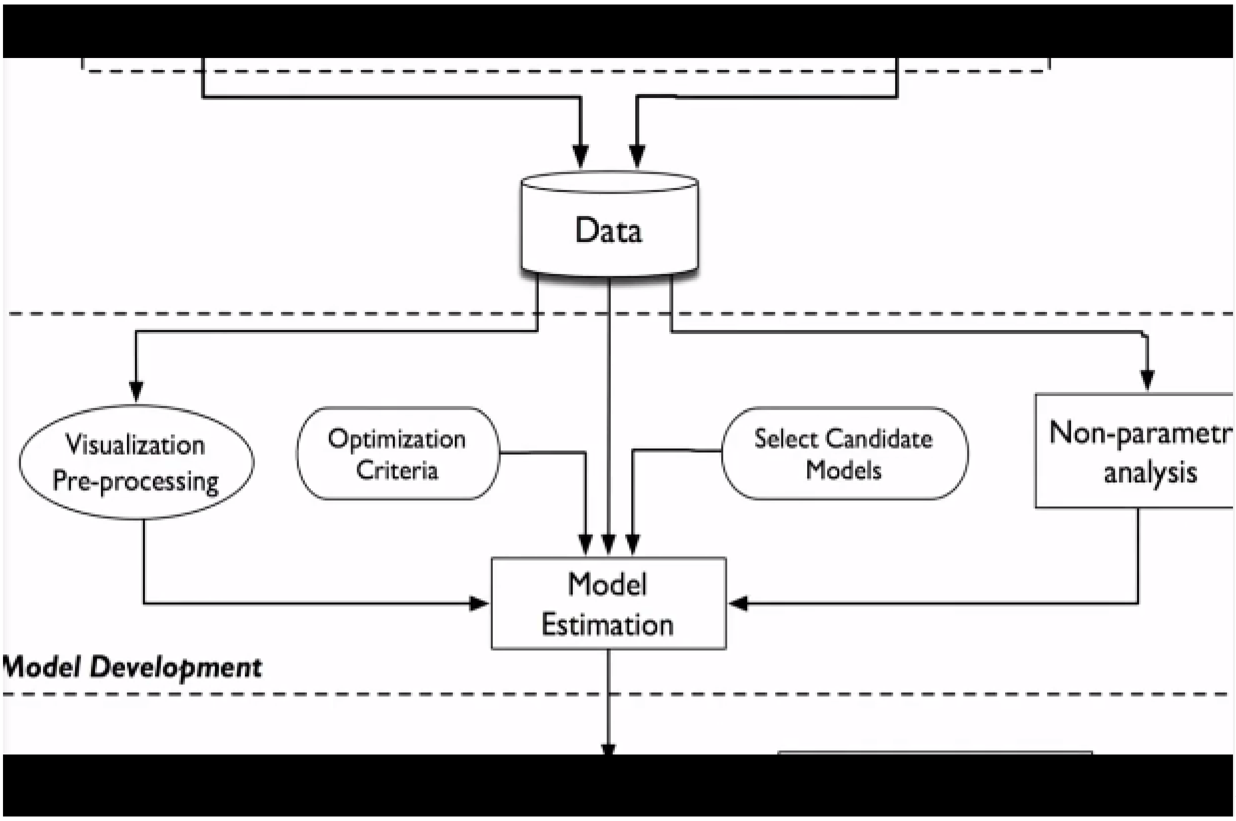
CH5230: System Identification

Journey into Identification 2

ARUN: Then the next one is non-parametric analysis. Where again I said, in a non-parametric analysis. Let me actually zoom in for you so that you can probably see it better. Okay, so on one hand you have visualization and on the other side you have non-parametric analysis. What constitutes this non-parametric analysis? Again, here we use a mathematical method by making minimal assumptions on the process. We don't assume for example, that the process follows the first order dynamics or the second order dynamics or anything. Whatever minimal assumptions I have to make, I'll assume. And typically you assume for example, the process is linear or non-linear, time invariant or not and so on. A very generic set of assumptions that you make and then see if you can with this minimal assumptions, draw some inferences, about the process that I can eventually use in my model. And that is what you will see also in this case study that really go through, if not today but in the next class for sure.

(Refer Slide Time: 00:27)



)

This constitutes non-parametric analysis and the technical definition of what is non-parametric we'll follow later on. But this point, don't be under the misconception that non-parametric means there is nothing to estimate. Generally we use the word parameter in a very generic sense as a substitute for unknown; it's a technical word for unknown. But if you think of it this way then that way then non-parametric analysis may sound as if you are not going to estimate any unknowns at all, which is not at all true. In non-parametric analysis also you estimate some unknowns. The reason why it's called non-parametric analysis will become clear at least, when I discuss a case study it will become clear. A stricter, a better term would have been non-parameterized analysis. Unfortunately, whoever has coined this term has decided to use non-parametric and it has stuck on since then. So it's a misnomer in that sense.

Once you go through this visualization, you know, visual analysis, non-parametric analysis then you are in a much better position to make certain decisions that are key for developing the model. Ultimately, I want to develop a model and a model means to us a mathematical form, a mathematical description of the process. And when I write down a mathematical description like if you recall on the first day, I wrote four different models. We talked about Delay. Then we saw that there was some order to the equation whether it's a first order differential equation, second order differential equation and so on. So there are some features of those models which change with the process and as a user it I have to make certain decisions on watch it, what is that delay? What is appropriate order? How many parameters? Is it a linear model? Non-linear model? All these decisions can be made at once you have done a thorough job of visual analysis and non-parametric analysis. And then you select some candidate models. For example, I may say I'll consider all models up to second order and see if-- in which among this candidate models essentially explain my data. It's more or less like an interviewing process. You have shortlisted. You have used the visual and non-parametric analysis as means for short listing the models. Before you begin, the set of possibilities, when I give you data and I ask you what is a model that explains, the model set is huge. It's much bigger than this. These are the set of possible candidates. The moment I apply certain criteria, then which is through non-parametric and visual analysis, maybe I'm now looking at a smaller subset of models. All right? But within this itself the set of possibilities is very high. Now, with the help of data and with the help of a suitable estimator, when I say estimator, it's an algorithm. I'm going to really be-- hopefully, I'll be able to pick a model among this set. And the bad news is that I will never be able to pinpoint, I will never be able to say accurately that this model is the one that is suitable.

Typically, because of uncertainties in the data, I may say that this subset, the model within this set here, any model within the set qualifies to be a good model for the process. So I will never be able to accurately that is a fact, an undeniable fact about data driven modeling, that I will never be able to pinpoint a particular model and say, yes, this is a model for the process. I will always end up with interval models. All right? How you construct this interval? There are two approaches, you learn in estimation theory. One approach is to first construct a point estimate and then draw the region of uncertainty around it, which is a classical approach. The other approach is directly estimating interval models and you pick the point that you want.

For working purposes you need a point estimate. What I mean by working purposes is, when you want to make predictions and so on. You need some point here. Definitely, you'll work with only a single model. But for giving bounce on your predictions, whenever you're predict there are always bounce that are given plus or minus something and so on. You need that interval. So, one more aspect of identification is revealed to us today. All along we have been saying will identify a model from data but the fact is we don't identify a single model; we identify an interval of models. Okay? And how large is this region going to be? What do you think are the factors that will actually govern this region of uncertainty or, you know, that interval. What do you think are the factors that come into play?

Essentially you want to partition this into a very fine space, right? It's like you have your birthday cake and you want to, you are on a big diet and you want the smallest piece. Sorry.

STUDENT 1: Algorithm Efficiency?

ARUN: Okay, so you think that algorithm plays a significant role, all right. Anything else?

STUDENT 2: Number of parameters to estimate.

ARUN: Number of parameters to estimate. Okay? Anything else?

STUDENT 3: Noise.

ARUN: Good. So let's quickly write it down. So one is algorithm, which is estimation algorithm. Second is number of parameters. And three, noise. Anything else?

STUDENT 4: Acquired output or response.

ARUN: I'm sorry.

STUDENT 4: Acquired output or response. Expected output R response.

ARUN: What do you mean by expected response?

STUDENT 4: Just model, when you are initializing we are excepting something to get.

ARUN: That's okay, but what-- see, I have data and I'm asking, and I have data and I have an estimation algorithm together they will help me. Obtain this interval set of models for a process. Question is what aspects of the data or the estimation algorithm or in your system identification workflow determine the width of this interval? Do I want it to be narrow or wide? I wanted to be narrow. I would like to be able to pinpoint ideally but that's not possible. At least from finite data, let me actually complete that statement. When you have finite number of observations, it is not possible to obtain both, an accurate and precise model. That means, I will not get to the truth, number one, and I will not be able to shrink that interval width to zero. Both are not possible from finite number of observations. We will prove later on that when you have infinite observations it is possible to get an accurate and a precise model. In fact, that will be an important requirement of any estimator or estimation algorithm that I choose. That if I were to supply infinite number of observations, I will not be able to, in an engineering sense, large number of observations. If I were to do that then the estimate should give me a more accurate and more precise, in the asymptotic case it should fetch me the truth and there should be no uncertainty about it. That we'll talk about in estimation theory. At this stage, what do you think affects this interval?

STUDENT 5: Data Size.

ARUN: Data size, okay.

STUDENT 6: Optimization criterion.

ARUN: That everything goes into the algorithm. The algorithm is not coming out of the sky. We are formulating it, right? And a part of the algorithm development is this optimization. And you have to be clear in your mind, why am I talking about optimization? Because there are many possibilities and I want to pick the optimal one. I'm surprised you are not mentioning a very, very key aspect.

STUDENT 7: Data Quality.

ARUN: Yeah, but that's a very vague term, part of it is noise, you've already said. Anything else?

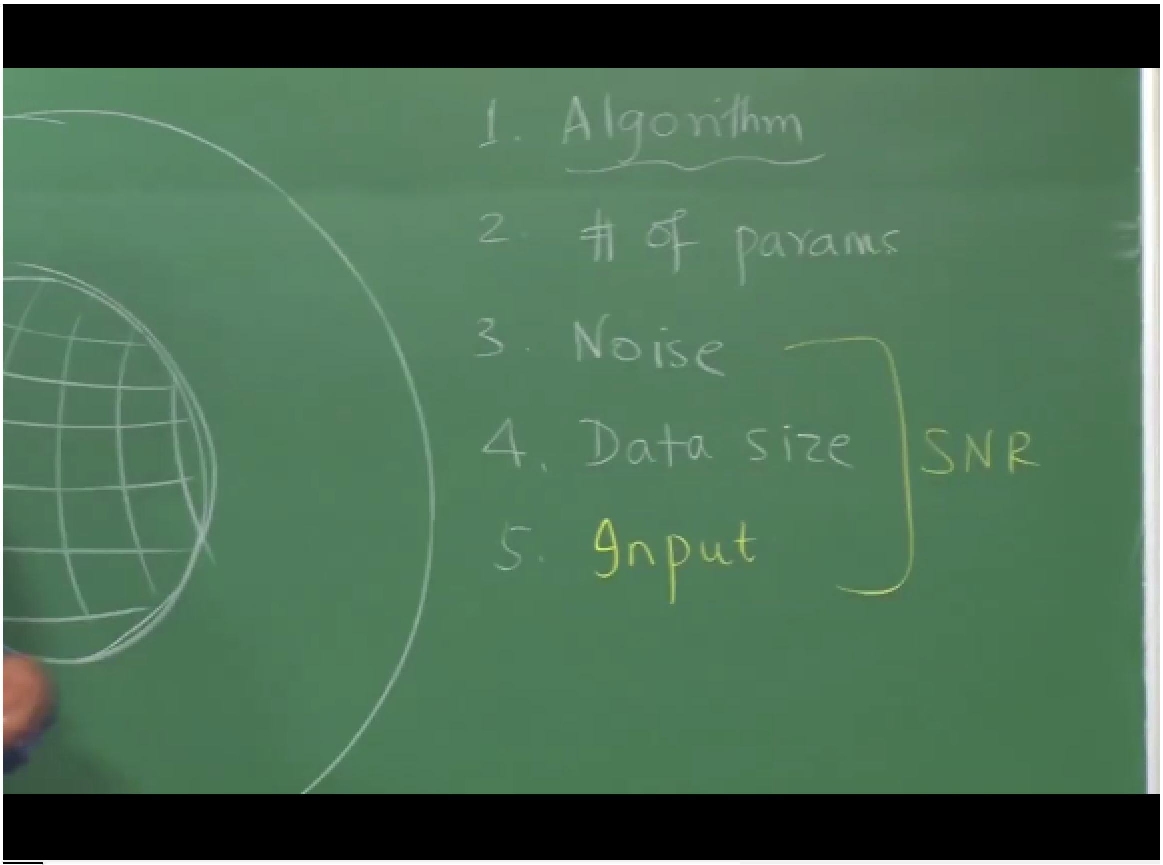
STUDENT 8: Input.

ARUN: Sorry. Correct. Input is an extremely important thing. See, when you are in doubt, when you do not understand identification, go back to the analogy of an interview process. If you think of an interview, what are you actually trying to do? You're trying, you have candidates that are shortlisted, right? So already the short listing has been done by whoever is doing the job, right? It's a tragedy, but somebody has done the job. But the job is not yet done. The interview is being held to find a suitable candidate. Correct? Typically, at least in interviews we hope and we would like to convince ourselves that I have picked the best candidate. But the fact is there would have been other candidates also equally qualified. Let us say that is a case. Now in order for me to pick as an interviewer, the right candidate, don't I have to ask questions. Right? And my ability to select the right candidate depends on the questions that I ask. If I were to only ask, what is your name? And conclude the interview. Do you think that's a very informative interview? Unless, you have made your decision on who the candidate should be. Right?

You would spend some time. What'd you do? How did you come? Was the bus crowded or did the cab driver haunt you? I can ask many questions and make it look like a very elaborate experiment, interview process. But none of the questions I'm asking has no, as any bearing on the information that I required to select the right candidate. So together we call this an identification, such an experiment which generating the amount of information, and quality of information, we call such an experiment an informative experiment. So, one has to conduct informative experiments. And a big factor informative experiment is an input. And that's why input design is extremely important. If I do not ask sufficient number of questions, I will not be able to select the right candidate.

Or I may end up saying , well, four or five candidates are qualified. But that is not what is needed. In the end, the company wants me as an interviewer to select one candidate. For working purposes, one candidate is necessary. You can see on paper there are other candidates and you put them in waitlist. Correct? So, the ones that are surrounding are waitlisted. But for working purposes, I need one candidate and that my ability to select good or the so-called best candidate depends on the kind of questions that I ask. And that translates to input design. So it's very important to design the right input. All of the other factors also play a significant role. I'm not denying, but if your input is not properly designed, if you're not on the right number of questions, or right kind of questions, it's not just a number; I have to ask the right kind of questions. What does it mean in identification, we learn later on.

If I do not fulfill the criteria here, these are the two things and of course, once you have done this together, you have the signal to noise ratio, playing an important role.
(Refer Slide Time: 13:30)



You have to ask the right kind of questions, the right number of questions and with the right amplitude. Suppose I ask a question, you can't hear it. You only hear the AC noise and I'm asking you, what answers will you give me? Your answer will also be noise. That is you will be silent, I'll also hear in-turn the AC noise. So that's not going to get me a good model. So there is amplitude of the input, type of input, the number of questions that I'm going to ask, all of that matters. And that has a huge bearing. Of course, as you rightly pointed out the algorithm and the number of parameters and data sets also governance. So there are this many factors that I have to worry about, when I want to get a good model. Look at the world of difference between what we learn in high school, when it comes to curve fitting It's okay, we are presented with some data, fit a straight line, reported it and that's it, the job is done. That's what we learned in high school. But there is so much to curve fitting. Even the curve that you're fitting is a model. What is-- first of all we in high school, we are only taught how to estimate this point curve, just a single line. But actually there are an interval of lines that can be drawn.

And that's because of the uncertainty. We are never thought that. So system identification, yes, at a very crude level it is curve fitting but there is so much more to it. Okay? And once you have kind of selected, obtained the model, estimated the model, then you have to subject it to validation. It's a very, very important state. Without letting the model go through this gate of quality assessment you should never accept the model. And if you do it, then you have to accept it with a bucket or truck of salt, because you're not sure whether this model is actually correct or adequate for the process. And that involves certain steps like one of the steps that we talked about is, I'm going to see-- look at the residuals, whatever the model could not predict, whatever it is left behind, does it have any effect of inputs? If it does, then that means I have not captured fully the effects of the input. So I have to go back and refine my model structure. And then I may make some assumptions on the noise model. That is I may assume that it is unpredictable or I have fit a noise model and I assume that whatever is leftover is unpredictable even in a stochastic sense.

One is unpredictability with respect to input, the other is by itself it is unpredictable. So there is a second test that involves adequacy of the noise model. Thirdly and very importantly, I should not have over fit, where a third order polynomial perhaps suits the data, I should not end up fitting a fourth or a tenth order polynomial. Right? And always-- also remember this analogy, training a model is equivalent to teaching a student, certain concepts. They presenting-- so what happens is the model learns the data in presence of an estimation algorithm. All right? So you can think of estimation algorithm as a teacher, data as a concept and model as a student. It's very important for the teacher to ensure that the student does not over learn. What is over learn? Is it possible to over learn? Is it possible? Can you think of over learning from a student's viewpoint? Is there something called over learning? Is it possible? I can give you an example. And we have done that, but for good reasons. What is overlearning? When I give you a question in the assignment or any instructor gives you an assignment, the purpose of that question is to help you understand test for yourself, apply the concepts-- a particular concept in this subject and see if you have understood very well. Once you have tested, the numbers in the question do not matter. They are very specific to those questions. Right?

Now you probably understand what I'm trying to get at. If you start memorizing the numbers in the question, and you simply say, well, these are the numbers in the question exact verbatim and let's say this is the answer that is called over learning. You have to let go the numbers in the question, because if you start remembering the question that way, then you're going to be in trouble. What kind of trouble you're going to be in? I change one number in the question, right, in the exam, and that's it, your ability to answer is gone for a task. So the same is the story with over fitting a model. What we mean by over fitting in model is, fitting a structure more than what is necessary to explain the data. What happens when you over fit? It may explain the data very well. So if I give you 100 points, you can free the 99 degree polynomial and come back and say, "Look I have explained the data accurately."

But then I give you a 101 point, your model will produce a disastrous prediction. And there is a case study again that will go over next class, which where the concept of over fitting is highlighted. So you had to watch out for that. And in estimation, you can say one of the symptoms of over fitting have a very prominent symptom of over fitting is large errors in parameter estimates. Over fitting has got to do with how many parameters should have been included in your model and how many you have included. When it says should have included, it's got to do with, what the data is demanding. All right? Maybe the data required only a three parameter model to explain and you have used a very complicated model maybe a 50 parameter model. And remember we always keep saying this, data is food for identification. If you have more parameters than necessary it is as bad as inviting more guests than you can actually feed. You have prepared food only for three guests. Imagine that you are invited 50 guests. What happens? All these guests will go out hungry. They may get only a spoon to eat, right? That is not what you want. You want to feed the guests very well. And you want to feed as many guests as you have prepared the food for.

The same is got to do with over fitting. When you have fit a model with more parameters than there is actual necessity, then all of these parameters will go out hungry. What I mean by hungry is they will have large errors in that. And you don't want that because that means you have a unreliable model. That the next time you ask them to come, they will come but they'll also come with certain ideas in mind, okay? Likewise here, when you use a model that has parameters with large errors, you use it for prediction, it'll show its colors at that point in time. It'll give you a prediction with very large interval and that means you have very unreliable predictions, sometimes these models can give you unstable predictions, unrealistic predictions. So you don't want models with parameters estimates that have large errors in them. Anyway, so you go through this-- you subject the model through this quality assessment and then, determine if you decide satisfactory.

The big challenge is always is when it is not satisfactory. And that typically happens that why itsaysit's iterative. But when you learn the theory of identification, you not only learn how to execute each stage of this identification properly, but also be in a good position to figure out what could have gone wrong, at least a better guess than blind user. You may know as you are doing through the identification if the model goes wrong or it's not satisfactory, this is probably where I have to come back and fix certain things, right? And that is why one should go through a formal study of identification. So that is the workflow of identification.

Of course, we don't have time to go future. But remember, all models should be subjected to critical validation test. And final model quality depends largely on the quality of data. Quality and quantity also. Although I don't mention here but quality is extremely important and we have already discussed what is meant by quality.

So when we come back on Tuesday, we will get started with some very important key concepts in fact, these are key concepts in identification and one of the foremost the foremost concept is identify ability. What is identify ability? It is our ability to identify a model uniquely. You may think that is contradicting what I have just said okay? But I have also said that given infinite observations if I can get to the truth that's also great. So this identify ability constitutes that requirement as well. It demands that when I'm given infinite number of observations, I should be able to get a unique model. That means the algorithm should converge to it to the truth. To one model and that should be the truth. But there are two aspects to identify ability. One is data driven that is taking data into account, the other aspect is a model itself. There are two aspects that is one is the experimental aspect and other is a model that you have chosen. What we mean by that is, just to give you something to think about when you go back.

Suppose, I'm fitting a model of this form and we'll discuss this more in detail. But suppose you're fitting a model of this form, you've decided to fit the model of this form where theta 1 and theta 2 are m parameters to be estimated from data. Forget about data at this moment that is what data, how the experiment has been performed and so on. Just if you look at this model alone, we should ask if there exists a unique model whether the truth itself is unique. Right? What we mean by truth itself is unique is, do they exist, does it exist a unique set of parameters that for the given model. How do I know that? Well I should ask the question whether there is more than one combination of parameters that would give rise to the same prediction.

So I give you an input. What do you think? Do you think that that is more than one set of parameters that will give the same value of y? At this moment keep aside noise, dataand everything. Is it possible? That means there is no unique answer itself. Then what are you going to ask expect of the algorithm? So there is an identifiable issue straight away. This is what I meant by model identify ability. First of all the answer should be unique. There should be a unique point in the parameter space, a single point that will give rise to the one particular prediction. If I find more than one set of parameters that give rise to the same prediction, then there is a problem with identify ability. That means upfront itself there is a uniqueness issue. Why is it important? Because only when I have a unique answer, I can talk about errors in my parameter estimates and so on. Always we talk about errors with respect to the truth, right? So in this case if I were to fit this model, I would have an issue. What is the solution? Solution is to take away this theta 1 and theta 2, perhaps introduce a single parameter which is a product of that. Now, this model has a unique point in the parameter space. For one value of beta you'll get only one prediction. There is no two values of beta that will generate the same prediction. So this model is identifiable. And then I have to ask how I should conduct my experiment?