

Download the Movielens dataset 1 (the 100K dataset) from <http://www.grouplens.org/node/73>

Build a simple user-based and item-based recommender system (papers have been provided).

The dataset will have available and missing ratings. For evaluation you should use 5 fold cross validation. In each run, use 4 parts for training and the remaining 1 part for testing. Use the training set to predict the ratings of the test set.

For user based approach (table 1) see how the MAE changes when the threshold for neighborhood selection varies: use thresholds 0.4, 0.5, 0.6, 0.7

For the item based approach (table 2) see how the accuracy changes when the number of similar items are restricted to K most similar items. Vary K: 10, 20, 30, 40

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

For measuring similarity you should use the **Cosine Similarity**

Table 1. MAE values User based

Fold #	$\tau=.4$	$\tau=.5$	$\tau=.6$	$\tau=.7$
1				
2				
3				
4				
5				
Average				

Table 2. MAE values Item based

Fold #	K=10	K=20	K=30	K=40
1				
2				
3				
4				
5				
Average				

- 5+5 marks