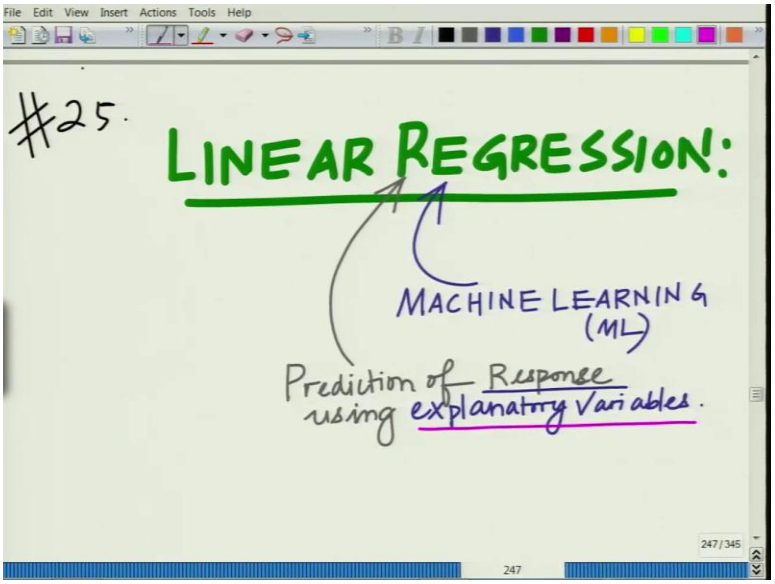
Applied Linear Algebra for Signal Processing, Data Analytics and Machine Learning
Professor Aditya K Jagannatham
Department of Electrical Engineering
Indian Institute of Technology Kanpur
Lecture 25
Machine Learning Application: Linear Regression

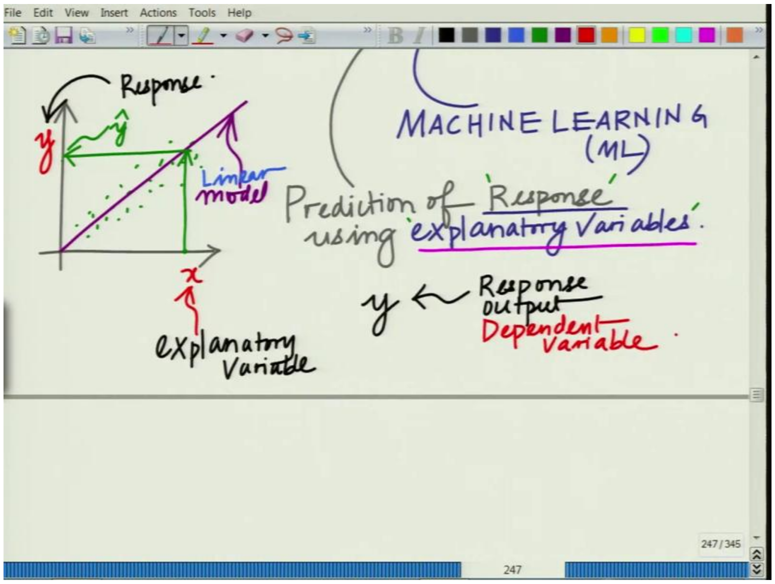
Hello, welcome to another module in this massive open online course and in this module, we want to look at yet another important application of least squares, and that is in machine learning in the context of what is known as a linear regression.

(Refer Slide Time: 0:32)



So we want to look at another very relevant and modern application of least squares in the context of machine learning, which is essentially, what is termed as linear regression and as I have told you, this is a very important application in the context of or in the area of machine learning or ML. Now, what is linear regression about to put it simply linear regression is basically prediction, prediction of a response using, these are technical terms, prediction of a response using what are known as explanatory variables. What is the meaning of this?

(Refer Slide Time: 2:08)



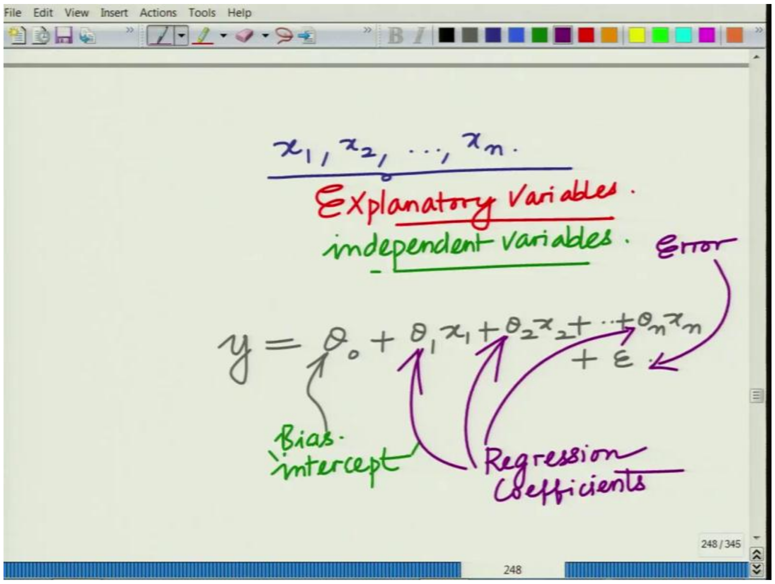
So let us say, let us illustrate it using a simple plot. So let us say, you have a lot of data and clearly you can see this data are correlated. So what I can do is, so this let us say I denote by x and y. This is what I am calling as my explanatory variable and y is what I am calling as response.

And, clearly you can see that the explanatory variable and the response are highly correlated. One that is knowledge of the explanatory variable can help you predict the response of the system. And this can be done as you might have already guessed by fitting a model through this and this is essentially your model or what you are terming as your linear model. So we fit a linear model, so that given any new exclamatory variable, one can predict, one can obtain an idea or estimate of the response y had.

This is essentially what regression is that is prediction of the response using these explanatory variables and this is essentially trying to fit a model or in this case, what is a linear model and this is essentially termed as linear regression. You are fitting a linear model to the explanatory variables to predict the response of the system. And You are trying to learn this linear model and that is the example and that is essentially what the machine learning aspect is about.

So this is essentially the linear model as we have already stated. So this is a linear model. So put it more precisely. You have a response, which is why we are calling this as the response of the system. There are many things or you can call this as the output. You can also call this as the observation or you can also call this as the dependent variable. All of these are essentially.

(Refer Slide Time: 5:17)

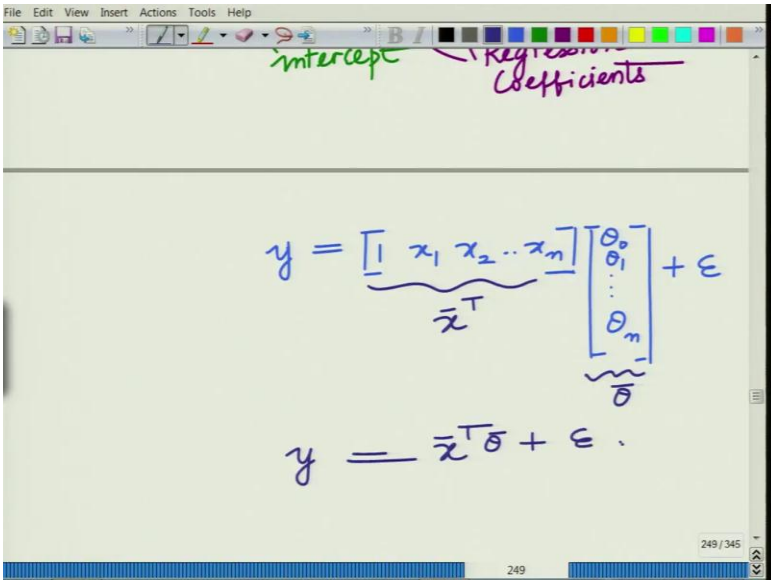


And then you have a set of explanatory variables. So you have x1, x2, up to xn and these are your explanatory variables. These are the explanatory variables. And, now or these are also your independent variables. So these can vary independently. That is the idea. So you have the dependent variable. That is why these are essentially your independent variables. So we have the dependent variable and these are the independent variables.

And now your regressor is essentially or predictor is given as y equals theta naught plus theta 1 x1 plus theta 2 x2 plus theta n xn plus. So this is your regression model. So this theta naught, this is a fixed term. This is essentially a constant. This is termed as essentially your bias term or essentially this is also termed as the intercept. So this is, this is termed as the intercept and, these theta 1, theta 2, theta n including you can also call this as theta naught. These are your regression parameters or these are your regression coefficients and this epsilon is, this is your error.

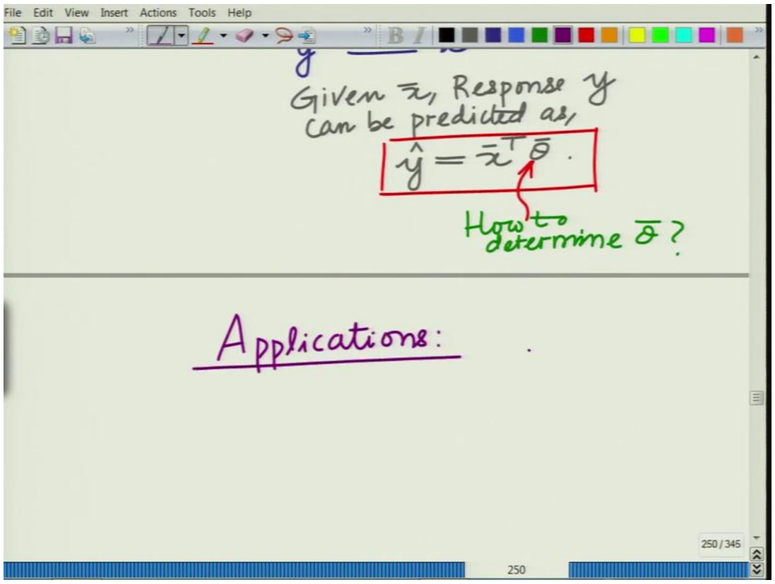
You can term this as your error term or your or essentially your disturbance term. So essentially your y is given as the sum, the linear combination, weighted linear combination of your explanatory variables. So, theta naught, which is the intercept plus theta 1 times x1 plus theta 2 times x2 so on up to theta n times xn plus your epsilon where this quantity epsilon, this is essentially your modeling error or this is your disturbance and these thetas are essentially the regression coefficients.

(Refer Slide Time: 8:27)



So I can write this essentially as y equal to, it is not very difficult to see 1 x1 x2 xn cross theta naught, theta 1, so up to theta n plus epsilon or the error. And this essentially you can denote this by x bar transpose and this I can denote by the vector theta bar. So I can write this model in a compact form as x bar transpose theta bar, plus epsilon.

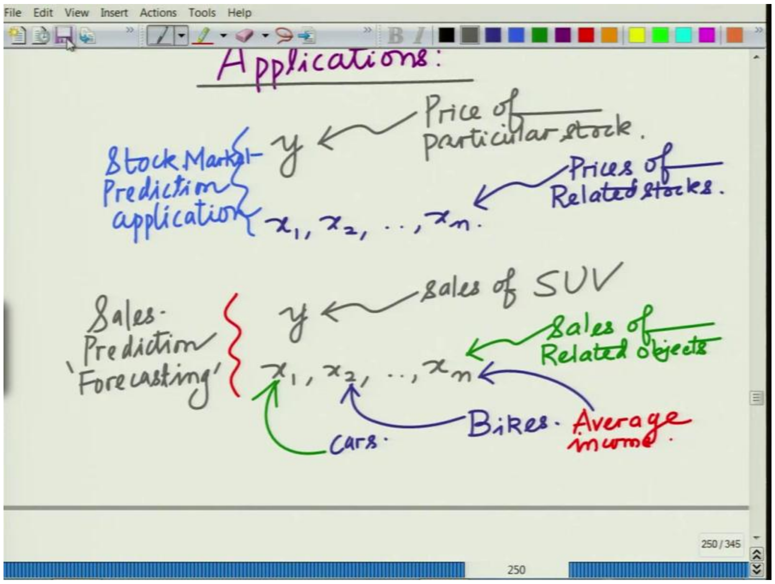
(Refer Slide Time: 9:37)



And this again, as I already told you, this theta bar, this contains the regression coefficients, and naturally now the prediction response can for any unknown, for any unknown. So given x bar, given the independent variables x bar or the explanatory variables x bar, once you know the model, response y can be predicted as y hat equal to x bar transpose theta bar. So this is essentially your prediction. So this is essentially the prediction corresponding to the, a new.

So once you have, so once you fit the model, once you determine the regression coefficients, theta bar for any set of explanatory variables or for any set of, for any particular explanatory variable vector x bar, the response can be predicted as y hat equals x bar times theta bar, but it is important to first determine these regression coefficients. So that is, that is the thing that has to be done, because if theta bar is unknown, then it is not possible to predict. So one has to determine these regression parameters theta bar. So how to determine, That is the question. Let us look at some examples, some applications of this.

(Refer Slide Time: 11:25)



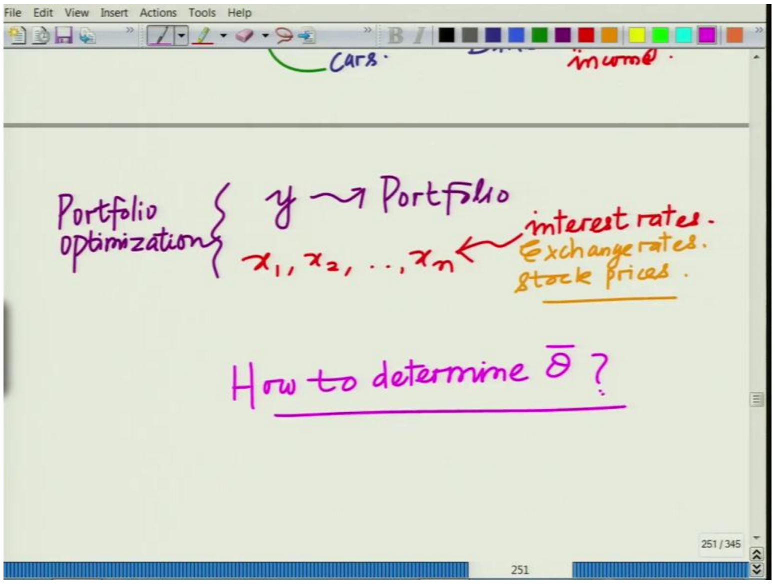
As I have already told you this is very important, linear regression, it has several applications. For instance, you can have y, in the context of y, this can be the price of a certain stock. This can be price of a particular stock, y can be the price and x1, x2, xn, these are the explanatory variables. These can be the prices of related stocks. These can be prices. So determine y that is the price of a particular stock that you are interested in from, or predict the price of y that is a particular stock from the prices of x1, x2, xn, which is essentially you can think of it as a stock market index or a basket of related stocks.

So this is essentially a stock market or stock prediction, stock price prediction problem. This is one of the examples of applications. So you can, one can determine why using a linear regressor in terms of the explanatory variables, x1, x2, xn, which are the prices of related stocks. So this is a stock market prediction problem. This is a stock market prediction application.

So another interesting application could be a sales prediction. So y can be for instance sales of related objects, such as for instance, you can think of sales of SUV vehicles, and x1, x2, xn, these can be for instance, sales of related objects, such as for instance, x1 can be sale of cars, x2 can be sale of bikes, and it can also be other things, such as for instance, xn can be the average income or GDP, or so on average income of an area or average income of a particular city or town.

So if the average income is higher, naturally the demand for SUVs or demand for vehicles is going to be higher. So all of these, so there is also one has to also choose these explanatory variables carefully so that these are related, sufficiently correlated and can be sufficiently good indicators of the response. So this is essentially an application of your sales, sales prediction, or it is also termed as forecasting, sales prediction or forecasting.

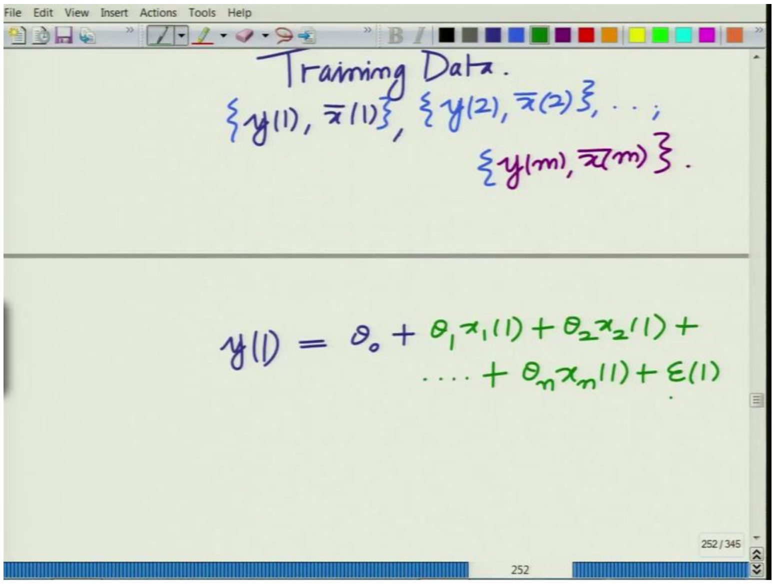
(Refer Slide Time: 15:32)



And, another interesting application again, a finance application, this can be for instance, your portfolio risk, what is the risk in your portfolio and your x1, x2, xn, these can be various related parameters or various related quantities such as, for instance, interest rates or these can be for instance your interest rates, exchange rates, or stock prices, etcetera. Now, the question as we already said, so this is again, a portfolio optimization problem, or a risk optimization or a portfolio risk optimization.

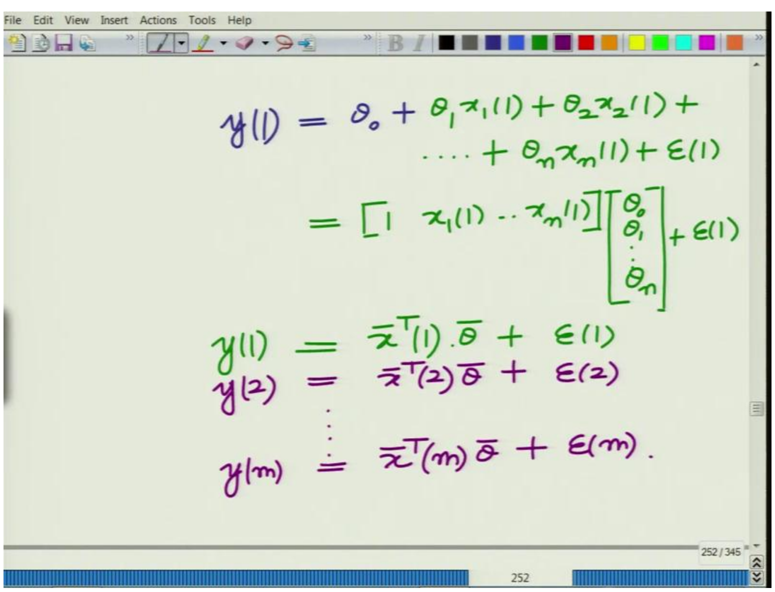
Now, once again, we can ask the question, remember how to determine the regressor coefficient’s theta bar. Because if you know theta bar, then given any explanatory variable vector, x bar one can predict the response y. Now, for that, we start with a set of training data that is responses and corresponding explanatory variables.

(Refer Slide Time: 17:19)



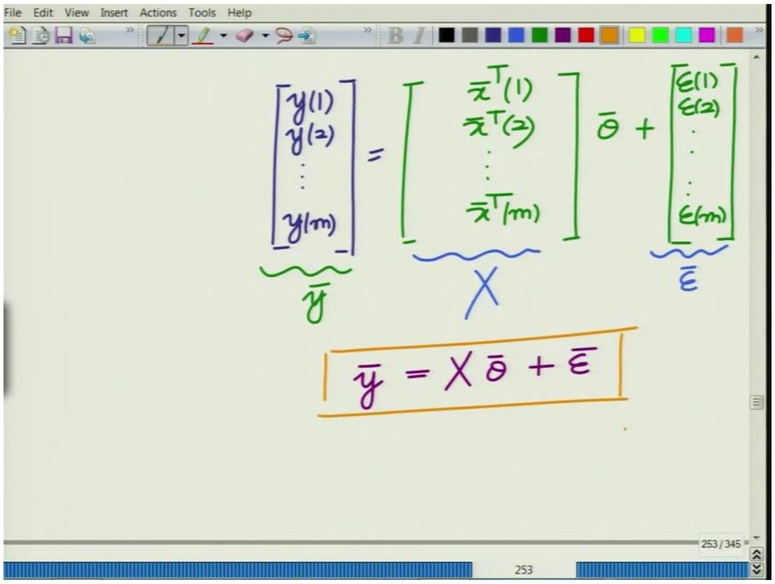
So we start with, the answer to this is start with training data that is, you have a bunch of y bar y1, x bar 1. One set of training data and then you have, then you have y2, x bar 2 so on and then finally you have ym, x bar m. So this is your m points, or your set of m training data vectors, or training data samples. And now one can again construct the model. So I can write y1 as well theta naught, this is the first training vector, theta, theta naught plus theta 1 x1 1 plus theta 2 x2 1 plus so on and so forth, theta n xn1 plus the error.

(Refer Slide Time: 19:05)



Which I can write, of course, in compact form as once again, 1 x1 1 so on, xn 1 times theta naught, theta 1 so on up to theta n plus epsilon error 1, which is essentially, you can think of this as x bar transpose 1 times theta bar plus the error. Similarly, one can express y2 as x bar transpose 2 theta bar plus the error and ym, the mth point as x bar transpose m. So these are your regression.

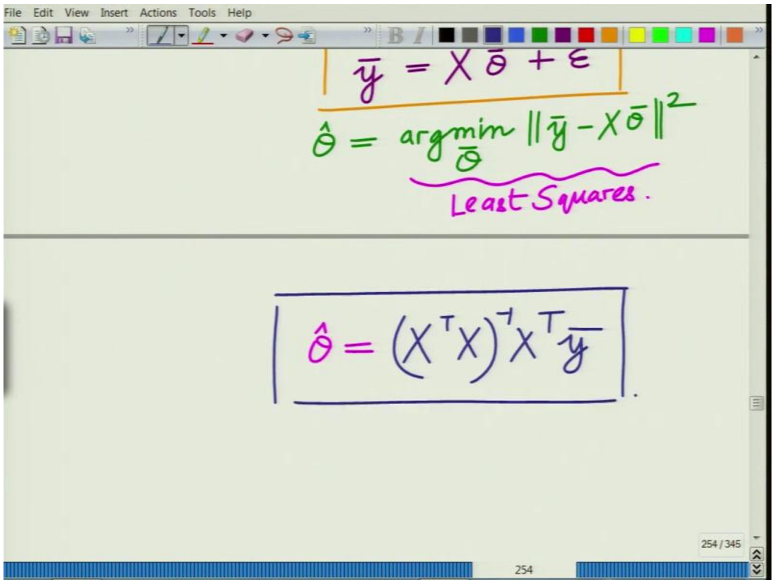
(Refer Slide Time: 20:28)



These are your regression models or these are basically the corresponding model for the output. So now you can write this as the vector, y1, y2, ym as we always do, write it in compact vector notation, as I have already told you and this is where the linear algebra comes in. X bar transpose 1, x bar transpose 2 so on x bar transpose m times theta bar plus epsilon bar, which is the error comprises of the errors epsilon 1, epsilon 2, up to.

So this is essentially your y bar. This is essentially your matrix x, which contains the training vectors of explanatory variables. This is your vector x epsilon bar, which is the error. And therefore, I can write this as y bar equal to x theta bar plus e bar. So this is the model, or this is your model for the training data. So these are basically the regression outputs corresponding to the explanatory, training explanatory vectors, the explanatory training vectors and, the training responses.

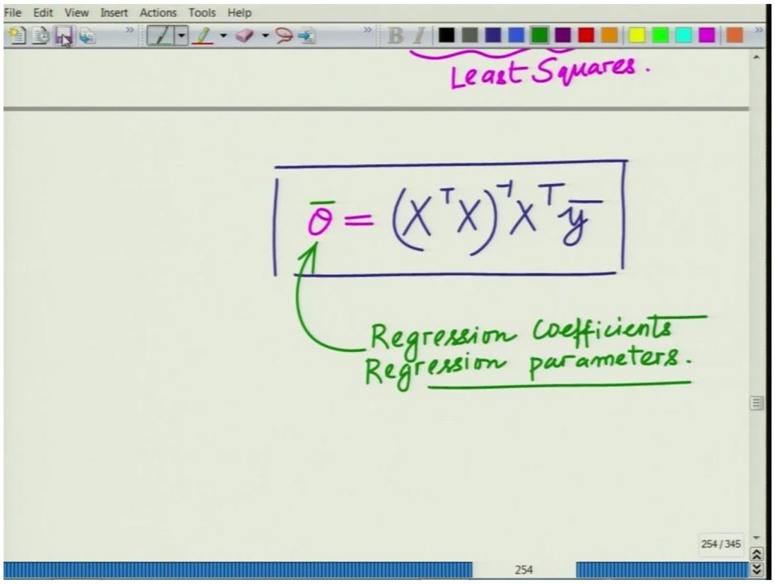
(Refer Slide Time: 22:40)



And now the idea is to determine the theta bar, which gives the best fit for or which gives the best fit for y bar that is determine theta hat as the best vector theta bar, the best set of regression coefficient that minimize the error between the training responses and the explanatory or the training explanatory vectors.

And now you can naturally see, this is again a least squares problem. This is again and you can now immediately see how powerful the applications of least squares can be, it can be applied anywhere. Why are this linear regression model can be applied now anywhere else, sales predictions, stock market forecasting, portfolio risk optimization, so on and so forth. So theta hat again is given where the solution to the least squares problem that is it is given by x transpose x inverse x transpose y bar.

(Refer Slide Time: 23:56)



So this is essentially your, this is essentially your theta hat, which is your or we can simply call this theta bar, which is your regression coefficients, or you can also call this as the regression parameters. So therefore thus essentially one can use once again, the least squares paradigm, that you have developed so far and determine the regression parameters or the regression coefficients theta bar, which essentially includes, which essentially is basically the vector, if you remember, this is the vector theta naught, theta 0, theta naught, theta 1, theta 2 up to theta n. So it also includes that intercept term theta naught.

And therefore, now you can see regression or linear regression. This is a very important aspect of machine learning, which in turn has several applications in various areas such as for instance stock predictions, sales forecasting, portfolio, housing price prediction, traffic prediction, wherever you can think of prediction, forecasting, or basically determining a response or getting an idea of response based on a set of explanatory variables, which are highly related to this response.

One can use linear regression and naturally in all such areas, the least squares paradigm that we have looked at. Of course, linear algebra has a very important role to play and in particular, the least squares can determine the particular regression model by essentially determining the regression coefficients, or the regression parameters, and once you define, I mean, once you determine the regression coefficients or the regression parameters, that in turn gives you the regression model.

So essentially least squares can be used to determine that regressor. So therefore, it is again, one of as I have already told you a very, very important concept that is this problem of least squares and solution, the least square solution has wide ranging applications. We have looked at yet another application of the same in the context of machine learning and particular linear regression. So let us stop here and continue this discussion in the subsequent modules. Thank you very much.