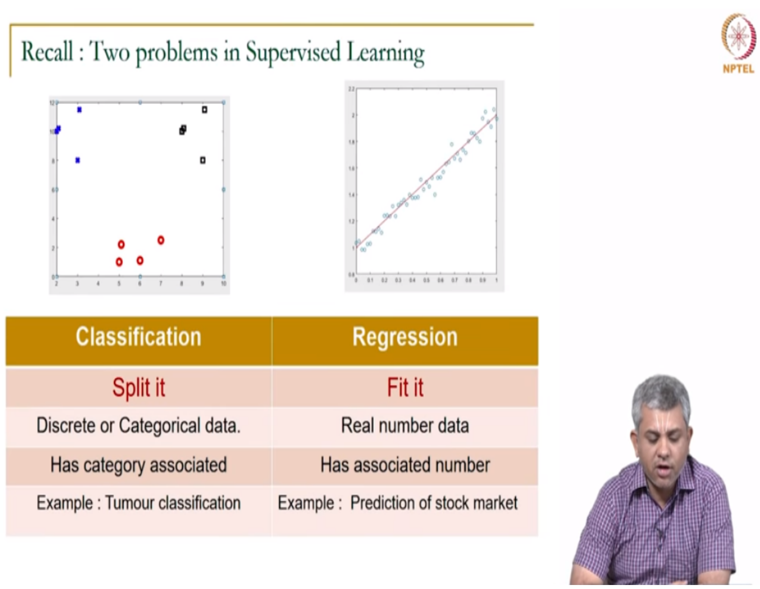
Machine learning from Engineering and Science Application
Professor Dr. Balaji Srinivasan
Department of Mechanical Engineering
Indian Institute of Technology, Madras
A Linear Regression Example

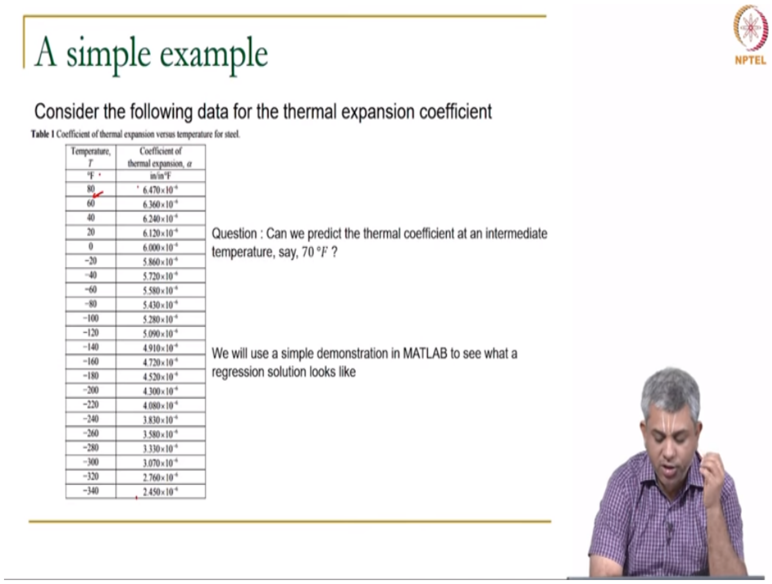
In this video we will be looking at the first model that we have for this course which is a linear model and we will be doing a simple regression many of you would be familiar with this even from school days and even college days but we have slightly different take though initially look very very similar later on we will see how this fits into the whole machine learning idea so please do not be careless with this because once you understand this idea a lot of deep learning automatically become accessible because most of the issues that occur in deep learning do occur even in this very simple case so it's actually an advantage if you already understand linear regression from before.

(Refer Slide Time: 00:56**)**



So remember that we were discussing two problems in supervise learning back in the first week there is classification and there is regression and we will be looking at regression which is simply the idea that you have lots of data points and you want to predict Y at a particular X which is not yet available, so if you have X new you want to predict new Y.

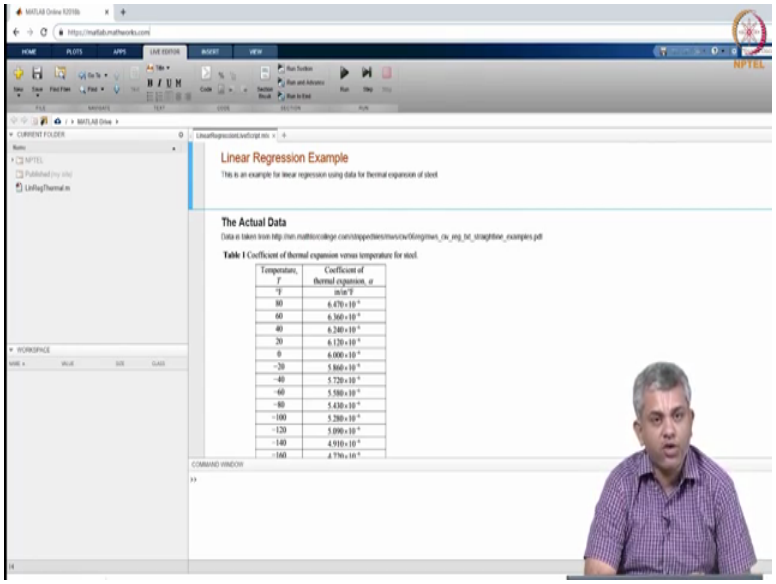
(Refer Slide Time: 01:26**)**



So let us take a simple example, this example is about the thermal expansion coefficient. So the thermal expansion coefficient as you know metals expand as you increase the temperature but this thermal expansion coefficient is actually not a constant it even varies with temperature, so here is some data about the thermal expansion coefficient of steel with temperature this is taken from a source where this thermal expansion coefficient is actually given in Fahrenheit instead of Celsius, so you can see a reasonable amount of variation between 6.47 to 2.45 for large variation in temperature remember this is temperature in Fahrenheit which is by you get minus 340 etcetera, in Calvin you would not get that so the question is let say we want you have all this data in a tabular form and you want the thermal coefficient at some intermediate temperature let say at 70 degrees.

So let say somewhere here you want is that our several ways of doing it a very ways to simply say I will interpolate it will be between 50 percent of this and this, but suppose you want to use all this data you don't want just use the nearby data, you want to use all this data and see if you can come up with a better module this is usually what happens in data signs we might have one or two data points but linear interpolation might not always be a good idea, So we will see a simple demonstration in MATLAB and see what a regression solution actually looks like for this problem.

(Refer Slide Time: 03:07)

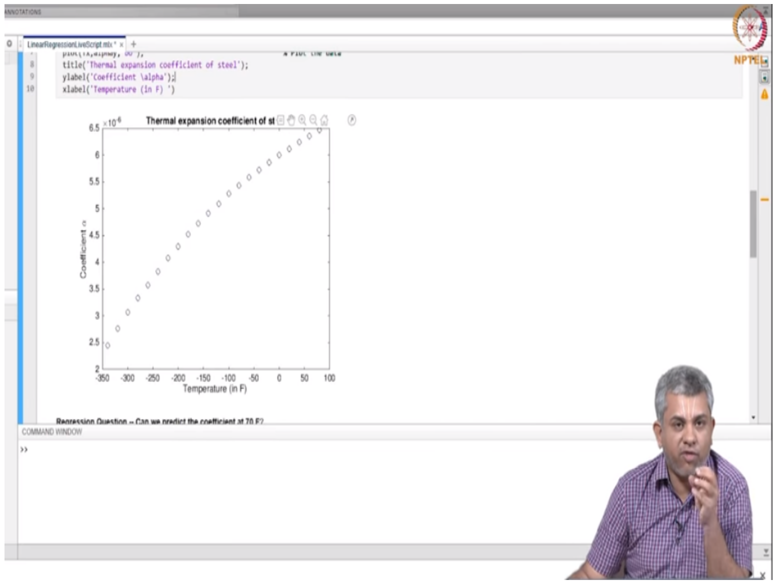


What you are seeing here is my account in the MATLAB online website its accessible to all of you, you are welcome to use this same thing we will also be sharing a copy of the code that I am showing right now.

So that you can try it out for yourself, you are not constrained by this you can also try the same thing anywhere else in a JUPYTER notebook you will see that the format of what's called the MATLAB live editor is very very similar to the JUPYTER notebook if some of you are familiar with JUPYTER notebook while using python, so I have this example of linear regression for the problem that we just saw on MATLAB the same data that I showed you is here its actually taken from the source its actually in excellent source, so the source that I am showing here is an excellent source for both data as well as general numerical algorithm, I would highly recommend that people who are interested take a look at nm.mathforcollege.com.

So now here is the data i have codded this up here this is the MATLAB code we will be looking at various different ways of doing it as we go on with the course once again you have free to use any programming language that you would like it's just that as you can see both for demonstration and for initial testing out MATLAB is extremely convenient okay, so we look at this all I have done coded up this data right here, I will now execute this code and now alpha Y and TX which are my X and Y co-ordinates, T is for temperature alpha is the thermal coefficient those are now placed, now we can look at what the data looks like.

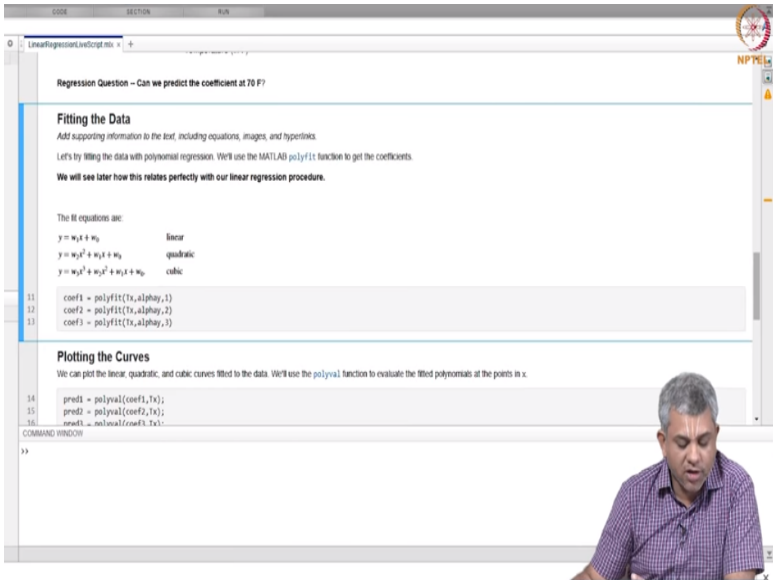
(Refer Slide Time: 05:03)



You will see now that MATLAB has plotted the data and you see X temperature Y is the coefficient and there is a slide curve now we don't know what sort of fit is good for this kind of data is it a linear fit which is good, is it a quadratic fit which is good, is it a cubic fit which is good and now it turns out that we can systematically try this out, I am going to use certain in build MATLAB functions there are several options for this specially when we come to later parts of this peak I will tell you what other options exist in MATLAB which are in build functions we will also be programing this from scratch because this is a simple thing that we can program from scratch using gradient descent will be showing examples for that too but from now just to get a physical picture of what is happening I will just use some in build function within MATLAB for you to see what it looks like.

So what we are going to do is what is called Polynomial regression all that means is I am going to find out remember the best fit, if I fit a line for this data all of us know that the line is not going to fit properly it's impossible to fit all of these with a single line but once I try a line I could try a quadratic curve, I could try a cubic curve, I could try all that for this set of data so we will first try that step by step.

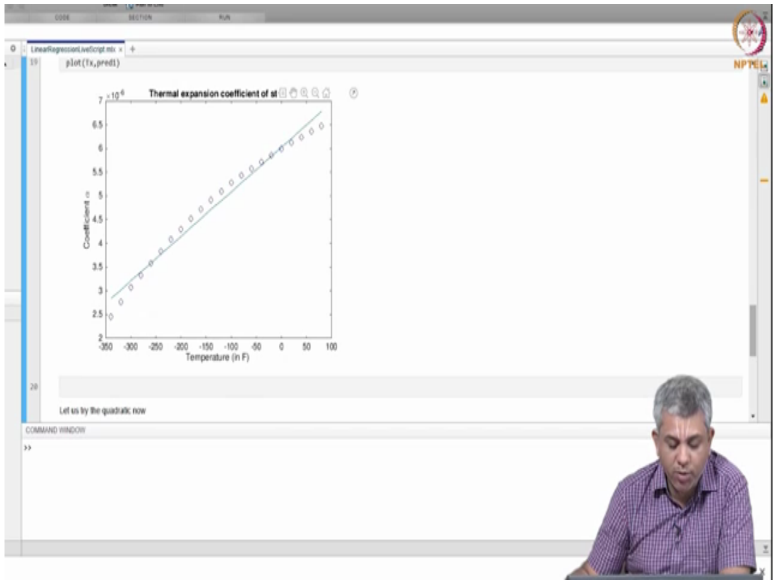
(Refer Slide Time: 06:29**)**

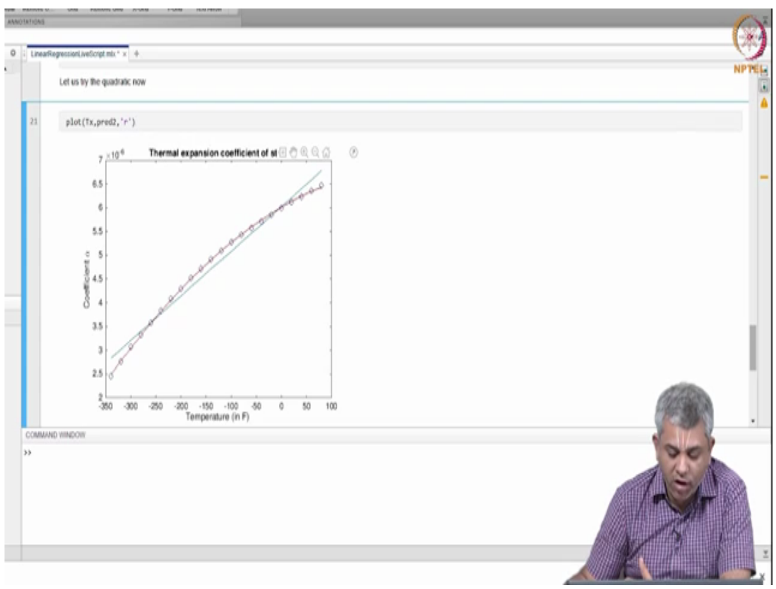


There is function called polynomial fit poly fit within MATLAB we are going to try the following three equations remember these are now our hypothesis equations. So I have written there mean exactly the same form that I have showed you in the previous video which is Y is W1X plus W not you can think of this as W not X par 0, so WN X par N that's basically what we are trying so this is a linear fit we will try a linear fir, we will try a quadratic fir, we will try a cubic fit and the coefficient are stored in this three variables coefficient 1, coefficient 2, coefficient 3 poly fit simply does the order of polynomials.

So this is a first order polynomial the linear fit is called coeff1, quadratic fit is called coeff2 and cubic fit is called coeff3, so let's run this two and you see there some coefficients have come out so the first coefficient second coefficient W0 W1 W0 W1 W2 etcetera, so all these are sitting here now but for us we are visual creatures we would like to know how good is this fit run this and we will fit it for this whole thing.

(Refer Slide Time: 07:53)





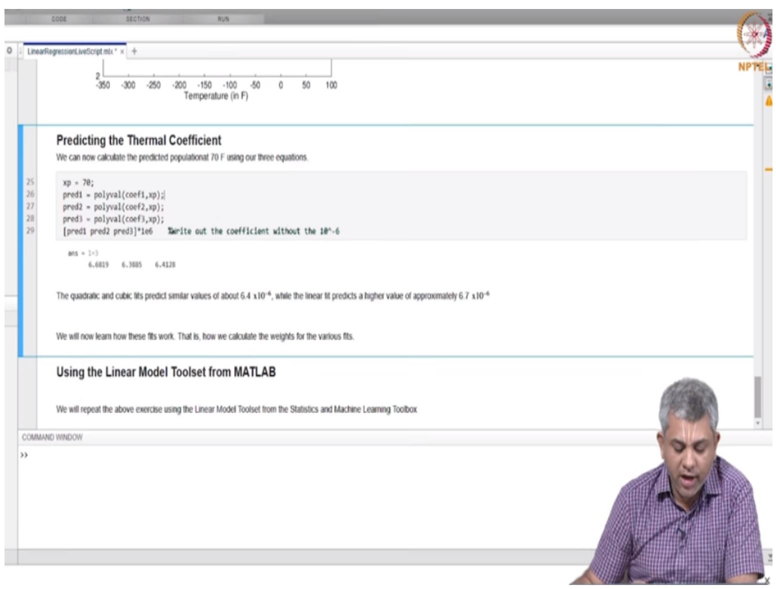
And now we see that the linear fit looks like this, it’s not unexpected in some sense this is what is called the best fit line remember we were trying to find out the optimal W’s. So what poly fit has done is indeed found out this optimal W it’s kind of a fit weather it’s a great fit or not we will find out but remember suppose I am predicting at 70 I will predict my value somewhere here using this fit using this hypothesis function that I have just found out.

We will now try the quadratic, the quadratic is now the red curve you will see this fit’s this much better than the linear does it’s always happen that the quadratic fit’s better than the linear you will see later on that this is not necessarily true.

But in this case it does happen that the quadratic fit was better so it will give a much prediction for the temperature the thermal coefficient of expansion at 70 degrees, we can now try a cubic fit also and here is the full data summarized this line here is the linear fit the red line is the cubic fit and the black dotted line which is almost indistinguishable from the red line but you will see at the end that the cubic is actually better.

So, for the data set that we were given we find that the cubic fit looks much nicer than the quadratic fit and the quadratic fit looks much nicer than the linear fit and our looks always important we will see that's also not necessarily true in life or in machine learning that it's not necessarily a good idea to makes something that fit, so well in this case it is actually a good idea but we will see how to formally calculate whether it's a good idea or not later on, so we can also calculate our predictions.

(Refer Slide Time: 09:58)



So you see different predictions here the quadratic and the cubic fit about predict about 6.4 into 10 power minus 6 whereas the linear fit as we saw predicts much higher and that about 6.7 into 10 power minus 6. So in the next video we will actually see how this fits work we will also see later on later videos how to use the linear module tools at from MATLAB but that's not relevant right now but we learn how this fits actually work what is actually going on behind the scenes when we do poly fit or when we do any kind of linear module fit so that will be see in future videos.