Essentials of Data Science with R Software - 2
Sampling Theory and Linear Regression Analysis
Prof. Shalabh
Department of Mathematics and Statistics
Indian Institute of Technology Kanpur

Linear Regression Analysis
Lecture - 52
Variable Selecting using LASSO Regression
Introduction and Basic Concepts

Hello friends, welcome to the course Essentials of Data Science with R Software-2, where we are trying to understand the topics of Sampling Theory and Linear Regression Analysis. In this module on Linear Regression Analysis, we are going to begin with the new chapter on Variables Collection using the LASSO Regression. What is this LASSO regression? That is the first question we have to understand.

You know particularly, when you are trying to work in the area of data science. The usually the data is collected in an automated way. And the cost of data collection is very less. For example, you can imagine that, whenever shopping websites want to collect the data they will try to create the login details.

So, any customer who want to do some shopping from the site, the person has to first give all the information about herself or himself. And based on that there can be many more question from time to time, which can be answered by the customer for that website. So, now, the data is collected automatically. The cost is very very less; the administrative difficulties are very very less.

So, now, in this process what is happening that sometime the experimenter become very enthusiastic and they try to collect the data on large number of variables. Definitely if a model has got large number of variables the model will be good. But there is a condition the condition is this, the variable have to be important variable. If you are trying to collect the information on those variable, which are not contributing in explaining the variation in Y, the model will not remain as good.

So, the question now here is, that how to choose the important variables? So, for that we have considered the aspect of test of hypothesis, but definitely when you are working in a such a huge setup, where you have millions and billions of data and hundreds of

variables, these things may not work very well. And then in that case we have to look for some other alternatives.

Well, I am not saying that those things are bad, right remember. I am saying that because the data is so huge though. So, for example, obtaining the regression coefficients by inverting the matrix X'X might be very difficult. You cannot obtain the inverse of a matrix of say 1 million by 1 million observation that is computationally very challenging.

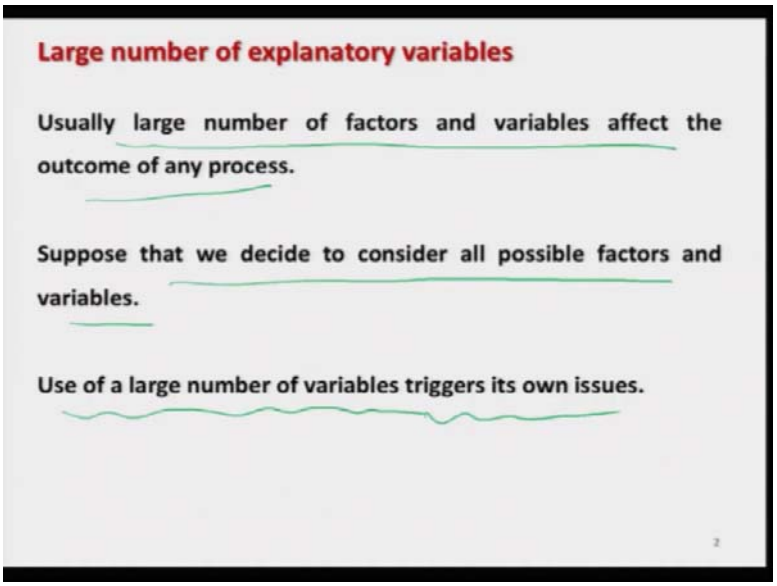
So, the question is we have, when we have this type of setup, we have large number of variables. How do we find out the important variables? Then only the important variables can be selected, they can be used in the construction of the model. So, this is achieved by LASSO regression.

So, now, in this chapter we are going to discuss, that what is LASSO regression? But it is very important for you to first understand that how LASSO was developed and what is happening inside the LASSO regression. So, that you can take a correct out correct interpretation.

So, in this lecture I am not going to give you much mathematical details. But I will try to connect couple of things together. I will *t with the linear regression modelling then, I will try to come on the independent variable, then I come to the problem of multi collinearity, then I will try to come to the ridge regression and finally I will jump into LASSO regression.

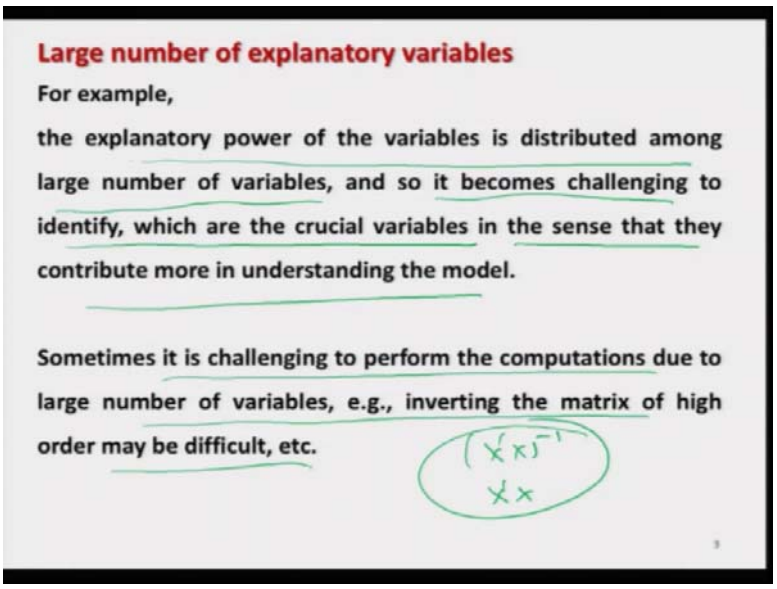
So, this is how I am going to give this lecture. You please try to have some patience so, that you can interconnect all the things at the end. So, we begin our lecture, ok.

(Refer Slide Time: 04:36)



So, now, we know that usually large number of factors and variables they affect any outcome of any process. And suppose we decide to consider all possible factors and variable. So, that means, we are using large number of independent variables and this will trigger its own issues.

(Refer Slide Time: 04:58)



There will be many complications which may arise when we are trying to consider large number of independent variable. For example, the explanatory power of the explanatory variable is distributed among large number of variables.

You have done it by sum of regression in the analysis of variance, where you obtain the sum of a square due to total that was divided into sum of a square due to regression and sum of a square due to error.

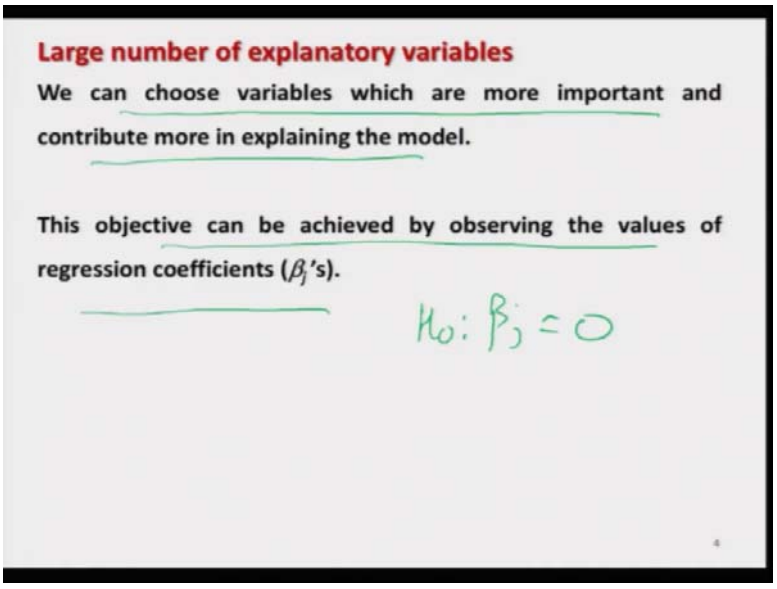
Now, the sum of a square due to regression is divided into large number of components. So, it is possible that every component becomes so small, that you cannot even judge that, which of the contribution is small and which contribution is large.

For example, if there is a value say 20, which is now divided into suppose 100 components. So, every component will become very small. So, in such situation it becomes challenging to identify, that which are the crucial variable, which are the important variable in the sense that they contribute more in understanding the model.

And sometimes when you have large number of variables, suppose the number of independent variable $X_1, X_2, \ldots, X_k$ they become very large, then your matrix like $(X'X)^{-1}$. This is what you have to find.

Now, the matrix X'X become so large, that it is difficult to get an inverse of this matrix and if you cannot get the inverse of this matrix even using the algorithm, then how you will estimate the parameters. So, these types of mathematical challenges also come into picture, ok.

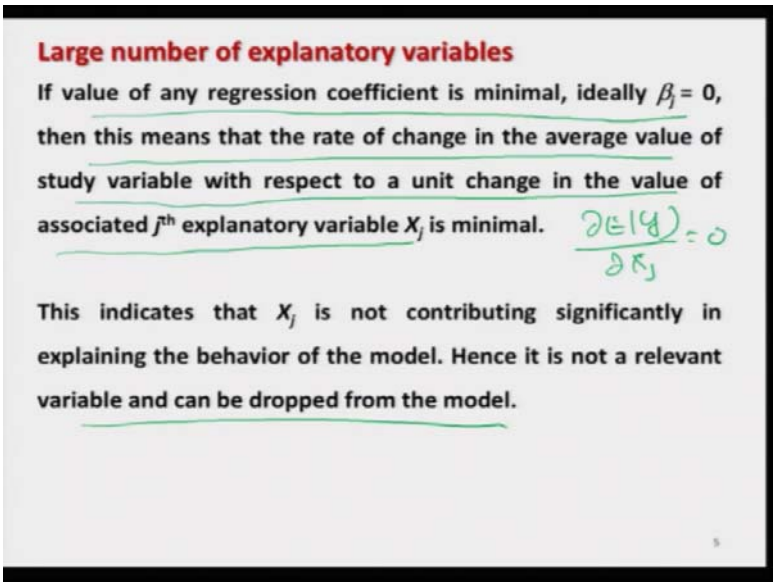
(Refer Slide Time: 06:33)



So, now, we our objective is this, we have to choose those variable which are more important and they contribute in the in explaining the model. So, this objective can be achieved by observing the values of regression coefficient that we already have discussed.

If you remember we had discussed that, if $H_0 : \beta_j = 0$ is accepted then the then how the model is revised, right.

(Refer Slide Time: 07:02)



So, that we know that if the value of any regression coefficient is very small say, ideally $\beta_j = 0$, then this means that the rate of change in the average value of study variable with respect to the unit change in the value of associated jth explanatory variable X_j is very small.

You have done that partial derivative of expected value of Y with respect to $X_j = 0$, right. So, this indicates that X_j is not contributing significantly in explaining the behavior of the model. And hence, I can say that it is not a relevant variable and this variable can be dropped from the model.

So, think about a situation where you have large number of variables. So obviously, when you have large number of variables there is a very high chance that these variables might be inter correlated. And you have a one basic assumption in multiple linear

regression model that your all X_1, X_2,…, X_k are linearly independent, the rank of X matrix is k.

So, now, this assumption is violated. The independent variables are correlated. Once they are correlated then this creates the big question on the non singularity of the matrix X'X, finding out the inverse X'X becomes difficult. When the inverse of X'X is difficult to find or it gives you a wrong value. Then the standard errors of the ordinary least square estimator of β, which are $\sigma_X^2(X'X)^{-1}$ they will also become very high.

So, it is possible that; if your variables are in are not independent, then the standard errors of the regression coefficient turn on turns out to be very high. In fact, this is one of the way by which we try to understand, that we are trying to assume that my independent variables are independent. But are they really 100 percent independent or not, this is how we try to see we try to look into the standard errors of the ordinarily square estimator regression coefficients, for the regression coefficients.

Now, the question is this once you have large number of variable there are very high chances that those variables may be inter correlated somewhere and it is very difficult to find for us. Under this stage now, ordinary least square estimator cannot be used. This will give us a very bad result, wrong results this cannot be used..

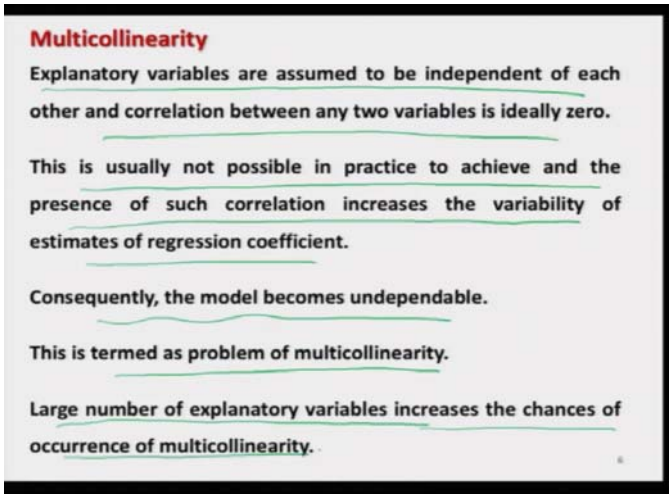
So, now, the first question is this; we have to put some condition on the ordinary least square estimation methodology. And we need to find out the estimators of the regression coefficient in a different way. So, that we can estimate the parameters correctly.

As far as this problem of correlation of independent variables is concerned, this is actually called as problem of multicollinearity. And how to obtain the estimates of the parameters under the problem of multicollinearity this is very difficult and no, 100 percents good results have been obtained up to now.

One of the very good method which gives us a good outcome is the ridge regression. So, ridge regression puts a penalty, on thus on the sum of square due to random errors and then it tries to find out the value of the regression coefficient. So, this concept of regression of ridge regression has been extended and which will give us the LASSO regression.

So, now I am going to give you a brief introduction to the multicollinearity problem and ridge problem, ridge regression issue, right, ok. So, let us begin once again.

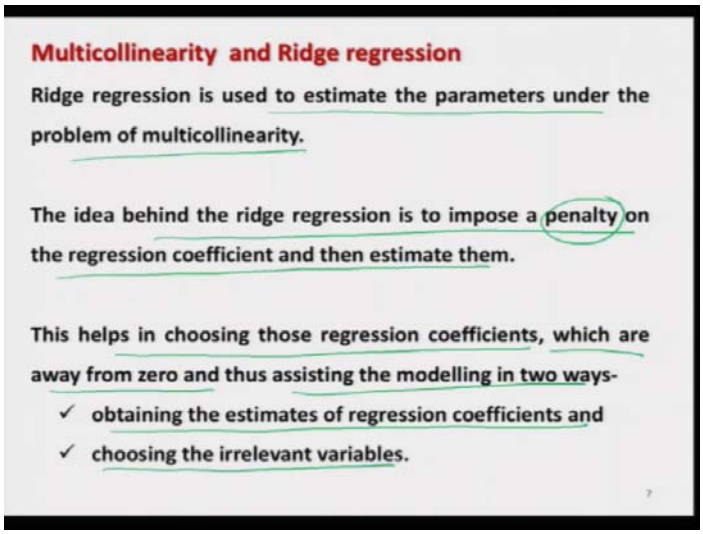
(Refer Slide Time: 11:29)



So, we assume that explanatory variables are independent of each other and the correlation between any two explanatory variable is ideally zero. But this is usually not possible in practice to achieve and the presence of such correlation increases the variability of the estimates of regression coefficients and consequently the model becomes undependable. This is called as a problem of multicollinearity.

And when we have large number of explanatory variable, then this problem increases and the chances of occurrence of multicollinearity becomes more, right.

(Refer Slide Time: 12:02)

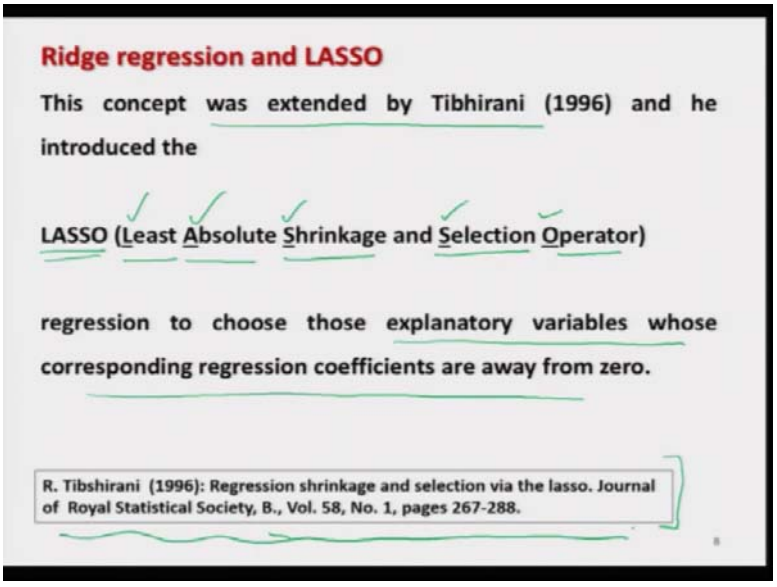


So, in order to solve the multicollinearity problem, one good solution which have been suggested in the literature is the ridge regression. So, ridge regression is used to estimate the parameters under the problem of multicollinearity in the data. And the idea behind the ridge regression is to impose a penalty on the regression coefficients and then estimate it. What is this penalty, we will try to understand in the next couple of slides.

But this imposition of penalty helps in helps us in getting a good value of regression coefficient and then this helps in choosing the those regression coefficients, which are away from zero. So, regression so, this ridge regression is going to help us in identifying those regression coefficient which are close to zero and possibly using the earlier concept, it will help us in choosing the important variable..

So, this assists in modelling in two ways; obtaining the estimates of the regression coefficient number one and number two choosing the irrelevant variable so that we can remove them from the model.

(Refer Slide Time: 13:19)



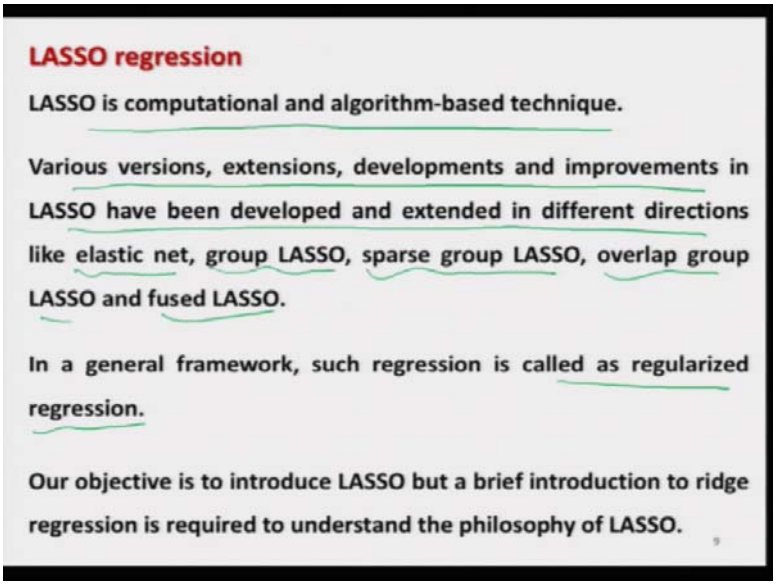
Actually, this concept was extended by Tibshirani. So, Professor Tibshirani, Professor RJ Tibshirani, he is at the Stanford University..

So, he has devised a different type of regression which is called as Least Absolute Shrinkage and Selection Operator regression. So, if you try to take the first letter of every word, this will briefly called as LASSO, L, A S S, O, right. So, this technique is called as

LASSO regression and it helps in choosing the explanatory variable whose corresponding regression coefficients are away from zero.

And if you wish, you can go to his webpage and this paper is there in which he has introduced this LASSO. This is a very good technique and nowadays, different versions have been appear.

(Refer Slide Time: 14:15)



So, this is actually LASSO is a computational and algorithm based approach. You cannot find out the exact expressions of the estimator, as you have obtained in the case of ordinary least square estimator, maximum likely estimation and so on.

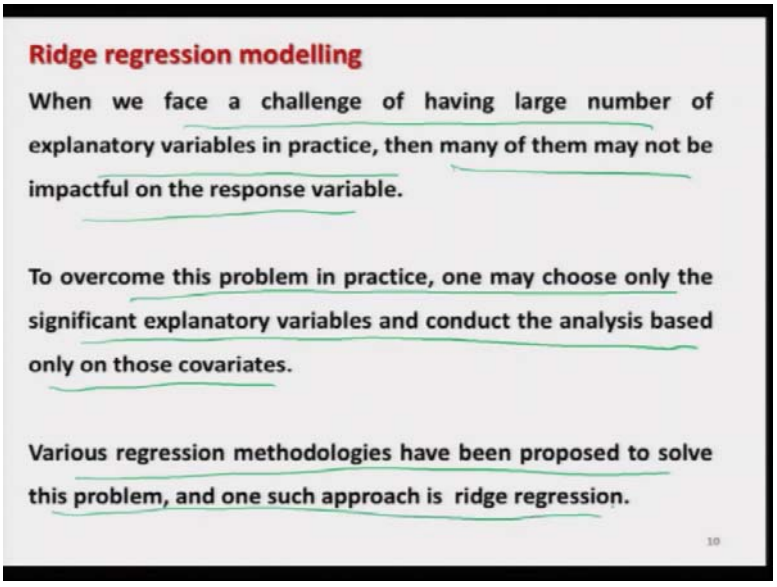
But for a given set of data, there is an algorithm which you have to employ and finally, the algorithm will give you the outcome from where you have to choose the important variable, ok. So, now, after Tibshirani, introduced this LASSO many people have *ted working and this LASSO has been improved in different directions.

And nowadays, various versions, extension, developments and improvements in LASSO have been developed and they have been extended in different directions. For example, nowadays you can see in the literature or they are available in the software's also like; elastic net, LASSO, group LASSO, sparse LASSO, sparse group LASSO, overlap group LASSO, fused LASSO and so on.

If you go to any recent book on the computational statistics possibly, you will get all these topics over there. And in a general framework, such regression is called as regularized regression. Our objective is not to understand all the mathematical details behind the LASSO, but our object is that we want to introduce LASSO with a brief introduction to the ridge regression..

So, that we can connect the LASSO to the ridge regression and we can understand the importance of LASSO, ok.

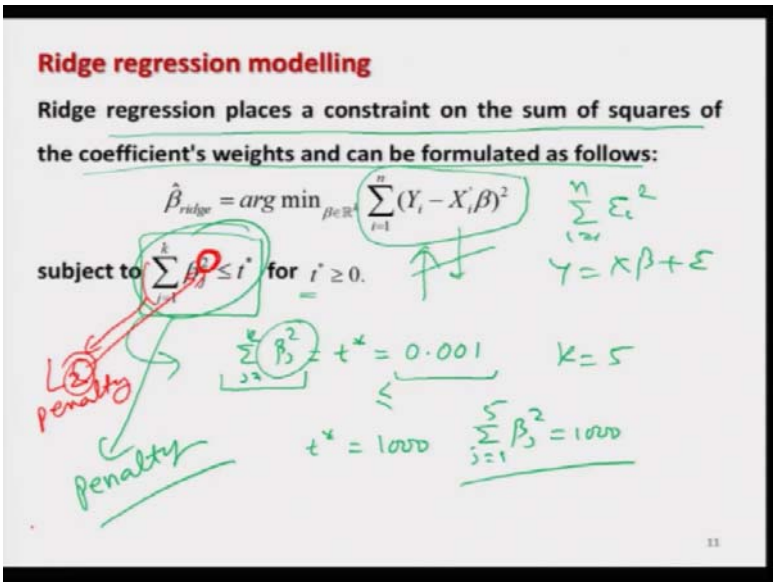
(Refer Slide Time: 15:58)



So, first we try to have a having some quick idea about the ridge regression modelling. So, whenever we have large number of explanatory variables, then many of them may not be impactful on the Y, the response variable.

So, to overcome this problem, if one can choose only the significant or important explanatory variable and conduct the entire regression analysis based only on those important explanatory variables or covariates. So, for that various regression methodologies have been proposed to solve this problem, and among them one such approach is ridge regression.

(Refer Slide Time: 16:40)



So, in ridge regression what it does? That it tries to provide to minimize the same sum of square due to random error that was done in the case of ordinary least square estimator also, but it tries to put some penalty on it, right. So, ridge regression places a constraint on the sum of a squares of the coefficient weight and can be formulated as follows.

So, it is trying to say, you try to consider this quantity which is your something like; $\sum_{j=1}^{n} \varepsilon_j^2$ in the model Y = X β + ε, ok. And it is trying to say that you try to find out the value of β by minimizing this quantity. But, there is a condition this β has to be found under a condition such that a condition like this one; $\sum_{j=1}^{k} \beta_j^2 \le t^*$, right. This t * is a fixed quantity.

So, now, if you try to see you are trying to put a constraint on thus on the sum of a square due to random errors and the constraint is in the form of sum of a squares of the regression coefficients. So, what will happen? Now, you can think about if your $\sum_{j=1}^{k} \beta_j^2 = t^*$, suppose we try to first understand the basic idea.

Suppose t * is very very small, suppose 0.001. Now, you are trying to say that the sum of squares of regression coefficient is equal to 0.001. And suppose, if I take suppose k equal to here 5 for example, just to understand. So, $\beta_1^2 + \beta_2^2 + \beta_3^2 + \beta_4^2 + \beta_5^2$ is equal to 0.001.

So, now, you are trying to say find out the value of β from this function, such that this condition the summation of $\beta_j^2 = 0.001$ is satisfied, or if on you if you make it smaller than 0.001, then the condition becomes more strict. So, but you can see here what will be the possible values of β_js, those possible values of β_j will be very very close to 0.

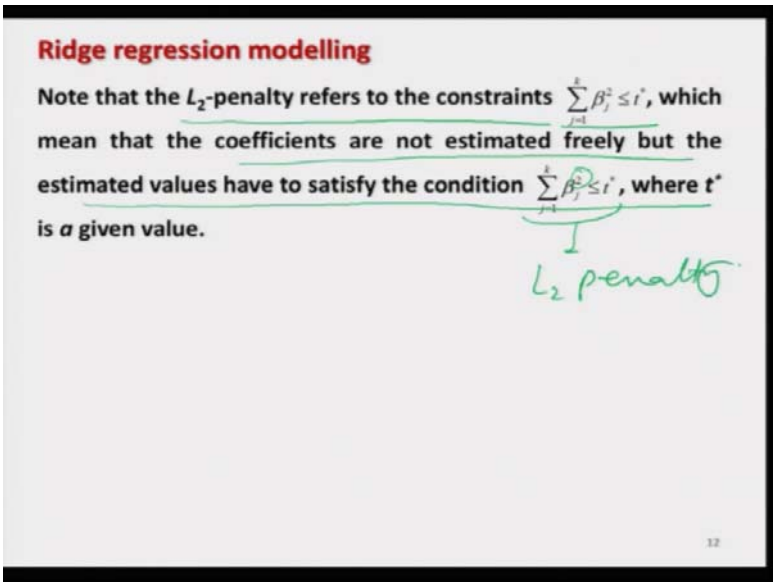
Now, on the other hand if you try to take here the value of your t * to be suppose 1000 and if you try to find out j goes from 1 to 5 β_j^2 should be equal to 1000, then what will happen that $\beta_1^2 + \beta_2^2 + \beta_3^2 + \beta_4^2 + \beta_5^2$ should be 1000. So, in this case you can expect that none of the β_j is going to be very close to 0.

So, now, these are two extreme conditions. That if you try to estimate your β_j in such a way such that this constraint is satisfied, then if you try to choose very small value of t * then most of the variables will turn out to be as if they are not important. And if you try to take very large value of t *, then most of the variable will turn out to be important. So, in both the cases you may not get a nice and correct outcome.

So, the challenge here is that how to choose this t *. So, that you can get a reasonable value of t * and hence we can get a reasonable value of this regression coefficients. So, this is actually, when we try to write down this condition here, like as $\sum_{j=1}^{k} \beta_j^2 \leq t^*$ is less than equal to t *, this is called as penalty, right. Because this is a constraint which has been given a new name which is called as penalty. And you can see here I am writing now in red color.

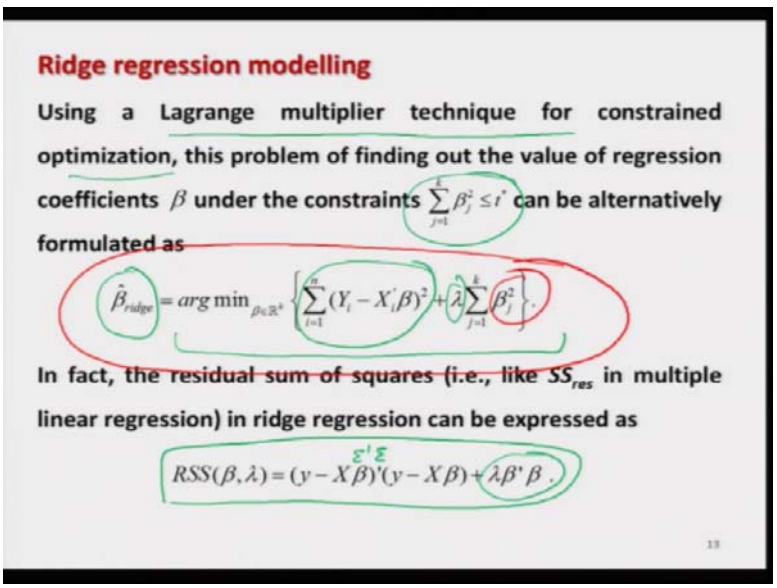
You can see here this square, because of this square this constraint is called as L_2 penalty, this 2 is coming from this 2 actually, this is square. That is what you have to just keep in mind, because when you try to read any book on LASSO regression they will be talking of L_2 penalty L_1 penalty. So, I will try to explain you all these words, right.

(Refer Slide Time: 21:07)



So, this is what exactly I have written here, that this L_2 penalty refers to the constraint this $\sum_{j=1}^{k} \beta_j^2 \le t^*$ which means that, the coefficient are not estimated freely. But the estimated values have to satisfy this condition, right and this 2 is, because of this two, we will call it as a L_2 penalty, right, ok. So, now, you have understood this thing.

(Refer Slide Time: 21:35)



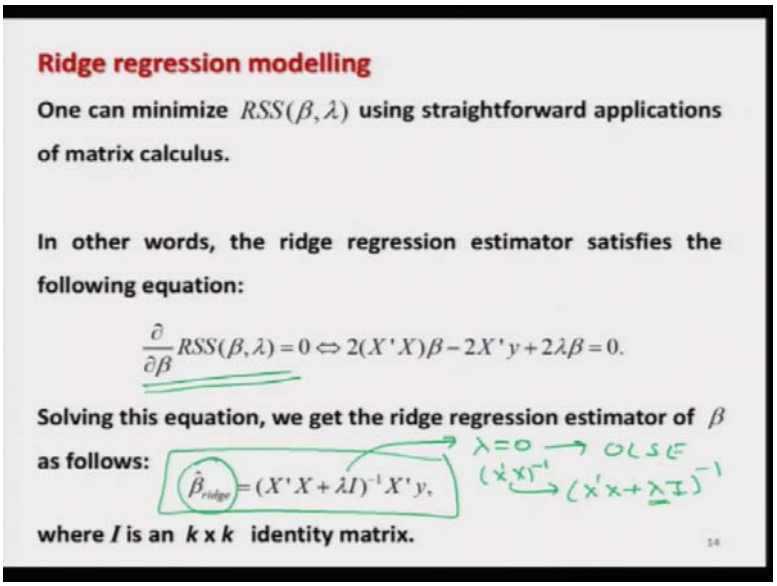
So, how do we obtain it? In order to obtain the values of βs under the ridge regression setup, we try we use the Lagrange multiplier technique for the constraint optimization

and we try to find out the value of β such that this sum of a square of random error is minimum and this condition is satisfied.

So, I can formulate the value of β which will be obtained after optimizing this thing as the solution of β and the solution of β will be called as ridge regression estimator of β denoted as β hat ridge. So, this is going to be obtained as a solution of this equation. So, you have to minimize the sum of a square due to random errors subject to this constraint.

And here this λ is a Lagrangian multiplier, right. So, what we try to do, that this problem can be written as a function which has to be minimized in order to get the ridge regression and this function is written here as a RSS as a function of β and λ. So, we try to write down the same thing if you remember this in nothing but, your ε'ε and this is a constraint on this β, ok.

(Refer Slide Time: 22:58)



So, now, if you try to simply differentiate this RSS (β) with respect to β and λ and try to solve those two equation. For example, you can just partially differences this RSS(β, λ) with respect to β and λ. And you try to solve the equation; you will get the ridge regression estimator here like this.

So, if you try to see here the structure of this one, this is $(X'X+\lambda I)^{-1}X'y,$. So, if you try to substitute here λ = 0, what do you get? You get same as OLSE. So, now, you can

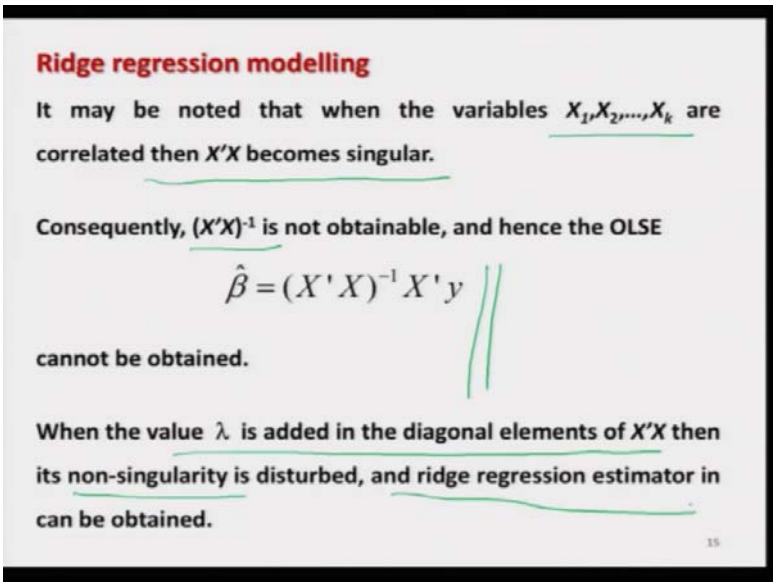
understand this thing very easily, since there were lot of variable lot of independent variable.

So, possibly they were becoming correlated and because of which you were unable to find $(X'X)^{-1}$ and this was creating the trouble in finding out the estimator as well as the standard errors. So, what I am trying to do? I am trying to add something on the diagonals of this matrix. And then I am trying to find out the inverse.

Now, how to find out this λ that is itself a challenge. So, at the moment you can assume that, ok one can find out a find out the optimum value of λ. For example, in the case of ridge regression this is one popular technique is to find out the value of λ by using the ridge regression, right.

So, this is what we are trying to do and this is how we are trying to obtain a modified version of the ordinary least square estimator, which will be called as ridge regression estimator of β, ok.

(Refer Slide Time: 24:41)

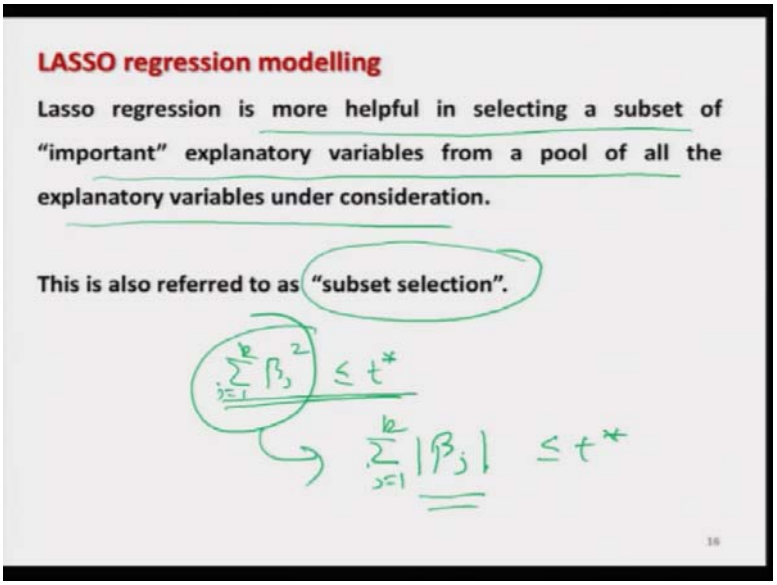


So, this is exactly what I am trying to show you here, that if X_1, X_2,…, X_k are correlated, then X'X becomes singular and then $(X'X)^{-1}$is not obtainable.

So, the ordinary least square estimator cannot be obtained. So, when we are trying to add λ values in the diagonal elements of X'X, then possibly it changes the matrix from

singular to non singularity and then possibly we can find out the inverse and hence we can obtain the ridge regression estimator, right.

(Refer Slide Time: 25:13)

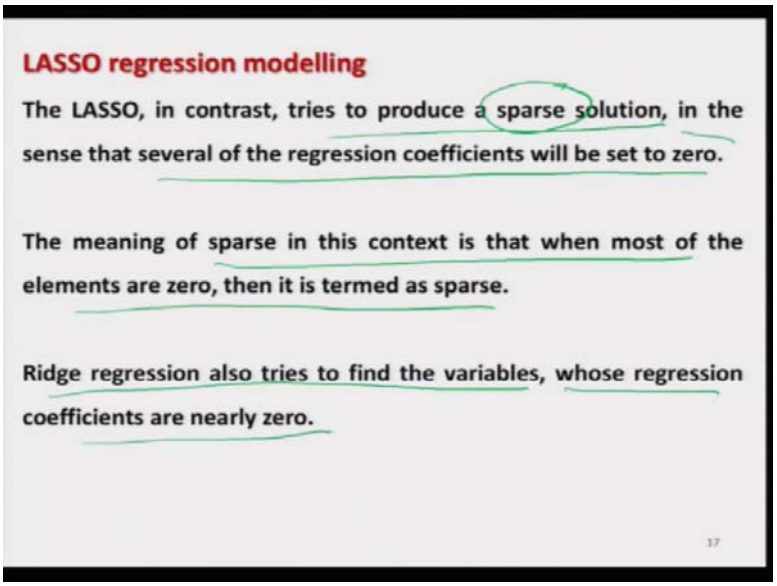


Now, this concept has been extended to LASSO regression modelling. How, this LASSO regression is actually more helpful in collecting a subset of important explanatory variable from a pool of all the explanatory variables under consideration. So, this is also called as subsets selection. How this will help? Means, you have seen that in the case of ridge regression you have put a constraint that summation β^2 is less than t *.

So, this actually constraint is trying to help us in identifying the important variables and since you are putting a constraint. So, there is a high possibility that you will finally, end up in collecting those β_js which are not close to zero, right. So that is the basic idea here. Now, what Tibshirani did in the case of LASSO regression? He changed this constraint.

And instead of considering summation β_j^2, he simply took the absolute value of $\beta_{,j}$ goes from 1 to k same as in both the case; j goes from 1 to k but instead of taking the square, Professor Tibshirani took the absolute value of β_j and same constraint was there. And this helped a lot in the selection of important variable.

(Refer Slide Time: 26:36)

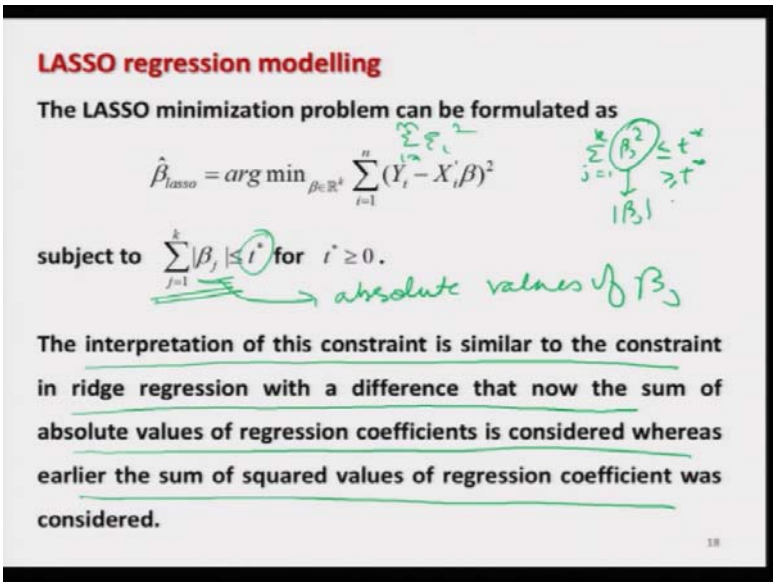


But, that there is a difference that regression, that in the case of ridge regression you have a very nice close form of the estimator. But in the case of LASSO you do not have a closed form of the estimator and the LASSO tries to produce a sparse solution, which is just contrast to the regression ridge regression. What is the meaning of this word sparse or what is the interpretation of sparse solution?

LASSO will try to produce a solution of β_j in which several of the regression coefficient will be set to zero and this is what we want. The meaning of sparse in the context is that most of the elements are zero and that is why it is termed as his sparse and this is exactly what we want. We want that we want a procedure which can inform us well, in this process these many regression coefficients are close to zero.

And hence after that I can take care and I can choose only those variable, which are not close to zero or the corresponding regression coefficients are not close to zero. Now if you try to compare the LASSO regression with ridge regression, ridge regression also tries to do the same thing, whose regression coefficients are nearly zero.

(Refer Slide Time: 28:07)



But definitely, LASSO works better and that is why LASSO has become very important. So, now, if you try to see now, in this slide I am trying to formulate the LASSO problem. You can see I have not done anything here; I simply have taken the same expression which is here in the case of ridge regression. Try to look at this expression, right. You have all the slides, so you can very easily compare them.

I am trying to take the same thing and just I am trying to play here, I simply try to change summation β_j^2 to summation β_j. So, I try to minimize the $\sum_{i=1}^{n} \varepsilon_i^2$, but I try to find out the value of β in such a way such that some of the $\sum_{j=1}^{k} |\beta_j| \leq t^*$, right.

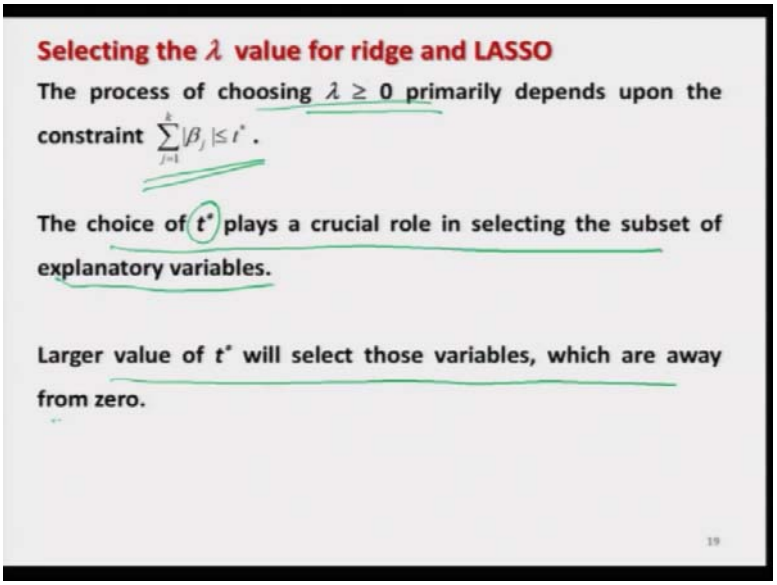
So, I have done only one thing here now, you can see here this is only the $\sum_{j=1}^{k} |\beta_j|$ that is all. But this small change made wonders, right and I wish if all of you can come up with such question small idea once in your life which can make you as popular as now, Professor Tibshirani has become, right.

Anyway, so, the interpretation of this constraint that $\sum_{j=1}^{k} |\beta_j| \leq t^*$ exactly is the same as in the case of ridge regression. The only difference is this there we took the sum of a

squares and here we are trying to take the sum of absolute division that is all. Now, you have understood that if I try to take $\sum_{j=1}^{k} \beta_j^2 \leq t^*$ or greater than t * what is the outcome?

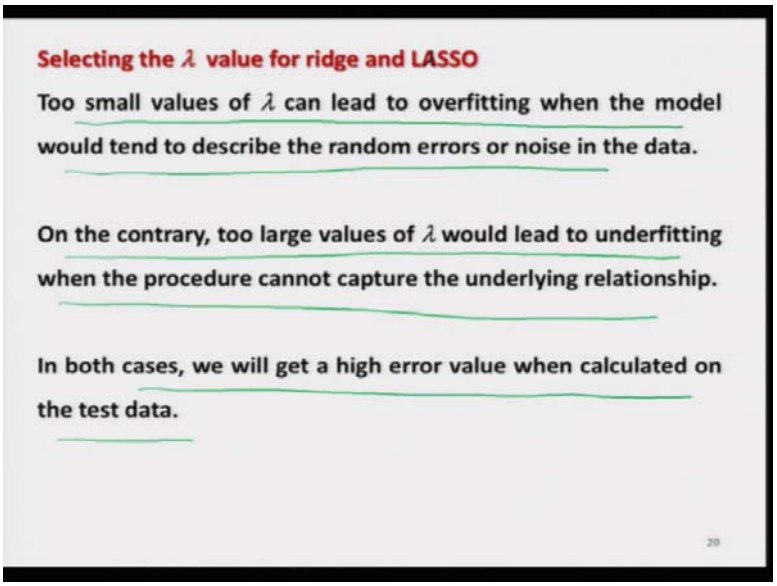
Means I have I already explained you, now, you are simply trying to say instead of this summation β_j^2 I will take $\sum_{j=1}^{k} |\beta_j|$, right. So, if you try to choose any value of t * you will have a similar outcome, right.

(Refer Slide Time: 30:16)



The question is this how you would try to choose the value of λ?, right. So, this value of λ primarily depends on the choice of this t*. So, I already have explained you what are the consequences of choosing the value of t*. If the value of t* is very small or very large what happens? So, this choice of t* plays a very crucial role in collecting the subset of explanatory variables. The larger value of t* will select those variables, which are away from zero.

(Refer Slide Time: 30:52)



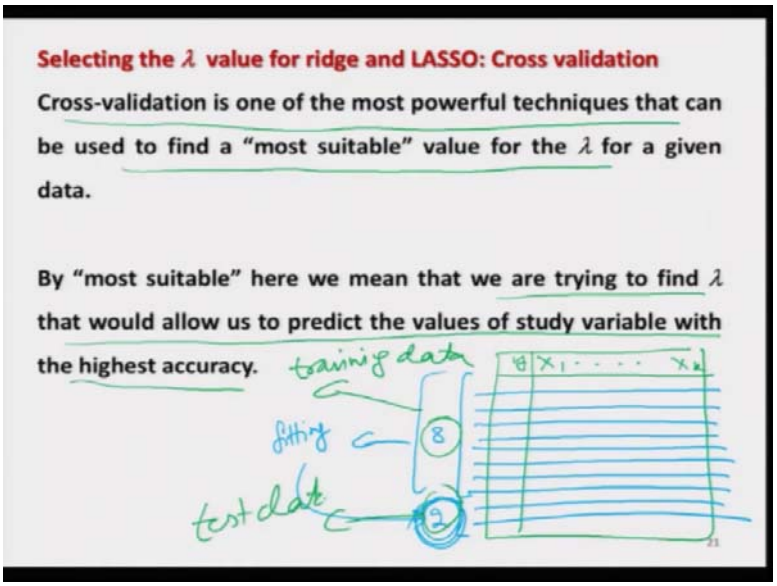
And too small values of λ can lead to overfitting when the model would tend to describe the random errors or noise in the data. So, that means, if you try to choose very small, try to choose very small value of λ possibly it will reject all the variables and it will appear as if the model is controlled only by the random error.

So, that is also not needed, you need to strike a balance between the two extreme values of t*. If t* become very very large, it is trying to show you as if all the variables are important this is also wrong. If the value of t* is very small, it is indicating that all the variable are worthless and that means, only the random error is going to control the process which is also wrong.

So, now, you have to strike a balance between somewhere so that you can choose the value of λ and this can be done using an appropriate algorithm, right.

So, that is what I am trying to say here in my slide, that the too large values of λ would lead to underfitting when the procedure cannot capture the underlying relationship. So, in both the cases the model will not be get, will not be good and we will get a high error value, when we try to experiment on the test data. What is this test data? I will try to show you.

(Refer Slide Time: 32:08)



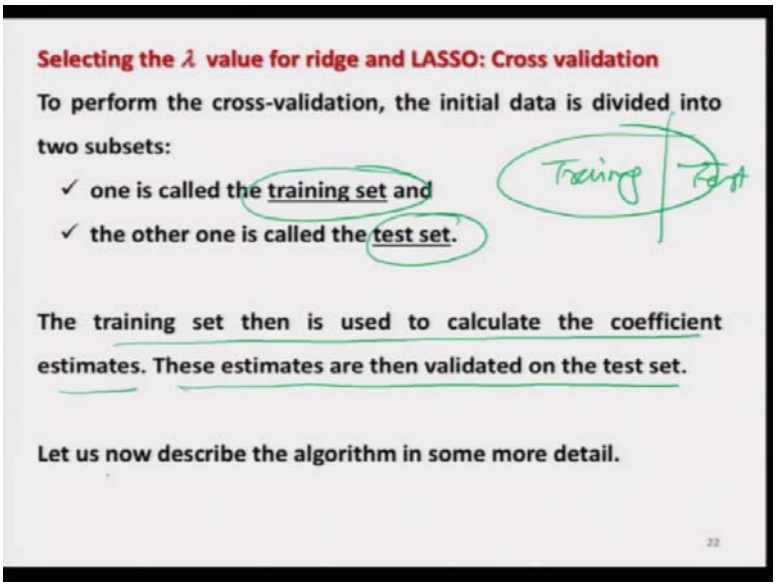
You see, whenever you are trying to get here a model, what happened you have got here a data set. This data set is something like here say here y, X_1, X_2, …, X_k, right. So, now, this is the only data set from where we have to do each and everything. I need to find out the sum of square, I need to find out the value of λ and I also need to find out the value of β. Nobody is coming from sky to tell us all these values.

So, what we try to do, we try to divide this data into several parts. For example, I can divide this data into suppose 1, 2, 3, 4, 5, 6, 7, 8, 9, 10. Now, what I will say here that I will try to divide them into for example, 8 into 8 is to 2. 8 part in one group and 2 part in one group.

So, what I would say that I will try to use this part say 8 parts for the model fitting and then I will try to use this model to judge the model fitting over the remaining 2 parts of the data. So, I will try to cross validate my model with this 2 parts of the data. So, this process is actually called as cross validation. So, cross validation is one of the most important techniques that can be used to find a most suitable value of λ for a given data.

And when I say more suitable, more suitable mean that we are trying to find out the value of λ which would allow us to predict the values of study variable with the highest accuracy. So, we are going to consider the criteria of prediction errors and then we will try to minimize the error in such a way such that we get good prediction.

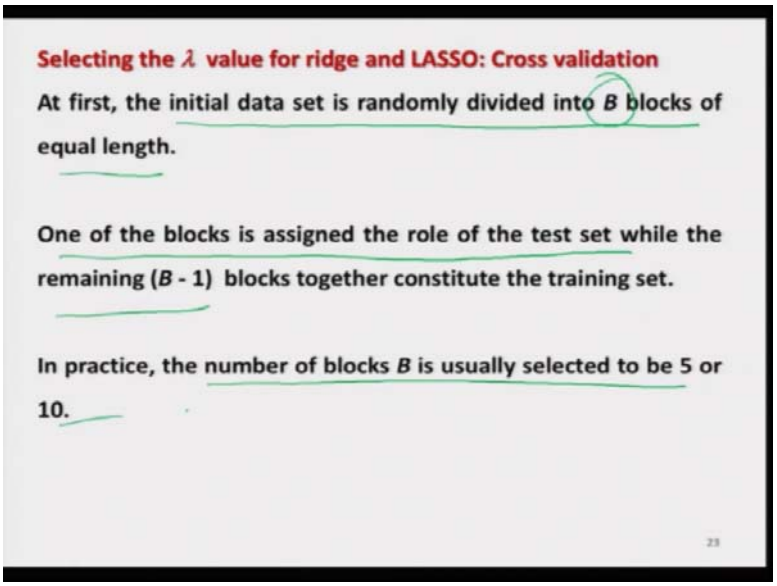
(Refer Slide Time: 34:08)



So, in order to conduct the cross validation as I said, we have defined we have divided the data into two parts for example, here you can see 8 is to 2, and this part is called as training data, right. And this part is called here see here, test data that you can see. So, I am going to explain you here that we have now, the some data set. Here, I am trying to divide it into two parts training data set and other part will be test data set, right.

So, the data on which you are going to fit the model this is training data set and the part of the data where you are going to cross validate your model that is called the test data set, right. So, this training data set is used to calculate the coefficients and then these coefficient estimates are validated on the test set. So, we try to understand this algorithm in a little bit detail so, that you can understand it very quickly.

(Refer Slide Time: 35:14)

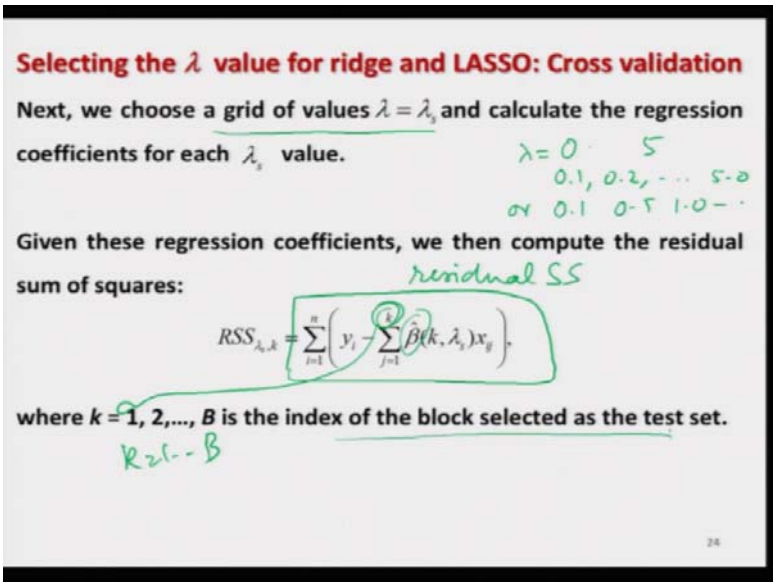


So, what we try to do? That suppose we get a data and we try to divide the data into suppose say capital B blocks and every block has got an equal length. And one of the block is assigned the role of test set while the remaining B - 1 blocks together can constitute the training set, right.

So, it actually, this is also an issue; means, sometime people try to divide use two parts, sometimes people try to divide into only use one part for the test data set, though that depends on the experiment also. And the complications in the experiment and complications in the data set also. But in practice usually, we try to divide the total data into 10 blocks or 5 blocks.

So, that is called as 10 cross validation or 5 cross validation.

(Refer Slide Time: 36:09)



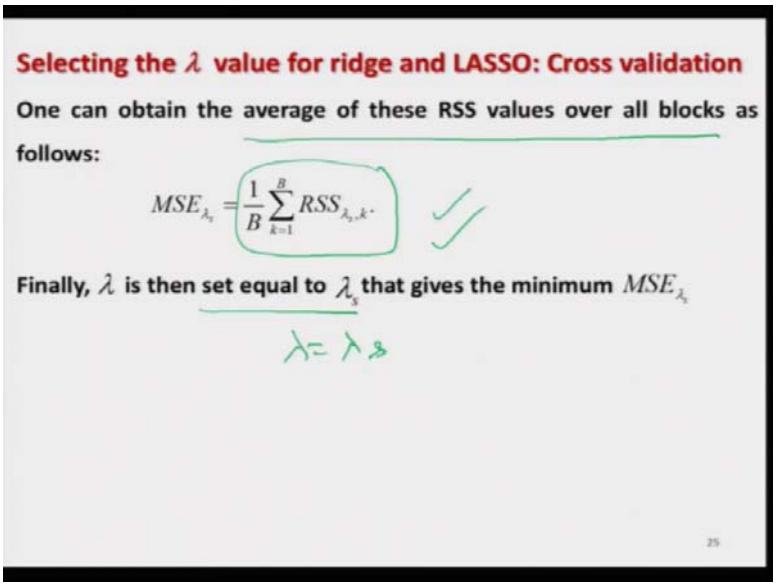
Then after this we try to choose a grid of values, say $\lambda = \lambda$ s. So, suppose I can take λ is equal to say between 0 and 5 and then I can say 0.1, 0.2, 0.3 up to 5.0 or even I can say alternative is this 0.1, 0.5, 1.0 and so on.

Whatever you want that we have to actually experiment to see where I am getting a good value. So, I try to choose a particular value of λ from this sequence and then using this sequence I try to minimize the sum of a square due to regression, which is defined here like this. So, this is your actually residual sum of a squares which I try to minimize.

So, whatever is the solution of after this minimization that will be the solution of LASSO estimator of β, right. And where here this here k that you have to remember this is the k is going from here 1 to B, which is here the blocks, blocks index of the blocks it is not the number of independent variable that you have chosen in the case of multiple linear regression model remember, right.

So, this is the so, it is trying to say that you have to choose the k, which where this k is going to be the index of the block which is selected for the test data.

(Refer Slide Time: 37:41)



Now, you can choose different values of λ which are defined here as say λ s and then try to obtain the average of these residual sum of a squares over all the blocks.

So, we try to take the arithmetic mean of all the residual sum of a squares over all the blocks and I try to find out their arithmetic mean. And then you try to see, whatever is the value of λ, which gives us the minimum value of MSE that is that equal to λ equal to λ s, right.

So, you try to do this calculation and then try to see that what is the value among all the calculation which is trying to give us the minimum MSE. And then that corresponding value is chosen as λ. So now, we come to an end to this lecture. Well, I must confess that I have not given you the theory of all these things which is not so easy to explain in half an hour or 1 hour time.

And for that, I would request you if you are interested then please try to look into the books and research paper which I have indicated here also, right. But, my idea is very simple, because we are now going to work in the area of data sciences. So, we are more interested in the computation also. Given a data set, given a large number of variables how you will you choose the important variable that is my objective here in this course.

So, I and then there are several concept which are linked together to give you a fair idea about the LASSO regression. So, I have tried my best to give you good overview, how

the LASSO works. There are various types of algorithm, people are coming with different types of algorithm and they claim that this algorithm is better than this and in R software also there is a package LASSO, which gives us all the computations.

So, my more interest in this chapter is to show you, how can you compute it and how you are going to interpret the results. So, you try to have a review of this lecture, try to connect all these concepts together, that how the correlation in the explanatory variable disturb the properties of ordinary least square estimator, which gives rise to the ridge regression estimator and which helps in the development of LASSO regression. And I will see you in the next lecture with a application of data set using the LASSO package in R software. Till then good bye.