CH5230: System Identification

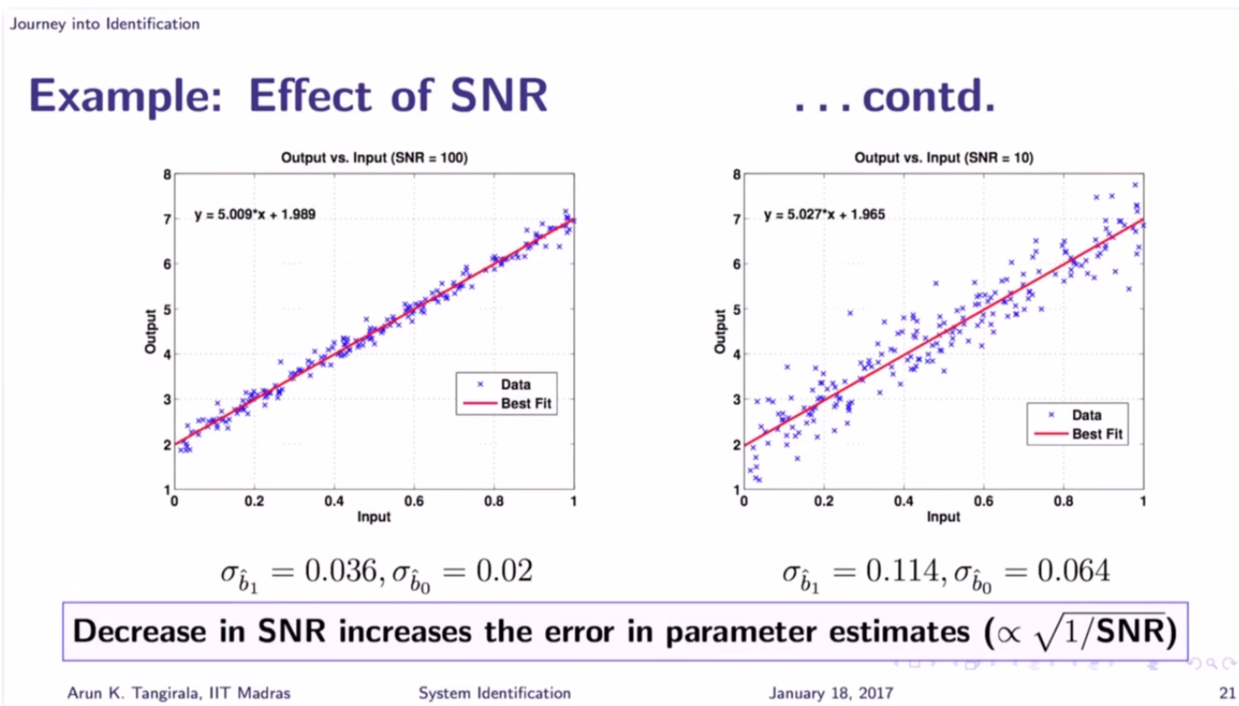
Journey into Identification

(Case Studies) 6

Okay, a very good morning. What we'll do is, we will continue our case studies that are highlighting different aspects of identification. In yesterday's lecture we learned, what is the effect of signal-to-noise-ratio, for example, that was the last item on a discussion and then, of course, we focused largely on identifiability. We learned what this model identifiability and particularly what is the role of input in fetching that model. And of course we briefly talked about the role of estimate at which we will expand a lot more in detail, when we talk of estimation theory. And what yesterday's case study particularly on this that demonstrated the effect of SNR tells us that noise plays a critical role particularly the relative amount of noise to the input.

Okay? So you may say that the noise in the data is very small in whatever quantities that you are thinking of, it is not about the amount of noise, but it's the amount of noise relative to the input which is characterized by this SNR

(Refer Slide Time: 01:36)

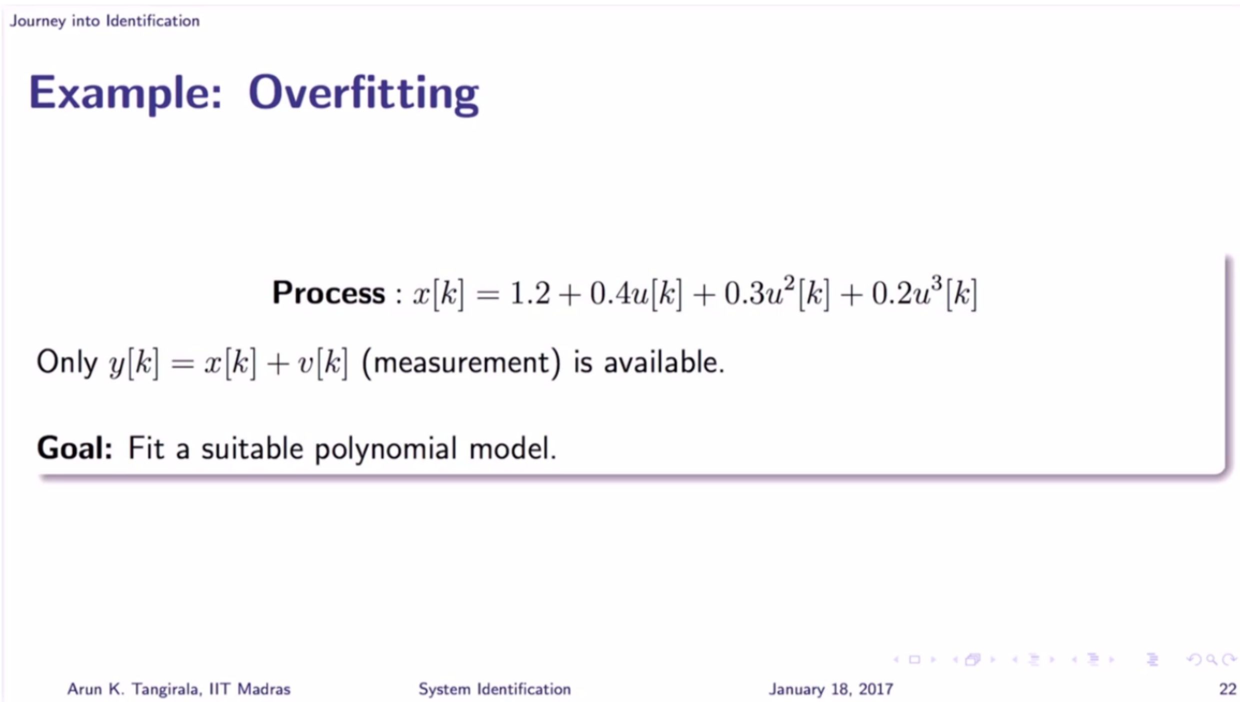


So it's a variability of the signal to the variability of the noise or the power of the signal to the power of the noise. That plays a very important role in shaping the errors in the estimates or even the standard error itself. Okay? So we will prove this result that the error and parameter estimate is inversely proportional to the SNR a later time, for now this is all for conceptual consumption I would say. The other role of noise is in fitting the model. That is of course the parameter estimates have also got to do with the fit. But here we are talking of over fitting.

So let us say I have guaranteed that the SNR is conducive for obtaining good parameter estimates. Great. I have done the experiment in that manner. Now at the modeling stage one has to be careful.So look at, at how many stages one has to exercise caution in system identification, from the choice of model to the way we perform the experiment and when we say where perform the experiment both the input design and the SNR.And then again back at the modeling stage where we are training the model for a given dataset. So this example here will hopefully illustrate this concept of or this phenomenon of over fitting where we train the model or we fit a model that is more complicated or more complex than what is actually necessary. We will revise the statement a bit later.

So let's look at the example here, the example again is a very simple linear process.

(Refer Slide Time: 03:23)



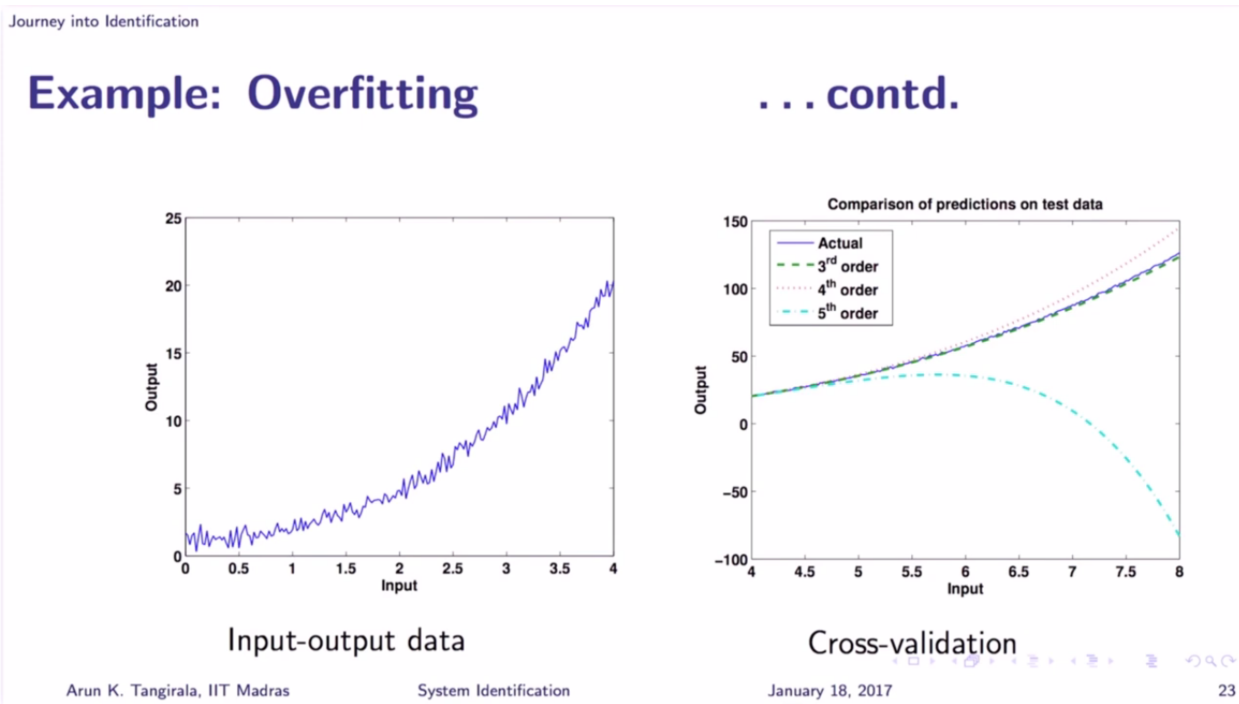
In fact here. It's a static model. We are not looking at a dynamic model as you can see.I said linear. In fact it is not linear. It's annon-linear model. Rather it is a static model than a dynamic model. And the non-linearity is in the form of a polynomial relation. Assume for now that we know it's a polynomial relationship but we do not know the order, just for the sake of illustration.

And of course once again the story is that the true response of the process is not available. We have a measured response that's available and we are just the noise level such that the SNR is healthy. Now the goal is of course to fit a suitable polynomial model from the given data. The first step always as I said is visualization. So we plot the model here on the left, I show you the--Sorry we plot the data, on the left I show you the input output data y versus u. Even if I were not told that there is a polynomial relationship,the plot itself suggests a polynomial relationship.

Now what you'll notice of course is that there are some mild fluctuations here at every point on the plot here and these fluctuations are of course due to the noise that we have introduced. Now the goal is to fit a polynomial relation between the input and output. Obviously one starts off with different degrees. Now there are two approaches here. One is called a bottom-up approach. Another is called a top-down approach. In a bottom-up approach you would start with the simplest model possible and then gradually raise the complexity of the model.

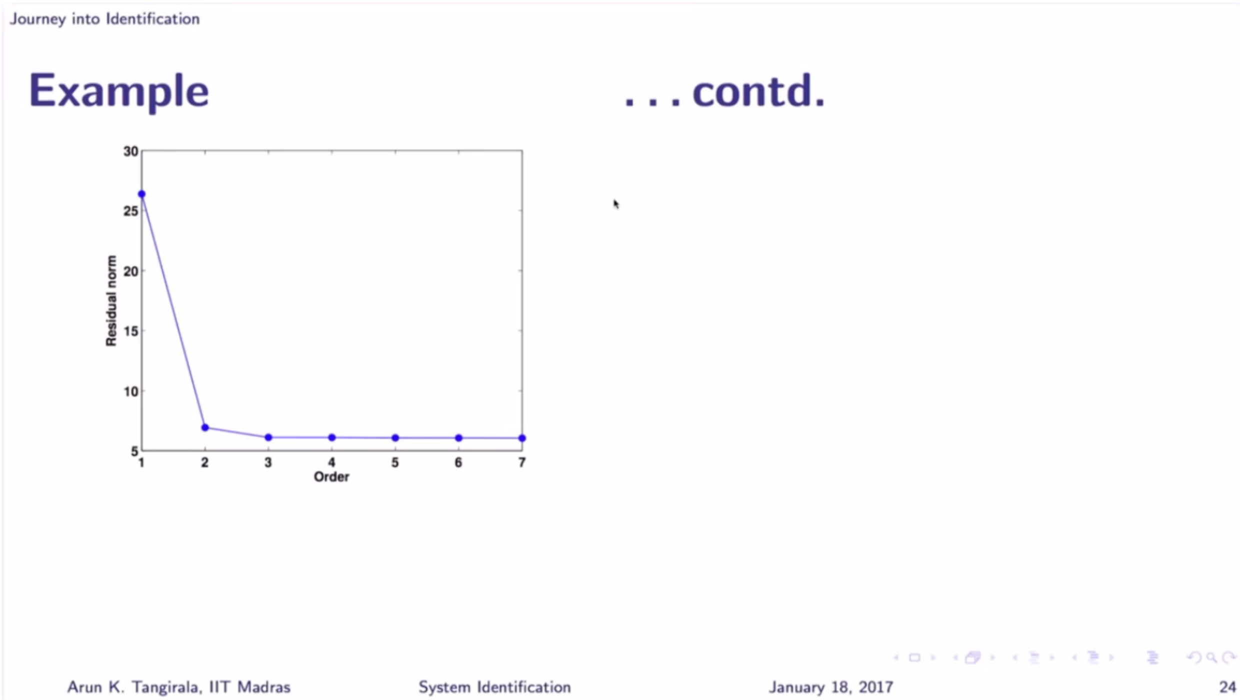
And then the top-down approach, you start with the most complicated model that you can think of. And if that has over fit then you start simplifying it. Typically in system identification and classical approaches you adopt the bottom-up approach, that is, you start with a simple model because the general recommendation is, if a simple model does the job for you then live it. Don't try to complicate life more than what it is already. Right? So naturally one starts with a simple model. In this example, we suppose that there is a quadratic relationship to begin with. Obviously a linear relationship is ruled out. Just from a visual analysis itself. So we start for example with the second order. But in fact, I am not showing the second order plot here. But typically you would start with the second order polynomial.

(Refer Slide Time: 06:11



What I'm showing on the right is the polynomials of third,fourthand fifthdegree. The predictions obtained by these polynomials. In fact, I'll return to this plot a bit later. Let me go to this plot first and then we'll come back to the cross validation plot. That plot is the prediction of the respective polynomial models and a fresh data. So what do we do? When I fit a second order or even a first order, second order, third order and so on, general common sense approach is to see how much the model has managed to explain in the data. That is, what is the error, what is a residual, what is left over? And what I am doing here is, I'm drawing a plot of the prediction error here starting from the first order polynomial.

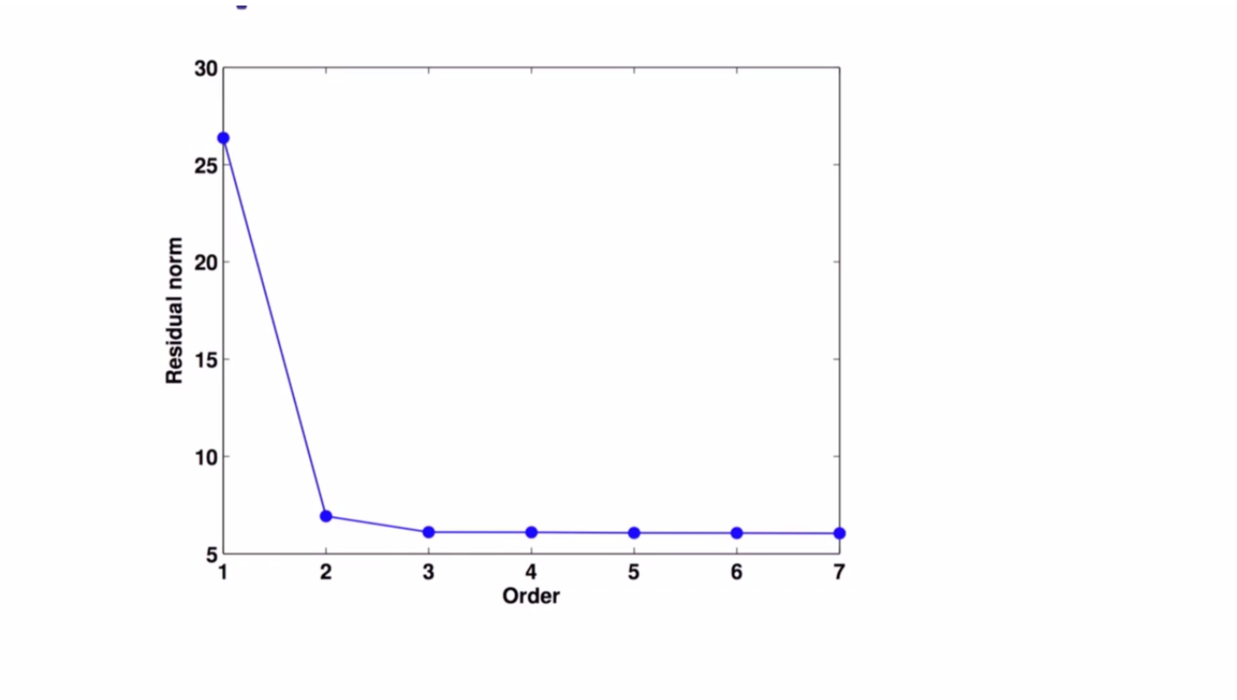
(Refer Slide Time: 06:58)



Just to keep things complete. I've also fit a first order polynomial, in other words I fit a straight line.And what I'm actually reporting here. I don't know, ifhow well you, can see. Although I say here residual norm. What? It is essentially some squares. When I say norm here, you have to understand that we are looking at the two norms, square to norm. So how is the residual is defined, epsilon k is y k minus y hat of k. This concept of residual and this definition is extremely important in model development. So you have to be very clear on this definition. It's a very simple definition. It's nothing much. It's the difference between the observed outputversus the predicted output.

Nowy hat of course comes from the model. And what you see on the y axis is the squared two norm. So notice now that I have written a vector here, this residual vector consists of the residuals at different instance. You can say from 0 to N minus 1 or 1 to N, assuming you have any observations. So what do you see on the y axis is a squared two norm, although I say residual norm. In some sense you can think of it as variants but it is sufficient to say it is some square errors. So what do you see here in this plot? Let me zoom out a bit, so that you can read the x axis as well.

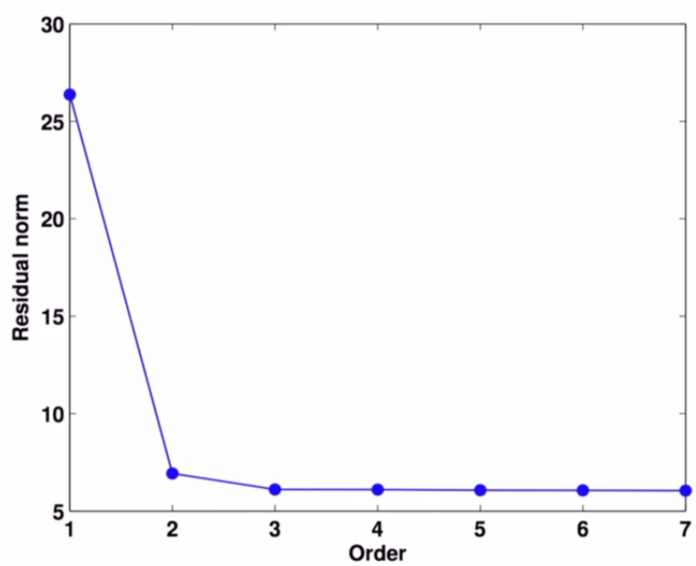
(Refer Slide Time: 08:39)



So if you look at this plot carefully, you observe that the first order polynomial which is a linear fit produces the maximum error, and there after the error falls down. Naturally, so because I'm fitting higher order polynomials. My ability to explain the data improves, fine.

What else do we observe? On the x axis you have the order of the polynomial that's being fit. What we observe is that, yes, there is a decrease in the residual variance which is good, which means that,I'm going the right direction. But what else can you infer from this blog?

(Refer Slide Time: 09:21)



Fourth and fifth order is kind of marginal. What about third and fourth? Slight decrease. Right? So you see some of you feel that third and fourth, there is some marginal improvement, but fourth and fifth it kind of flattens off.

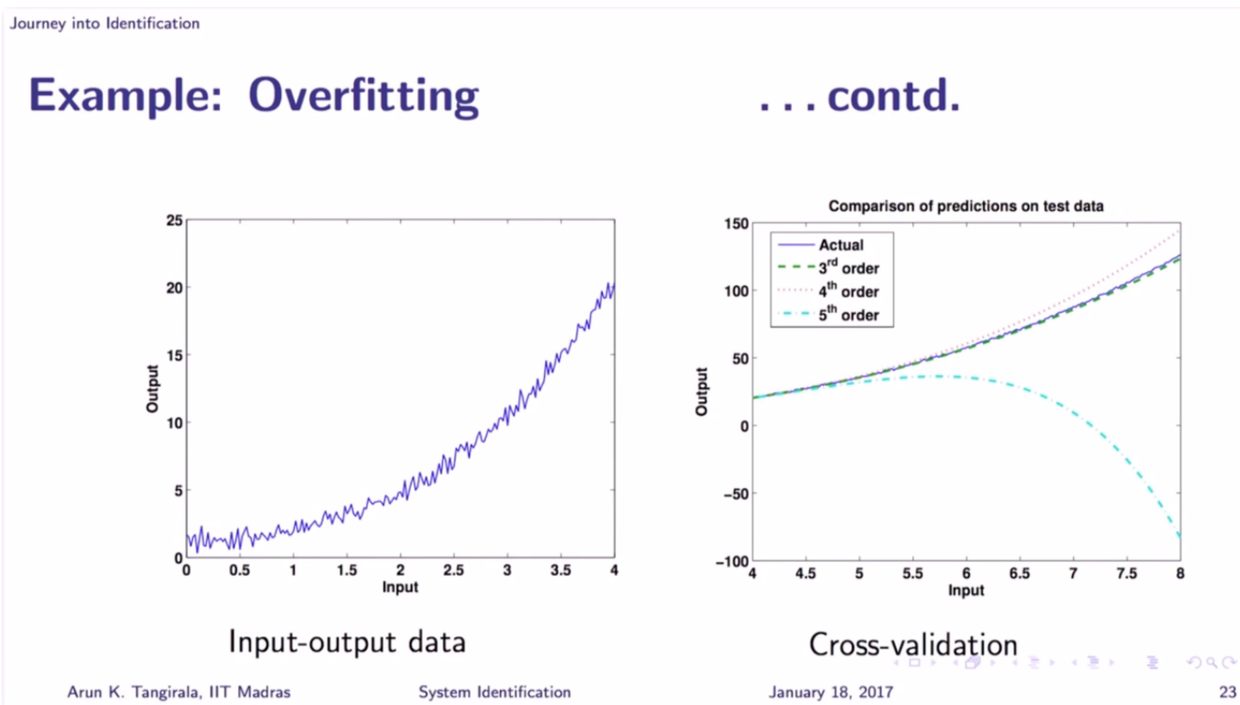
So what can we conclude from these observations with regards to the model that we want to fit? Third order is suitable. Fourth order, what do you expect? A better fit, of course, a better fit. Suppose I say no, you know, I want really a very low residual norm and they'll probably fit a fifth order or a sixthorder. What would happen? Is it a crime to do so or what do you expect? More parameters, fine. But let's say, you know, I have enough points, I have enough observations with me to estimate those parameters. What do you think happens as we go beyond the so-called adequate order?

So most of us agree that third order is a sufficiently good choice compared to second or fourth and so on. As we proceed to fourth, fifth and so on, yes, we do get some improvement on the fit but that improvement is highly marginal n the verge of almost nil. Right? Now this is going to be the typical scenario in any model development, as you keep increasing the model complexity, the fit gets better and better. But at some point you will reach, I mean, at some point what you will see is that, the improvement in the fit is only marginal.Fine, so what is the problem? I mean, I know I can fit the fourth or fifth starter model. Improvement is still marginal. Is there going to be some harm? I'm actually going to take a U-turn? Ultimately what am I going to use a model for prediction on a fresh data. Remember. Right? It's like I have been trained to understand the subject by solving the homework problems our project and so on.

And I've understood the concept. But in a bit of in a bit to understand the concept I have all also memorized every number in the homework problem. Right? That is what we mean by over parameterizing ourselves. When I am given a fresh question in the exam because I have over learned, I have attached, I have given undue importance to the numbers in the homework problems. I have unfortunately mistook those numbers also to be a part of the subject. That's a problem. And then you know what happens in an exempt, our ability to answer a fresh question goes on. And that's what we should expect with regards to model.

So now that is what you see in this cross validation plot. That's why I said we'll discuss later. So now is the time to look at the plot on the right which shows the predictions of the model on a fresh data.

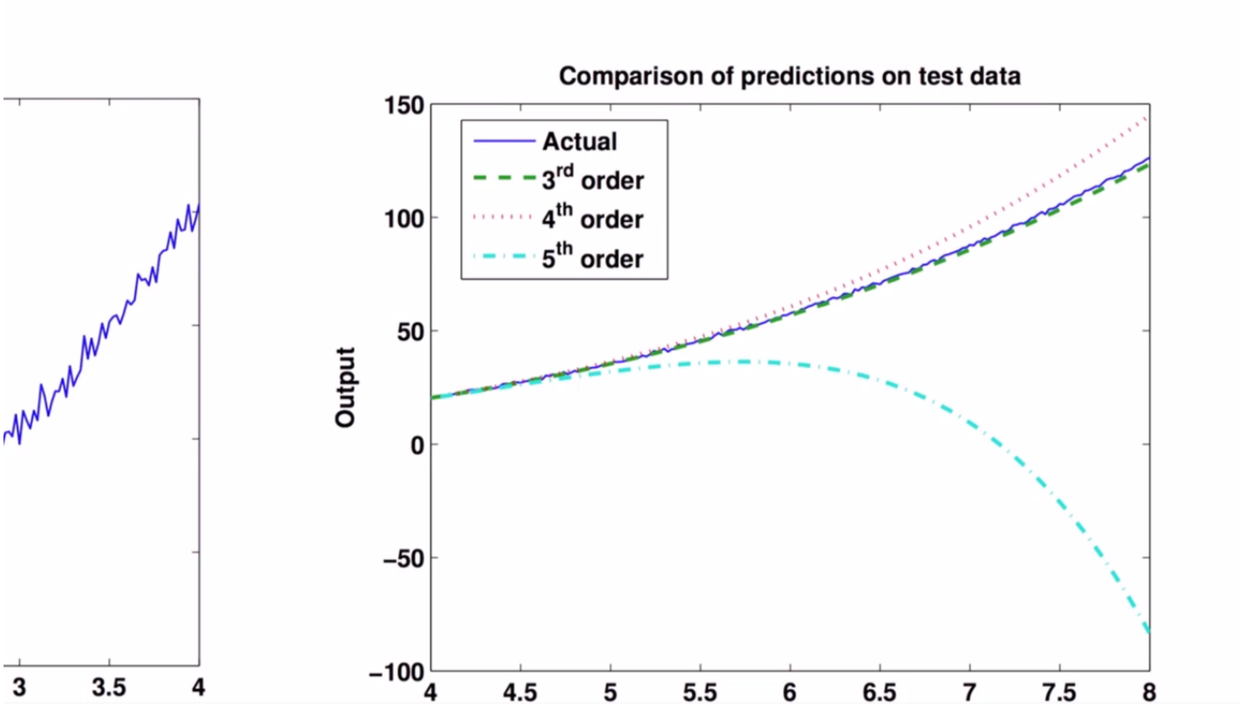
(Refer Slide Time: 12:45)



I'm showing you the predictions of the third, fourth and fifth order polynomials. What do you notice here? You notice that the third order does a good job. Right? The actually shown, the blue line is actual one and the third order, let me zoom in for you. For both of you who have difficulties sitting in the back. Okay, so the blue one is actual one from the fresh data and then you have this green dashed line as a third order and dottedred as the fourth order. And then there is a dash dot which is fifth order. Unfortunately that is almost similar of color, third and fifth order but you can distinguish.

So you can see that the third are the polynomial does a very good job. You know, in all of this we have pretended we didn't know the true relation at all. That model that I showed you earlier was only for generating data. Once data is generated you forget everything. You develop temporary amnesia and forget everything. And pretend that you didn't know anything. You're only given the data. So the third order does a good job.

(Refer Slide Time: 13:58)



The fourth order does relatively a bad job compared to third order. What do you see of the fifth order? The predictions have gone unstable. Right? They've taken a complete turn and they're just falling off somewhere. And of course this is not always the case but this is a typical scenario. When you over parameterize you can not only run into poor predictions but also unstable predictions. And that is the reason why one has to watch out for over fitting.

So here we say fourth and fifth starter models are over fits. What has actually happened? How do you explain this? Sorry. No. I'll give you a more data points. That's not an issue. That sometimes is partly the reason but it is not the reason. Any other reason? Why do you think the fifth order or even the fourth order for that matter.

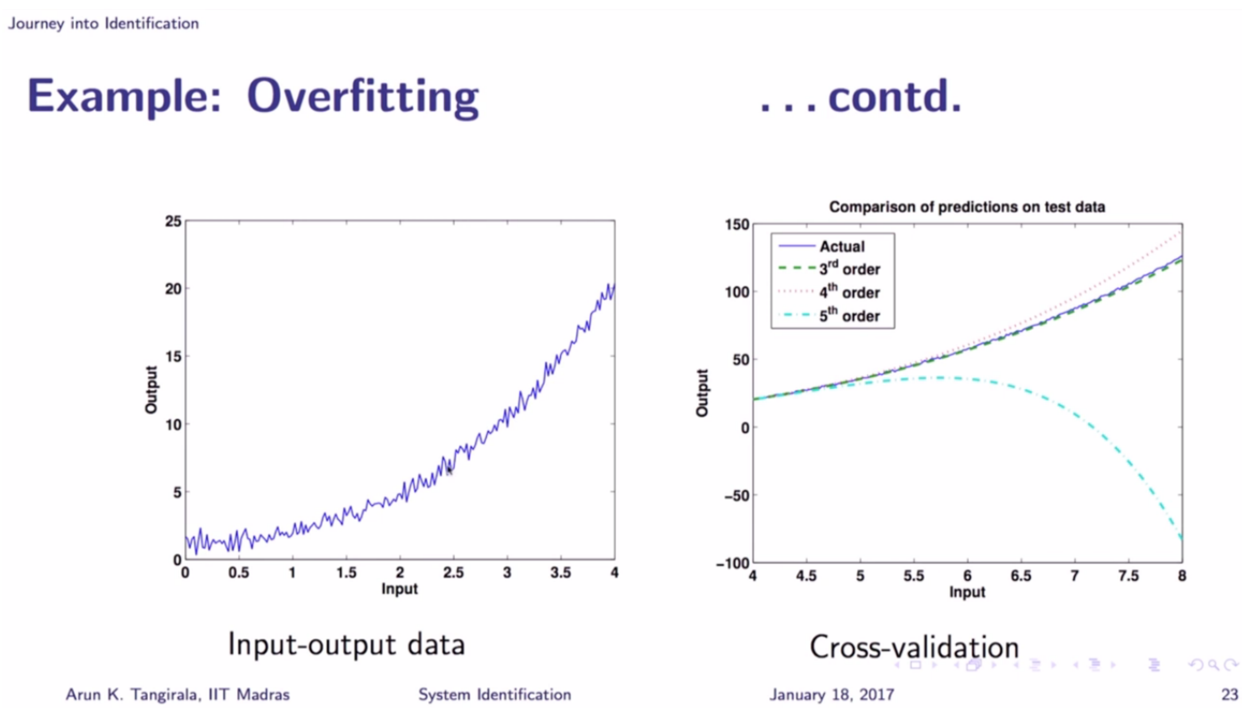
Okay fine. That's correct but that's a symptom of over fitting. But conceptually how do you explain this? When you're fitting a model to a data what you're trying to do is establish a map between the input and output data. Right? But remember the model that we are interested in. See this y k consists of two parts. Here is u exciting some process G. Of course, in our case G is not a dynamic one that doesn't matter. And then there is this observation error vk. We have access to y. We are trying to get this from data.

And I don't have y star. I have u and y. In other words a model is supposed to are the training algorithm. Model is just a mathematical entity. It cannot do anything on its own. It's your choice your total estimationalgorithm. I believe this is a map between you and y star. Although you do not have u and y star, you're saying that I believe this is the model, although I have written, that is, in all of this is, I said [16:21 inaudible] I agree. So you can say this is the model between u and y, in the next case study we will explicitly say this is a model between u and y star. In this example we are by postulating a certain order, third or fourth or fifth and so on. I am telling the estimation algorithm that I believe this is a map between u and y.

When I, suppose that it is a third order. So y hat is some polynomial of u is what I am postulating, when I assume F to be certain order or whatever that model may be, it is trying underneath the parameters that estimate it by minimizing this prediction error, sigma y k minus some square residual. So k is equals 1 to N or 0 to N minus 1. When we are doing this what happens is, the estimation algorithm is searching for parameter values that will minimize this objective function. And in that process what essentially we're trying to do is we are trying to drive y hat pretty close to y. But remember there is no mathematical model that can explain a part of y. That means, there is there is a part of life that cannot be explained by the input.

So you can only drive y hat close to right y to a certain extent. Beyond that if you start driving y hat very close to y, because my head is a function of u and there is a part of y which is not a function of u. Once you start overdoing things. What that model tries to do is, it tries to explain that part of y which is not a function of u which is a v noise that also it tries to explain with input. And what is therefore happening is that, the estimation algorithm or as user what you're doing is, you are not only trying to fit the global relation between y and u, but also the local variations.

(Refer Slide Time: 18:41)



So now you're confusing the global variation. The global variation is a polynomial relationship. And the local fluctuations are due to noise. So it starts now thinking that these local fluctuations are also a part of this function. And that is what we mean by overdoing. So those local ones are, in fact if I generate, if I perform a fresh experiment for the same process I would have different kind of local fluctuations but the global trend won't change.

So that is what we mean by remembering the numbers in the homework problems. Those are very local, I could have actually given the same problem with different numbers. You should not fail because if you have all but learned then you have memorized the numbers everything, not just a concept that is required to solve that problem. So over fitting has occurred here because the fourth and fifth order models have started to confuse are or to assume that their local variations that you see due to noise are also part of the deterministicmodel.

(Refer Slide Time: 19:38)