Applied Linear Algebra for Signal Processing, Data Analytics and Machine Learning

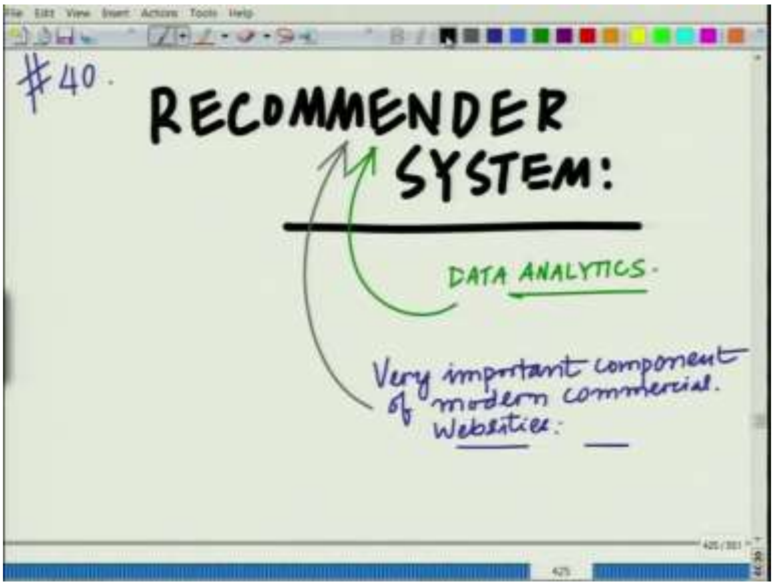
Professor. Aditya K. Jagannathan
Department of Electrical Engineering
Indian Institute of Technology, Kanpur

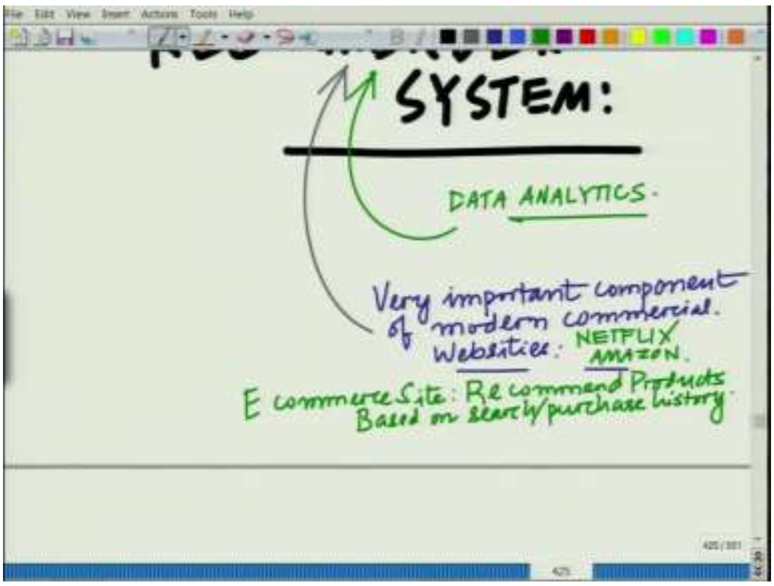
Lecture No. 40

Recommender system: design and rating prediction

Hello, welcome to another module in this massive open online course. So, let us continue our discussion on interesting applications on linear algebra in various areas such as machine learning, signal processing, communications, data analytics and in today in particular, let us look at another very interesting and upcoming application and that is in the area of data analytics, data analysis and what we call as recommender systems, which is as I am going to describe are very, very important.

(Refer Slide Time: 00:50)



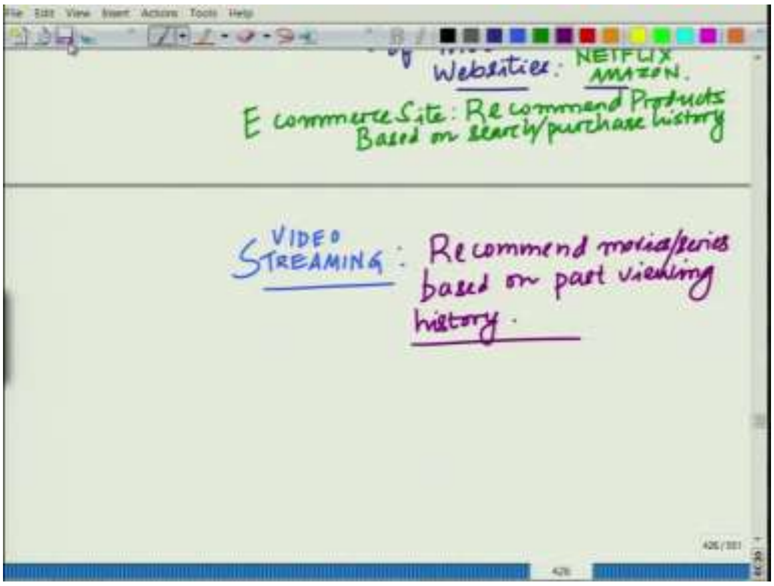


So, this is what is known as recommender system, this is a very interesting and novel application in the area of data analytics and this recommender system is a very important component of modern commercial websites for instance, if you look at websites such as either E-commerce websites such as Amazon or Flipkart or other streaming sites such as Netflix or Amazon Prime, which essentially recommend products, movies or CDs and so on.

So, that is essentially what the recommender system is about so, this essentially is very important comp1nt of modern commercial websites such as for instance, we have Amazon Netflix and so on. So, essentially what is the idea here? The idea here is for instance, if you have an E-commerce so, for instance, you have the commercial websites such as either Netflix or Amazon, Flipkart so on and so forth so, if you have an E-commerce site, how to recommend products?

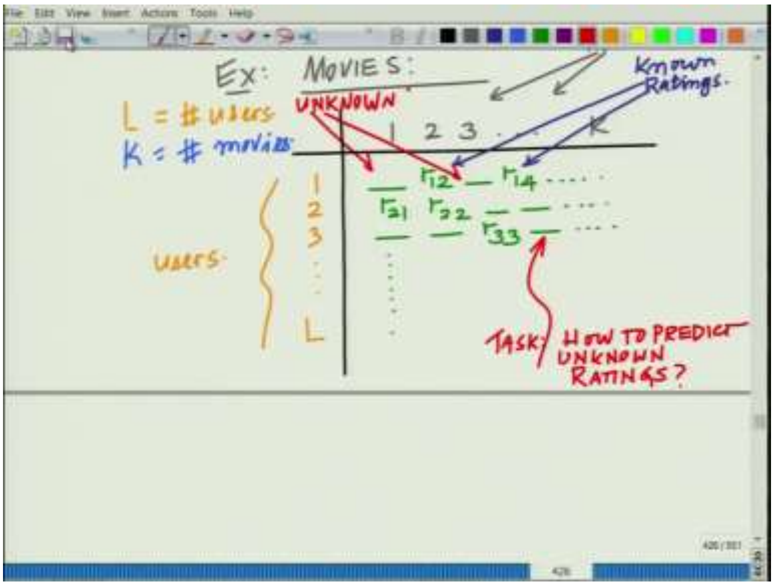
You would like to recommend products based on search, slash purchase history, the idea here is to mine the data to save to the data to look at what products you have purchased or what products a particular user has purchased or searched and look at what products other users with similar interests have purchased and then try to find a pattern and recommend the suitable product so, as to maximise the possibility that, the user is going to purchase a particular product. So, that is a very important comp1nt of modern E-commerce.

(Refer Slide Time: 4:00)



Similarly, for instance if you have a streaming site, if you have video stream then, what you want to do here, is to essentially you would like to recommend movies or series based on past viewing history that is essentially, what movies etc person has seen, what are the movies or series a person with related interests as in and then, somehow try to find these match these interests and try to recommend movies or series to a new user so, or recommend new movies to another user so, these are very challenging problems.

(Refer Slide Time: 05:20)



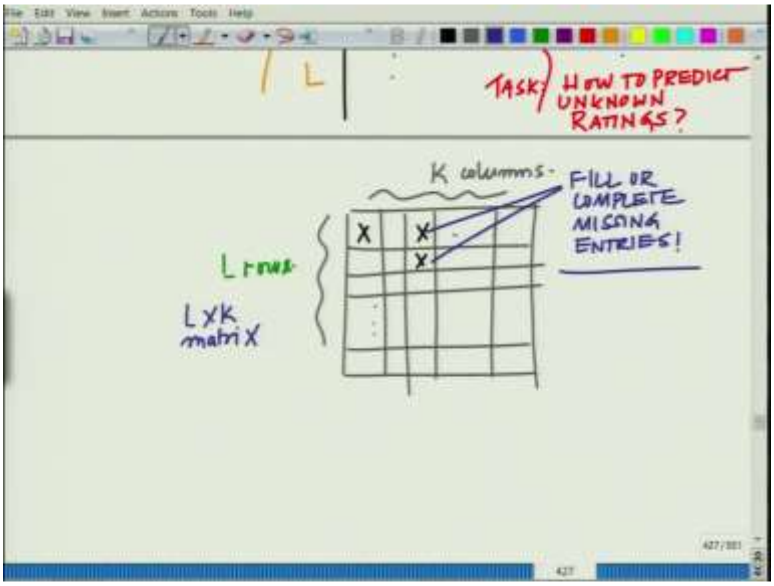
So, let us take a simple example to understand this thing for instance, let us take a movies scenario or movie streaming website let us say, we have this table in which the columns are, these are the movies and the rows are the different user. So, you have so, L equals number of users and K equals basically the number of movies and there are some ratings some users have assigned for instance, you have user 1 has assigned rating r12 to movie 2 no rating to movie 3 r14 to movie 4 and so on.

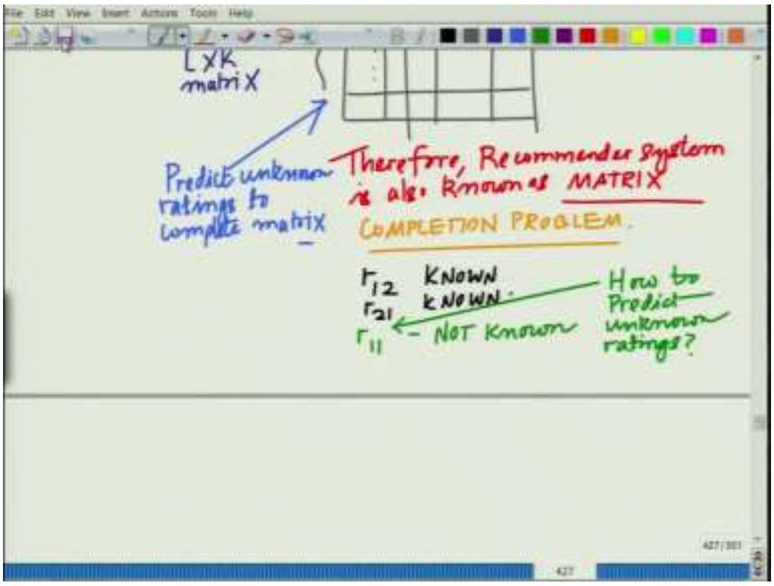
User 2 has assigned rating r21 to movie 1 r22 to movie 2 no ratings 234 maybe some other movies, there are ratings, so user 3 no rating to 1 to rating to r33 no rating to r4 so and so on and so forth. So, the essence here is now point is not all users have seen all movies. So, some users have seen some movies, for instance, user 1 has seen movie 2 and 4 and rated that user to has seen movie 1 and 2 and user 1 has rated movie 2 and 4 and rated and user 2 has seen movies 1 and 2 and rated those and so on and so forth.

So, essentially you have some users you have seen some (())(07:47). So essentially, so these are the known ratings, if you look at this, try to understand this problem in depth these are known ratings and these are essentially these where you have the blanks, these are unknown. Now, how do you predict, the task here is, how to predict? So, task is how to predict, which is if you think about this, how to predict these unknown ratings and that is essentially how to predict these unknown ratings.

For instance, user 2 has not rated not seen movie 3 so, if user 2 want to watch movie 3, what would the rating r2 be 3 r2 3 be how to predict that and based on that one can recommend based on the rating predictions therefore, one can recommend these products or movies to user 2 and so on for other users.

(Refer Slide Time: 09:32)





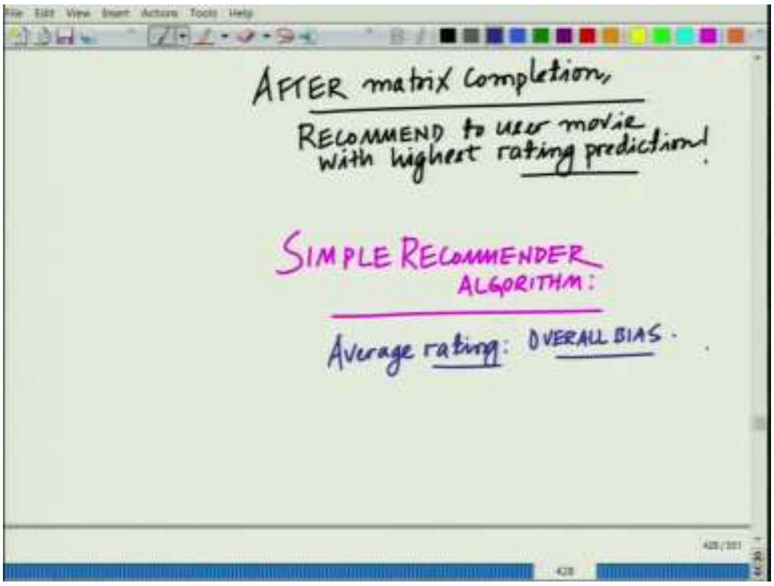
So, essentially as you can very see, this is a matrix, if you look at this, this is a matrix so, if you look at it at the heart of it, you can easily see and so, this is I think K is the number of movies so, you have K columns and you have L rows with some entries that are vacant for instance r11 is vacant r13 is vacant so, these are for instance, these are vacant r23 is vacant.

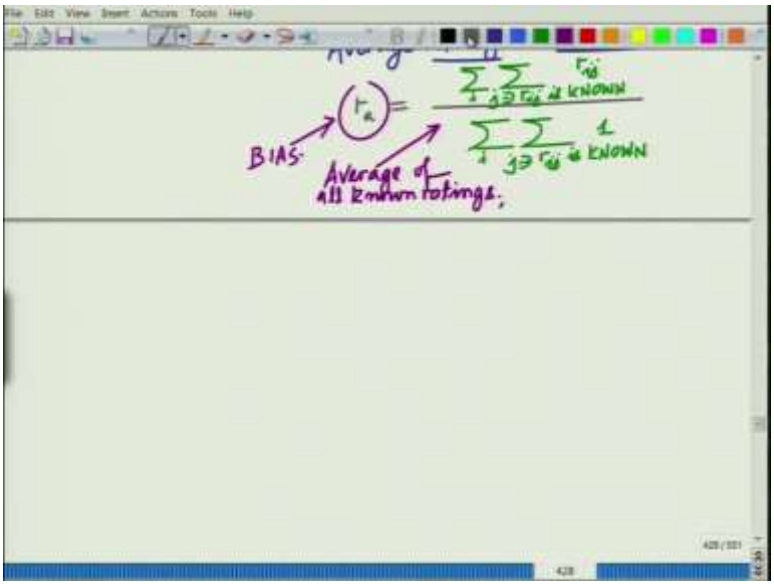
So, now you have to fill this vacant so, this is a K cross so, you can see this is an L cross k matrix and we have to fill or complete the missing entries and therefore, this is also known as this recommender system is essentially also known as a so therefore, is also known as your matrix completion problem so, this is bariums to broad class of problems known as matrix completion problem or broad class of algorithms to complete matrices.

So, matrix so, you can think of these as matrix completion techniques, the essential idea is there are some ratings some ratings are missing, how do you predict these unknown ratings to complete this meeting matrix that is the idea of the matrix completion problem. So, the idea is predict these unknown ratings to complete the matrix that is the idea behind this matrix completion problem. So, for instance not all ratings are known, r12 is known so, you have r12 know r21 this is also known and then, you have r11 this is unknown or not known so, how to predict, that is the idea so, how to predict these unknown ratings?

And therefore, this is what made a matrix completion problem once you complete the matrix after matrix completion, you simply recommend the movie now, how does this work once the matrix is completed, you have the predictions you have the ratings that the user has given and you have the predictions of the ratings for the movies which the user has not seen now, for the movies with the user has not seen simply recommend the movie with the highest rating prediction so, that is essentially or recommender system so, the matrix completion problem naturally leads to a recommender system.

(Refer Slide Time: 14:22)





So, once so, after matrix completion command to use a movie or product with highest so, once you complete this matrix, the only thing that is left to do is essentially recommend a movie and what a movie or you can go into recommend the movie you can recommend is only of course, you cannot recommend a movie that, the user has already seen does not make any sense, you can only recommend a movie makes sense to recommend a movie that, the user has not seen and how do you recommend the movie that not the user has not seen?

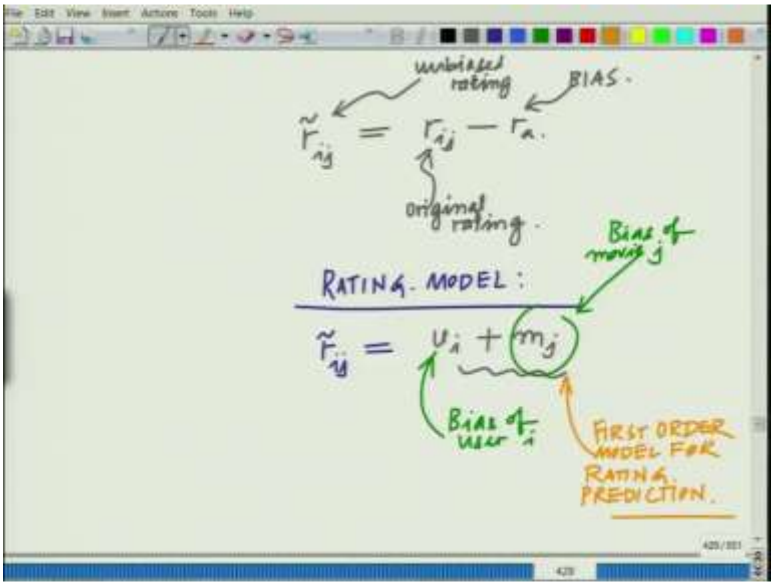
Simply choose the movie that the user has not seen with the highest predicted rate and that would be your guests of the most preferred movie buy but, that particular user and so that you maximise the chance that the user probably purchases and watches the movie or subscribes to that streaming service or website and so on and so, all the user purchases that particular product and so on so, that is the essential idea.

And how do we implement and if you look at it, it is very interesting how linear algebra comes into this so, let us look at a simple recommender system what a simple recommender let us look at a simple recommender algorithm for instance, this would be rA let us take all the ratings. Now, let us compute the average rating or you can say this is basically your overall bias.

So, similar to several techniques that, we have seen such as PCA and so on I can faces etcetera, one can look at all the data and remove the bias, compute the mean sample mean and remove that treat that, as a bias and remove that so, the overall bias or the average can be computed as follows.

So, I can use all the existing ratings so, sigma i sigma j such that, rij is known, you take this that is you choose all the ratings that are known and simply take the average j such that rij is known and this is basically, what this is doing is this is simply taking the sum of all the known ratings and taking their sample that is average of all known ratings so, if you look at this would this is performing average there is performing average of all known ratings and this is what we are calling as the overall bias.

(Refer Slide Time: 18:35)



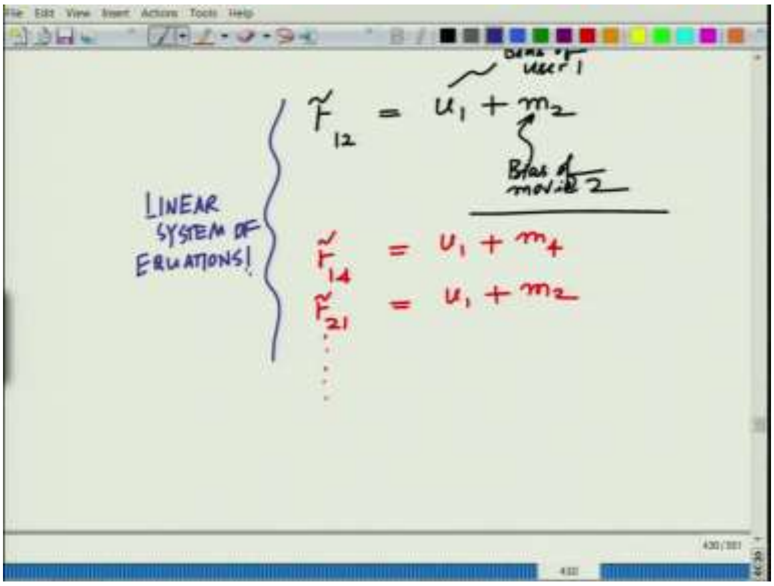
So now, you remove this to form the unbiased rating r tilde ij equal to rij minus so, this is your and this is your bias and this is the original rating now, we have to construct a model for the rating, what is our rating model? Of course, here one can construct a very complex model but, the idea is to keep it simple let us try to come up with a first order prediction for the purpose of this lecture and of course, in research and of course, in further analysis one can come up with more complex models.

And in fact this recommender system is the its confidential, you can clearly see, if a website has a good recommender system then, that naturally has a big advantage in the market so, many of these things are unknown, it was people the different companies or websites would like to keep these algorithms confidential and either would like to keep them confidential unknown or maybe would like to patent them so, that no one else can use the abuse it without paying a royalty.

So, let us look here at a simple ratings model, let us say this r tilde ij and this is reasonable can be expressed as a sum of ui plus so, a simple model that is where this ui is the bias of user i and this mj is bias or essentially characteristic of movie so, there are 2 things here, each rating can be reviewed as a sum of 2 things one is, if the user has a particular bias maybe some users read all movies high some users are very critical some beat all movies so, there is a certain user bias and then, there are some movies for instance, which are good which are rated well by all the users some movies which are rated poorly by all this, there is some inherent qualities of the u of a movie naturally.

So, we can express this as a first order model as ui plus mj that is bias of user i plus bias corresponding to movie j this is the r tilde ij so, this is you can think of this as essentially a first order model for rating you can think of this as a first order model for rating prediction.

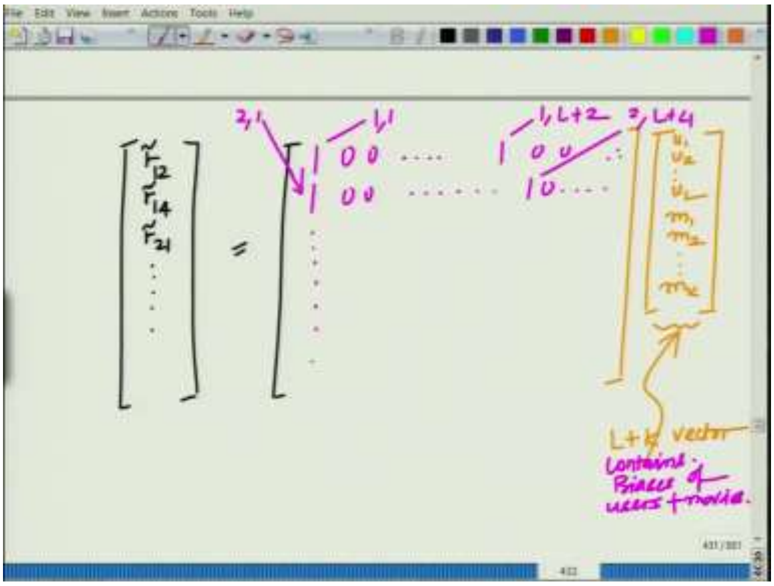
(Refer Slide Time: 22:00)

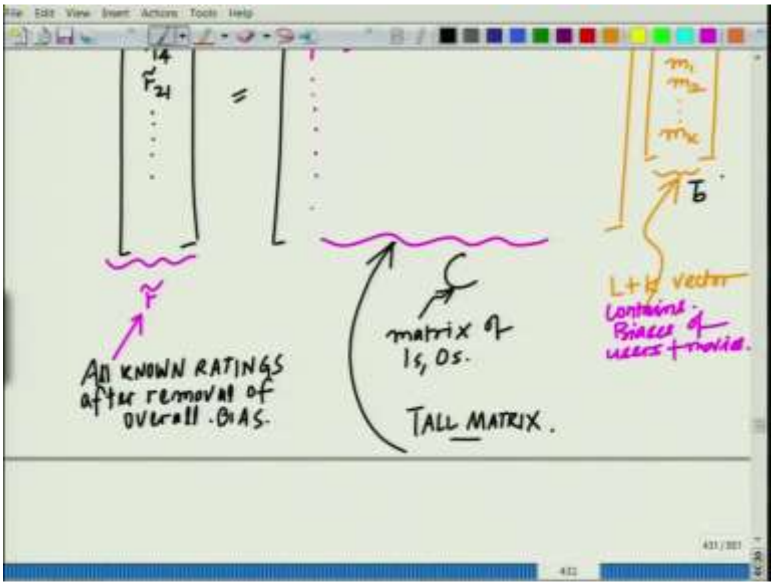


And there for instance, you have your r tilde 12 that can be expressed as for instance, let us say that can be expressed as u1 plus m2 that is bias of user 1 and this is bias of movie 2 so, you have r tilde 12 equals u1 plus m2 now, we are in business now, we have therefore, r tilde 1 2 equal to u1 plus m2 and then, you can also write for instance r tilde 1 4 equals u1 plus m4 r tilde 2 1 equals u1 plus m2 and so on and so forth and then, therefore for all the known ratings you have a nice linear system of equations and what are the unknown quantities?

The unknown quantities are these biases corresponding to the users and the movies that is ui and mj these are unknown quantities and we have to determine this from this linear system of equations. So, we have a very interesting problem and now, you can see the application also these are essentially all these quantities, these are essentially your unknown and naturally, when you look at this something like this, you should be immediately tempted to represent this in matrix form that is the whole idea behind this course, to use the principles of linear algebra matrix algebra to get valuable insights.

(Refer Slide Time: 24:09)





And therefore, I can naturally write this as a system, as the system which is of course, in practice you can guess this is going to be very, very big r tilde 1 2 r tilde 1 4 r tilde 2 1 this is equal to I have matrix now, let me first write this vector over here, these are the biases corresponding to L users and these are the biases corresponding to the K movies so, these are, this is an L plus K size vector contains biases of users and its contains biases of users and movies.

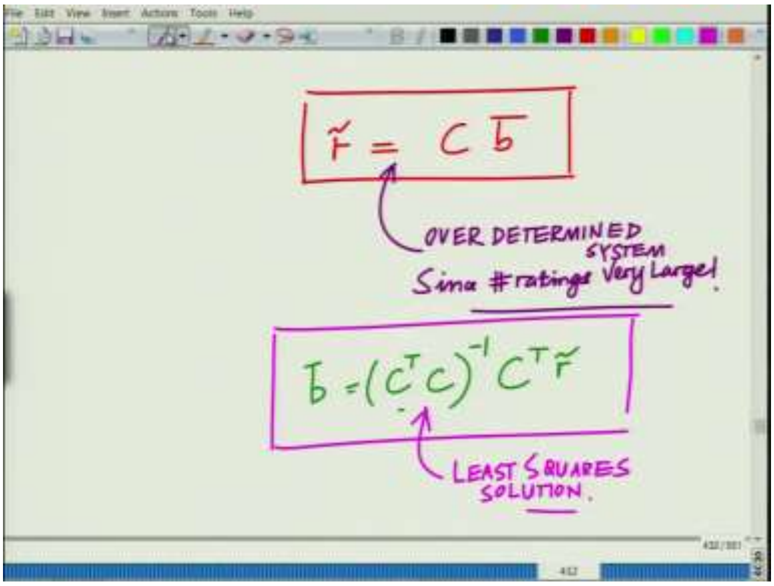
And here, you naturally have this for instance, r tilde 1 2 this is r u1 plus m2 so, you will have a large number of 0s and at some point you will have m2 so, this will be your 1 1 entry and this will be since you have the L users and then this is m 2, so, this will be the 1 comma L plus 2 entry corresponding to m2. So, you have 1 comma L plus 2 and similarly, r tilde 1 for this will be u1 plus large number of 0s plus m4.

So, this will be again, your 1 comma L plus 4 entry and this will be the second row so, this will be 2 comma L plus 4 and this will be to comma and so on and so forth and of course, rest will also be so on and so forth, what you can find this matrix, in this matrix, you can call this c this is basically you can call this as your vector r tilde, which contains all the unbiased all the after removal of overall bias.

All ratings all known ratings after removal of overall bias so, this and this is the matrix c which essentially matrix of 1s and 0s, you can clearly see this is a matrix of 1s and 0s and presumably, this is going to be a tall matrix, the idea is that you have a large number of ratings because, remember each user is probably seen hundreds of movies so, I mean let us say you have about 100 users and let us say you have 1000s of movies, each user has let us say seen and rated about 100 movies.

So, we can clearly see there is going to be each user, each user has seen 100 movies so and you have 100 users and you are going to have 100 into 100 it is 10,000 such. So, 10,000 such rows in this matrix corresponding to the known ratings so, you can see this is going to be a very tall matrix. In fact, the number of unknowns will be very small compared to the number of known because, the number of ratings that is available is typically huge and of course, one has to come up with techniques to solve this efficiently and so on and so forth and therefore, now you have and if I call this as the vector b bar, this is the bias vector b bar.

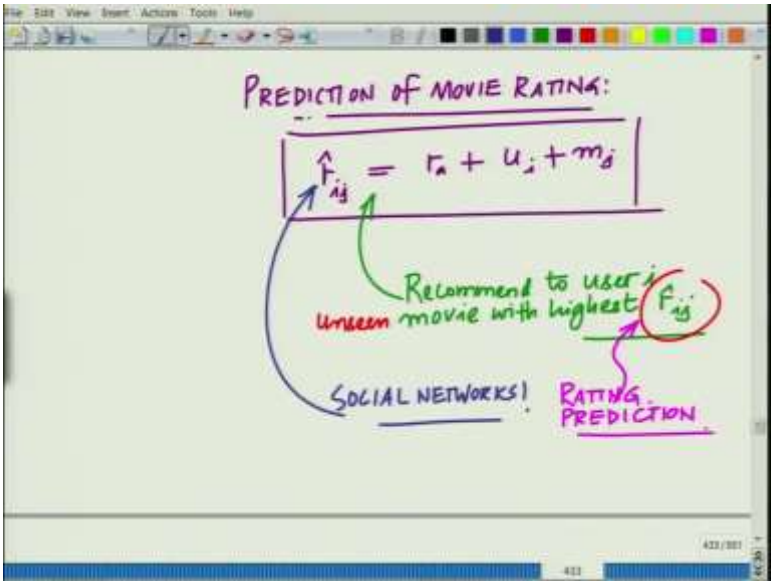
(Refer Slide Time: 28:48)



Then, you have a very interesting system that is r tilde equal to cb bar where, this is essentially an over determined system because, you have a large number of ratings. Since, the number of ratings is very large and naturally, you know what is the solution for this, at this stage, we have to be very comfortable solving over determined systems and therefore, we can write b as c transpose inverse c transpose c inverse c transpose r tilde, let me just write it, you can write b bar equals c transpose c inverse c transpose r tilde.

And this is essentially the once again you are very familiar and popular and well loud solution throughout all of engineering and sciences this is what we call as the least squares solution and therefore, once you have the b a bar.

(Refer Slide Time: 30:38)



Now, prediction for an unknown movie can naturally be determined prediction of movie rating can be obtained as now, r hat ij equals this is the overall bias plus ui plus mj this is what gives you the prediction and now the recommender system is to recommend to a particular user i the movie j which, that particular user has not seen and has the highest rating now, to recommend.

Now, the next the last step is to recommend to user i movie with highest movie with highest r hat ij of course, I am not saying unseen movie because, the natural if you have a rating prediction then, it would be a movie with the user has not seen otherwise, if the user has seen the movie, you would have a rating, you would have actually have a rating so, you can we can add this but, it is not just recommend to user i movie or you can see unseen movie with highest r hat ij which is essentially the rating prediction.

So, this is essentially a schematic or an outline of a typical recommender system and as I have already to so, this is a very interesting and novel application, there is not something that was there let us say even 30 40 years even 10 years ago, the all these algorithms have become very popular and becoming more and more popular every day with 2 things, one is the large amount of data, social networking, this is a lot of relevance for social networking.

You can see this depends on the kind of mean wherever, you have a social network large number of users, users liking posts, users rejecting posts, users rating posts, and then, you can take all this information and come up with patterns of which user rate is going to rate which tweet or which post or whatever, what are the going to be the predictions and as you can see, clearly this is something that is very relevant to the current era of social networking and E-commerce and web streaming services.

So, another very interesting application of this could be to social networks and ratings of posts and so on so forth. So, this is these are in fact, very interesting applications that are very relevant to the current era dominated by social networks E-commerce, video streaming services and so on, where you have a large number of users participating rating, choosing different objects, products rating them and that information can be aggregated process using linear algebra to come up with meaningful patterns and information that is once again useful to everyone who is a part of this big let us say internet, society or the social networks.

So, let us stop here on this note and continue with other such examples, interesting examples and modules, interesting examples in the subsequent modules. Thank you very much.