Applied Optimisation for Wireless, Machine Learning, Big Data.

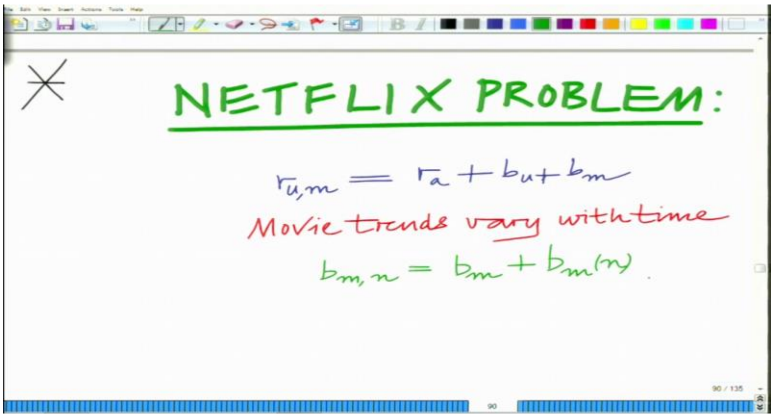
Prof. Aditya K. Jagannatham
Department of Electrical Engineering
Indian Institute of Technology, Kanpur

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

***Keywords**: Matrix Completion Problem*

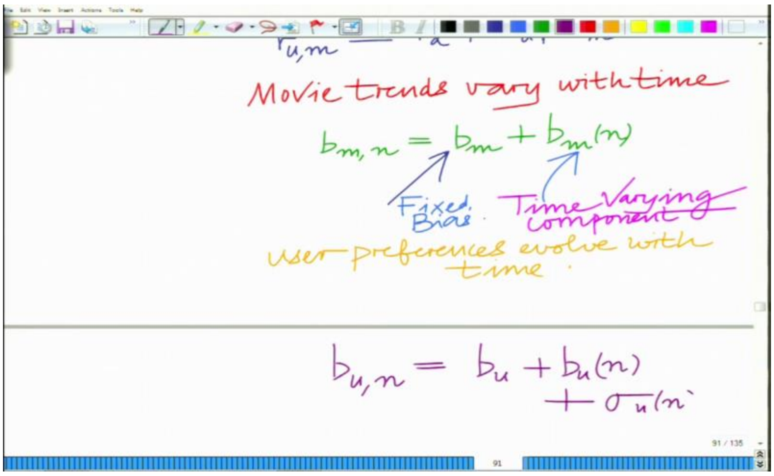
Hello, welcome to another module in this massive open online course.

(Refer Slide Time: 00:29)



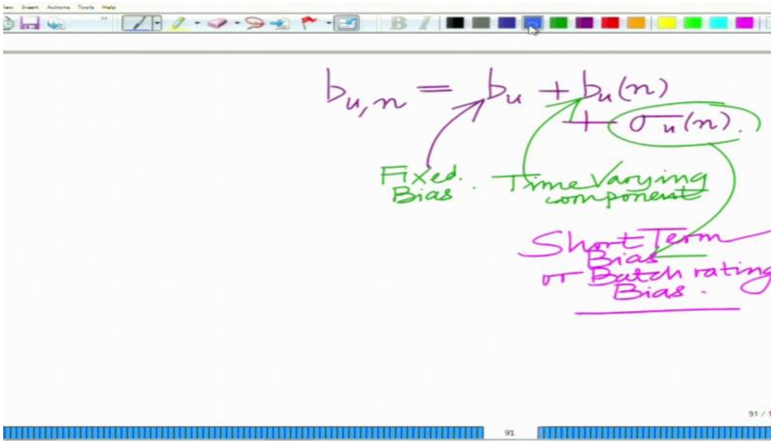
So we are looking at the Netflix problem that is how to predict the user ratings for movies which he or she has not seen. So this is basically a specific case of matrix completion problems. And we have seen that you can express the rating of each movie as shown in slide. Now let us refine this a little bit because the movie trends change with time. So a more refined model can be expressed as a fixed bias along with a time varying bias as shown in slide.

(Refer Slide Time: 01:55)



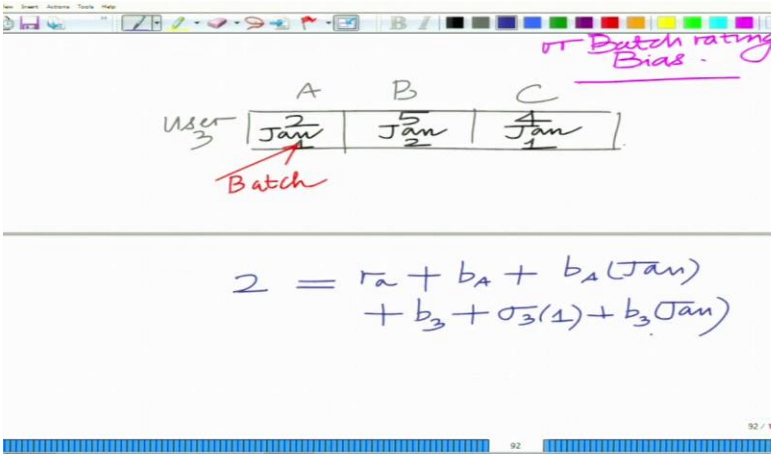
So user preferences are also varying with time.

(Refer Slide Time: 03:25)



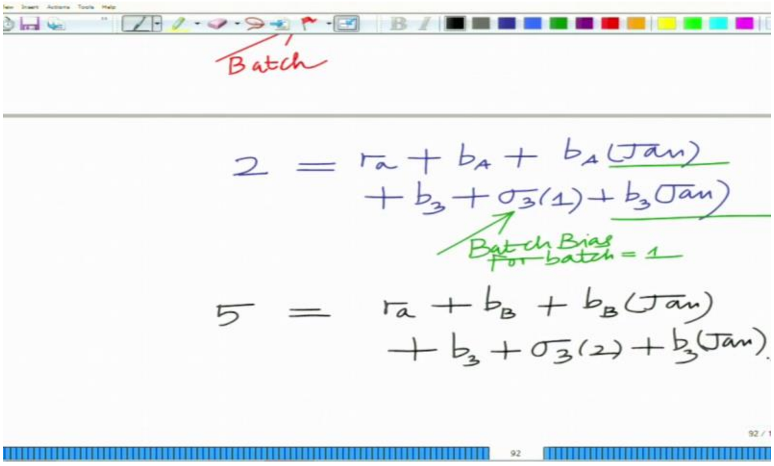
So we have $b_{u,n} + b_u(n) + \sigma_u(n)$. For any rating system, the same account can be used by multiple users. So one set of ratings can be given by a male member of the family, another set of ratings can be given by a child, another set of ratings can be given by a female member of the family and so on, different users of the same account have different preferences. So this accounts for that short term bias resulting from different users giving ratings in a batch.

(Refer Slide Time: 05:05)



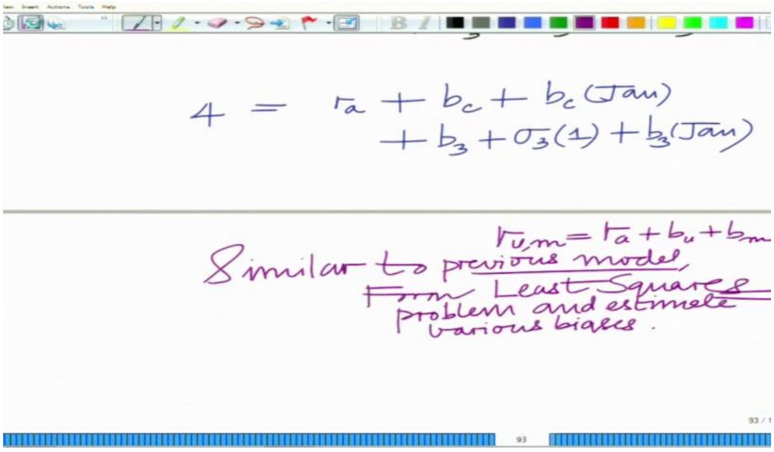
Let us look at for instance a simple example lets refine our previous model itself comprising of 3 movies A B C with a person's ratings if you remember 2 5 4. Now let us say he or she rated those movies in January. And let us say this corresponds to batch 1, so these are the batches. So some person who 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. So now you can develop the model again as shown in slide.

(Refer Slide Time: 06:51)



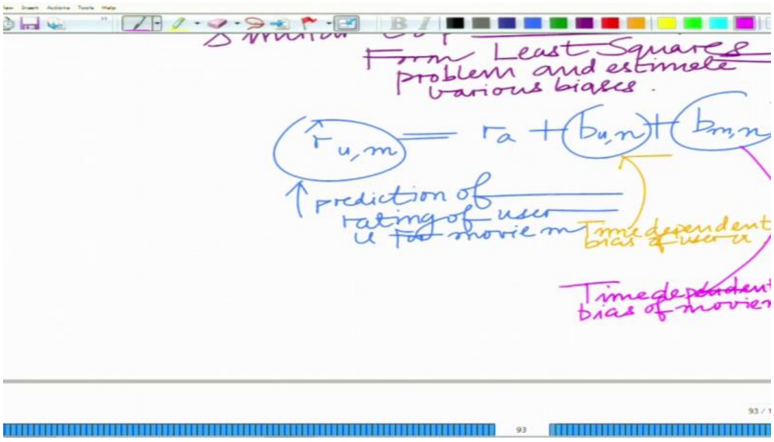
So this is done as shown in slides below.

(Refer Slide Time: 07:49)



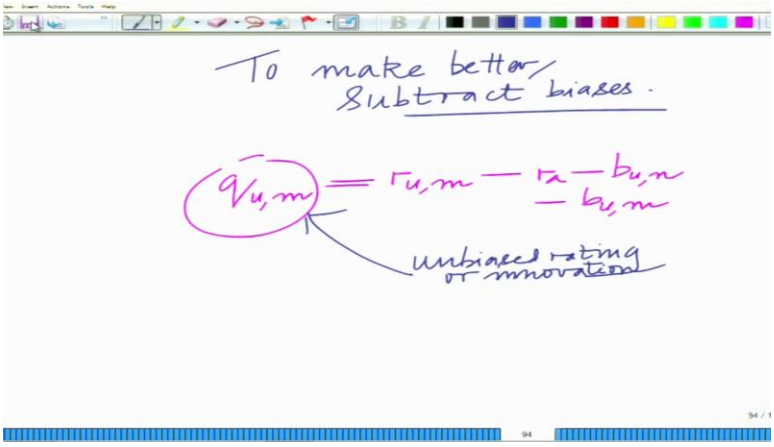
So now you have many more biases, so you can form a least squares again similar to the previous one and solve for the biases.

(Refer Slide Time: 09:42)



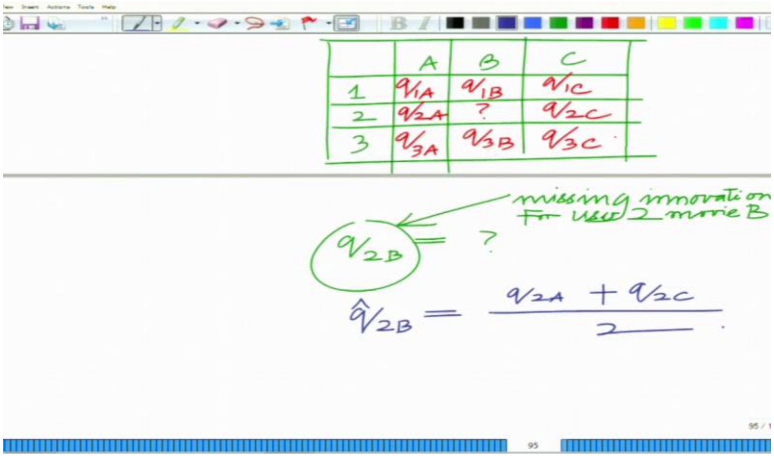
Now, at this stage you can once again predict substituting the biases and the missing prediction $q_{u,m} = r_{u,m} - r_a - b_{u,n} - b_{u,m}$.

(Refer Slide Time: 11:18)



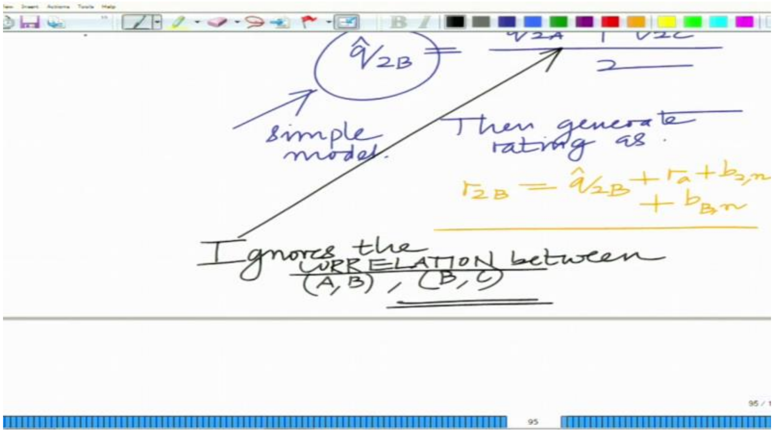
Now, we can refine this model slightly further by subtracting the biases. So after removing the biases we can call this as the unbiased rating or the innovation, something that we have not been able to predict.

(Refer Slide Time: 12:46)



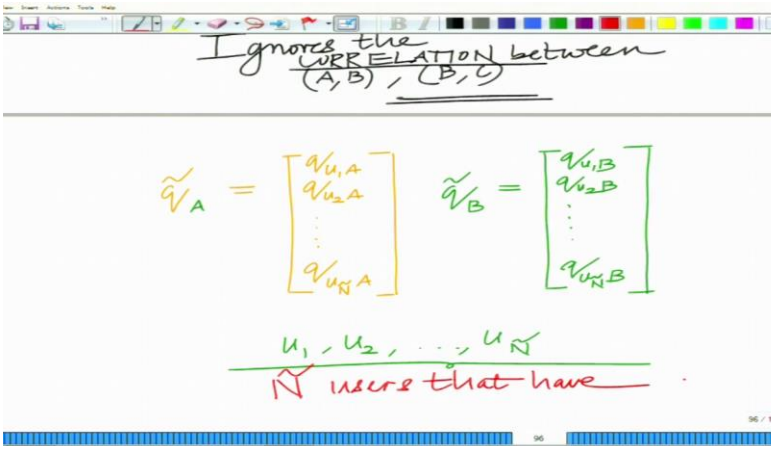
And now you can form the innovations corresponding to the different ratings that are available as shown in slide and compute the missing innovation. So we have $q_{2B} = \dfrac{q_{2A} + q_{2C}}{2}$.

(Refer Slide Time: 14:27)



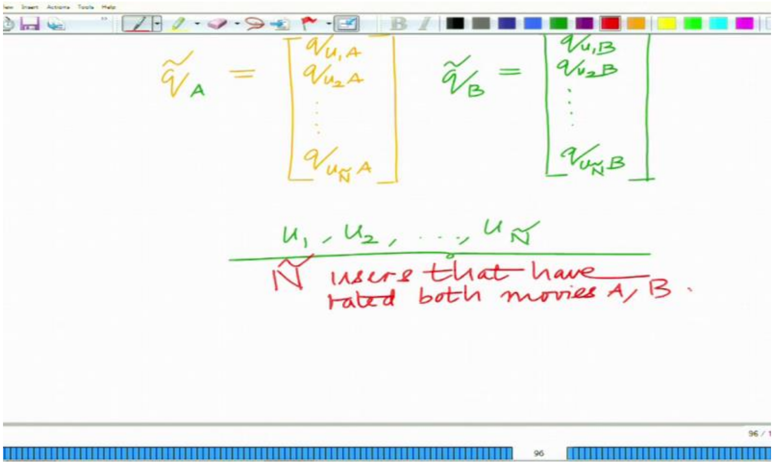
This is a simple model, then once you do this you generate the rating as $r_{2B} = q_{2B} + r_a + b_{2,n} + b_{B,n}$ and this is your prediction model. However, this model ignores the correlation between movies A, B and movies B, C. Because if A and B are very similar then you have to give higher weightage to A, on the other hand B and C are very similar then you have to give a higher weightage to B. So one has to give adequate weightage depending on the correlation.

(Refer Slide Time: 16:37)

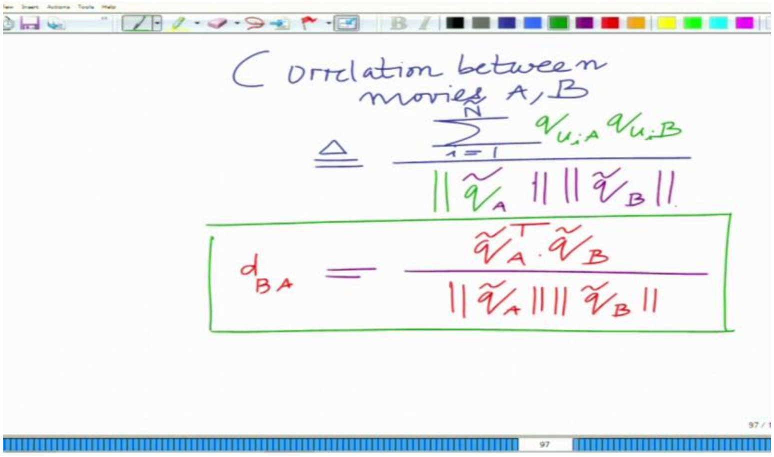


Let us consider now the innovation vectors as shown in slide.

(Refer Slide Time: 18:56)

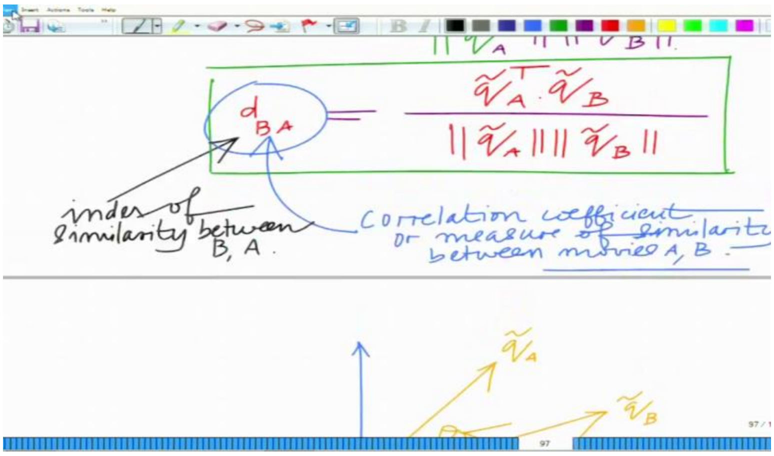


(Refer Slide Time: 19:13)



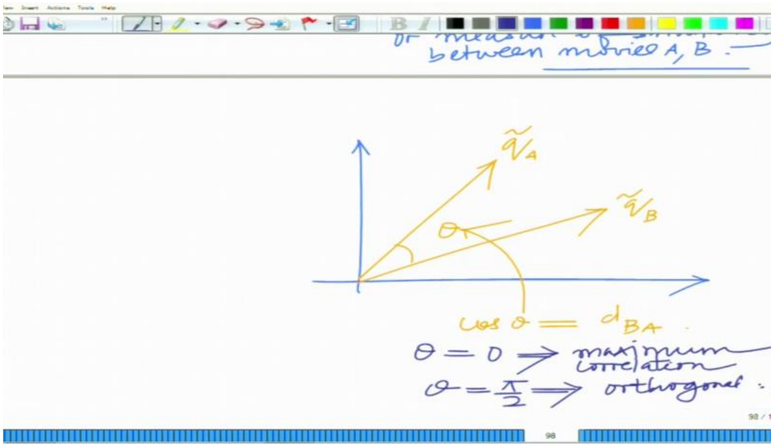
Then the correlation between these two can be defined as $\frac{\sum_{i=1}^{N} q_{u_i A} q_{u_i B}}{\left\| q_A \right\| \left\| q_B \right\|}$. So we call this as the similarity coefficient or measure of similarity, so what we are doing is we are taking the ratings of users who have rated both A and B, computing the inner product between them and dividing them by their norms of these two vectors.

(Refer Slide Time: 20:5)



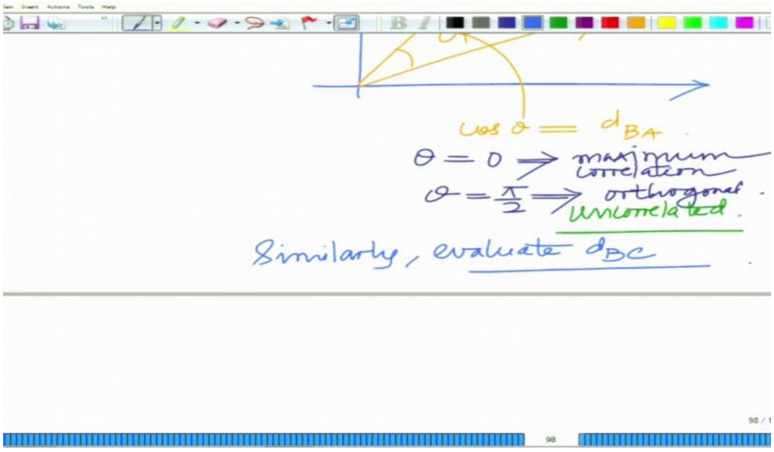
In fact, this is exactly the cosine of the angle between the two vectors.

(Refer Slide Time: 21:43)



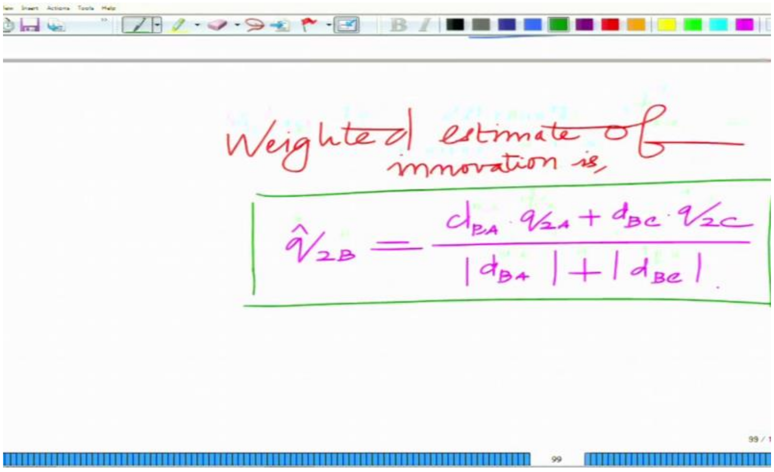
So cosine of the angle between these two vectors is nothing but what we have defined as the similarity coefficient. And you can clearly see that if $\theta = 0$ the vectors are aligned, movies A and B are similar, as θ increases movies A and B are dissimilar. If $\theta = \frac{\pi}{2}$ they are perpendicular in fact, A has no bearing on B implies they are orthogonal or no correlation, implies uncorrelated, the innovations are uncorrelated.

(Refer Slide Time: 22:59)



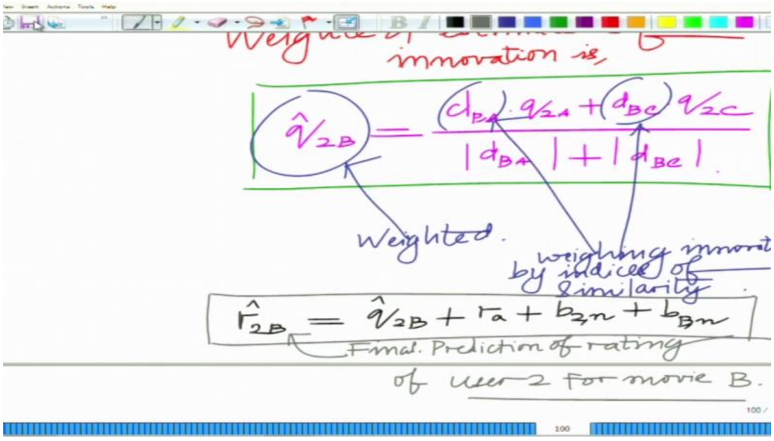
And similarly now evaluate $d_{B,C}$, this is the correlation or measure of similarity.

(Refer Slide Time: 23:50)



Now, the weighted innovation or weighted estimate is $q_{2B} = \frac{d_{BA} q_{2A} + d_{BC} q_{2C}}{|d_{BA}| + |d_{BC}|}$ and this is the final step.

(Refer Slide Time: 24:51)



And now once you compute this innovation estimate you can add it to the biases as $\hat{r}_{2B} = q_{2B} + r_a + b_{2,n} + b_{B,n}$ and that is basically the final step in this procedure. So this is your final prediction of rating of user 2 for movie B. So this brings across various ideas in both linear algebra as well as optimization. So we will stop here and continue in the subsequent modules. Thank you very much.