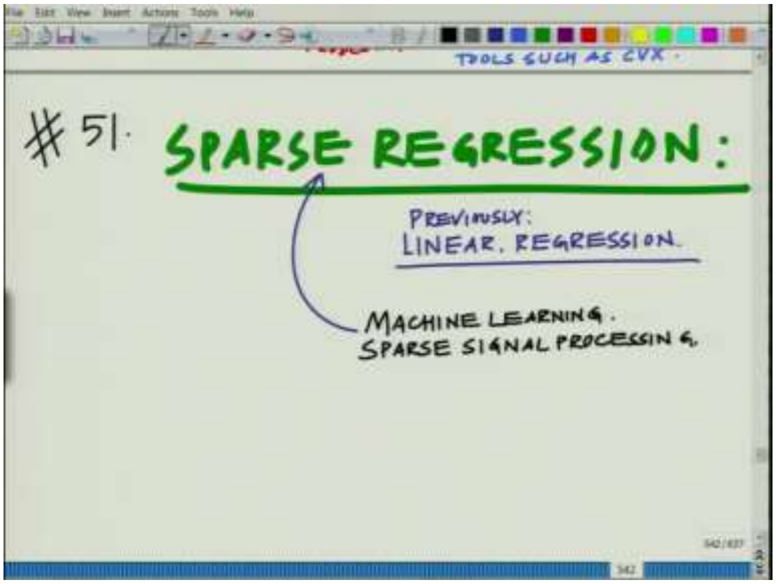
Applied Linear Algebra for Signal Processing, Data Analytics and Machine Learning
Professor Aaditya K Jagannatham
Department of Electrical Engineering
Indian Institute of Technology Kanpur
Lecture 51
Sparse Regression: Problem Formulation and Relations to Compressive Sensing (CS)

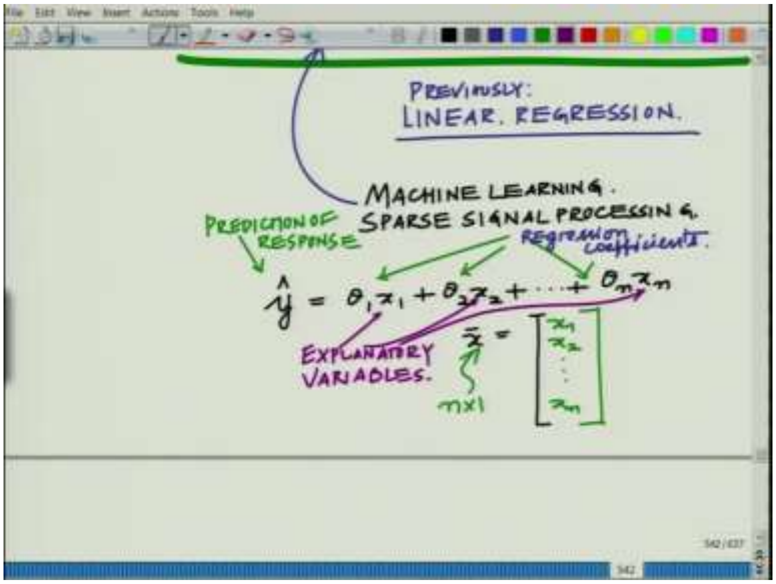
Hello, welcome to another module in this massive open online course. So in this module let us start looking at one, yet another interesting application of the principles of linear algebra. That is the context of sparse regression. So far we have seen linear regression, which is of course an important concept or an important technique in machine learning. Let us know extend it to the area of sparse regression.

(Refer Slide Time: 00:40)



So what we want to start looking at in this module is sparse regression, and what we have seen previously is basically linear regression. This is what we have seen previously and this sparse regression naturally, this also is an important technique in machine learning and as well as single processing, especially what we call as sparse signal processing.

(Refer Slide Time: 01:51)

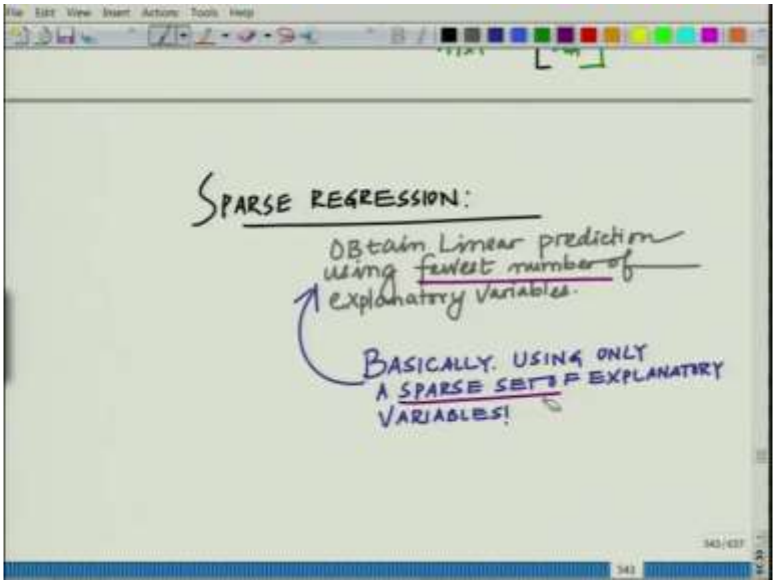


Now what happens in sparse regression, remember linear regression, sparse regression is a form of linear regression so let us start with our linear regression model. So you have y hat equals theta 1 x 1 plus theta 2 x 2 plus theta, I'll say theta n x n where these are the components of the vector that is your x 1, x 2, x n, these are the components of the n dimensional vector.

So this is your n cross 1. Now if you remember this is the prediction of the response and these are essentially your regression coefficients and these are x 1, x 2, x n these are your regressor or these are also essentially what you call as the explanatory variables.

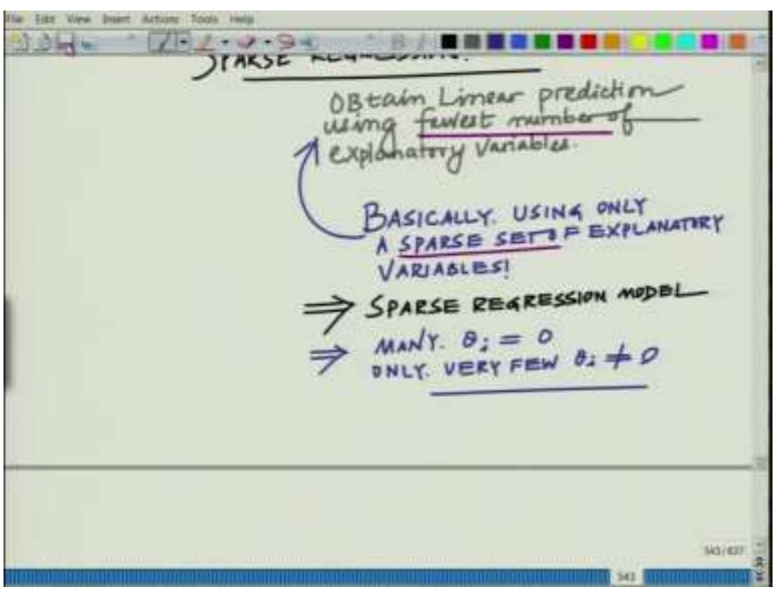
Now the point is this is your general linear regression model. Now sparse regression is a special form of linear regression wherein only a few of the explanatory variables are used to obtain a prediction. That is to obtain the fewest number of... that is to obtain the prediction using a linear combination of the fewest number of explanatory variables. So what that means is essentially, so what is sparse regression?

(Refer Slide Time: 04:10)



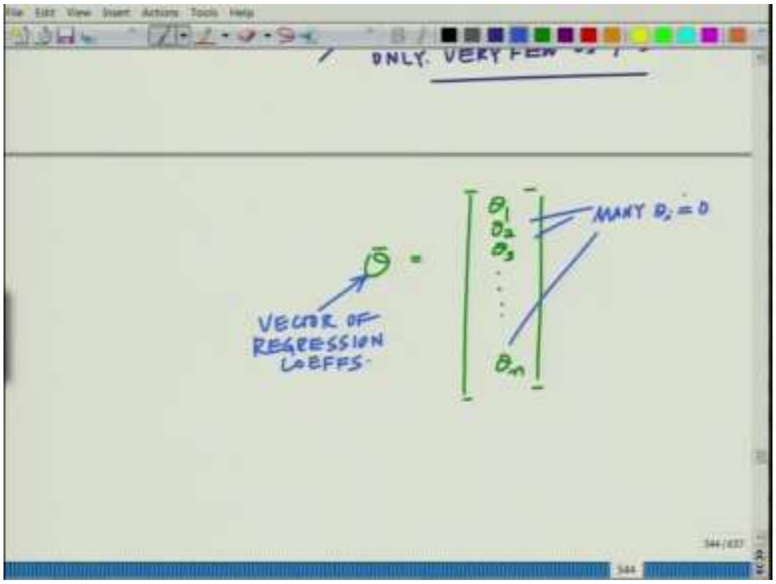
Sparse regression is to basically obtain the linear prediction, using the fewest number of explanatory variables or essentially what it means is using a sparse set or essentially, basically using only a sparse set of explanatory variables. That is we want to use only very few explanatory variables to explain the response y.

(Refer Slide Time: 06:05)



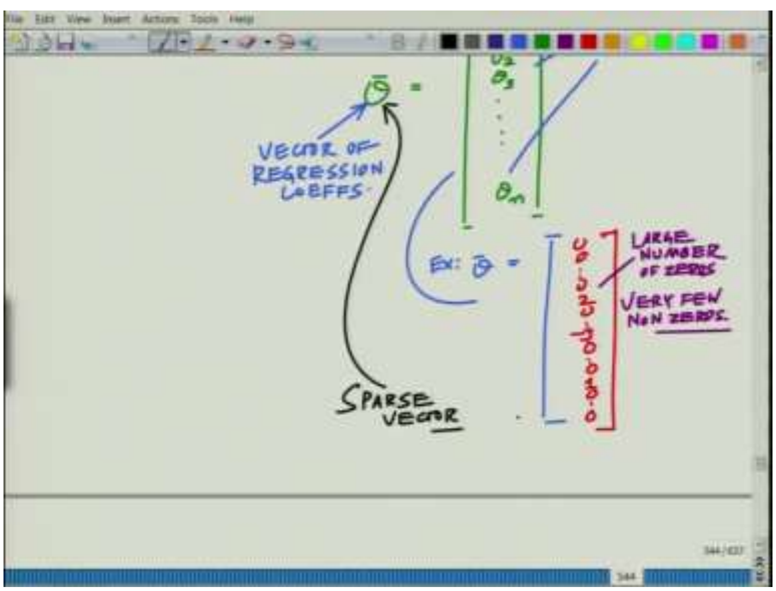
And therefore we wish to, this implies we have to determine a sparse so this basically is your essentially a sparse regression. So this basically gives your the sparse regression model in which this implies many theta i not all, but many theta i are 0 only very few theta i not equal to 0. Only very few theta i's that is only very few, remember the theta i's are regression coefficients so only very theta i's are not equal to 0.

(Refer Slide Time: 07:04)



So if you look at the vector of these regression coefficients, this parameter vector, so this is the theta bar which will contain the regression coefficients. You will have n regression coefficients. So this is the vector of regression coefficients or the parameter vector you can say. So what happens is many theta i, so therefore this looks something like this.

(Refer Slide Time: 07:46)

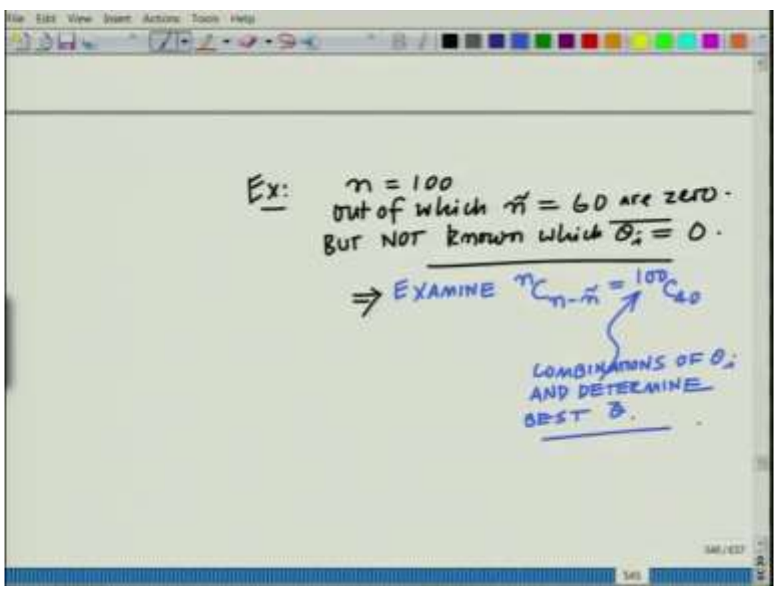


So this vector looks something like, so if you look at an example so you will have theta bar equal to, for instance you will have a lot of 0's. At some point maybe 2 0, maybe minus 3 0, 0, 0 maybe 1 0, so on. So large number of 0's, very few non 0's. And this such a regression vector, this is known as a sparse vector. So this contains large number of 0's so this is essentially what is termed as a sparse vector.

So this has an interesting property that is this theta bar vector of regression coefficients the parameter vector is sparse which means out of these n, let us say, take a typical number. For example you have 100, n equal 100, very few, maybe around five or six are non 0 right and large, the rest of the 95 or 94 values of theta i are 0.

Now what is interesting it is not known a priory which theta i are 0, if it is known that which theta i is 0 then the problem obviously as you can see becomes very simple. If it is not known which theta i are non 0 then the problem is the following.

(Refer Slide Time: 09:39)

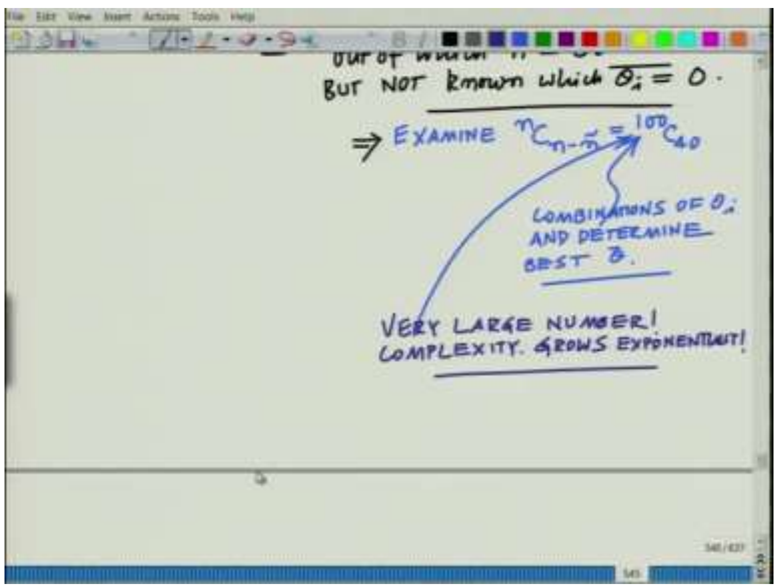


So let us take a simple example, let us say n equal to 100 out of which n tilde equal to maybe let us say 60 are 0 but not known which theta i equals to 0. So implies we have to examine all the possible combinations, examine n choose n tilde equals, or you can choose, you can say n choose n minus n tilde, which is the same, where we choose, 100 choose 40 combinations of theta i and determine the best theta bar.

So you have to choose a large number, you have to choose it for each possible combination of 40 theta, defect the linear regression model then determine, fit all such linear regression models and determine which one is the best. And natural complexity is going to be very high because 100 choose 40 is a very large number.

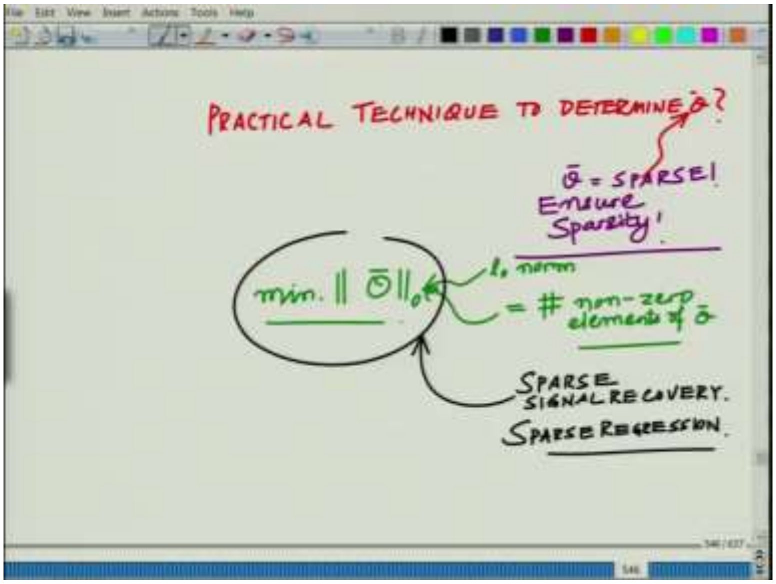
Even if it is not, if you make this number slightly larger maybe of 1,000 regressors and let us say 500, roughly 500 are 0, 500 are non 0 it becomes un-intractable right? The problem is #### 00:11:42 hard because the complexity grows exponentially, so the complexity.

(Refer Slide Time: 11:46)



So this is a very large number and therefore the complexity grows exponentially, now therefore what is a feasible technique? So this is not a feasible technique determine the sparse vector theta bar. So what then is a feasible technique? So how then to determine the sparse regression that is the vector, the sparse vector theta bar of regression coefficients.

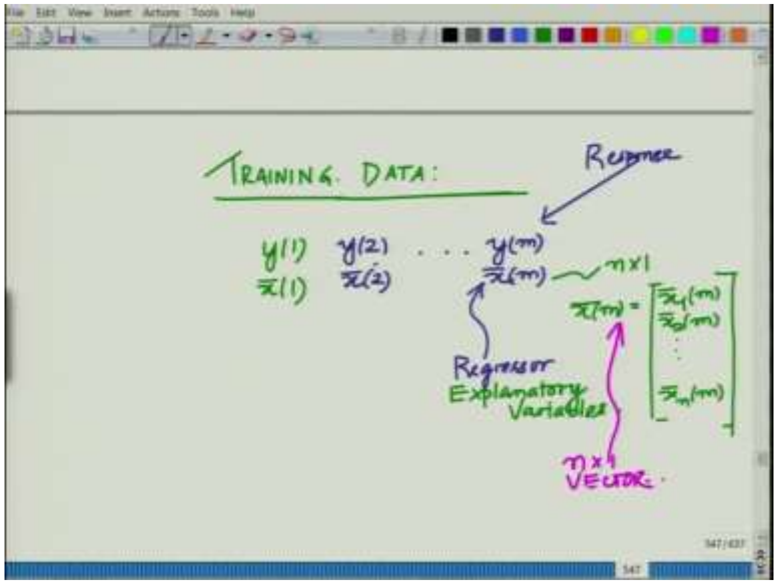
(Refer Slide Time: 12:40)



So what is a practical technique? So what then is a, what is, such that theta bar is sparse. That is you have to ensure sparsity. This can also be stated as, you can look at this theta bar, if you look at this one can determine what is known as the l 0 norm of theta bar, this is also known as the l 0 norm, which is basically equal to the number of non 0 elements of theta bar and we want to minimize this l 0 norm, that is basically what is the problem of sparse estimation, this is basically a problem of sparse estimation or you can say sparse signal recovery or you can also say in general sparse regression.

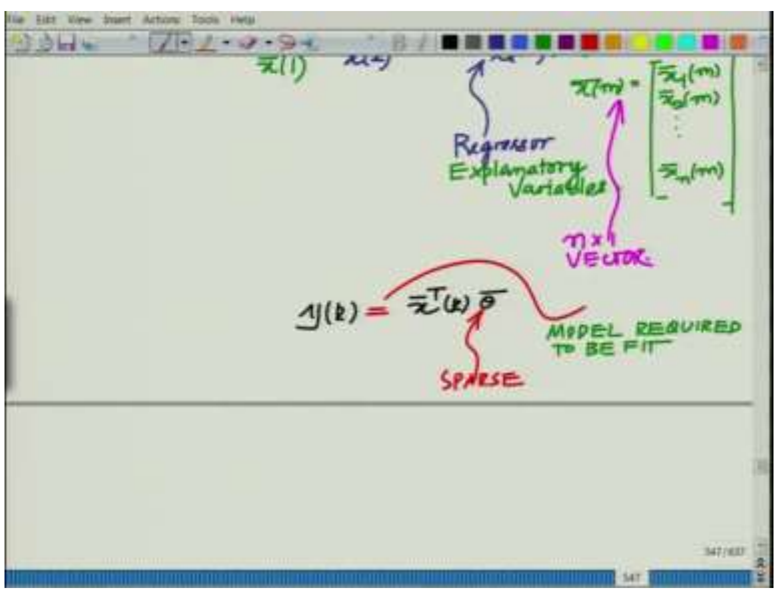
Essentially the problem of sparse regression. Now we consider to do this for this sparse regression, remember we always start with the training data. So let us start with the training data, again let us say we have these m training samples.

(Refer Slide Time: 14:58)



So let us start with the, or training data is comprises of y 1 the corresponding regressors x 1, y 2 these are the responses and these are the corresponding regressors. So each of these is the response and these are the regressor or basically or explanatory variables and this is basically your n cross 1 vector I can denote x bar, the m, the sample as x bar 1, x bar 2 of m, x bar n of m. So these are basically your n cross 1 vector.

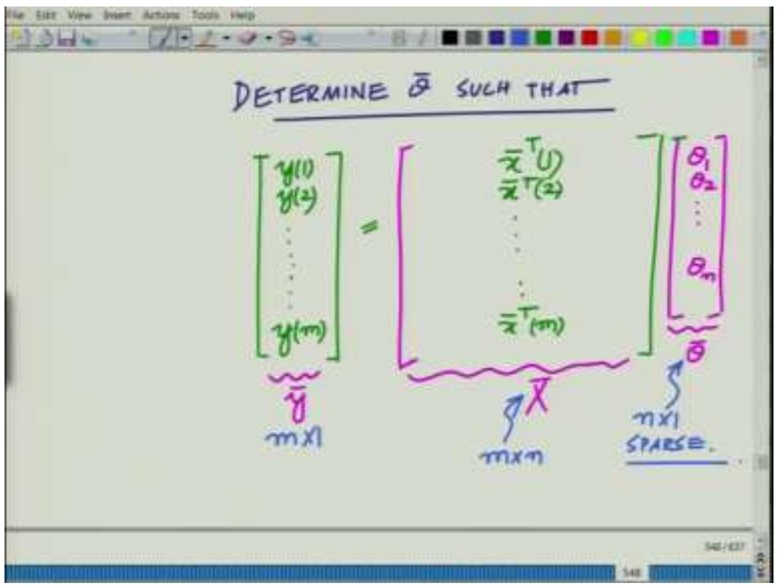
(Refer Slide Time: 16:30)



Now, we have the model. We have to fit the mode essentially the model that we want to fit is that if we call this y k minus x bar k transpose theta bar, if we call this as the error, or you can also say we want to fit the models such that y k equal to x bar k transpose theta bar.

Either there is an error which you want or minimize or if possible you want to fit this model exactly. So this is the model that we are required to fit. Once again do not forget the fundamental constrain that this model has to fit while ensuring that theta bar is a sparse vector.

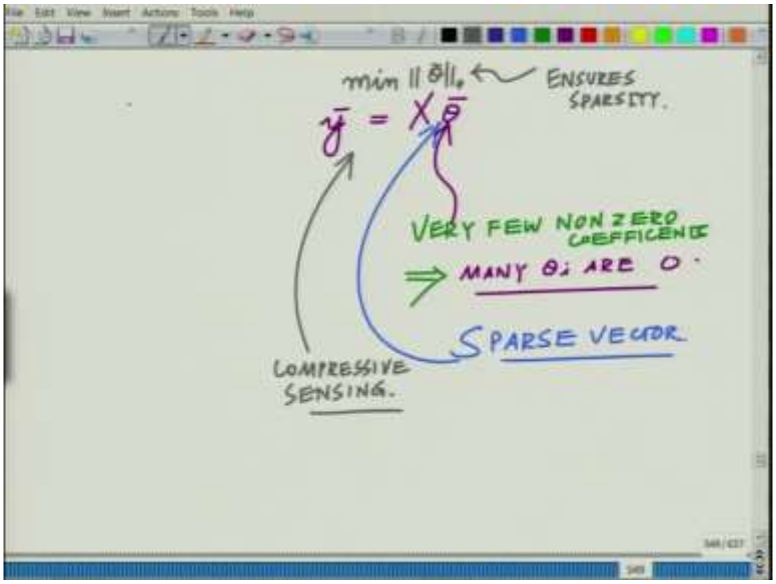
(Refer Slide Time: 17:48)



So essentially we want to find, determine theta bar such that you have basically, you have y 1, y 2, once again putting these as a vector you have y 1, y 2, y m this is equal to, similar to what we have in the linear regression x 1 bar transpose or x bar transpose 1 times you have your vector of the parameters, which is your, except that this is going to look a little different. I am going to describe that later.

So this is your regression model which essentially looks as, so this is your y bar, this is your matrix of regressors x bar and this is your theta bar and remember this is basically your m cross 1 vector. It is good to always note the dimensions. This is your m cross n matrix and this is your theta bar which is n cross 1 and this has to be sparse, all right.

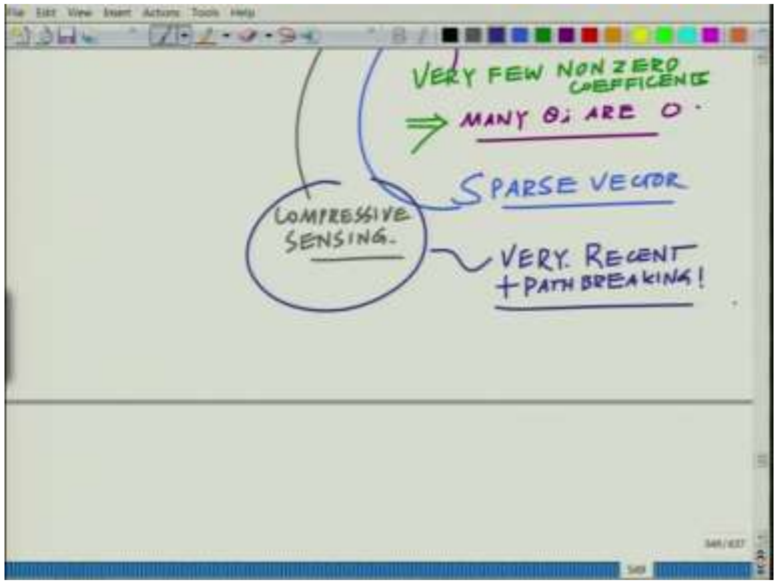
(Refer Slide Time: 19:48)



And therefore, I can formulate this as y bar equals to X times theta bar and as you already know theta bar has very few non 0 coefficients which implies many coefficients are, which implies many theta i are 0, and we have already seen an example and such a theta bar this is known as a sparse vector. And this problem also has an interesting name, this is also termed as compressive sensing that is when we are trying to do recovery of the parameter vector theta bar, such as theta bar is sparse.

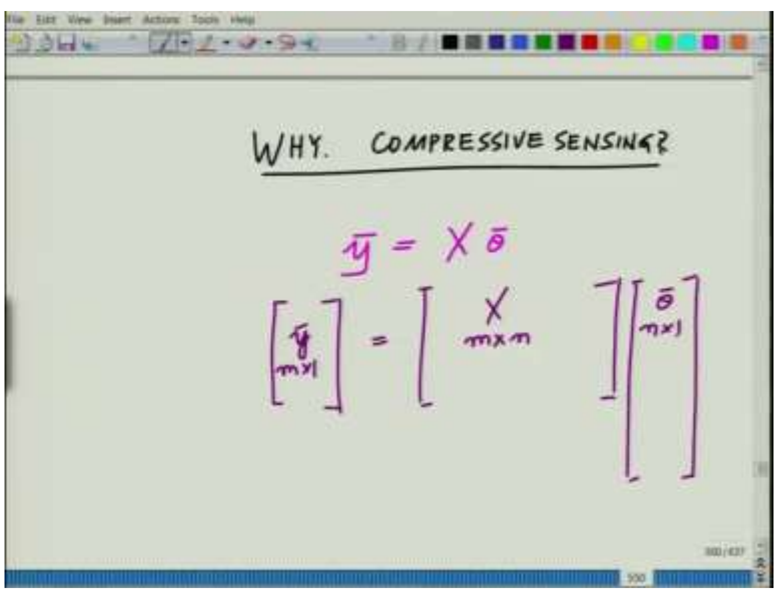
This is also has a name, this also in fact, it is a field rather, which is termed as compressive sensing in area that responds several algorithms and many interesting techniques, which is also termed as, this is also termed as, this particular problem with which we can call as including to remember theta bar 0 that is, which ensures sparsity. This is also termed as compressive sensing.

(Refer Slide Time: 22:05)



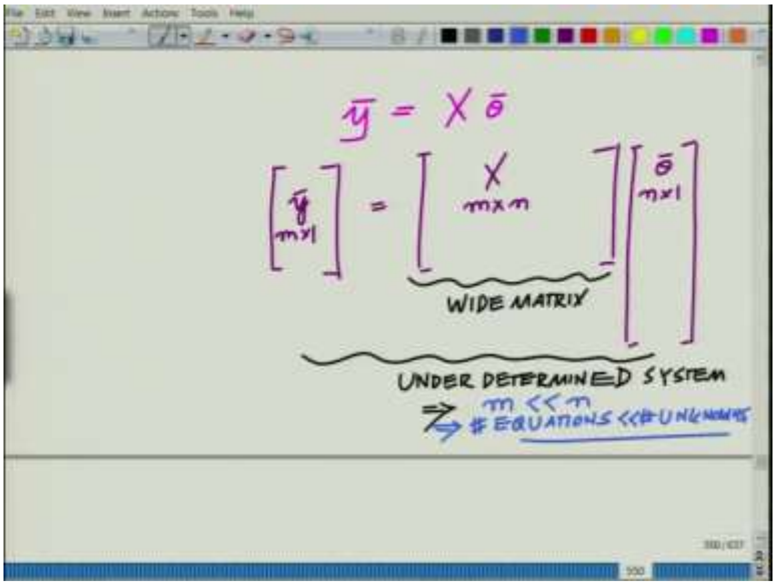
Now, why is this termed as compressive sensing? This is a very big field, this is a very recent and sort of path breaking field with several algorithms, with several innovations across many areas such as signal processing, communication, radar, tomography so on and many significant implications for a broad variety of fields wherever there are problems that are similar to for instance regression, least squares, signal recovery and so on and so on.

(Refer Slide Time: 22:57)



Now, why is this termed as compressive sensing? Why compressive sensing? Now if you go back to this, you have the problem y bar equal to x theta bar, this typically looks as follows, the vector y bar looks as this with X looking like wide matrix like this and your theta bar looking typically like this, not what we had drawn earlier but rather like this. So this is your y bar which is m cross 1 and this is X which is m cross n and theta bar which is n cross 1.

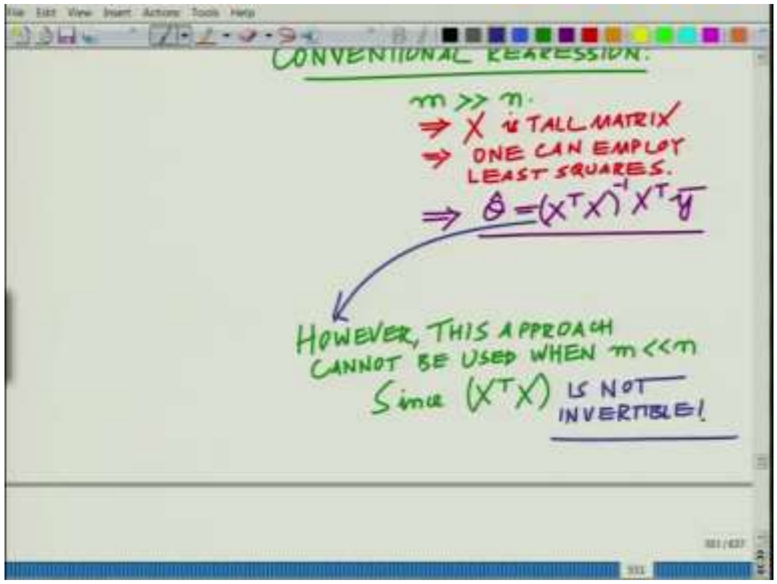
(Refer Slide Time: 24:03)



So typically what happens is this matrix is not at all matrix but this matrix is a wide matrix. That is this is basically, if you look at this, this is typically an under-determined system. This implies that the number of equations m is significantly lower than n.

This implies number of equations much less than number of unknowns. This is typically the problem one faces. And in such a situation remember the conventional linear regression. Conventional linear regression has m greater than equal to 10, right? There you have the matrix X bar is a tall matrix and then one can employ the least square solution, right? You remember?

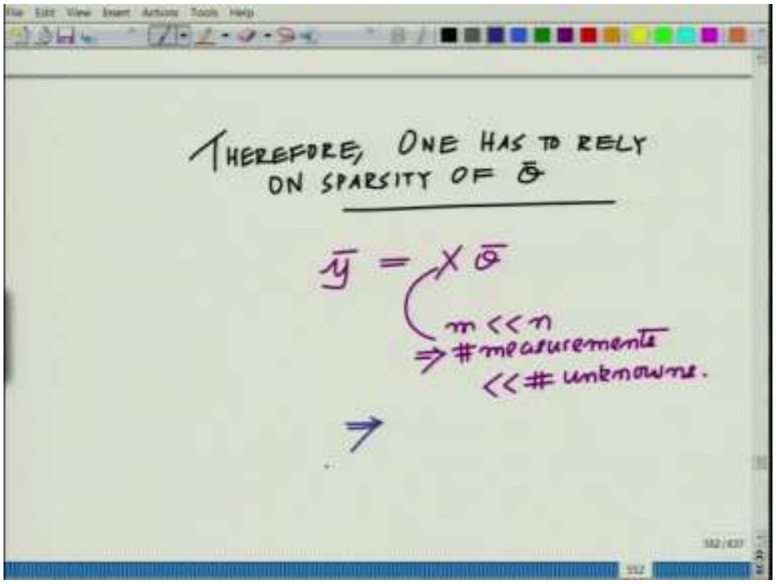
(Refer Slide Time: 25:22)



So conventional regression m greater than equal to n implies that basically X is a tall matrix and implies one can employ least squares, which is basically, which basically implies your regressor theta hat, what we call as theta hat, this is given as X transpose X inverse X transpose y bar. However, this approach cannot be used here when m less than n, since X transpose X, you can so easily, in this is not invertible.

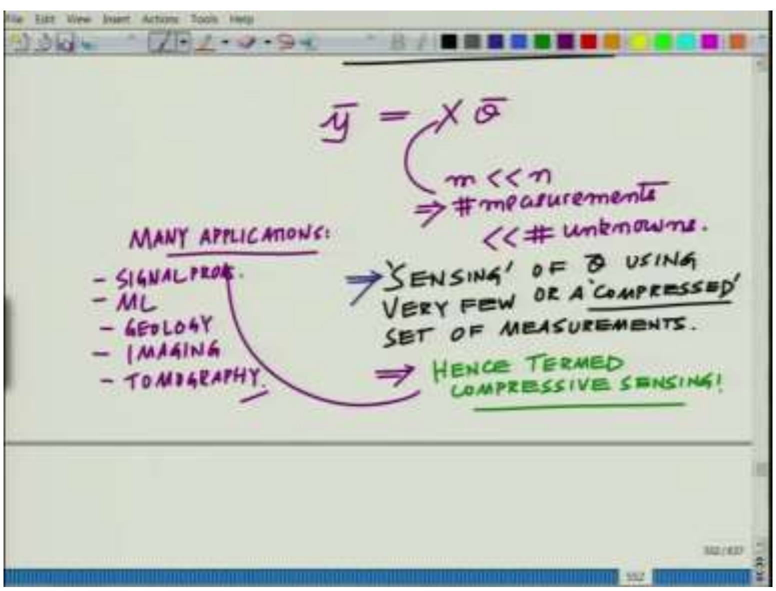
When m is less than n for that matter when m is much less than n, your X transpose X is not invertible. So you cannot use the conventional least square solution. Then what is the property that has to be used, one has to rely on the sparsity of the vector theta bar, there is only possible to reconstruct theta bar employing the property that theta bar is a sparse vector.

(Refer Slide Time: 28:16)



So in such a scenario, therefore, otherwise the reconstruction is not possible and therefore this is also known as sparse, and now therefore if you look at this problem and where m is much less than n implies number of measurement much less than the number of unknowns. Implies your sensing, that is sensing theta bar using very few measurements or essentially a compressed set of measurements. So this is sensing of theta bar using a compressed sense of measurements that is why this is known as compressive sensing.

(Refer Slide Time: 29:47)



So this implies that this is basically sensing of theta bar using very few or a compressed set of measurements and thereby this is termed hence compressive sensing. That is basically we are sensing theta bar using very few measurements or a significantly reduced set of measurements or a very compressed set of measurements hence this is termed as compressive sensing and this has significant application.

As I told you this is radically new field, I think it is, most of the techniques or most of the innovations have happened in the last 10 to 15 years. And significant innovation has been achieved and it has many applications, signal processing, imaging, tomography, geology, radar so on and so forth, so many applications. So this has such as for instance signal processing, ML, geology, imaging, tomography, etc.

All right, so there are significant applications of this field, of this newly emerging field of compressive sensing, which is radical, because remember conventional signal processing, signal estimation requires a number of measurements to be more than the number of unknowns that is we have seen in the conventional regression, the liner regression that we have, the solution for which was given using the least squares.

But in sparse regression, the number of measurement is significantly fewer and therefore one has to rely on the sparsity of the vector theta bar, otherwise the reconstruction is not possible and therefore the techniques that are developed are very normal and significantly different from the earlier generation of conventional techniques, which use the l 2 norm.

Remember this is, the compressive sensing, the area of compressive sensing is based on l 0 norm and in fact also as can be shown the l 1 norm minimization. Hence the set of techniques are radically different from the previous generation of techniques, and this naturally can lead to, has been shown to lead to improved performance and as applications in several fields be it the image processing, signal processing, tomography, so on and so forth and it has revolutionized several fields.

So let us continue discovering this and let us, let us continue, let us stop here and continue this discussion by formulating, by demonstrating the solutions, algorithm to solve those sparse regression problem in the next module. Thank you very much.