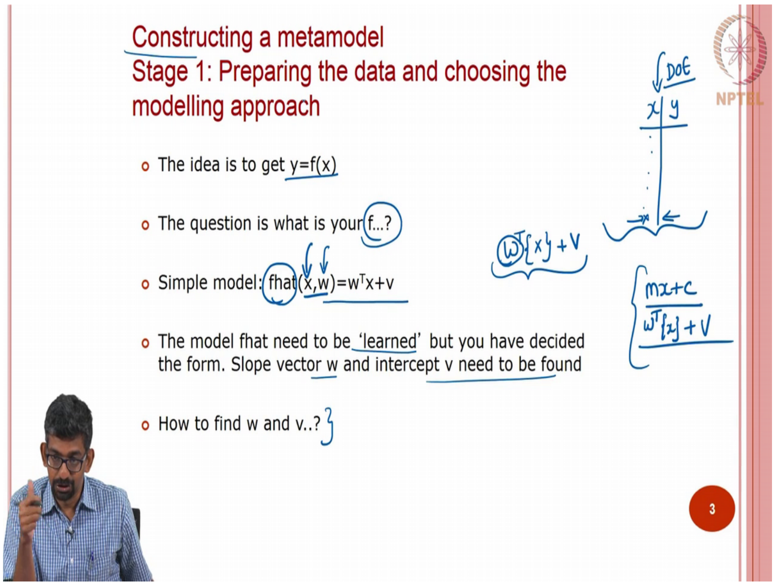
Surrogates and Approximations in Engineering Design

Prof. Palaniappan Ramu
Department of Engineering Design
Indian Institute of Technology, Madras

Lecture – 10
Introduction to Surrogate Modelling

So, we are entering into the important part is; not the important part the meta models part the surrogates part ok; design of experiments is also equal important. So, few slides in this particular presentation is borrowed from Professor Ramana Grandhi from Wright State University and many of the pictures are adopted, the second guys name is Sobester Andras, and the third guy is Keane Andy it is Saveli publication ok, it is Engineering Design visa Surrogate Model, correct one.

(Refer Slide Time: 00:49)



Constructing a metamodel

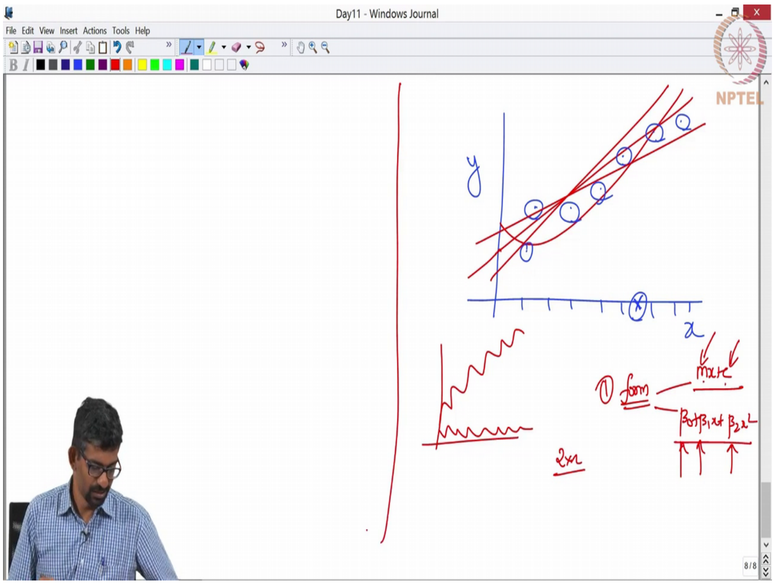
Stage 1: Preparing the data and choosing the modelling approach

- The idea is to get y=f(x)
- The question is what is your f…?
- Simple model: fhat(x,w)=$w^T x+v$
- The model fhat need to be 'learned' but you have decided the form. Slope vector w and intercept v need to be found
- How to find w and v..?

3

Let me just even before getting into the details of this I will just give you an idea, so that you will be able to immediately relate to what we are going to talk about.

(Refer Slide Time: 01:01)


Day11 - Windows Journal
File Edit View Insert Actions Tools Help
NPTEL
y
x
① form
mx+c
$\beta_0+\beta_1 x+\beta_2 x^2$
2xn
8/8

So, you have a set of points.

So, these points came out of expensive simulation let us say, you can given numbers. Now suddenly I wanted to know what is y value at this particular x. I go and run the experiments again that is not the wise way. This we already discussed probably we will do something like this that we will see.

But, there are a couple of decisions that I need to do to even draw this red line, what was that? It can be more decisions I just said couple of sorry. First thing is it is not even clear why I drew a line it could have been a curve ok. So, first you will have to make a decision on what is the form; that is the first thing. Do I look alike like him; you do not know the actual function, but we assume on all the cases that we are going to discuss we are going to validate we are we hope and we built, but in all the cases that we will discuss we know the original so will be able to.

So, you hope based on the pattern. So, look alike that is something that you want to figure out ok, so that is the first thing is the form. Once the form is decided, then you will have to decide where this guy is going to be whether it is like this, or it is like this, or it is like this; those become sizing parameters right. So, the form will tell you whether it is m x plus c or it is beta naught plus beta 1 x plus beta 2 x squared you can keep on doing this ok. But this is a quadratic equation this is a linear line, if it is a linear line I just need to find two coefficients if it is a quadratic I need to find three coefficients that will influence a number of points that I have. I mean those are influenced by the number of points that I have.

So, it is a chicken and egg problem ok. So, you say that I have 3, so the usual thumb rule is 2 times n where n is the number of coefficient that you need to find, you need to have at least 6 samples.

So, let us imagine that you have only 5 samples then you cannot do much ok. So, you can only do a linear fit. Then what happens you say no I should have done, but then this was done using a Latin I per cube sample. Now you cannot just go and add one more sample, because the space willingness property is lost. You understand what I am saying. So, you need to answer all these questions of front: what is your budget, what is my likely. So, you do not even have these points, you need to be able to tell me what is the likely response variability is: it going to be linear or will it way.

So, then what it that is why they say that you use only 25 percent of your total sampling to understand the problem first, then you worry about a doing a detail analysis. So, first people want to know. And sometimes what happens is you do not have control on this data also, you put a sensor, you do not have control on data meaning you get the data that is the good part, but you do not have control on the data ok. You cannot say- boss tomorrow you run at 40 kilometres per hour day after tomorrow you ride only at 60 kilometre, no you cannot say that. You sell it to him you say both I put the sensors, so I will get all the data of your test right that is all, and we pay you this much monthly for that or that is all. But you do not have control on saying that guy you do this, you do this you cannot say that.

So, sometimes these blue data are not under your control, that problem is also there. So, you will have to live with the data that you have,. So, the first point is I need to decide on the form of my function then I have to go and estimate these guys ok. Then there is an inherent question that comes; you have got this fit, but how good is this fit with respect to the original points, why should I even believe you are fit.

You remember this curve that we are talked about yesterday original and then your model, that problem is also there, ok. This is just to set the context now will go back to.

So, I have just taken up from forester’s book and I am calling it as a first step basically. You want to construct a Meta model that is your overall goal, I want to construct a Meta model. So, first thing is you have some data, just like what I plotted just now you have some x you have some y. Sometimes you have control on this x where you generate your y, sometimes it is just a data. If you have control then DOE comes into picture but not always, sometimes it is experimental data given by some other team and all that. But the underlying assumption for this module is you have the data. If I have the data my first thing is I want to get y equals f of x.

So, the first question is what is your f, how does the f look like? I am not asking for the values, how does the f look like. So, there is a general recommendation to use this structure, not necessarily it is linear please understand; it is a generic structure. What it says is; use a weight vector times your input variables plus some constant equivalent; it could be an error also ok. You should immediately be able to relate this to something what is that, just now that we draw.

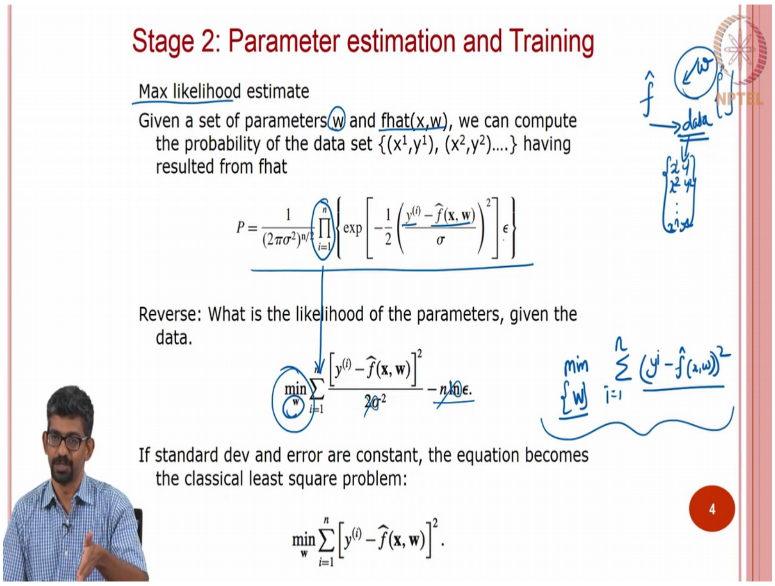
It is nothing but, but please understand this is not limited to a linear ok. I am just replacing m with w a generic representation of my slope as weight times x, when I put that parenthesis it says sorry brackets it says it is could be x 1 x 2 x 3 anything plus I am just calling this guy as a v non just the constant good enough for us to start ok.

I do not call this guy the f, I call this guy the fhat because he is only a representation not my actual function; I do not know the actual function. Please note this there are two things: of course, x is an input. So, for give me any new x I will be able to find your y that is what this one says. But, it is also a function of my parameters ok, this guy you change this my f hat will change ok. What does the w mean it is your m and c correct, it is your m and c that is what your that is basically the coefficient. So, coefficient vector is what is written here as w transpose.

The model fhat needs to be learned; this is comes from the machine learning kind of a learnt ok. So, basically you are going to build a model and the model will learn; I will learn from him on that particular stuff, he will learn from me on that particular stuff ok. So, only those data points, I am not going to know everything about goal only that data point particular data point. So, it needs to be learned, but you have to decide the form that is important. If this guy is trained on the same stuff he met probably do better than me, but he cannot go in this guy's of Sundar Pichai and present there is a problem. So, the form should match number one; the form should match look alike should be there. Then I have to be trained that is what it does ok.

The slope vector whatever w and the intercept v need to be found; how do you find w and v? Now, you need to find w and v because x is known, how do you find w and v. We already discussed this once, but we will introduce a more generic way of looking at. So, the first step is preparing the data and choosing the form, you need to choose the form that is the main part of this right; are the modelling approach whatever.

(Refer Slide Time: 11:07)



Stage 2: Parameter estimation and Training

Max likelihood estimate

Given a set of parameters w and fhat(x,w), we can compute the probability of the data set {(x¹,y¹), (x²,y²)....} having resulted from fhat

$$P = \frac{1}{(2\pi\sigma^2)^{n/2}} \prod_{i=1}^{n} \left\{ \exp\left[-\frac{1}{2}\left(\frac{y^{(i)} - \hat{f}(\mathbf{x}, \mathbf{w})}{\sigma} \right)^2 \right] \epsilon \right\}$$

Reverse: What is the likelihood of the parameters, given the data.

$$\min_{\mathbf{w}} \sum_{i=1}^{n} \frac{\left[y^{(i)} - \hat{f}(\mathbf{x}, \mathbf{w}) \right]^2}{\cancel{2}\sigma^2} \cancel{- n \ln \epsilon}.$$

If standard dev and error are constant, the equation becomes the classical least square problem:

$$\min_{\mathbf{w}} \sum_{i=1}^{n} \left[y^{(i)} - \hat{f}(\mathbf{x}, \mathbf{w}) \right]^2.$$

$$\hat{f} \quad w \quad \to \text{data} \quad \{x^1 y^1,\ x^2 y^2,\ \ldots,\ x^n y^n\}$$

$$\underbrace{\min_{\{w\}} \sum_{i=1}^{n} \left(y^i - \hat{f}(x,w)\right)^2}$$

The second is once you have decided the form you need to decide what are the coefficients or what are the parameters. One usually use parameter estimation technique is called the Maximum likelihood estimate. It is a very interesting concept it is also used in other places, but widely used for parameter estimation. It is a very interesting concept just focus here, what it says is it is it is an inverse modelling kind of an idea ok.

Given a set of parameters w; so I know the parameters and fhat I give you the model ok; meaning I give you the f hat and I am also giving you the parameters w. Obviously, you can find out the data; meaning the outcome of the f hat you can do right. If I give you these two you can compute see meaning there is some data that I have ok, if I give you the f hat and w you can tell me what is the probability that this data came out of this information; that is a straightforward problem. That is what we have written here ok.

Given a set of parameters w and the f hat I know as a function of x and w we can compute the probability of the data set this is the data set x 1 y 1 x 2 y 2 x n y n ok, having the resulted from that f hat without giving this w it is not possible I need to give you the w and I am also giving you the f hat, and I am asking what is the probability that they given data set comes from this f hat with that omega or w.

So, if I assume that it is a Gaussian distribution with an error epsilon, this is what it is ok. You should know the Gaussian distribution to appreciate this it is just a Gaussian distribution function the normal distribution function ok. And just replacing that x by this one, where y I is the actual estimate and f hat comes from the model that I have built that is all. This is the small error that we are talking about ok. Now just note that this is a product ok.

Now just invert the same understanding the equation should whole good; what I am gone ask you is currently what we did is f hat is given, omega is given, and I give you the data, and I am asking what is the probability that this data came out of that. But I can also ask you another question which says that if this data came out of this model what was the w? Ok.

So, what you can do is this you can assume some w I am sorry; you can assume some w and you can pose the same type of question what you can say is given the data and the f hat can you tell me what is the probability that the data came out of this w? And I will maximize that probability so that meaning like I will play around with the w and then I will maximize the probability and I will take that w for with the probability is maximized. You get the point this is simple stuff.

So, you take f it is a function of omega I give not omega w I also give you the w and I am asking you what is the probability that this data came out of that model. The inverse question is I give you the data, I give you the f hat, I am asking you what is the probability of that w? Ok. So, the reverse is what is the likelihood of the parameters given the data? So, what you do; the error still remains the same it is y thing minus that is squared, but please understand we are doing a log of this one ok.

So what; I mean this is a maximum likelihood because this is a product it is easier to deal with a log because it becomes a it becomes a linear sum ok. And then since this guy is an epsilon meaning the error he becomes epsilon n every time you do a product. So, it will become n long epsilon ok. And this sigma comes from this sigma.

So, here you are going to play around with the w and you want to minimize this quantity ok. If you did not know the normal which are this equation is difficult for me and all that do not worry about these guys. If you are not very comfortable about your Gaussian norm just ignore these guys. The problem is only this much; minimize w meaning find the set of w the weights such that for all the points that you look at; can you tell me this is the original value and this is the predicted value. This is the error that you are computing and penalizing you for more errors so I am squaring there is another reason also because the error could be positive or negative, but for that you can do an absolute value also there are two ways of looking error ok.

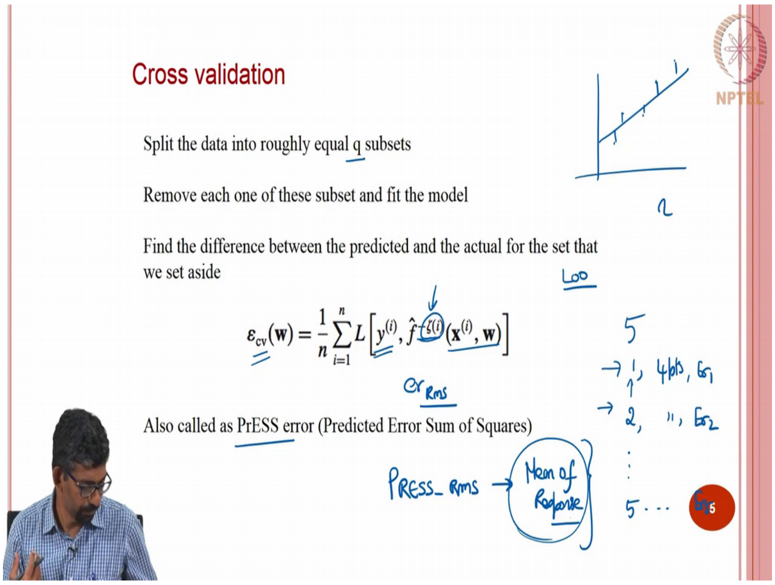
So, I am squaring the error and I am going to change; when I write it like this I am going to change the w. When I change the w what is happened is f hat will change. So, for every iteration I will go and change the w not for i; every iteration means every time the w is fixed and then I will compute this. Then the sensitivity information I will go and change the w, then I will again compute this information ok. But you seen the somewhere else not in the equation sense, but we have discussed the somewhere else where did we discuss.

Student: (Refer Time: 17:17).

This is nothing but your least square ok. What you are going to do? Least square error, this is the error that you are talking about. The error is squired and by minimization what you are doing, you are trying to find the least error. So, find the w that will give me the least error ok. This is the idea of minimum least squared. If the standard deviation and the error are constant the equation becomes a classical; least square error problem that is all.

So, this is a more generic way of writing or finding the parameters least square is a simple simplistic case of that that is all.

(Refer Slide Time: 17:43)



Cross validation

Split the data into roughly equal q subsets

Remove each one of these subset and fit the model

Find the difference between the predicted and the actual for the set that we set aside

$$\varepsilon_{cv}(\mathbf{w}) = \frac{1}{n}\sum_{i=1}^{n} L\left[y^{(i)}, \hat{f}^{-\zeta(i)}(\mathbf{x}^{(i)}, \mathbf{w})\right]$$

Also called as PrESS error (Predicted Error Sum of Squares)

So, we will talk about minimum least square and all that I will leave it more in detail ok. But, the point is now I guess now you appreciate why we spoke about optimization and all that. If you see that in Latin I per cube also you need to know how to minimize maximize, here also you need to know how to minimize maximize. So, you should at least appreciate that you might not be able to immediately solve, but you know that if you spend another 5 hours 6 hours you will be able to solve such problems. So, minimizations you need to have an appreciation ok.

So, now, the question is; I fit the model, but how good is this model how you know. So, you need to worry about how good this fit is, ok. So, then there is something called cross validation. So, you are going to cross validate with respect to your original values. But please understand because you have data which is what you consider original, and you want to train your model on the original data, but after the training also you want to, because the trained model need not predict the same value as your original. That is why that is what we showed in your in that graph line I may not have it here.

So, what I am trying to tell you is: if these are your data my line need not pass through the data. So, there is always an error positive or negative. So, will see what criteria you will take to get that minimum error anyway. So, the deal is in cross validation what you do is this you split the data; the entire data into equal q subsets meaning you can divide it one at a time or you can take 2 points at a time it does not matter 3 points n points whatever it has ok. And what you do is; is you remove one subset at a time, you fit the model with the remaining points, and using the fitted model you estimate what is being computed at the left out point. This is important, because press error or cross validation is one of the most important things for you to understand how your model works.

So, the question here is there is n data points I am going to divided into q sets. So, the subsets can be value it can have one point or you can have two point or it can have n points ok. Then what you are going to do is you are going to drop one of these subsets; at a time at a given iteration you drop one of these points you fit the model to the remaining points, you use the model to predict at the left out points. So now, you have the original value at the left out point, you have the predicted value at the left out point that will give you an error.

So, like this you can keep doing it for all sets. We will have a set of error metrics that will tell you how good your model is in a local sense as well as in a global sense, because you can take an RMSE value for the global ok; one simple question in a simple point ok.

The point is let us say that you have 5 points ok. So what you do is, you leave for the first iteration you leave point1 you use the remaining 4 points, you fit the model you use the model to predict at point 1. So, you know the original value at 1, you also know the predicted value. So, you have error 1. Similarly you go to point 2 you have other 4 points 1 3 4 and 5; so 4 points are there, you fit the model, you come back and computed point 2 you will get an error that is error 2. Similarly you do for all the 5 points. So, you will have error 5.

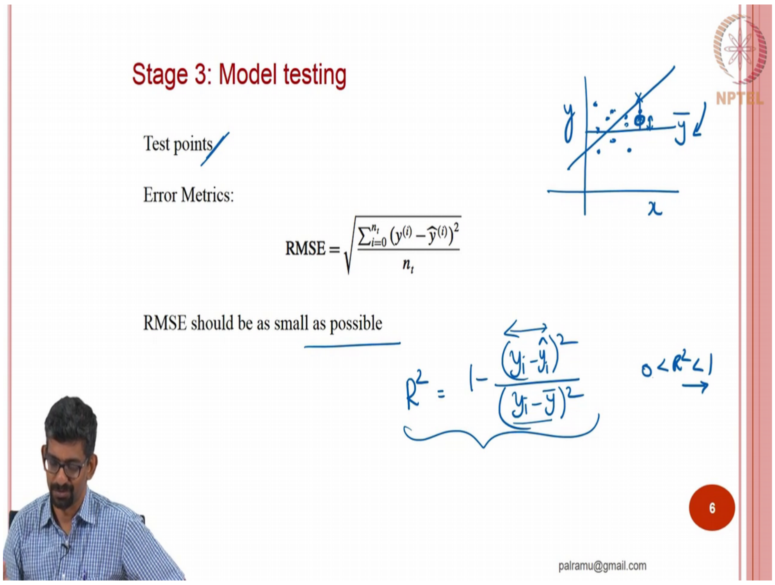
So, this is an interesting metric we will see how to, but take care I mean how to use this metric, because usually in a in the machine learning or the computer simulations kind of a scenario what people use is immediately they divide the sample into training data and test data. So, you lose the information, meaning lose information in the sense you have only 20 points, but you let it divide that into 15 and 5 that is not a great way of doing, because 20 itself is not enough for me. And then now you are going to take only 15 out of that to fit and then 5 of that to test the model ok. And there is always an argument on how much should it be, should it be 50 50 60 40 70 30 all those discussions are always there.

So instead, what this gives you an advantages it says that you are going to use all the points for fitting as well as for testing. So, that is a good idea. So, you are not losing on the point at, but here there are other questions. So, how many points do you leave it? Do you leave 1 point at a time, do you leave 2 points at a time or leave so that is why it is called K fold validation. And people have figured out that you can leave and look at your data because these are all only computational right you are not going to go run an additional simulation or experiment this. So, you can use it on your computer and you can figure out.

They usually saturate at about 6 points ok, but if you are using only 30 data point maybe 2 at a time is the best that you can do ok. So now, I compute all these errors. So, what this equation says is very simple ok; what it says is it says L stands for the likelihood kind of a stuff. So, it is says y i is your original data this f hat is not the original f hat it is the origin f hat minus the eta that i have dropped, that eta could be one point or it could be the two points that is what it says; f hat with the drop the data that is why it says minus of eta ok. So, if I left out one point it is called a leave one out analysis. So, leaving one f hat, because that f hat is not the same f hat the model is still the same see, please see. I mean still use a quadratic model, but I am fitting without the first data, I am fitting without the 3rd data. So, this is what it says- for I for each data that I am leaving out I am leaving out that particular data and x of i whatever x of i and w are the same.

So, then you want to; so basically this and this prediction will give you an error and you will have to minimize that error sorry not minimize that error you should just take the error and then you can have point wise errors like this or you can take an RMSE error. You can say error RMSE ok. This cross validation is also in structural engineering is also called press error which is predicted error sum of the squares, ok. This is error cross validation that is what it says they are just using a sum here,.

(Refer Slide Time: 24:43)



Stage 3: Model testing

Test points

Error Metrics:

$$\text{RMSE} = \sqrt{\frac{\sum_{i=0}^{n_t} (y^{(i)} - \widehat{y}^{(i)})^2}{n_t}}$$

RMSE should be as small as possible

$$R^2 = 1 - \frac{(y_i - \hat{y}_i)^2}{(y_i - \bar{y})^2}$$

$0 < R^2 < 1$

So, now I have fitted the model cross validation can also be used as a metric ok. It gives you an idea about the fit, but it can also be used as a metric. There are other sorts of metrics based out of test points using other additional test points. You can come up with the root mean square error it simple at the test point you find out what your y is and then you know what your y hat from your f hat is and then total number of samples, you just take an RMS. RMSE is what root mean square, you take the square of the errors, you take the mean value, you take the root value.

So, it is a root mean squared error that is all The idea is RMSE should be as small as possible. What about your press error? Should it be less, should it be more? Might be difficult for you to answer, but usually what you do is your press RMSE is usually compared to your mean of your response. There is some meaning to this. This is something similar to your R squared error. Do you know what your r squared error is?

Student: Y square.

Y i minus y hat.

Student: Y square, divided by y i minus y bar which is average that is much.

That is all?

Student: (Refer Time: 26:12).

That is all?

Student: 1 minus was that r something (Refer Time: 26:17).

That is it?

So, if this was 0 when you get a; this was f this was entire then you fine that is correct. So, this is the expression for your R squared ok. So, what it tells is it is a simple idea if we look at it. If you have a bunch of data without looking at what is the form we did not take the Meta models cause you do not know about y equals m x plus c ok. The moment I give you a bunch of data x and y what is the least thing that you can do. The moment you know that you do not get the same value for x every time I mean you do not get the same value of y for different x's. What is the easiest the least thing that you can do when I give more than one data, which is the least approximation that you can do.

The moment you have more than one data the easiest thing that you can do is the first moment which is your mean. In one sense what this metric R squared tells is that is the least approximation you can do. Now what I will do is this I will compare your model to this; that is what it is doing here. What it is saying is y i is the actual value. So, let us say that you have predicted this curve, you say that this is the curve is the best.

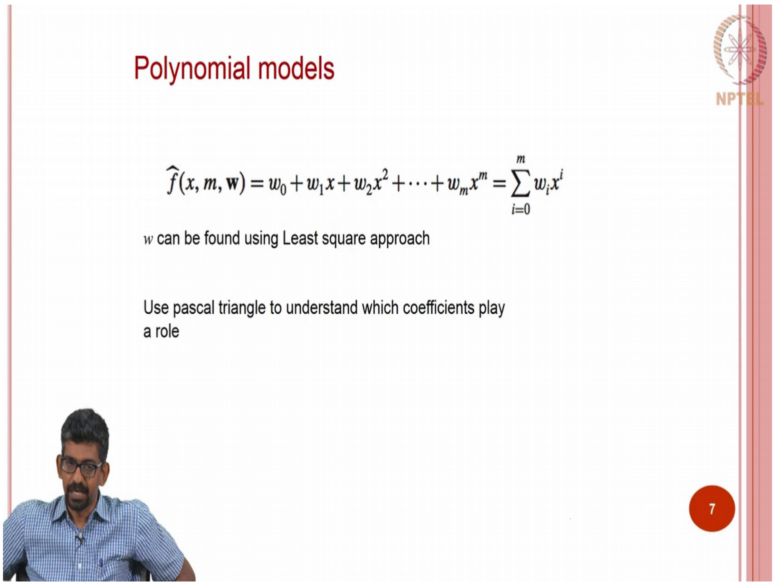
So, what it is saying is if your curve was as good as only y bar it will be kind of equal to 0, ok. Because if this is as good as y bar then this ratio will become 1 and this will become 0. R squared usually varies between 0 and 1, you want to be closer to 1 for a good fit. So, it is normalization.

So, it says if it is only as good as y bar it is 0, if it is slightly better than y bar then I will slowly start increasing it there values. So, what it says is at this point this is a value that is predicted and this is the actual value. So, it gives you some error value and this is the actual value and this is the mean that takes this value ok. So, you take a ratio and then you take one out of it ok. So, this is the R squared metric that is usually used. And people say this is that, that also we will talk about variance explained residual sum of the squares, regression sum of the squares, total sum of the squares; we will talk about all that.

So, this R squared is also one of the matrix that you can use. In a similar fashion press does not have something like this, it does not mean that if it is greater it is good if it is lesser; I mean the lesser it is it is good, but you do not know whether 10 is lesser 50 is less it depends on the mean of your response ok.

So, usually press error is compared with the mean of your response to understand how much of the variability it captures ok.

(Refer Slide Time: 29:42)



Polynomial models

$$\hat{f}(x, m, \mathbf{w}) = w_0 + w_1 x + w_2 x^2 + \cdots + w_m x^m = \sum_{i=0}^{m} w_i x^i$$

w can be found using Least square approach

Use pascal triangle to understand which coefficients play a role

Now for breeding any Meta model there are three steps: that was the ones that we discussed just now. The first one is you kind of need to get some model information which model, the second is meaning the form of the model. The third; second one is each model has some parameter; might not be as simple as your slope and, but in one sense they are all coefficients ok. So, you need to find the corresponding coefficients the figured out that it can be done using the maximum likelihood sometimes it can also be done using a regular regression. But regression is only a special case of Maximum likelihood. The third one is I want to know how good my fittest you can fit something to the data, but how good is my fit.

These are the three steps that you need to widely use, that is all ok.