The dataset you will be using is Movielens 100K. Regular 5 fold cross validation needs to be reported, i.e. the NMAE on each fold should be reported.

You have to implement collaborative filtering as a matrix completion problem and solve it via nuclear norm minimization.

If X is the full ratings matrix and Y is the observed one: $Y = R \cdot X$. Here '.' Is the binary product between the mask R and the ratings matrix X.

What you need to do is to recover X by solving the following:

$$\min_{X} \|Y - R \cdot X\|_F^2 + \lambda \|X\|_*$$

Here $\|X\|_*$ is the nuclear normof the matrix and defined as the sum of singular values.

The algorithm for solving this has been discussed in class.

You must submit in google classroom –

1. A report of the NMAE for different folds.

2. Implementation code / script.

Marks distribution

7 – For implementing. If you do not implement the exact formulation you will get ZERO, i.e. no marks for implementing matrix factorization solution. You are allowed to use off the shelf implementations of nuclear norm minimization available online in Python / Matlab.

1 – For tuning the parameter λ to yield the best results. For bad NMAE's you will lose this ONE mark.

2 – For proper report. If you do not report NMAE for each fold, but some other metric like MAE or RMSE, you will get ZERO. If you report overall NMAE but not for each fold separately you will get ZERO.