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

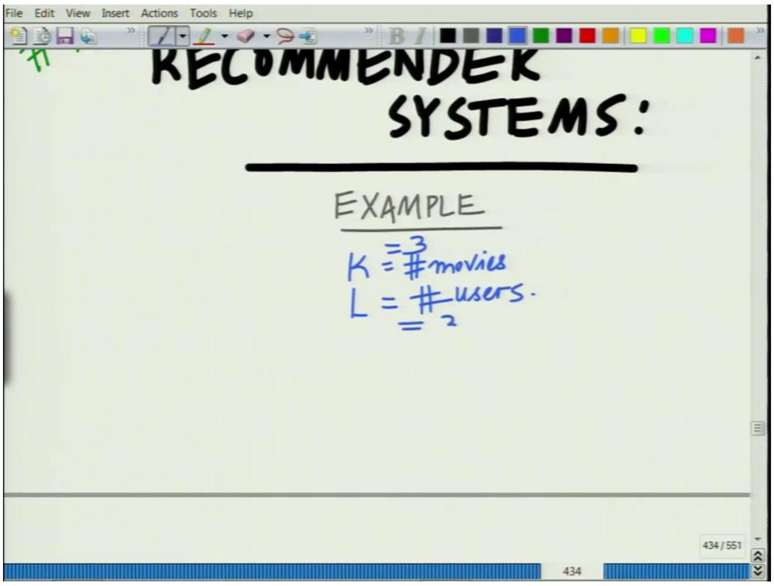
Professor Aditya K. Jagannathan
Department of Electrical Engineering
Indian Institute of Technology, Kanpur

Lecture 41

Recommender system: Illustration via movie rating prediction example

Hello, welcome to another module in this massive open online course. So, we are talking about recommender systems specifically, the relevance of linear algebra and the implementation of a good recommender system which has significant applications as I have already told you making recommendations on any commercial website be it E-commerce or video streaming, movie streaming and so on, such as even YouTube, which recommends videos along with commercial sites such as Netflix and so on. So, let us look at an example to understand this in action, a simple paper and pen example of a small system to understand how this such, a system can be implemented in practice.

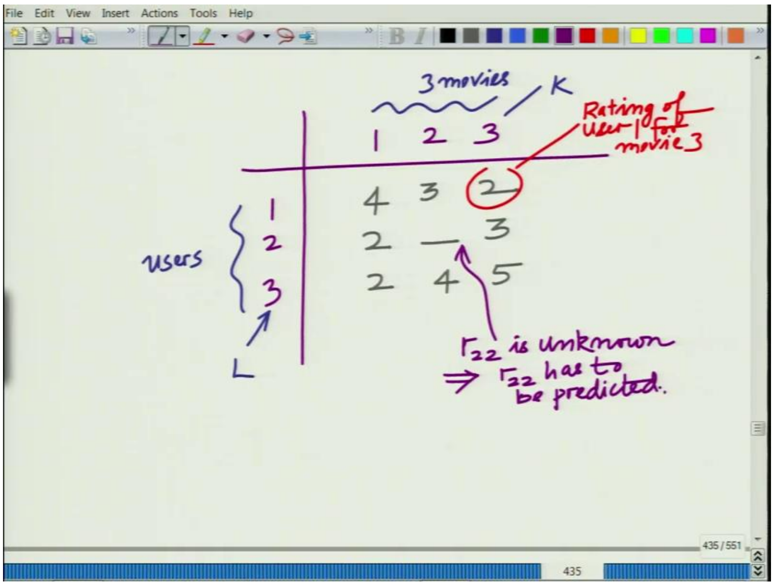
(Refer Slide Time: 00:58)

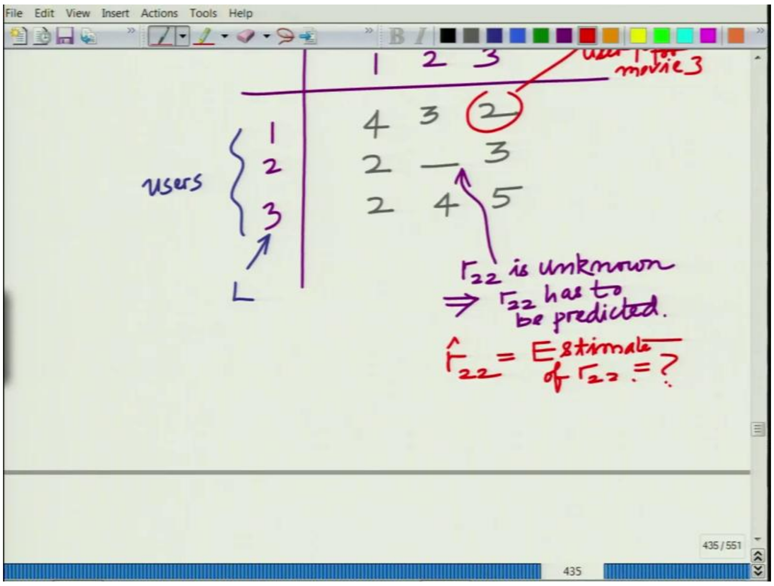


So, we are looking at recommender systems or a recommendation system and now, let us look at an example to understand this better an example of a recommender system for as we have said similar to what we have seen yesterday, a movie recommender system, which you are the kind of the kind that, would be that could be employed on a movie streaming website.

So, we have let us consider a simple example let us set K remember is the number of movies let us set K equal to 3 and L which is the number of users you will remember L is the number of users let us also set, let us set this also as 3 and remember, we have this recommendation of movies that is the table with users and their recommendations.

(Refer Slide Time: 02:26)



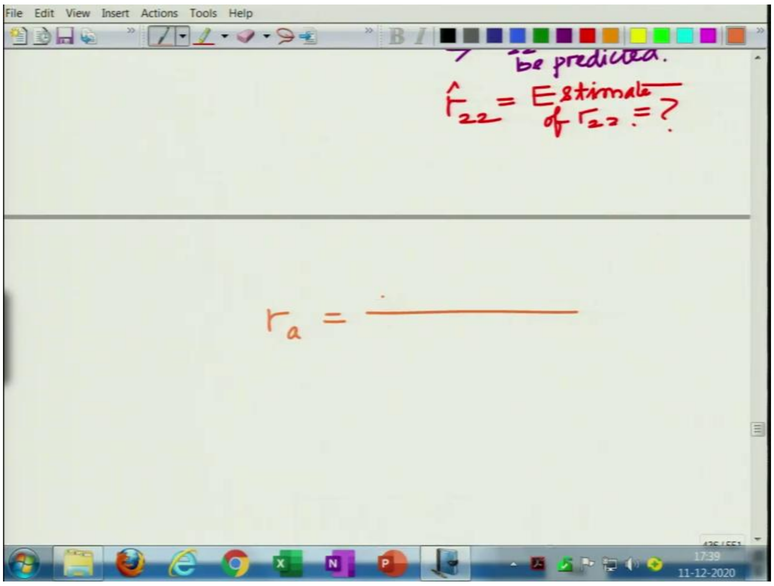


So, the columns are essentially the movies and the rows are the users so for instance, we have 3 users and 3 movies this is your K, this is your L let us put in some sample recommendations in this example, let us say the recommendations are as follows, so we have 4 3 2 2 blank 3 and user 3 has recommended 2 4 5 so you understand the meaning of this, which is basically the recommendation of the rating so 2 this is the rating of user 1 for movie 3.

So, ij we look at this as a matrix i is the rows are the users columns are the movies so, ij that is 1 comma 3 denotes the rating that user 1 has given to movie 3 and user 1 has given a rating of 2 and similarly, user 3 has given a rating of 2 to user 1. Now, this in between that is r22 the user 2 has not seen movie 2 and this rating has to be predicted. So, r22 is unknown this implies r22 has to be predicted that is we have to have a former estimate r22 hat.

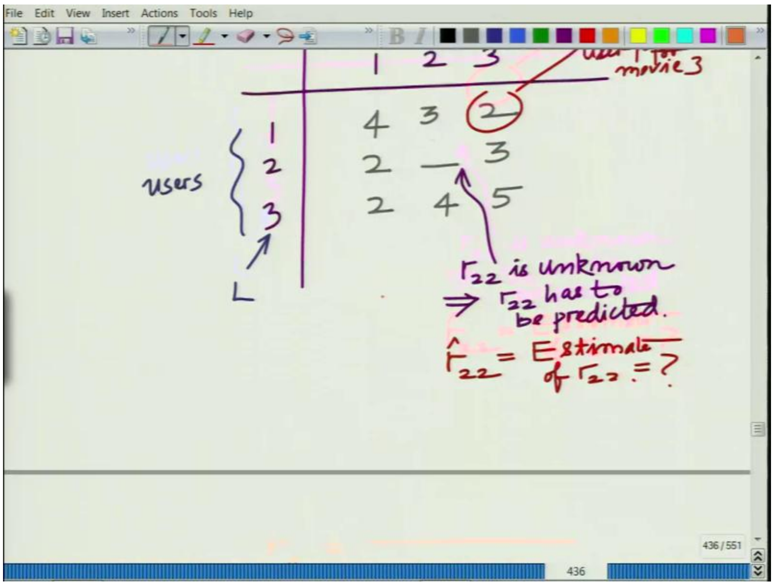
So, this is the estimate and we want to ask the question, what is r22 hat, what is the best possible estimate r22 hat of r22 which is the rating user to probably give to movie 2 had he watch movie 2, based on and look at this very interesting, based on the ratings that user 2 has given to other movies and also based on what other users have rated similar movies and so on so it is a very, it is a, there are layers of complexity in this problem then which makes it a very interesting and a challenging problem of course, a lot of applications as I told you in the modern era.

(Refer Slide Time: 05:30)



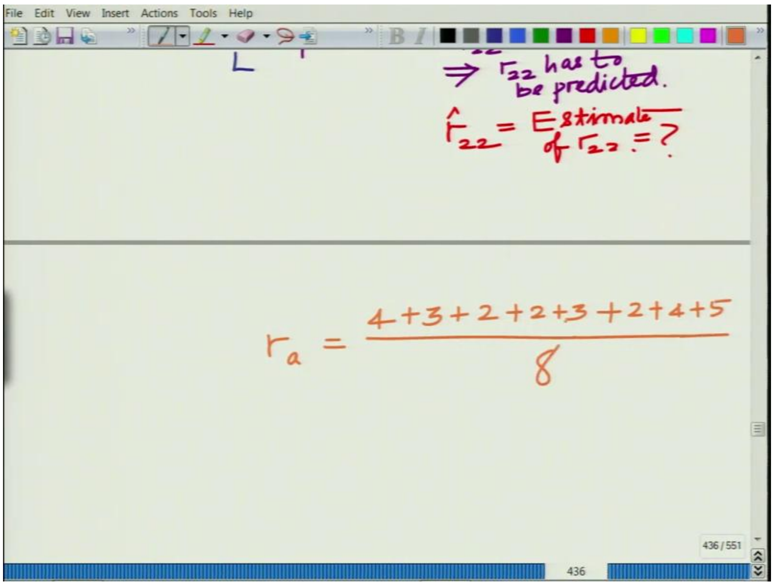
Now, the first step is as I told you to find the overall bias that is your ra which is equal to for instance, take all the ratings that are there which is basically.

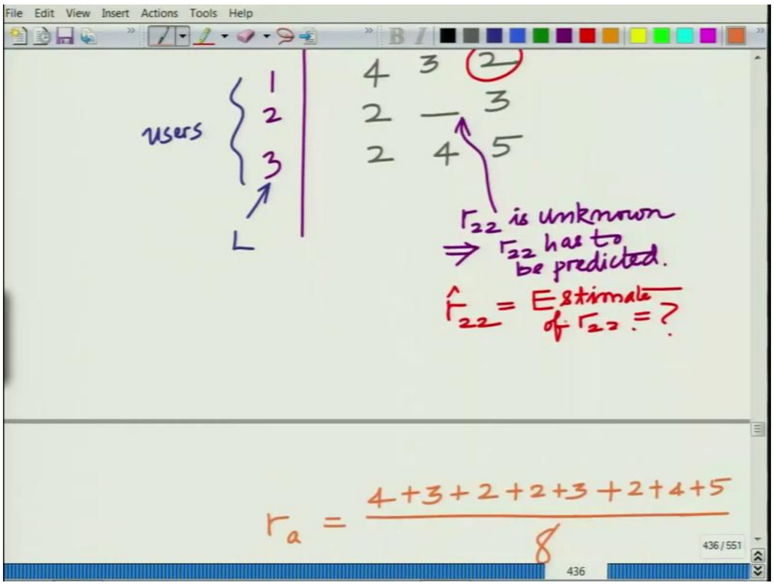
(Refer Slide Time: 05:51)



The sum of all the ratings and divided by the total number of ratings which you can see is 8.

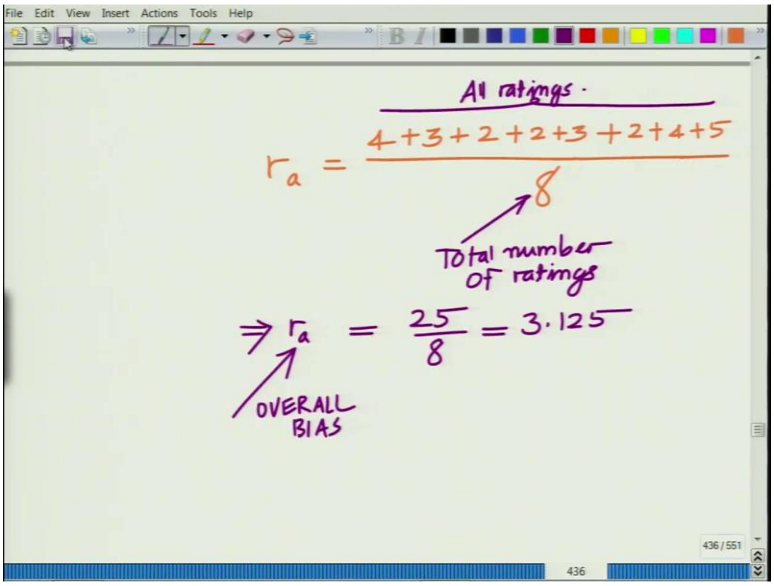
(Refer Slide Time: 05:54)

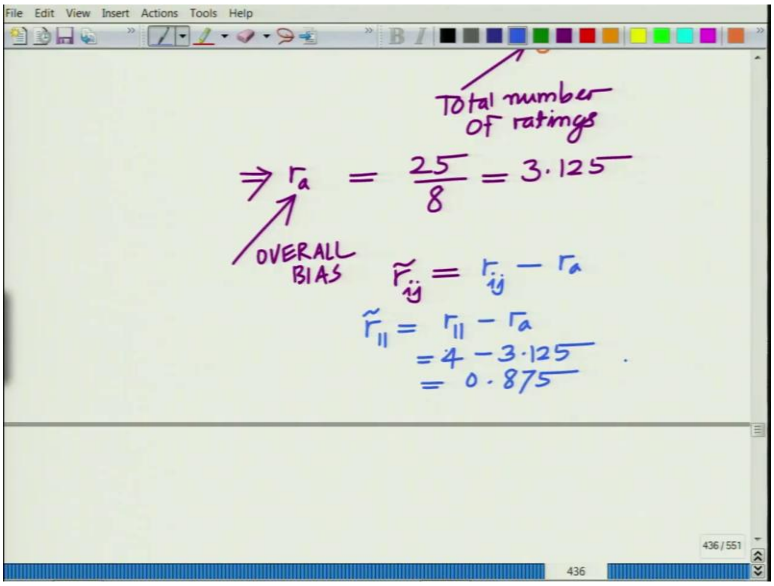




So, this would be 4 plus 3 plus 2 plus 2 plus 3 plus 2 plus 4 plus 5 divided by 8.

(Refer Slide Time: 06:17)

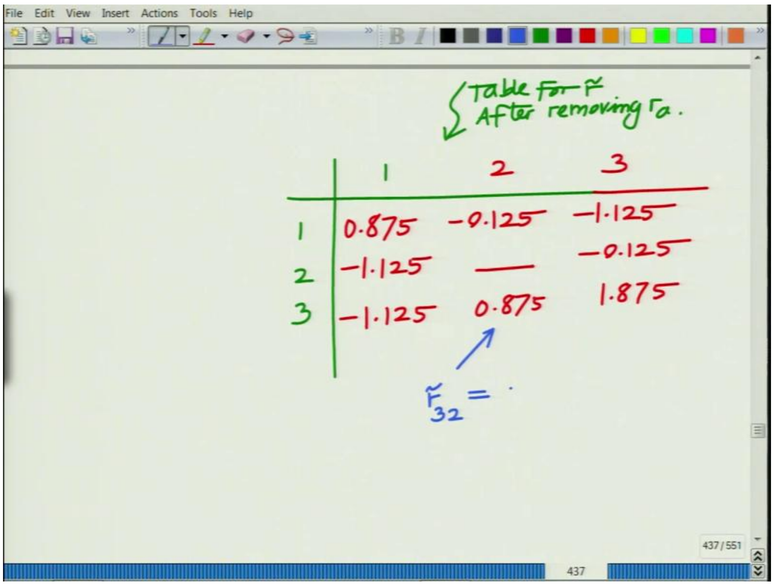


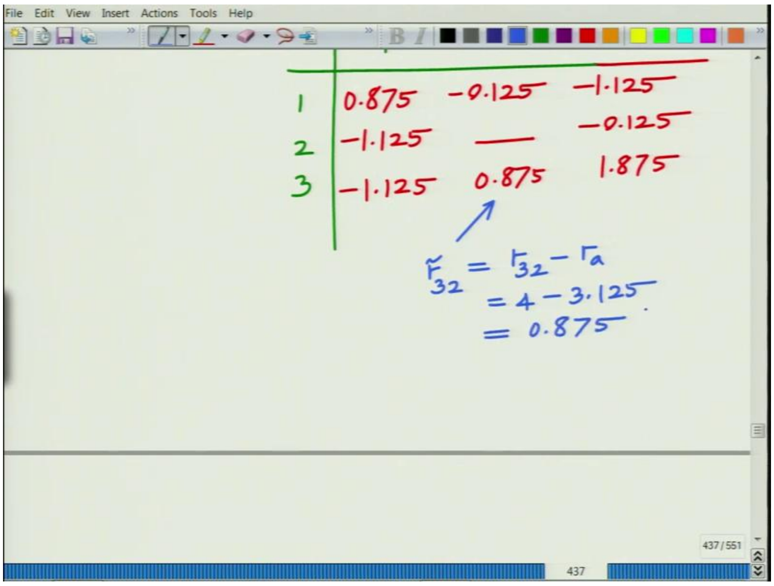


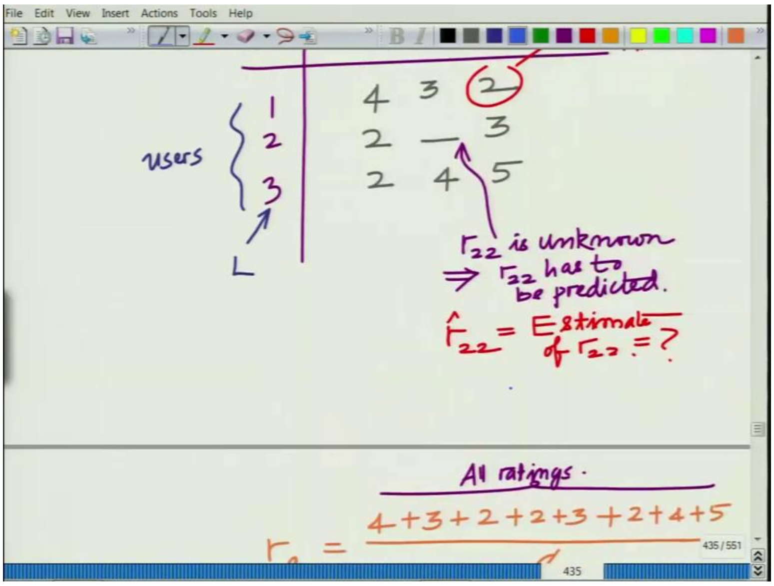
You can see 8 is the total number of ratings this is basically your, this is basically the total number of ratings and these are essentially all the ratings and this ra this overall bias this ra you can evaluate it, this comes out to be 25 divided by 8, which is equal to 3.125 so, this is your, this is the overall bias.

Remember, that is basically the sample mean of all the available ratings, subtract the overall bias to get the unbiased ratings so, you have r tilde this is basically rij minus ra for instance, r tilde 11 this would be r11 minus ra which is r11 as you can see, this is 4 minus 3.125 this is 0.875 so, r tilde 11 is 0.875 and so on and so forth.

(Refer Slide Time: 08:13)

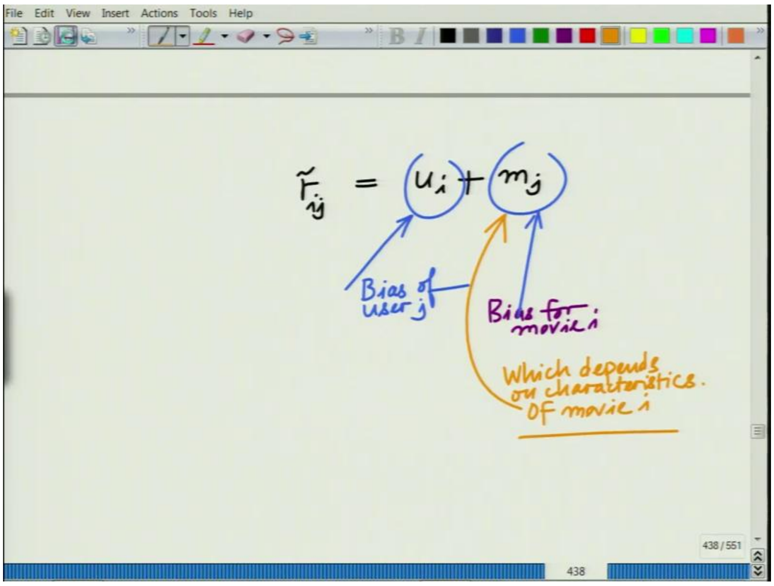


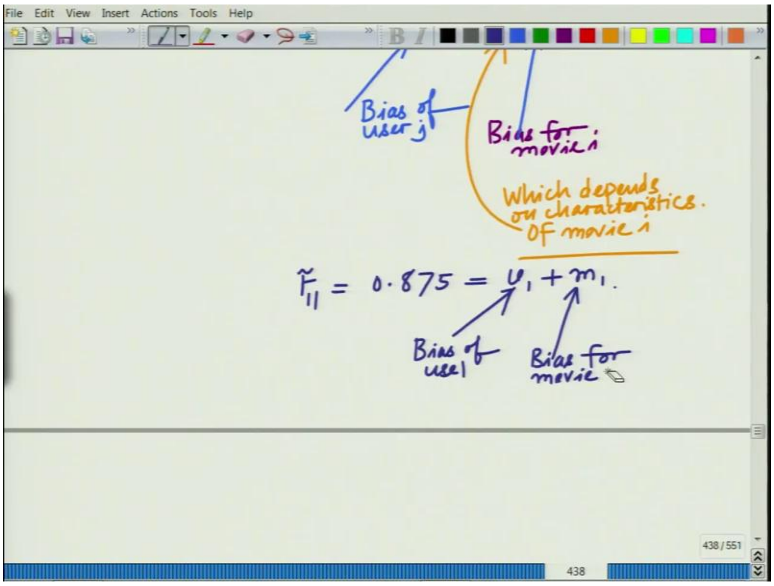




And once again you can form this table here, for these r tilde that is after removing bias after removing ra and these quantities, you can see these are given as 0.875 minus point so, let me write this down let me take a little bit of space 0.875 minus 0.125 minus 1.125 minus 1.125 and this is of course blank minus 0.125 minus 1.125 and this is 0.875 1.875 so these are after removing the bias for instance, this you can see this is your r tilde 3, comma 2. Which should be rating of 3 for user 2 that is 4 minus ra so this is r32 to minus ra that is 4 minus 3.125 which is 0.875 and so on.

(Refer Slide Time: 10:20)

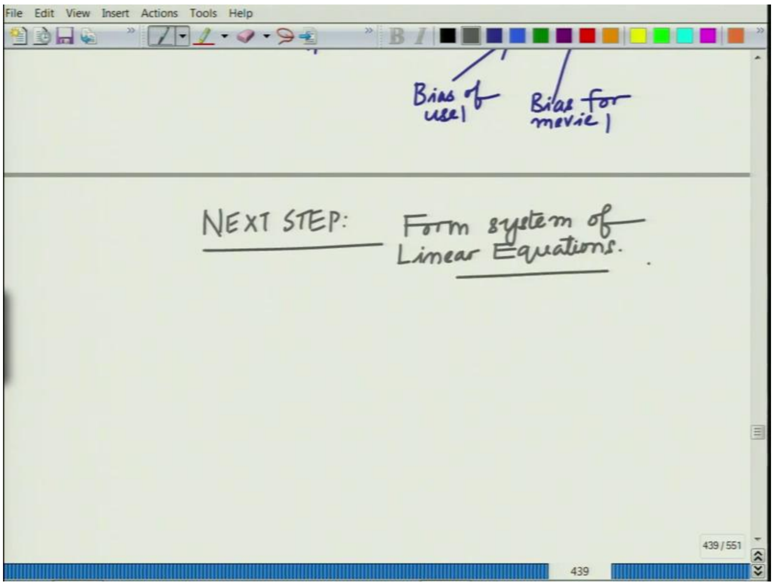


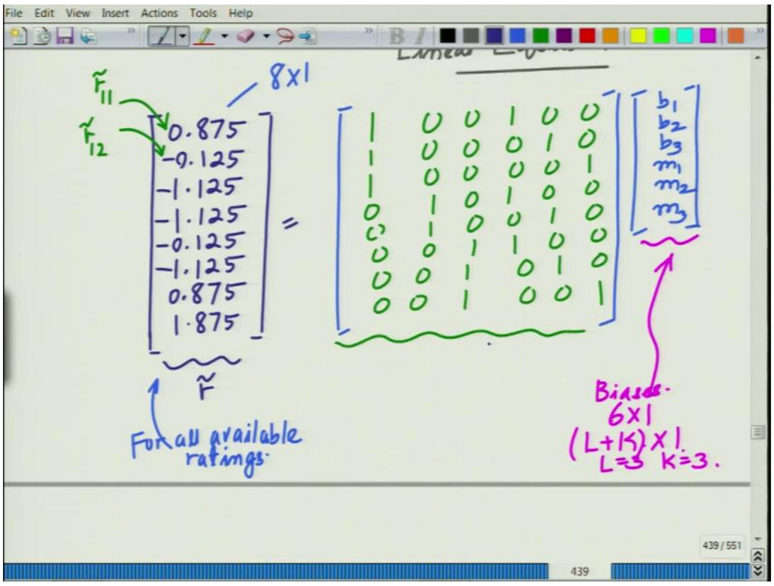


Now, the interesting thing here, is as we have seen yesterday we consider a simple model, where r tilde ij equals ui plus mj the bias of user i plus the overall the bias for movie j which you can say depends on the inherent qualities of movie j. So, you can say r tilde ij equal to ui plus mj now, what is ui?

This is the bias of user j and this is something that is depends on the overall bias for movie i which depends on the characteristics of movie i and therefore, for instance now, I can write r tilde 11 which is 0.875 this I can write as u1 plus m1 bias of user 1 and this is bias for movie 1 and this is movie 1 and therefore, now we can form our system of linear equations.

(Refer Slide Time: 12:20)





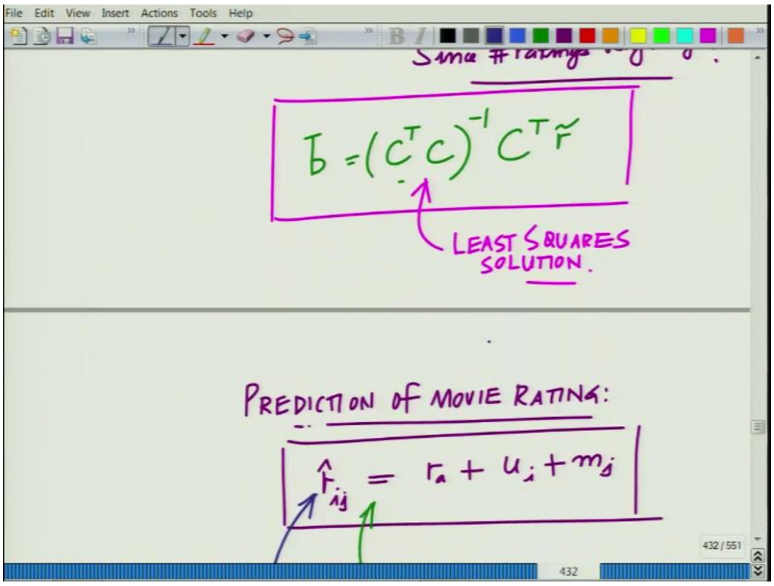
So, next step is form the system of linear equations, which explains and in this you have the vector which is all the available ratings after removal of bias so, you will have 0.875 minus 0.125 minus 1.125 minus 0.125 minus 0.125, I am sorry, this has to be minus 1.125 0.125 minus 1.125 0.875 1.875 this is your, what we are calling as the vector r tilde, we can write this as the matrix so remember, this is an 8 cross 1 vector.

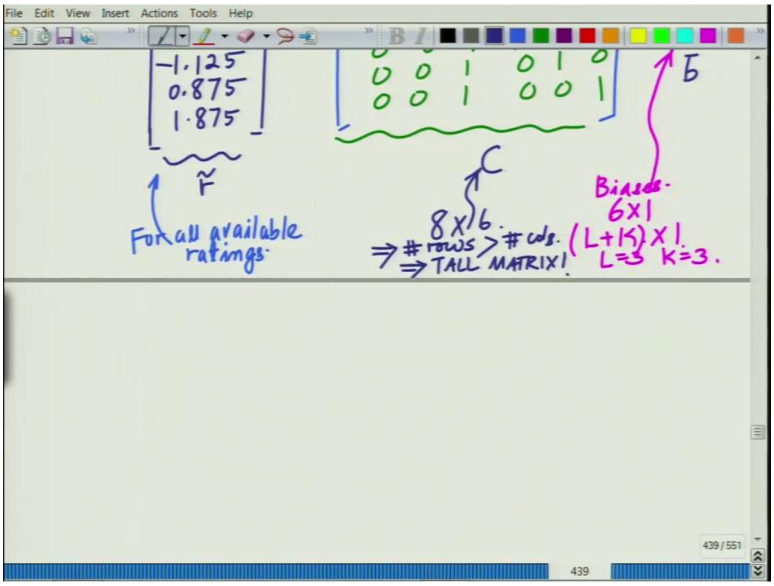
So, 8 is the number of available ratings that is, remember this is for all and this you can write this as this matrix C which we are calling as a matrix C and let us write this appropriately this I can express it in terms of the movie biases, the user biases and movie biases that is b1, b2, b3 m1, m2, m3 these are the biases so this is 6 cross 1 that is you have your, remember this is

L plus K size vector L equal to 3, K equal to 3 so, this is remember L equal to 3, K equal to 3 so, this is a 6 cross 1 vector and you have the matrix which releases, which relates these ratings these r tilde's to these biases.

So, essentially you have 0.75, 875 equal to b1 plus m1 so, I will have 1 0 0 1 0 0 and this is for instance, your r tilde 11 this is for instance, the next entry is r tilde 12 so, this would be b1 plus m2 that is so, this will be 1 0 0 0 1 0 next is your r tilde 13 so, this will be b1 1 0 0 plus m3 0 0 1 and so on, you can form the other rows and it is not very difficult to see 4th row will be b2 plus m1 so, this is 0 1 0 1 0 0 0 1 0 0 1 0 then, you have 0 0 1 1 0 0 0 0 1 0 1 0 0 0 1 and this is basically equal to your matrix not sure what we call this, let me just look at the nomenclature.

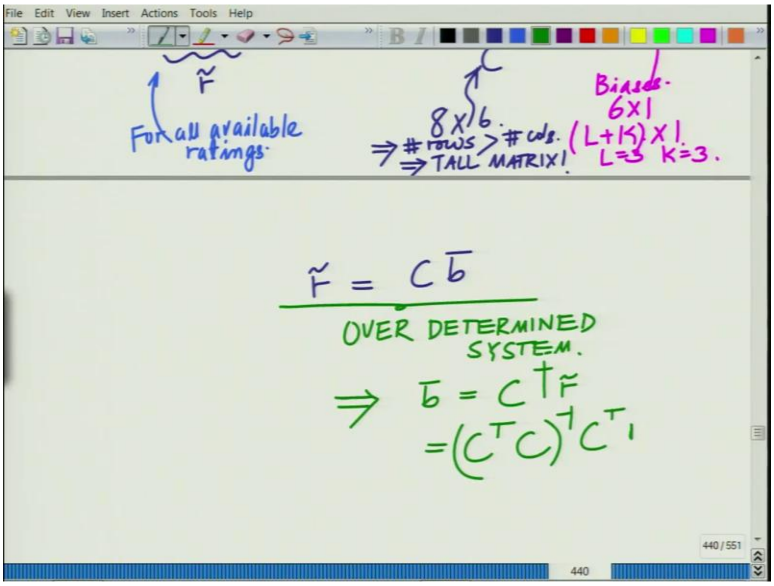
(Refer Slide Time: 16:51)

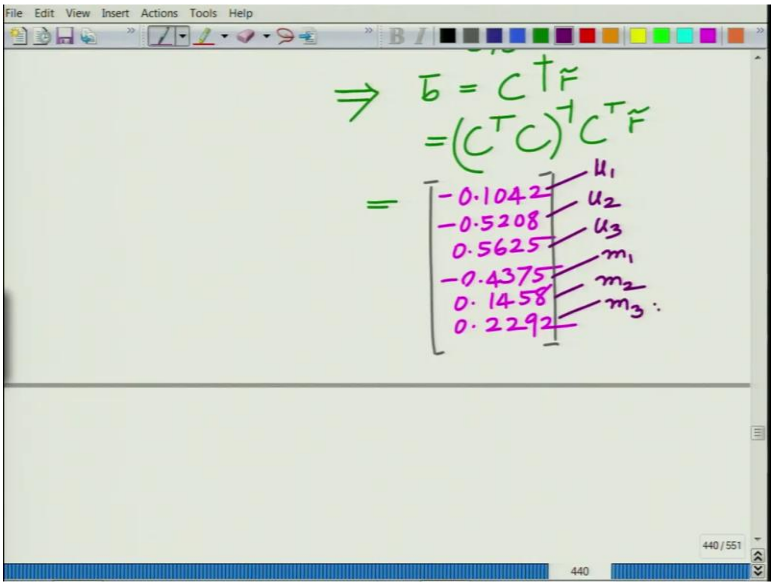




So, I think we call this as the matrix C so, we will call this as the matrix C this is your matrix b bar and you can see this C naturally this is an 8 cross 6 matrix implies number of rules greater than number of columns this implies this is a TALL matrix that is what, we have seen yesterday remember, in the previous module that, the number of rows is much larger than, the number of columns typically because, columns represent the biases which are much smaller compared to the number of ratings, the number of ratings available is basically huge and therefore, you have a Tall matrix.

(Refer Slide Time: 17:40)

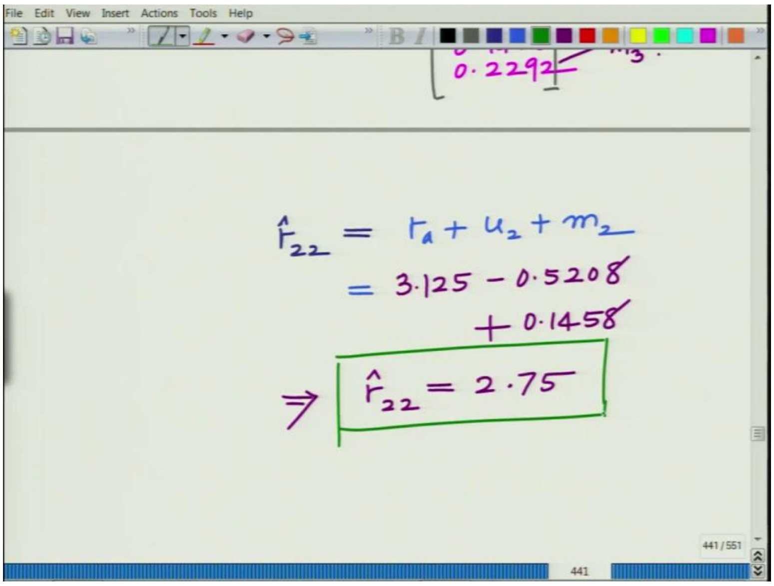


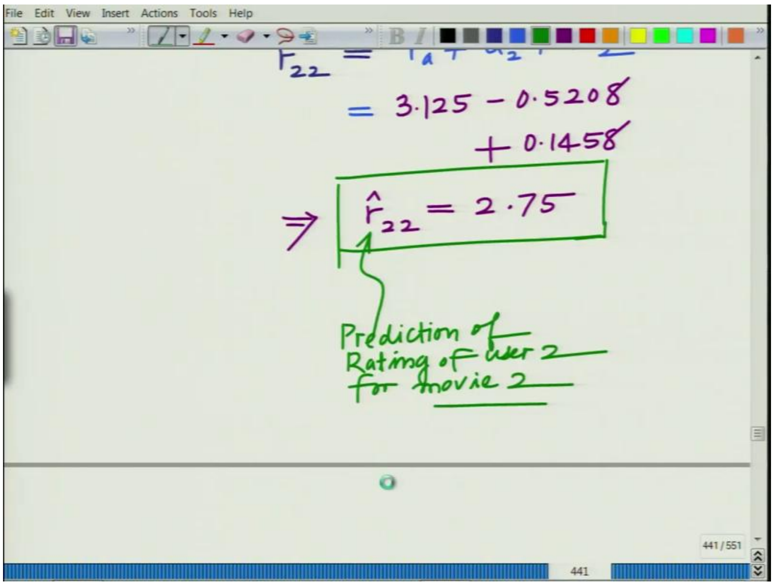


And which now, you have r tilde equal to Cb bar this is the bias vector therefore, this is essentially go back to this is an over determined system so this implies your b bar equals C pseudo inverse I hope you remember this notation C dagger that is C pseudo inverse which is C transpose C inverse C transpose r tilde and if we do that what we get is essentially you get the vector or let me write it down minus 0.1042 minus 0.5208 minus 0.5625 minus 0.4375 0.1458 and 0.2292 and therefore, and now you can see these correspond to the values this is for instance, your u1, this is u2, this u3, this is m1, m2 and this is m3.

And therefore, now how to get the prediction now the only thing that is unknown remember we said we have to predict r hat 22 predict r22 that is r hat 22 and r hat 22 is u2 plus m2 plus of course not to forget the overall bias so, do not forget that.

(Refer Slide Time: 19:59)





So, r hat 22 equals u2 user 2 bais plus m2 bais for movie 2 plus ra so, you can write this as, or let us put ra in the front ra plus u2 plus m2 which is equal to 3.125 plus u2 which is minus 0.5208 minus 0.5208 plus m2 that is 0.1458 which implies r hat 22 equals 2.75 so, this is our prediction of user 2 prediction of rating so what have we achieved?

This is the prediction of rating of user 2 for movie 2 and remember, the next step is of course, to compute these predictions of the ratings of unseen movies for all the users and ultimately recommend that movie which is not seen, not been seen previously by the user and that has the highest predicted rate that is the last step of course here, this is a simple toy example so, we have only 1 movie that is unseen by user 2 so, of course, we do not have other unseen movies for which the rating can be predicted and so on.

So, essentially, this example will stop here, where we have simply computed or predicted the rating of user 2 for movie 2 which he or she has not seen before, the logical conclusion would be of this for a large system would be the last step would be to compute such predictions for all the movies that, all the users have not seen and ultimately recommend to each user the movie with the highest rating from among the movies that particular user has not seen. So, let us stop here and let us continue discussing other such interesting applications in the subsequent modules. Thank you very much.