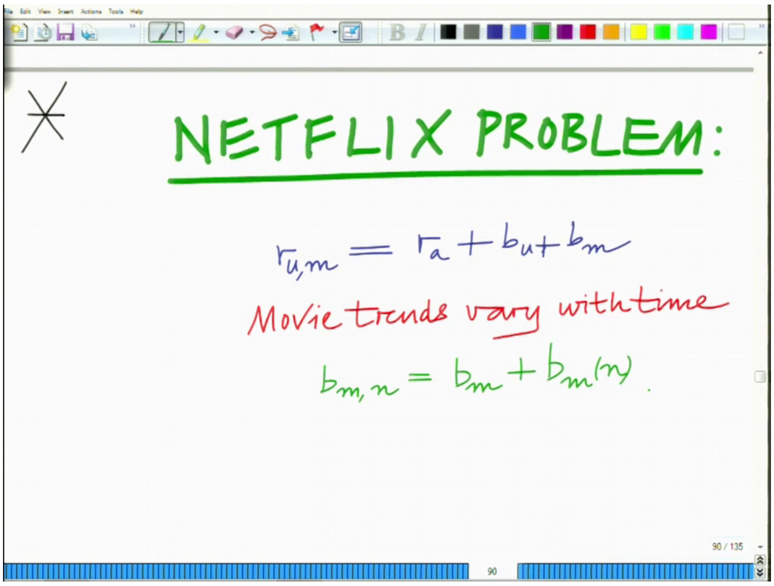
Applied Optimisation for Wireless, Machine Learning, Big Data.
Prof. Aditya K. Jagannatham
Department of Electrical Engineering
Indian Institute of Technology, Kanpur

Lecture - 79
Matrix Completion Problem in Big Data: Netflix-II

Hello, welcome to another module in this massive open online course.

(Refer Slide Time: 00:29)

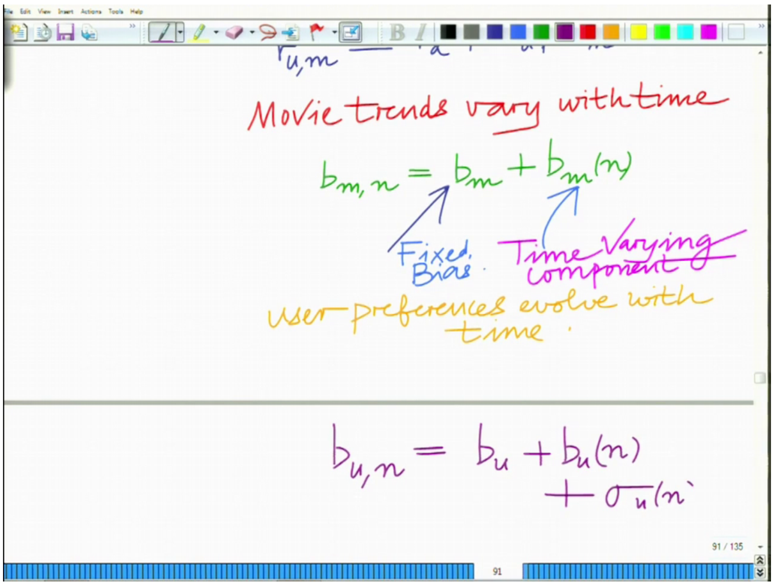


So, we are looking at the Netflix problem and how to solve this problem that is how to predict the user ratings for movies which he or she has not seen let us continue our discussion. So, you are looking at the Netflix problem which we said is basically a specific case of matrix completion alright or belongs to the general class of problems which can be termed as matrix completion problems.

And we have seen that you can express the rating of each movie r um user u for movie m as an average plus a bias for each user plus a bias for each movie. Now let us refine this a little bit because, what happens is the movie trends change with time correct. So, the feeling about a movie or appreciation of a movie some movie suddenly become popular.

So, movie trends vary these are time these are not static right so these are time varying these are dynamic. So, I can model them as a more refined model would be them as a function of the mood bias of movie m at time n can be expressed as a fixed bias plus bias for movie n at time n.

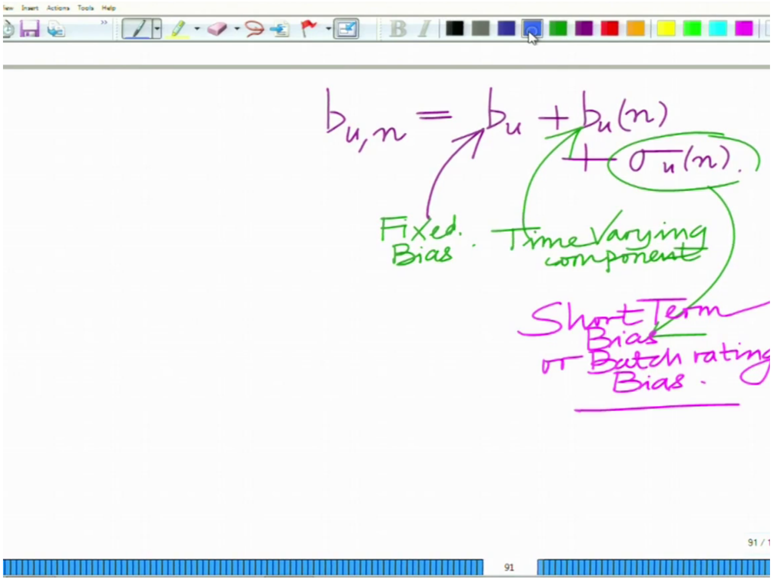
(Refer Slide Time: 01:55)



So, this is a time dependence where so, this is a fixed bias and this is basically your time varying component, time varying component of the bias. Similarly user preferences or user preferences evolve with its times for instance sometimes you hear the phrase that this movie was ahead of its time which means that is not appreciated by users or the movie were just currently. But, at some later time when the tastes are so on and forth of the users have developed adequately or changed adequately then this movie can be seen in your light alright. So, the user preferences user biases are also varying with time alright.

So, your b u which was fixed can now be a function of time which is again a fixed bias b u plus time varying component b u n; plus there is another component I am going to describe this is interesting and this is a quirk of the rating system.

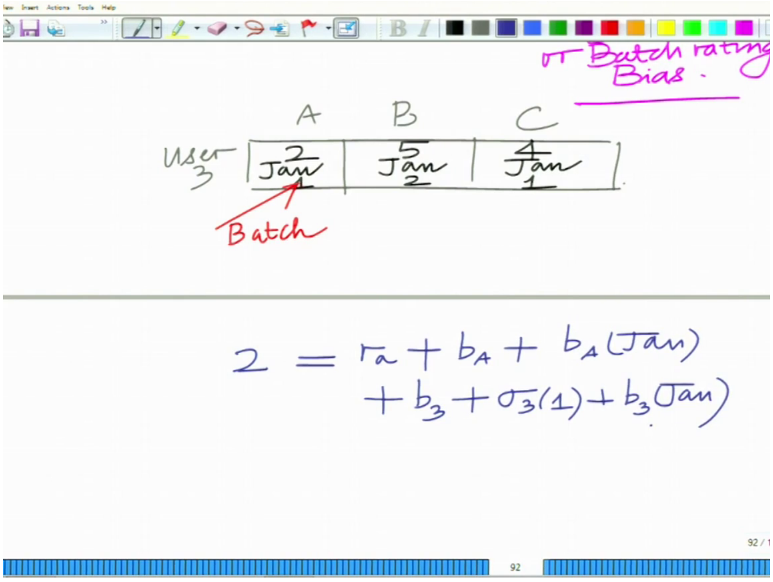
(Refer Slide Time: 03:25)



So, this is as previous this is your fixed bias, this is your time varying component and this sigma u n here this is basically something interesting. Because, specifically for an account like Netflix or any rating system they same account can be used by multiple users. So, one set of ratings can be given by a male member of the family another set of or another set of rate it can be given by a child another set of rate it can be given by a female member of the family and so on a different users of the same account have different preferences.

So, this account for this accounts for that short term bias resulting from different users giving ratings in a batch. So, this occur this can be thought of as a bias for a particular batch of ratings or a short term bias or batch more short term moderate is for instance different people use the account and the rate the movies in batches then each batch will have a certain bias ok, so this is an interesting aspect. And therefore, so this is a batch rating bias for ratings over a batch and therefore now let us look at its not very difficult to see how this can be modelled.

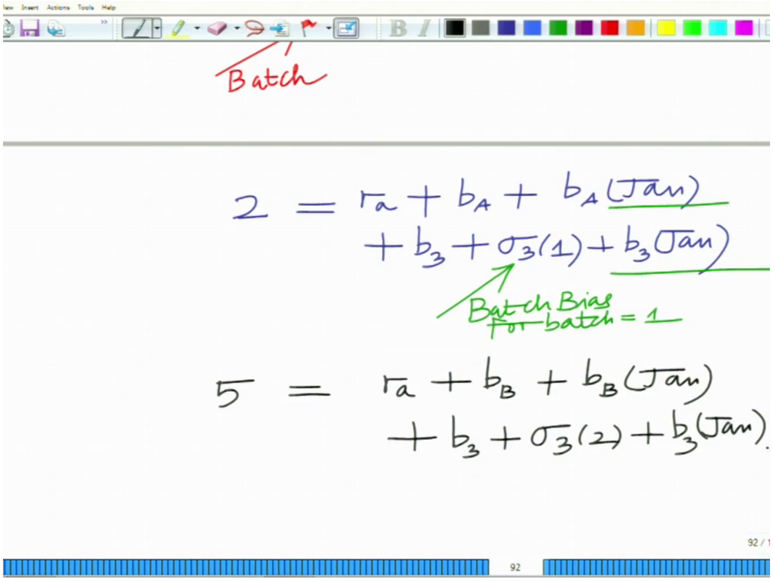
(Refer Slide Time: 05:05)



Let us look at for instance a simple example lets refine our previous model itself comprising of 3 movies A B C we have user 3 just for the purpose of illustration; his ratings were previously if you remember 2 5 4. Now let us say he rated those movies in January or he or she rated those movies in January so this is the time a January in a particular year. And let us say this is corresponds to batch 1, this corresponds to batch 2, this again corresponds so these are the batches. So this is batch 1, this is the batch number.

So, use a 3 rated or the some person who is using the same account as user 3 rated movies A and C in January in the first batch while possibly another person rated movie B in the same month, but in another batch ok. So, that is the meaning of this and therefore, now you can develop the model again. So, 2 which is the rating for user 3 movie A in the month of January in the first batch can be expressed as r a which is the average that of course, remains b A plus the bias of movie A for January. Might be the movie is very popular in January because it is recently released or so on plus b 3 plus sigma 3 1 the bias corresponding to batch 1 and plus bias of the time varying component of user 3 b 3 for January.

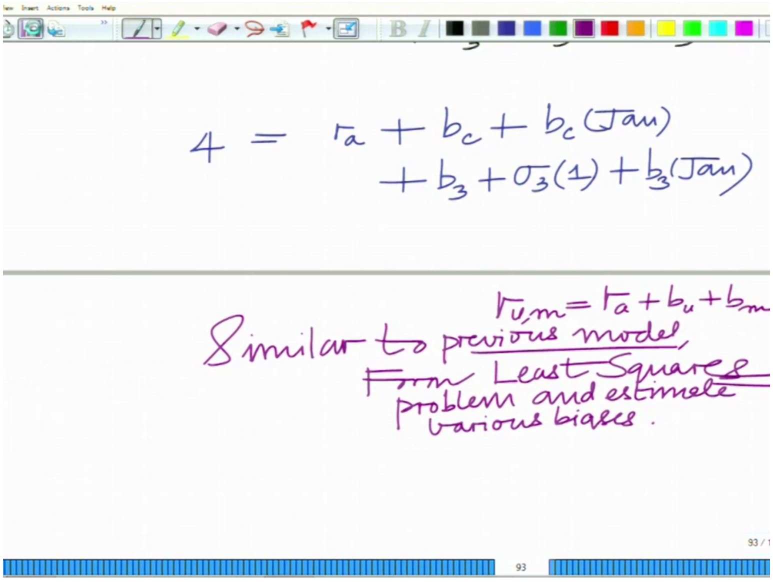
(Refer Slide Time: 06:51)



So, this is the bias batch bias for batch equal to 1, b A you can see is the bias of movie A for the month of January that particular time b 3 is the bias of user 3 for that particular month.

Again similarly you can write 5 is the average that is a rating of movie B for you given by user 3 in a month of January worse. The second batch this is the bias of B plus bias of B the time varying component for January plus the fixed bias of user 3 plus the batch bias of user 3 for batch 2 plus the batch, the bias of fake bias time varying component of the bias of user 3.

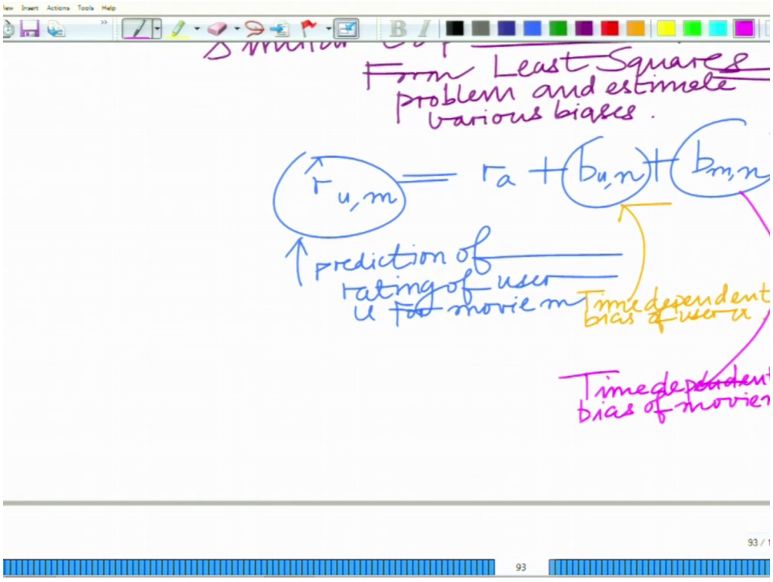
(Refer Slide Time: 07:49)



Similarly, you can have 4 which is equal to the rating of user 3 for movie C January in the month of January. And, batch 1 that will be r a plus b c bias of user movie see fixed bias of movie C plus the time varying component of the bias for the month of January plus bias of user 3 plus the batch bias for batch 1 plus the bias of user time varying component and the bias for user 3 for the month of January.

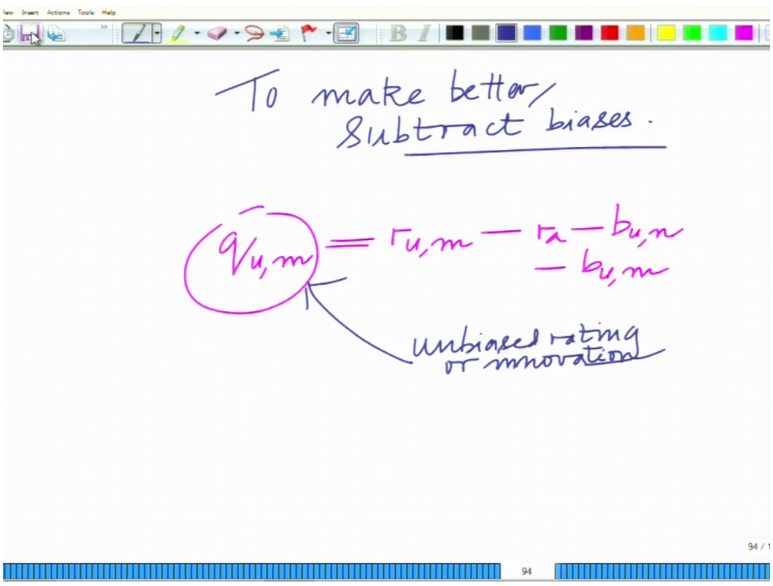
And so now you have many more biases, so you can form a least squares again similar to the previous from form a least squares problem solve for the biases. So, similar to previous problem or similar to previous scenario let us say or similar to previous model where you had remember what we mean by previous model is simply r u m equals r a plus fixed bias for user u plus fixed bias for similar to fixed bias for movie m. Similar to previous model form least squares we are experts now at least squares, least squares model, least squares problem and estimate various biases.

(Refer Slide Time: 09:42)



Now, again at this stage you can once again predict substituting the biases, the prediction the missing prediction q hat can now be obtained as average bias. Average plus or q hat u m the prediction of rating of or let us put it this way; the prediction r hat your prediction of user u for movie m can be obtained by adding the biases in r a plus b u. Now, this is the time varying component b u n plus the time varying bias of so, time varying bias of the time dependent bias of user u time dependent bias of movie m, time dependent bias for movie m.

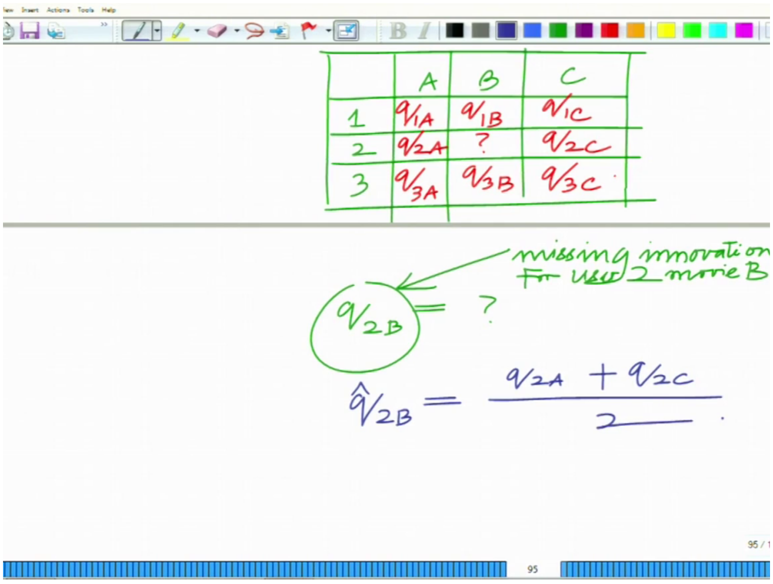
(Refer Slide Time: 11:18)



Now, if you we can refine this model slightly further and we can now, subtract the biases to refine further to make better subtract biases ok. So, we are going to improve we are not going to stop here either now, you subtract the biases that is if you remove these various biases. So, let us say q of u comma m this will be r of u comma m minus r a minus b of u comma n minus b of u comma m. So, this is after rating after removing the biases we can call this as the unbiased rating or also the unbiased, we can call this as a unbiased rating or the innovation something that we have not been able to predict.

So, far you have removed all the biases; now we have the unbiased rating or we can also call it in signal processing this sometimes also termed as the innovation ok. So, a certain model for prediction after you remove whatever you get from that prediction; what is remaining is something that you have not been able to predict or this is the novel quantity or the innovation ok.

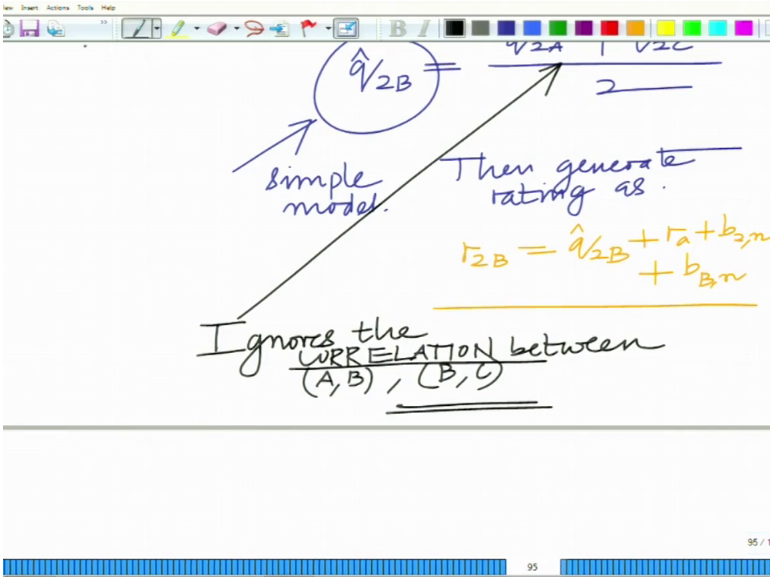
(Refer Slide Time: 12:46)



And, now you can form the innovations corresponding to the different ratings that are available for instance you have users 1 2 3 movies A B C. Of course remember user 2 has not rated movie B which we are trying to predict. But you can compute the innovations corresponding to others all others q 1 A q 1 B q 1 C q 2 A q 2 C q 3 A q 3 B q 3 C now this is missing. Now, how to compute q 2 B or q 2 B? How to compute this, that is the missing innovation for user 2 and movie B.

Now how to compute this, one simple way to compute this is simply take the average of the innovations corresponding to the movies user 2 has rated that is 2 A and 2 C. So, a simple model here again let us first look at a simple model, simple model is q 2 B the estimate hat is q 2 A plus q 2 C divided by 2 ok.

(Refer Slide Time: 14:27)

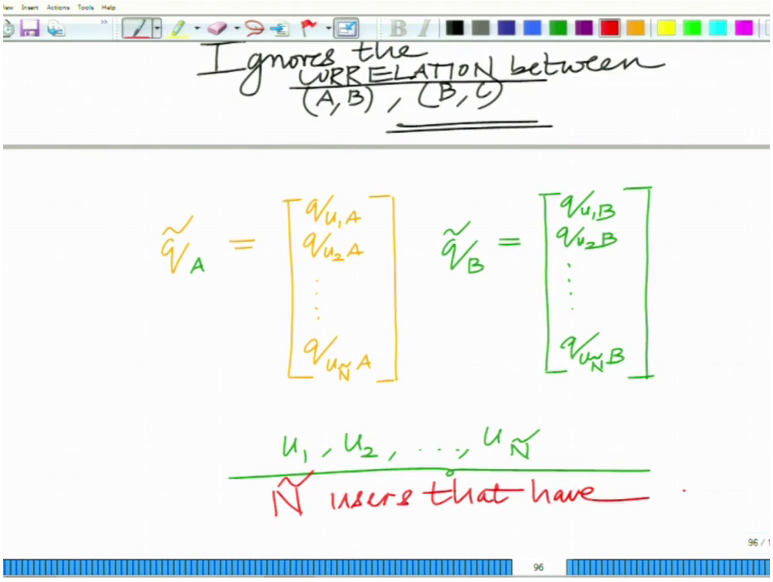


This is a simple model, then once you do this you generate the rating as r 2 B equals q 2 B or this q hat 2 B plus well r a plus b 2 n time varying bias for movie 2 plus b. Time varying bias for user 2 plus b B n time varying bias for movie B and this is your prediction model alright very simple.

However, this again there is a shortcoming of this ignores remember you are simply linearly you are adding this model ignores this is simple model agree this ignores, the correlation. And, this is an important aspect in any prediction between movies A comma B and movies B comma C, this is a very important this ignores the correlation that is a key here.

Because, the point is if some movies are very similar for instance romantic movies or there is a action movies are very similar. So, A and B are very similar then you have to give higher weightage to A, on the other hand B and C are very similar then you have to give a higher weightage to B alright. So, one has to give adequate weightage depending on the correlation. Now how do you get a measure of the correlation, from the user. So, there no there are some users who have rated both the movies from that you get a sense of the correlation so, now let us come to this aspect.

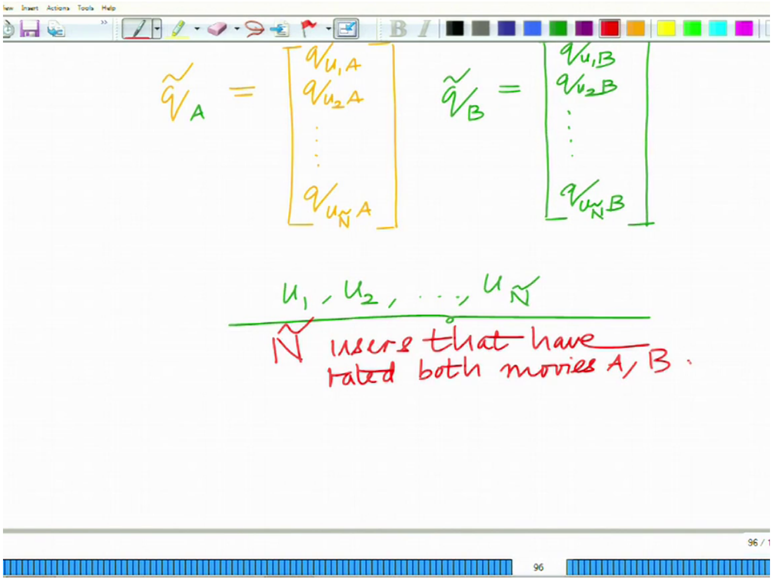
(Refer Slide Time: 16:37)



Let us consider now the innovation vectors q bar A equal to well q i 1 A or let us call this as q u 1 q u 1 A q u 2 A q u N tilde A and now I am just going to explain this notation in a little bit let me just write it down. And so let me just write it out again in a slightly elaborate fashion q 1 bar A or q tilde A I will not call it q 1 bar A q 1 tilde A equals q u 1 A q u 2 A q u tilde A and q tilde q tilde, I am sorry I am just calling q tilde A and q tilde B equals q u 1 B q u 2 B q u N tilde B.

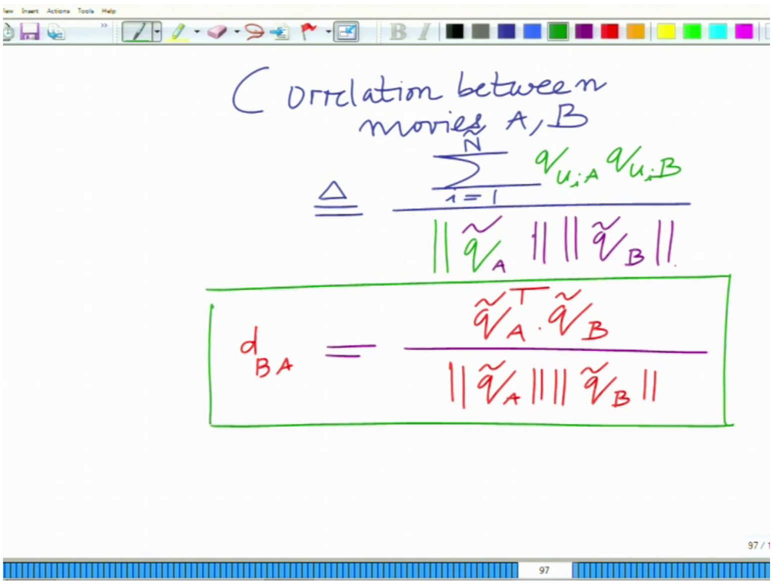
Now why am I using q 1 u 1 A q u 2 A because not all users have rated both A and B. Of course, we can clearly see for instance user 2 has not rated movie B, that is why we are that is indeed in the first place that is the reason why we are trying to predict the rating of user 2 for movie B. Therefore, now consider only those users of rated both the movies ok.

(Refer Slide Time: 18:56)



So, u 1 u 2 u N tilde these are N tilde users or N tilde viewers or N tilde subscribers that have rated both A and B that are rated both.

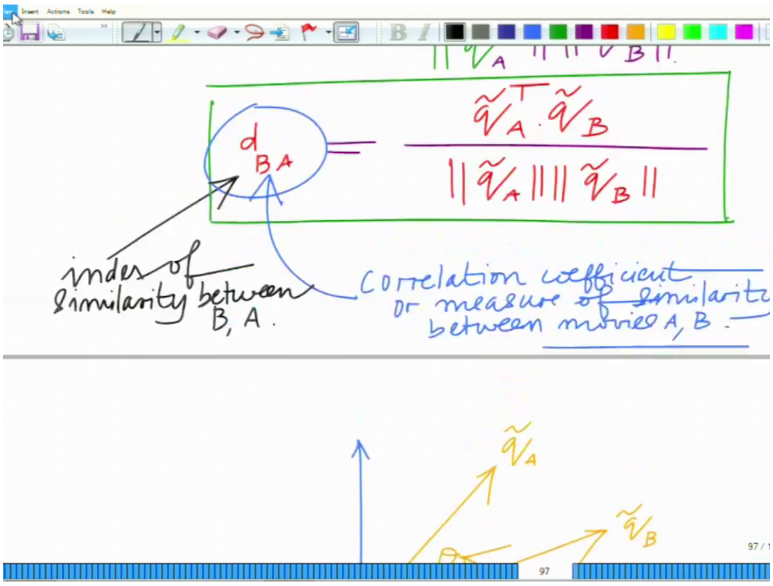
(Refer Slide Time: 19:13)



Then the correlation between these two can be defined as summation over i equals 1 comma N tilde q u i A q u i B. This is the correlation dot product between the vectors or q tilde u i A q tilde or q u i A q u i B divided by the norms of these vectors q tilde A times q tilde. Or in other words this is simply equal to the inner product or the dot product q of tilde A transpose q tilde B divided by norm q tilde A into norm q tilde B ok.

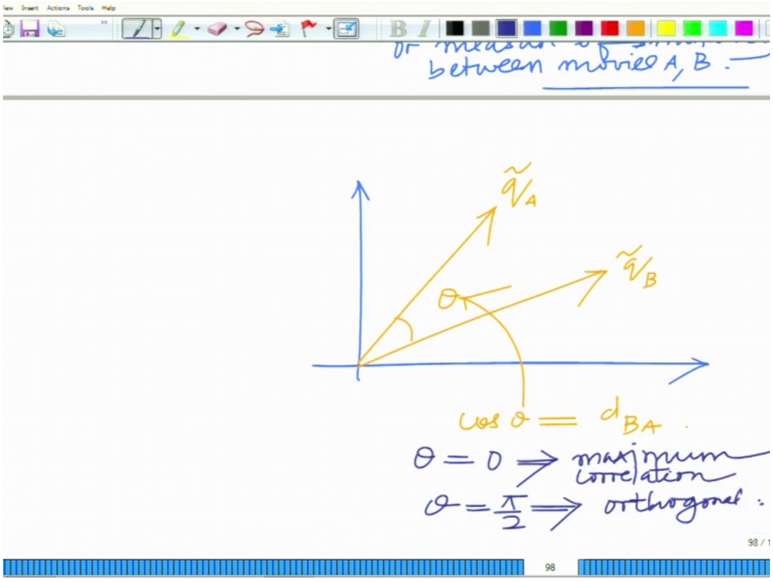
So, this we call this as the similarity coefficient, this is very important. This is the correlation coefficient or measure of similarity, this is the correlation coefficient this is very interesting.

(Refer Slide Time: 20:5)



This is the correlation coefficient or measure of similarity, this is so what we are doing is we are taking the ratings of users of rated both at the A and B computing the inner product between them and dividing them by their norms of these two vector. In fact, if you remember or if you can recall this exactly the cosine of the angle between the two vectors.

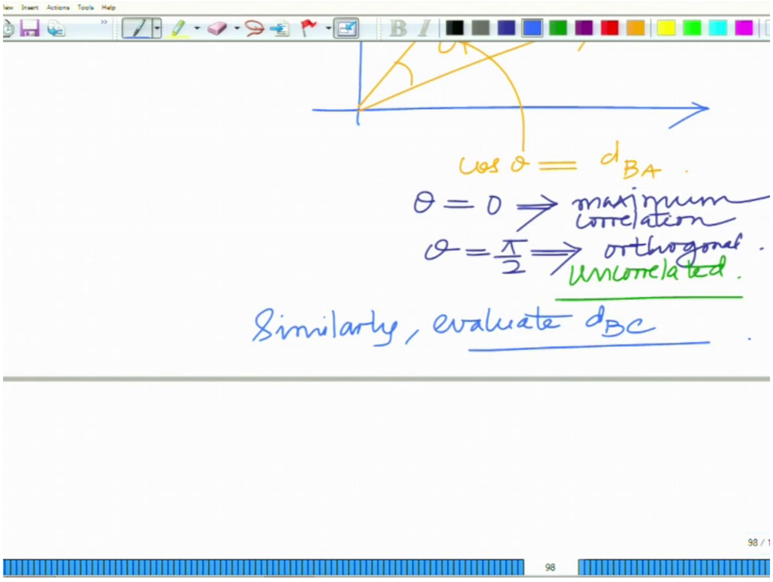
(Refer Slide Time: 21:43)



So, what we are observing is some an interesting application of the principles that you learned both in this course and as well as in your high school. If you have these two vectors q tilde A and q tilde B of ratings, the cosine of the angle between these two vectors remember this cosine of the angle between these two vectors is nothing, but what we have defined as the similarity coefficient.

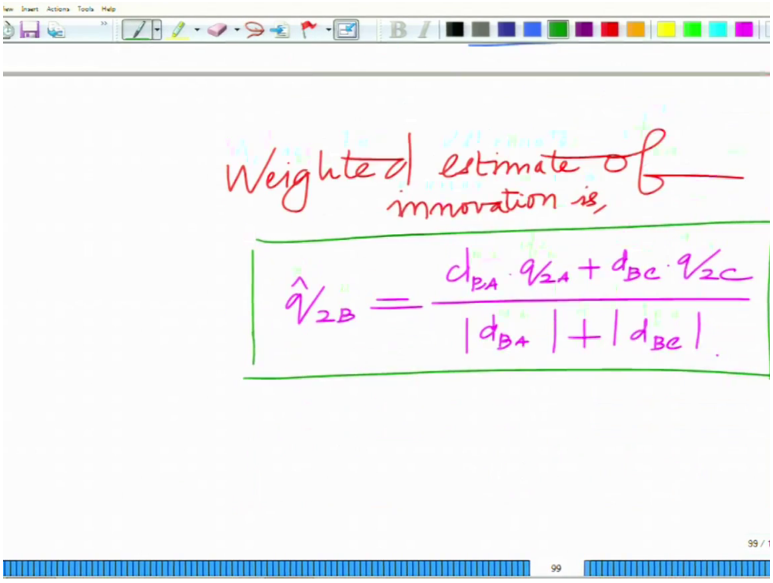
And you can clearly see that if theta equal to 0 the vectors are aligned movies A and B are similar, as theta increases the angle increases movies A and B are the similar. If theta equal to 90 degrees they are perpendicular in fact, A has no bearing on B, B has no that is a very interesting scenario ok. So, theta equal to 0 implies maximum correlation, theta equal to pi by 2 implies they are orthogonal or no correlation cosine theta equal to 0 a theta equal to pi by 2 cosine phi by 2 is 0 implies uncorrelated.

(Refer Slide Time: 22:59)



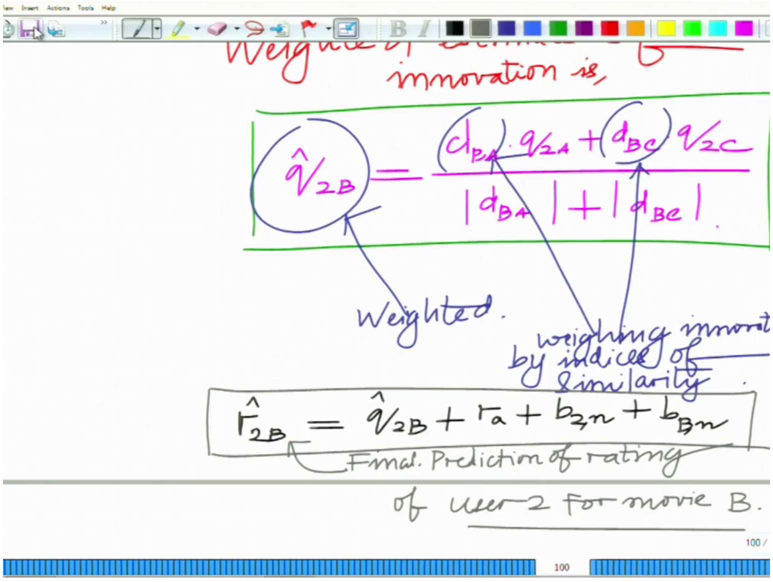
The innovations are uncorrelated. And, similarly now compute evaluate d B comma C this is the correlation or measure of similarity. You can call it correlation coefficient or index of similarity between movies, also index of similarity between movies B and A correlation coefficient of the innovations corresponding to B and A similarly evaluate d BC.

(Refer Slide Time: 23:50)



Now, the weighted innovation or rather weighted estimate, the weighted estimate of the innovation is q hat 2 B equals now you weigh each innovation by the correlation d BA times q A plus d BC correlation coefficient times the innovation q 2C divided by of course, normalized by the weights d BA plus magnitude d BC. This is the final step, this is the final step.

(Refer Slide Time: 24:51)



And now you can predict so, this is basically weighted and basically you can see weighing the innovations by the indices of similarity of the correlation. You are weighing innovations by the indices of weighing the innovations by then there is so similarity. And, now once you compute this innovation estimate you can add it to the biases.

Again you can add it to the biases that is r a plus b u m r a I mean this is q hat 2 B plus the average r a plus B 2 n time dependent by that is the time varying bias of user 2 plus b B n time varying bias of user or movie B to get the final rating. So, finally the rating will be r 2 B or predicted rating r at 2 B equals q hat 2B plus the average ratings plus time dependent bias of user 2 b 2 n plus the time dependent bias of movie B b p n and that is basically the final step in this procedure ok.

So, this is your final prediction of rating of user 2 for movie B. So, this is basically a very interesting after and then even see this clearly this brings across various ideas in both linear algebra as well as optimization that is first is the modelling. Modelling as a linear model then using least squares to estimate these, brings scenarios of estimation and also the least squares which is try to optimizes.

Then finally, you have this innovation and again the correlation between these two innovation vectors which is again tie to linear algebra and so on. So, this brings together a large number of interesting ideas to solve a practical problem which has a lot of relevance in modern systems alright. So, we will stop here and continue in the subsequent modules.

Thank you very much.