Factorization-Based Data Modeling
Practical Work 2

Umut Şimşekli, Mathieu Fontaine
Télécom ParisTech, Université Paris-Saclay, Paris, France
umut.simsekli@telecom-paristech.fr

Instructions: (please read carefully)

1. This homework can be done in groups of **maximum 2** people.
2. Prepare your report as a pdf file in English by using LaTeX or a similar software (Word etc). Do not submit scanned papers.
3. Put all your files (code and/or report) in a zip file: *surname_name_tp2.zip* and submit it to `https://www.dropbox.com/request/fFuzBEepyC9qOh7YhW9Z`. The deadline is **January 5th, 2018**. Late submissions will not be accepted.
4. One submission per group is sufficient.

1 Matrix Factorization with Stochastic Gradient Descent

In this section, you will implement the stochastic gradient descent algorithm for large-scale matrix factorization. The problem that we aim to solve is given as follows:

$$(W^\star, H^\star) = \underset{W,H}{\arg\min} \frac{1}{2}\|M \odot (X - WH)\|_F^2, \tag{1}$$

where $X \in \mathbb{R}^{I\times J}$ is the data matrix, and $W \in \mathbb{R}^{I\times K}$ and $H \in \mathbb{R}^{K\times J}$ are the unknown factor matrices. Here $\|A\|_F$ denotes the Frobenius norm of a matrix A and $\odot$ denotes element-wise multiplication. Finally, $M \in \{0,1\}^{I\times J}$ is the 'mask' matrix, denoting if a particular entry of X is observed or not: $m_{ij} = 1$ if x_{ij} is observed and $m_{ij} = 0$ otherwise.

2 Movie Recommendation

We will work on the MovieLens 1 Million dataset. This dataset contains ∼1 million ratings applied to $I = 3883$ movies by $J = 6040$ users, resulting in a sparse data matrix X with 4.3% non-zero entries. Our aim will be to decompose this matrix into W and H by only using its observed entries. Once we obtain estimates for W and H, we can then use them for predicting the unobserved entries of X, which will enable us to make recommendations.

3 Exercises

Now go to the file `matrix_factorization_template.m`

1. Complete the stochastic gradient algorithm.

2. At the end of each iteration, compute the roor-mean-squared-error, that is given as follows:

$$\text{RMSE} = \sqrt{\frac{\|M \odot (X - WH)\|_F^2}{N}} \tag{2}$$

   where N is the number of observed entries in X.

3. Play with the algorithm parameters, i.e. the step-size, the batch-size, initialization, and the rank of the factorization. What do you observe? How do the step-size and the batch-size interact?

4. After estimating W and H, use them to recommend a movie for a given user.