Assignment 2 Due Date – 30th March 11:59 PM 10 marks

Use the Movielens dataset 1 (the 100K dataset) from:

https://grouplens.org/datasets/movielens/100k/

Code a latent factor model for imputing the missing values. You cannot use existing codes from seniors or from the internet; that would tantamount to plagiarism. You cannot use in-built libraries for matrix factorisation; you need to code that part yourself. Either use alternating least squares (ALS) as discussed in class or use multiplicative updates (MU).

For the algorithm on MU, check

https://stats.stackexchange.com/questions/351359/deriving-multiplicative-update-rules-for-nmf

If you want to go deeper, you have to do your own research for MU.

Use 5 fold cross validation with the folds defined in the data. DO NOT define your own folds.

To test how good or bad your recommender system is, you should compute the Normalized Mean Absolute Error (NMAE) on the test set.

This will be an open-ended assignment. You will get marks based on how accurate you are. If you are using some kind of parameter, mention the set of parameters in the Table.

You have to fill two tables (one for user based and one for item based) of the following form

	Param 1	Param 2	…	Param n
Fold 1				
Fold 2				
Fold 3				
Fold 4				
Fold 5				
Average				

You have to submit a PDF file. You will also have to demonstrate the code in front of TA.

- 10 marks