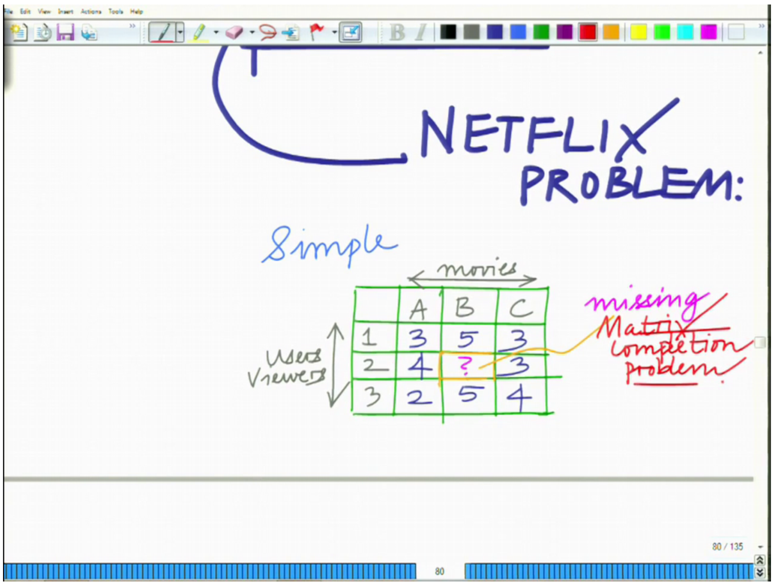
Applied Optimization for Wireless, Machine Learning, Big Data
Prof. Aditya K. Jagannatham
Department of Electrical Engineering
Indian Institute of Technology, Kanpur

Lecture - 78
Matrix Completion Problem in Big Data: Netflix-I

Hello. Welcome to another module in this massive open online course. So, we are looking at an application of Convex Optimization Big Data in particular to the Netflix problem. So, let us continue our discussion. So, we are looking at Big Data in particular the Netflix problem. Let us consider a very simple origin of that problem.

(Refer Slide Time: 00:48)



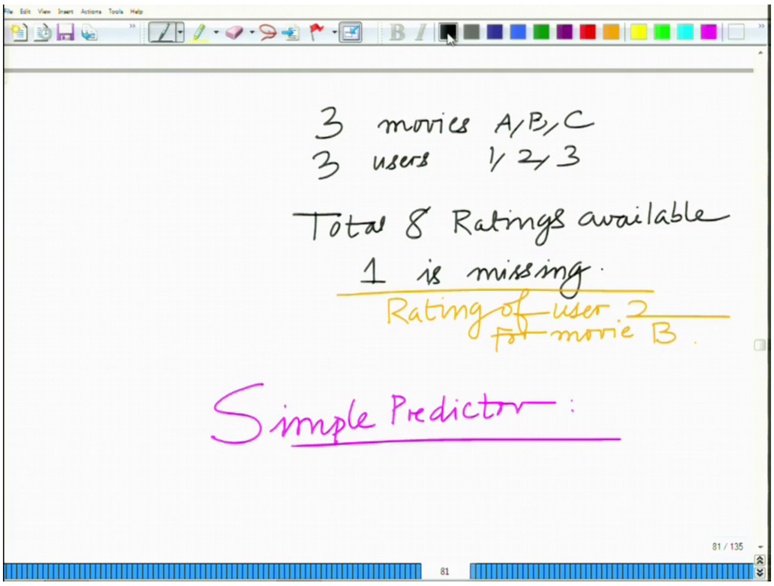
And to do that I will consider as very simple example, a very simple of course, the Netflix problem as I said involves about half a million users and about 20000 movies. We are going to consider like extremely simple origin of that which can be generalized of course, to any size since it basically illustrates the key principle. So, what I have here is a very simple example remember Netflix problem, there are users who are rating the movies so.

So, this is a simplified version of the problem. Let us say that you have 3 movies A B C. These are your movies and 3 users 1 2 3, similar to the table that we had seen yesterday. These are the users or these are the viewers, I am just going to call them as the users; users of Netflix or you can also think these are the viewers or you can also think of these are the subscribers in Netflix and let say different users have rated different movies. Let us consider again a simple example, where user 1 has rated movies A B C; the ratings are 3, 5, 3. User 3 has also rated movies A B C; ratings are 2, 5, 4.

But user 2 has only rated movies A and C and his ratings are 4 and 3. Now, what you can clearly see is that one rating is missing; if you look at this is a matrix, this is missing, this is missing ok. And therefore, we have to predict this rating to complete this matrix of users and ratings that is why it is also known as a matrix completion problem that is what we had seen previously ok.

So, this problem where you try to complete this matrix of users and their ratings for the movies. This also this is also known as matrix completion problem ok. So, this is matrix, this is a matrix completion problem.

(Refer Slide Time: 03:24)

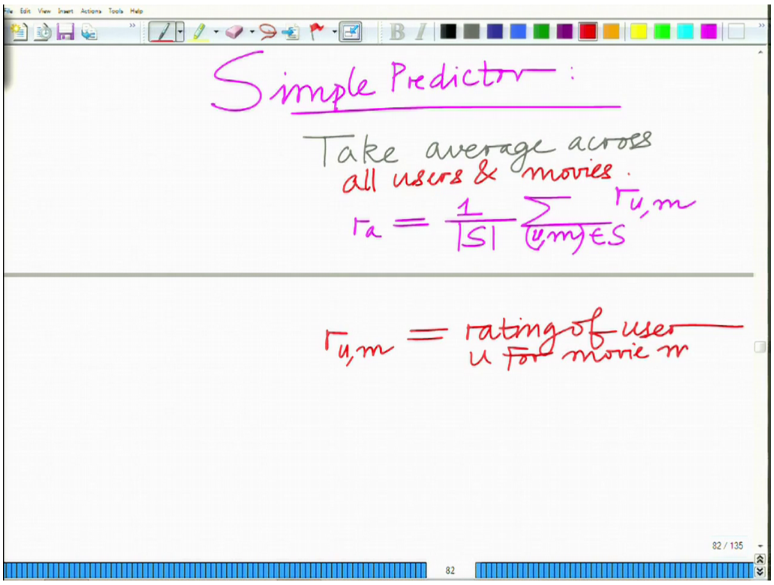


And how do we address this matrix completion problem? Well, there are 3 movies A B C; 3 movies, 3 users ok. How many ratings are there? Total 8 ratings, maximum is 9. Total 8 ratings are available. One is missing. What is that rating of user 2 for movie B? This is missing which we have to predict.

Now again, let us consider a simple predictor. What is a simple predictor? A very simple predictor. We can also think of this is a back of the envelope calculate. What is the simple predictor just looking at a very gross average; you are just looking at reducing this entire process to a single number. What is the best predict?

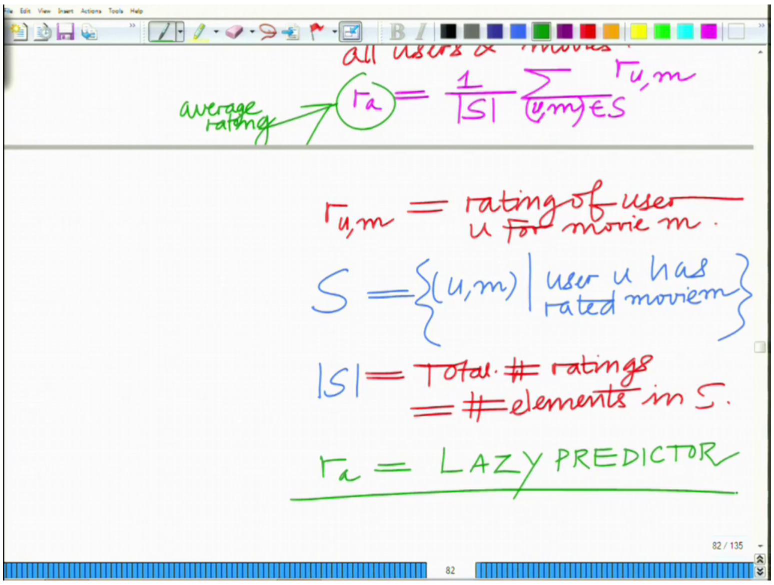
The best predictor is nothing but the mean; there is the average like the per capita income of a country will get the total gross domestic product divided by the number of people that is a simple matrix that shows what is the average rating of each movie per user, ok. So, we compute the average that is the simplest and the simplest print; might not be the best, but it is simplest ok.

(Refer Slide Time: 05:12)



So, the simple predictor is simply take average across, this is a gross oversimplification ok, but as I said it is predict an average across all users and movies. So, we have writing a which is 1 over let us say S is the set of all ratings that are available u comma m that is user u has rated movie m belongs to S; r rating r of user m ok. So, what is this r of u of m? r u, m equals rating of user u for movie m. Example r of r of 1 comma c equals 3; rating of user 1 for movie C ok; so, very simple notation all right.

(Refer Slide Time: 06:28)

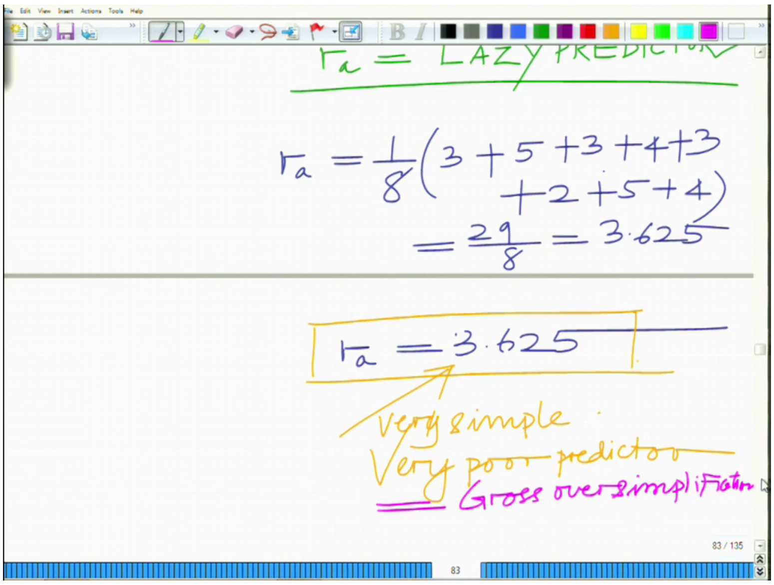


Now, S is the set of all ratings available or the S is the all the pairs u comma m or S contains all the pairs u comma m such that user u has rated movie m, very simple; I mean I can only consider the average of all ratings that are available that mean if user u has not rated movie m, I cannot include it in the average process. So, all the ratings that are available.

Simply take the average of that and this quantity if you look at it is standard this is basically the number of elements. So, I am dividing with it first number of elements, total number of ratings equals number of elements in S ok. So, now, this is also termed as a Lazy Predictor ok; you can think of this r a this is r average, this is you can think of this as average rating ok.

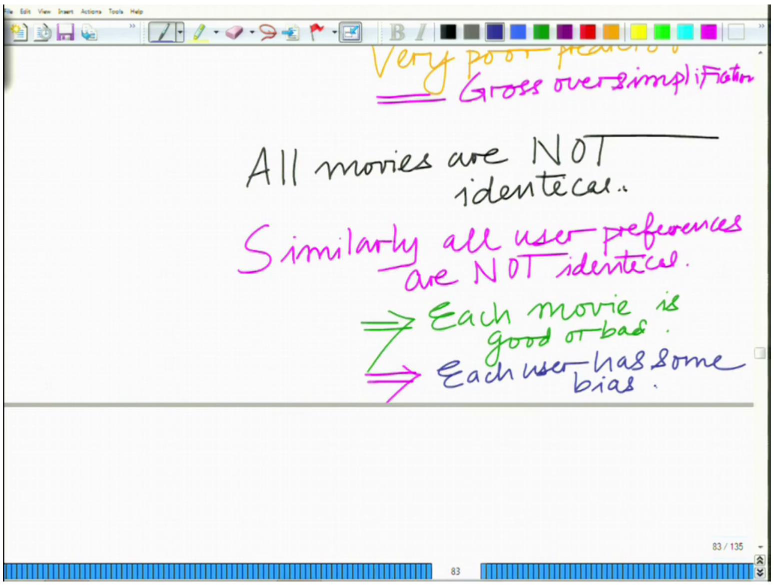
So, you can think of r a as a Lazy Predictor something that I said is a back of the envelope calculation that is something that you would do if you just grossly oversimplifying the entire process.

(Refer Slide Time: 08:01)



And for this example, the averages r a equals 1 over 8, total number of ratings is 8. Simply take the average 3 plus 5 plus 3 plus 4 plus 3 plus 2 plus 5 plus 4. This you can check is 29 divided by this part of 8. This is 11 in to 15 18 20 25 29; 29 divided by 8, 3.625. So, average rating of each user for each movie r a equals 3.625, but as I already told you, this is simple; very simple. Its performance is going to be very poor, but caution this is a very poor predictor because this is a gross oversimplification. This is a gross oversimplification. Why is that?

(Refer Slide Time: 09:17)



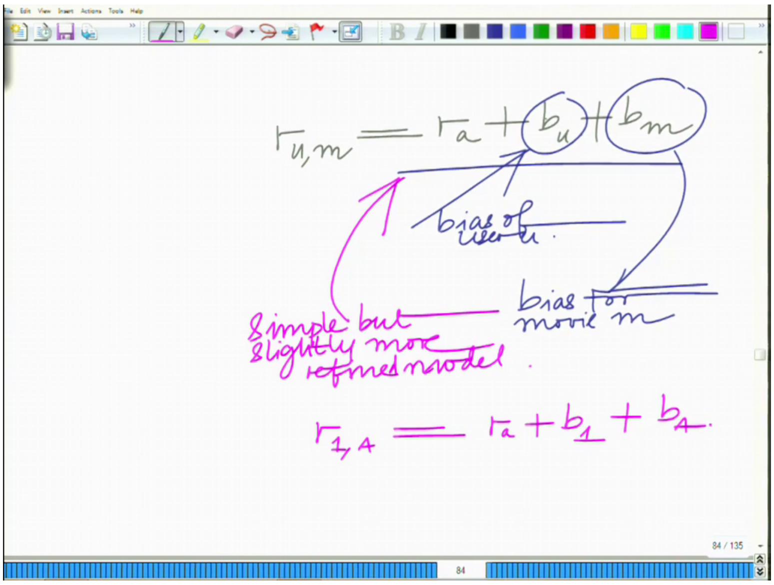
The reason for that is I mean we are not robots right, each user is identity each user is unique each movie is unique ok.

So, each user has a certain bias; each movie is good at remember mean if you look at a movie I mean everyone might not agree on the movie, quality of the movie, but by and large some people a large number of people think some movies are better than some other movies. If you look at the movie godfather a large number of people would think that godfather is probably the best movie that is hourly rate and so on.

So, some movies are consistently it the best, some movies are consider are some movies it of course, not everyone is agreeing and do not get do not mistake me do not everyone is agreeing on every movie, but some movies are better than others. Some users are same way more lenient towards their ratings while some user for instance like critics or writing newspapers might be very harsh usually ok. Now, so one has to take the vagarious of this process into account. So, to each user each movie is either good or bad I mean there is inherently a certain quality which makes a you movie good or bad.

Similarly, each user has a certain bias some of us like certain kinds of movie; some of us like comedy, some like romantic movies and so on and so forth. So, so bringing the human aspect into it both into the movie making as well as the movie rating aspect of it, each all movies are not identical not identical ok. Similarly, not all users are similarly all users similarly all user preferences or prejudices are not identical. Implies each movie is good or bad or each movie has a certain bias, each movie is good or bad and each user has a certain bias, each user has some bias.

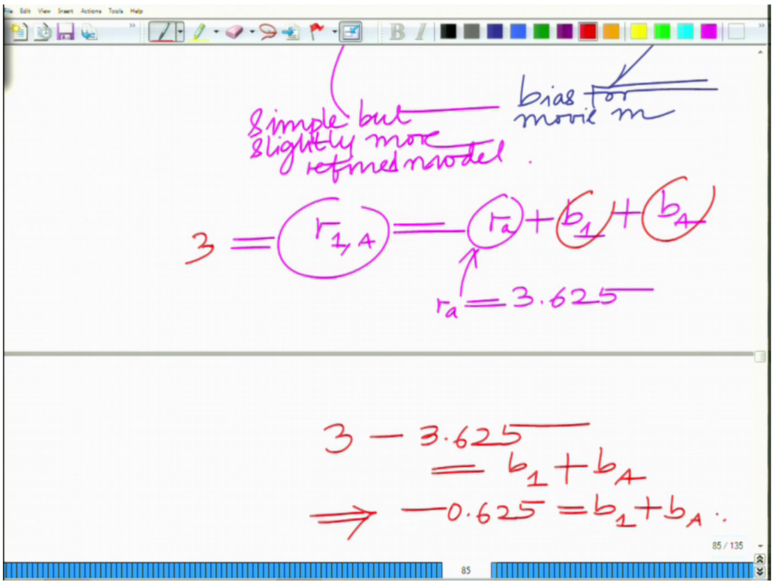
(Refer Slide Time: 12:09)



So, we can model the rating of each user r u of m; a slightly more sophisticated model is the average plus the bias of each user plus the bias of movie which captures the good or bad nature of the movie and b u capture. So, this b u is the bias of user u and b m is the bias that is shows if a large number of people think that the movie is good or bad. So, this is a simple slightly more refined model simple, but slightly more refined model.

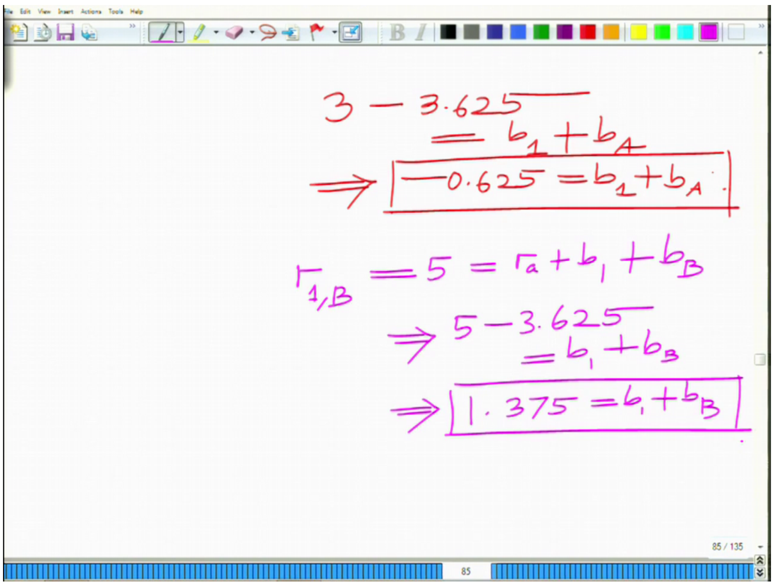
This is natural model because some movies are better or worse or some movies are worse and some users and differently different users have different biases. So, this is a slightly more refined and probably a more natural model in capturing the behaviour and therefore, now if you look at for instance, again let us go back to look at r 1 A that is the rating of user 1 for movie A. This can be split as the average plus the bias of user 1 plus the bias of towards movie A.

(Refer Slide Time: 13:46)



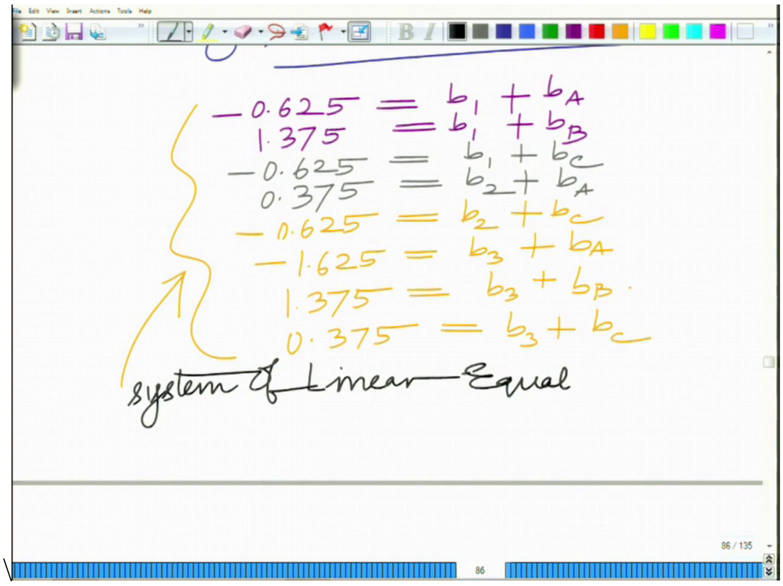
Now, we know what is r a. We know that r a equals 3.625 and we know r 1, A that is a rating of user 1 for movie A, if you go back look at our table. If you go back to look at our table that should be equal to 3; let me just check, yes that is indeed equal to 3 and therefore, 3 equals 3.625 plus bias of user 1 plus bias towards movie A bias of movie A. So, which implies 3 minus 3.625; this is equal to b 1 plus b A which implies well minus 0.625 equals b 1 plus b A ok. This is an interesting result that you have.

(Refer Slide Time: 14:46)



Similarly, if you look at this now let us consider rating of 1 for use rating of user 1 for movie C that is equal to well, what is that? You go back and look at it that is equal to 5. The, I am sorry rating of user 1 for C that is equal to again 3. Let us consider rating 1 of 1 B; rating of user 1 for movie B that you can see is 5. So, rating 1 for user B which 5 which is again r a plus b 1 bias of user 1 plus bias of movie B which implies 5 minus r a; 5 minus 3.625 equals to b 1 plus b B implies 1.375 equals b 1 plus b B. You can form all such equations.

(Refer Slide Time: 15:54)

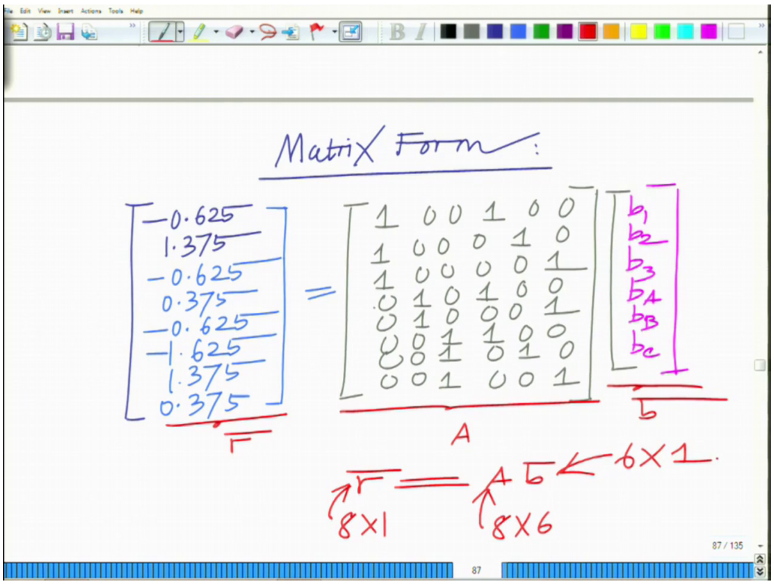


In fact, you will have 8 such equations. Since, you have 8 ratings, you will have 8 such equations. So, let us write down the equations.

You have minus 0.625. Let me just write the heat equations clearly. Two we have already seen; minus 0.625 is b 1 plus b B; 1.375 equals b 1. I am sorry this is b 1 plus b A; this is b 1 surface b 1 plus b B. 0.375 or minus 0.625, I am just going to write these things equals b 1 plus b C; 0.375 equals b 2 plus b A; minus 0.625 equals minus point 0.625 equals b 2 plus b C. B, I am sorry this should be B 2 plus b C; minus 1.625 equals b 3 plus b A; 1.375 equals b 3 plus b B and 0.375 equals b 3 plus b C ok.

This is the set of equations; you can see this is a system of linear equations. What is this? This is a system of linear equations. I can write this in the form of a matrix.

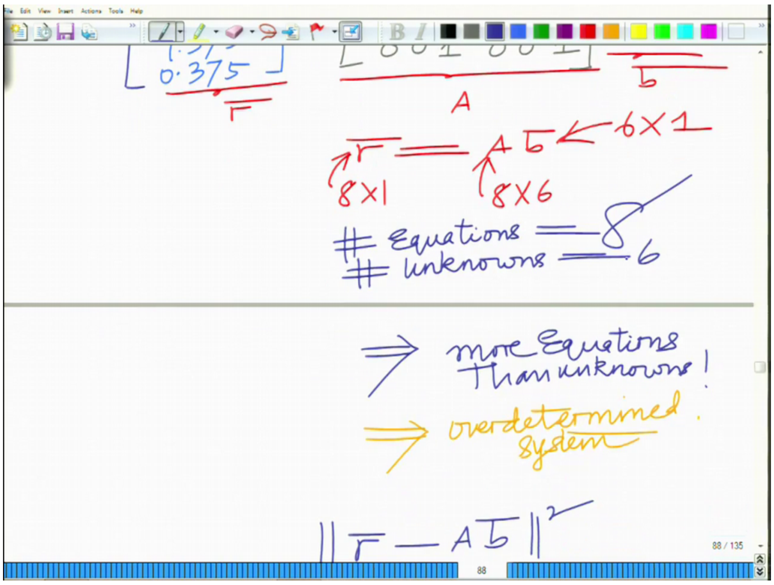
(Refer Slide Time: 18:29)



So, you write this in the matrix form. So, you will have vector on the right minus 0.625; 1.375 you can check this; minus 0.625; 0.375; minus 0.625; minus 1.625; 1.375 and 0.375 perfect. This is equal to the matrix 1 0 0 1 0 0 1 0 0 0 1 0 that is user 1 movie B; 1 0 0 0 0 1 0 1 0 1 0 0 0 1 0 0 0 1. You can see user 2 for movie B is missing; rest you have user 3, movie A; user 3 movie B; user 3 movie C that is it.

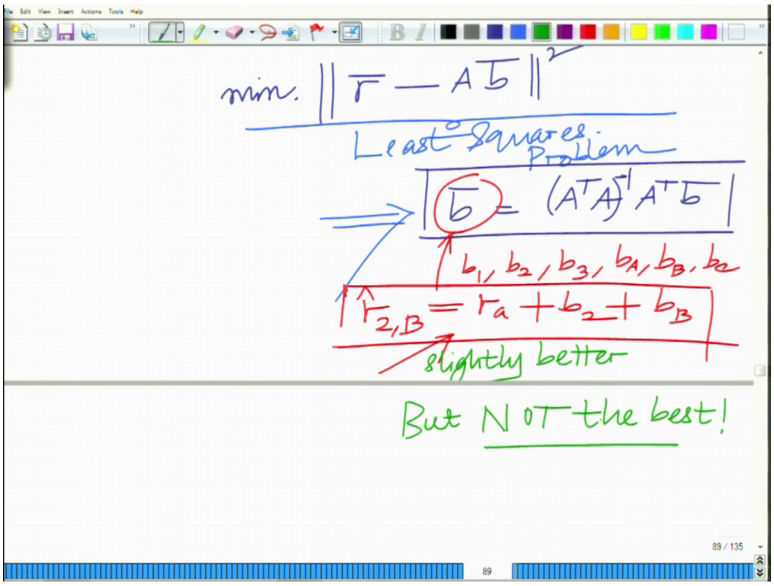
That is the matrix and then, you have the vector of lets write it down b 1 b 2 b 3 bias of movie A, bias of movie B, bias of movie C and now, you have this vector r bar. This is A; this is b bar. So, you have the system of equations r equals 8 as b bar. This is 8 cross 1; this is 8 cross 6 and you can see how many equations?

(Refer Slide Time: 20:38)



Number of equations equals 8, how many unknowns? Number of unknowns equal to 6; implies more equations than unknowns all right. We know we have seen this several times this is an over determined system. This is an over determined system and how do I solve this over determined system? We solve this over determined system by using the squares r minus A b bar square; we cannot solve this exactly. We have to solve it approximately. How do we solve it? We find the best vector b bar which minimizes the approximation error norm of the approximation error r bar minus A b bar.

(Refer Slide Time: 21:49)



So, you minimize r bar minus A b bar. This is your least squares problem implies b bar equals A transpose or A I am sorry b bar equals A transpose A inverse A transpose b bar that is it. So, from this you get b bar; this will give you b 1, b 2, b 3, b A, b B, b C. The predictor of rate use rating of user 2 for movie B will then be the average plus b 2 plus b B.

This is a slightly more refined better predictor. This is slightly better, but not the best we are not done yet; but slightly better, but not the best; but not the best ok, we are not done yet all right. So, we have got first stage, we started with the average. We are refined it, average plus bias of user bias of movie that gives us something all right. But we are not done yet.

We are going to refine this model further to predict it to get as close as remember, we have to be able to predict it as closely as possible and towards that and we are going to refine this model further and that we are going to run a subsequent module.

Thank you very much.