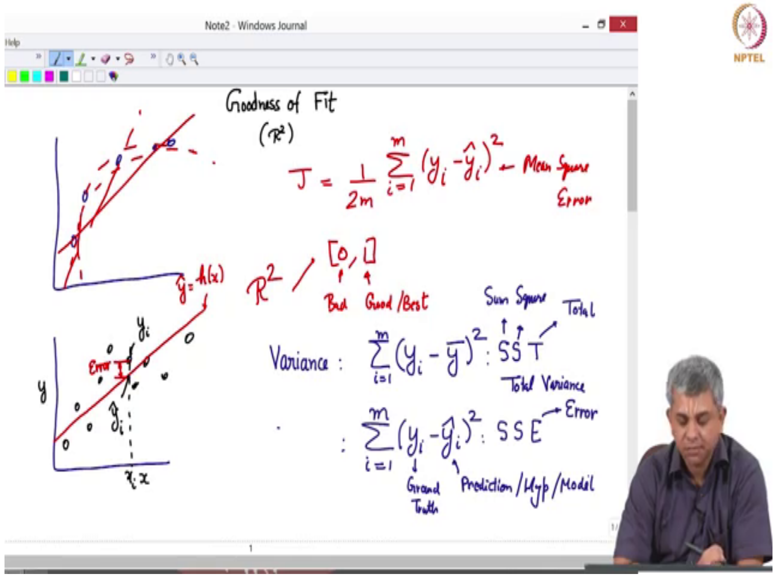
Machine Learning for Engineering and Science Applications

Professor Dr. Balaji Srinivasan
Department of Mechanical Engineering
Indian Institute of Technology, Madras

Goodness of Fit

(Refer Slide Time: 00:15)



Welcome back we have been seeing various types of Fit for the same data so for example for the same data we saw a linear fit, we saw a quadratic fit and we also saw a cubic fit. Now intuitively we could see that cubic fit is better in this case compared to a quadratic or a linear fit but sometimes it is not really visually obvious especially if you are dealing in high dimensions ok for example if I have this line versus this line which one is a better fit. So that is one question so immediately we might see or we night say that obviously I should look at J ok to find out how good a fit is and my J in that (ca) that case was 1 by 2 m sigma i equal to 1 to m y i minus y hat i square.

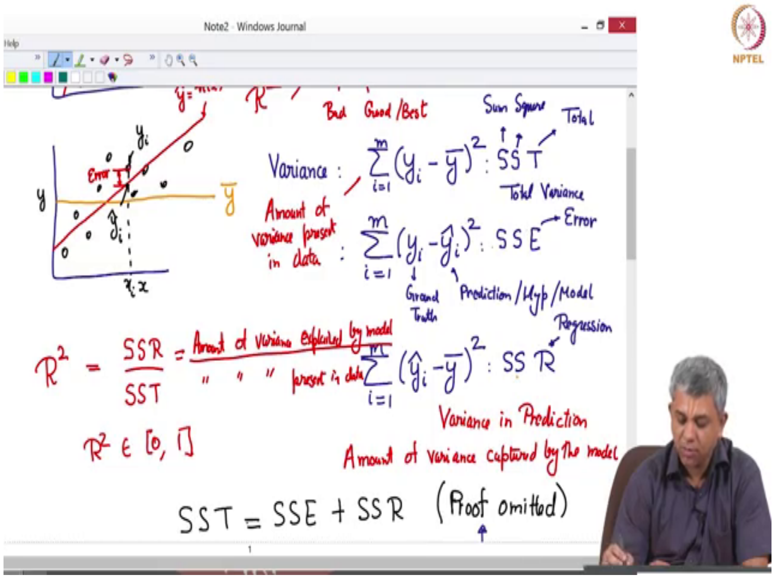
So basically this was root sorry not root but mean square error ok so this is one measure of how good the fit is but sometimes this is not a good enough measure for multiple reasons sometimes we just get a large value of J and we don't know whether this is a good fit or not typically we would like one number which lie between 0 and 1 where we can say something like 0 is a really a bad fit and 1 is a very good fit ok so we want to normalize this, this kind of thing will repeat again and again. You have a number you would like to non-dimensionalize it normalize it with respect to some denominator so that you get an idea between 0 and 1ok.

So we will try and do that a measure for that is something called R square ok so where are going to is R square will lie between 0 and 1 where basically 0 means really bad and 1 means great ok this probably the best fit that you can get ok. For his we need three different measures of sort of variance and data and (looks)let us look at this ok let us look at this three different statistical measures. So let us say once again that this is our original data and I have my hypothesis function this is h of x this is my original data x versus y ok. Now the ground truth at any particular point this is the ground truth y but what I am predicting is y hat ok.

So let us say this is the point x i this is y hat i and this is y i, I should really use super script but hits is a little bit convenient so I hope you understand. So the difference between this two is some error, this is the predictive error. Originally that data that is going here and there has some variance remember what is the variance? The variance is summation of y i minus the mean square. So this was you might recall from our probability week this is simply the definition of variance ok. This is sometimes called SST where S Stands for sum the second S stands for square T stands for total. So this term is known by various (term) terminologies sum square total or basically this is the total variance.

What does total variance mean? Mean before we even had a model there was some amount of variation in the data ok there are some amount of variation in the data and this term actually calculates the total amount of variance in the data before we even had a model ok. Now we also have this previous term closed to the previous term anyway so let us look at the second measure of error y i minus y hat i square this is our error in prediction and it is known by the term SSE where E stands for error. So the meaning of this term is fairly clear this is simply the difference between y i and y hat i where y i is the ground truth and y hat i is our prediction or our hypothesis or our model.

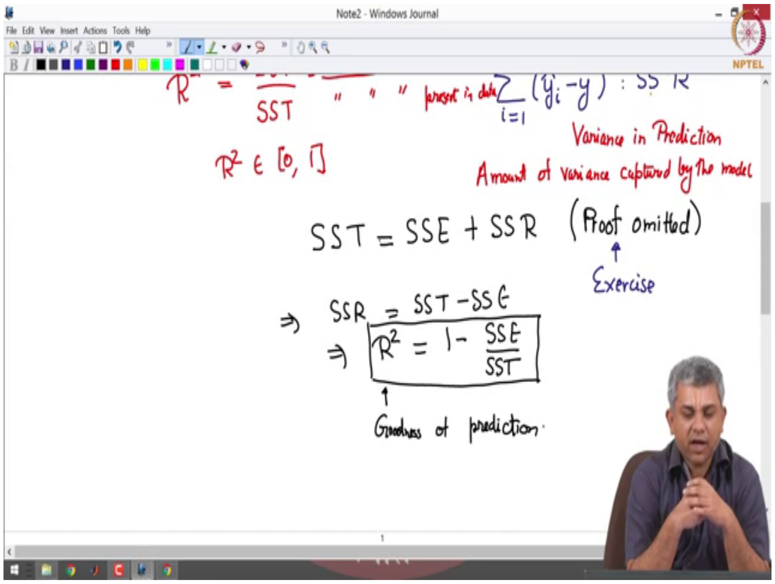
(Refer Slide Time: 05:44)



So this term is known as SSE there is a third term or a third measure of error this is y hat i minus y bar square I will name this term and then explain what it physically represents once again SS is sum square R stands for regression, now what does this denote? So suppose I mark y bar here what this says is just like variance told you how much this y i vary from y bar this term SSR or sum square regression tells you how much does your prediction vary from y bar ok. So this is the variance in prediction, statistically we will go into much detail about this this is called the amount of variance captured by the model and the first term is amount of variance present in data. So R square is defined as SSR by SST physically this means amount of variance explained or captured by model divided by amount of variance present in data and it can be shown R square will always be between 0 and 1 ok.

In the best case scenario your model actually predicts all the variance which is actually present in the data and in that case it will be equal to 1. Now it turn out that there is a nice relationship between SSR, SSE and SST it turn out and I am going to just say this without proof you can try the proof as a exercise it is a slightly tricky proof I must mention that in case you wish to try it out for your own edification so this is you can try this as an exercise ok. So obviously you can use this to realize that SSR is equal to SST minus SSE which gives you that R square is the same as 1 minus SSE by SST ok this is usually the form in which it is implemented because we are anyway calculating this term. The sum total error or the square sum error because that is really just the non-scaled part of j ok.

(Refer Slide Time: 09:34)



So what we have seen in this short video is that we use R square I should not really use the term Goodness of Fit because it has several technical meanings but atleast one number which is often used this something called R square in fact if you use many of MATLABS in built routines they will tell you after you make a particular model even in neural network model because this is really general if you use a neural network model it will tell you your models R square is so much ok. So if you have something about 0.95, 0.96 that is a great model that is a very good model because a lot of data actually has been captured by or the lot of variation in the data has been captured by what your model is and occasionally we will be talking about R square for various problems with respect this course thank you.