Kernelized Probabilistic Matrix Factorization

Nikita Alexeychuk, Kristina Belikova, Milena Gazdieva, Mikhail Kuzin,
Ivan Maksimov

December 23, 2018

Abstract

The primary purpose of this project was to realized matrix completion algorithm — Kernelized Probabilistic Matrix Factorization. ompared to existing probabilistic approach PMF, which KPMF efficiently used appropriate side infortmation. We have implemented KPMF along with PMF and SVD decompositions as baseline and compared these approaches through application of matrix factorization for recommender systems.

1 Background

One of the biggest problems of many research areas, e.g., recommender systems and image restoration is the problem of missing values prediction. That is, the goal is to complete missing values of some matrix, such that these values are coherent with existing data — either observed entries of the matrix itself or some side information (e.g., users' social network in recommender systems). There already exist some matrix completion techniques. In these algorithms partially observed matrix is factorized, and, therefore, latent vectors for each row and column of the matrix are obtained. Then the prediction of each missing entry is the inner product of latent vectors of the corresponding row and the corresponding column, obtained from the factorization. One of the main issues of such techniques is that they fail to make value prediction based on small amount of data. That is, they suffer from the data sparsity, which is quite common. On the other hand, these techniques ignore available side information, which may be useful for evaluating of the underlying model. In order to overcome mentioned problems [1] proposed Kernelized Probabilistic Matrix Factorization (KPMF) model. By this technique side information is incorporated through kernel matrices over rows and over columns. The matrix is modeled as the product of two latent matrices, which are sampled from two different zero-mean Gaussian processes (GP). Covariance functions of these processes, which encode covariance structure across rows and columns respectively, are obtained from the side information. It is important to note one more advantage of the KPMF over the existing Probabilistic Matrix Factorization (PMF) techniques. PMF methods assume an independent latent vectors for each row, while KPMF deals with vector, spanning all rows. Therefore, if, e.g., an entire row of the data matrix is missing, PMF fail to make prediction for that row. KPMF, however, can still make predictions based on the row covariances alone.

2 Problem formulation

The aim of our project is to apply theoretical results, presented in [1], to some real task. In order to do so, we need firstly to evaluate full covariance matrices K_U and K_V for rows of initial matrix R and columns of R respectively. Then generate R_{nm} for each of the observed entries in the target matrix R given latent matrices U and V, and finally perform gradient descent to learn the latent matrices U and V, which maximize the log-posterior $\log p(U, V|R, \sigma^2, K_U, K_V)$. After the estimations for given latent matrices are derived, the maximum likelihood estimation for each of the missing values of R is the inner product of the corresponding latent vectors.

3 Data

We are going to run experiment on public datasets that were used in [1]. We are going to use dataset Epinion, which contains customer review and ratings on varios items, such as electronics, movies, companies.

4 Related work

The problem of missing values prediction is encountered in many research areas and in many cases it is solved by factorization based algorithms. A probabilistic framework for matrix factorization was proposed in [2], and generalized to a full Bayesian model in [3]. This approach was reviewed in [4], [5] for construction of recommender systems.

5 Scope

Below is presented the scope of work with task allocation.

- Downloading and preparin g the data; (Alexeychuk)
- Evalution of covariance matrices K_U,K_V; (Alexeychuk)
- Performing gradient descent to learn the latent matrices U and V; (Belikova)
- Performing maximum likelihood estimations for each of the missing values of R; (Gazdieva)
- Programming and evaluating baseline models - SVD and PMF; (Makimov)
- Evaluating the results of work; (Kuzin)
- Preparing presentation and report. (Maksimov)

6 Evaluation

The work will be evaluated by performing predictions for ratings, that already exist, and counting the explained dispersion. Also the correctness of the solution can be evaluated by comparing it with results, that were obtained in [1].

7 Methodology

For KPMF method we took 2 different Kernel functions:
Diffusion:

$$K_{Diffusion} = e^{-\beta L}$$

Regularized Laplacian:

$$K_{RL} = (I + \gamma L)^{-1}$$

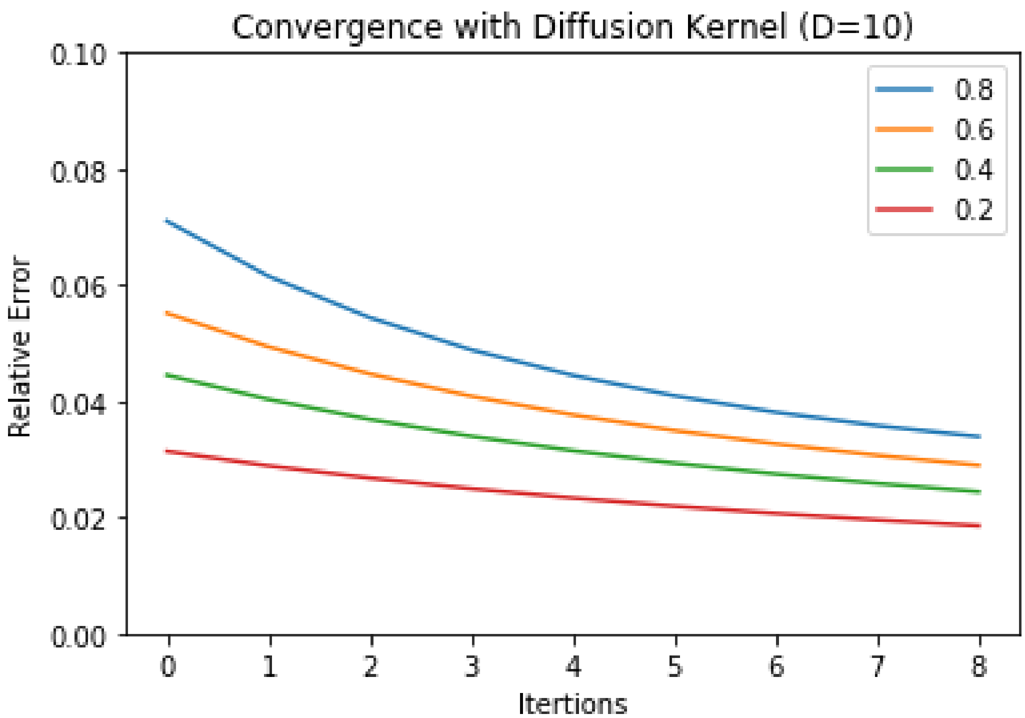
For each kernel we calculated Covariance matrices (K_D, K_{RL}). And implemented gradient descent scheme for estimating latent matrices U, V such that $R = UV$ where R is original matrix.

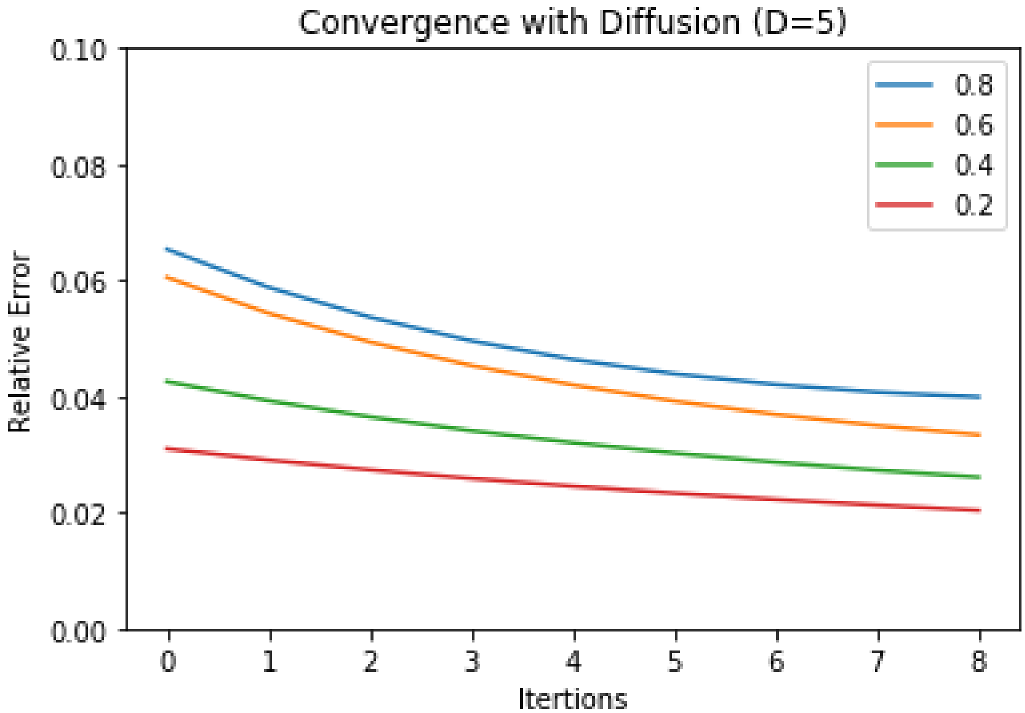
Matrices $R \in R^{m \times n}$, $U \in R^{m \times D}$, $V \in R^{D \times n}$.

We measured convergence speed (relative errors from iteration to iteration) and RMSE for $(R_{original} - UV)$.

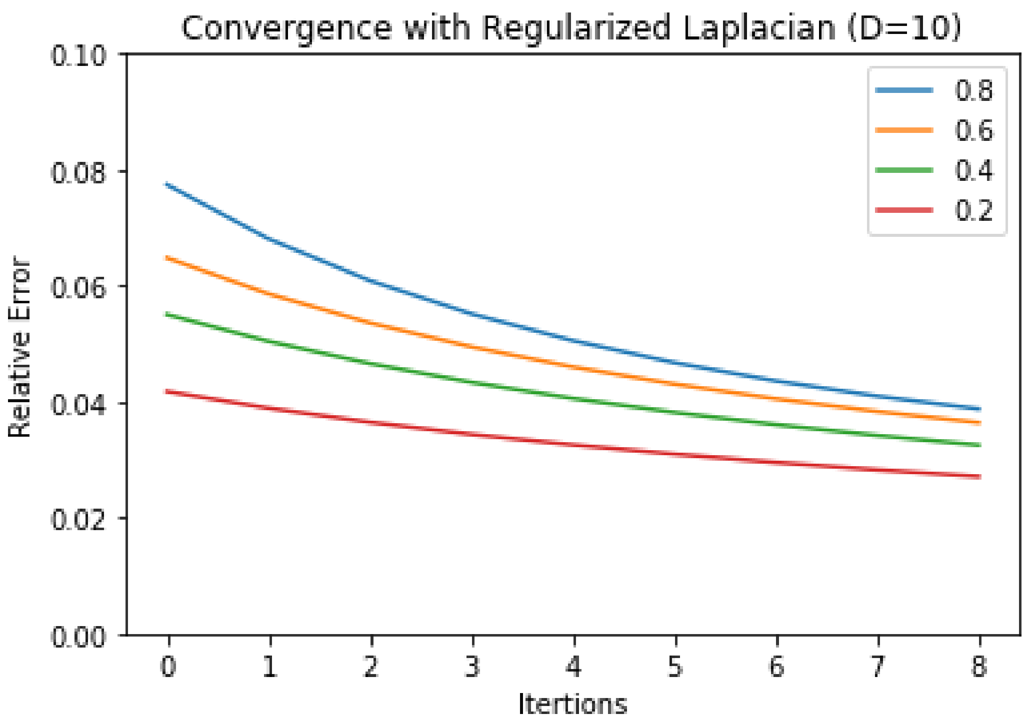
8 Results

We propose our results here.

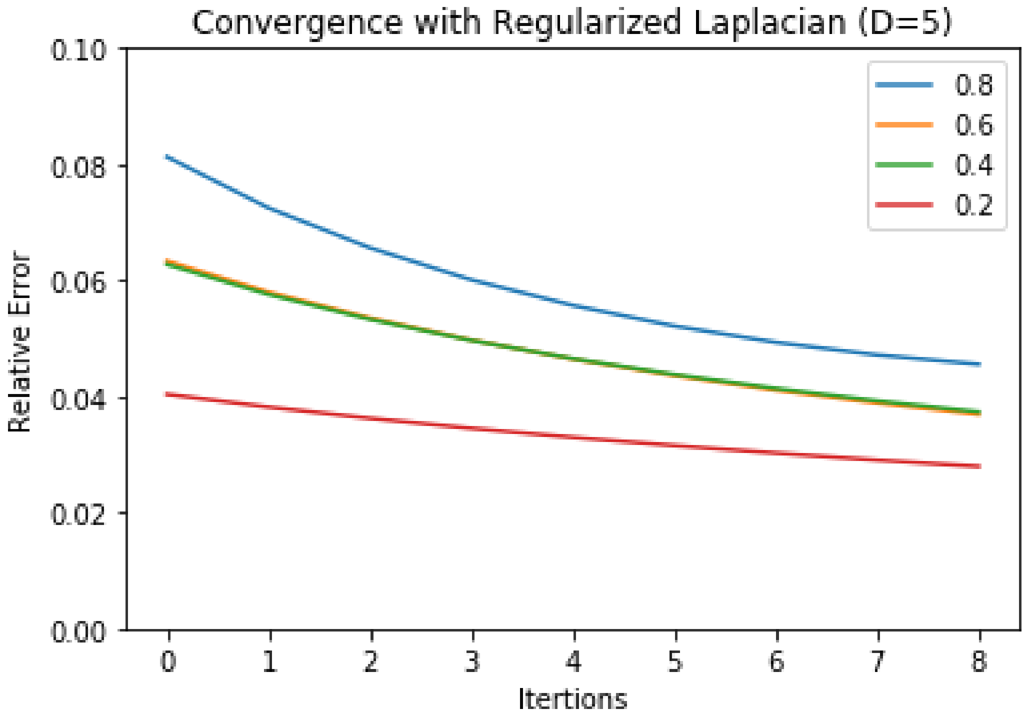


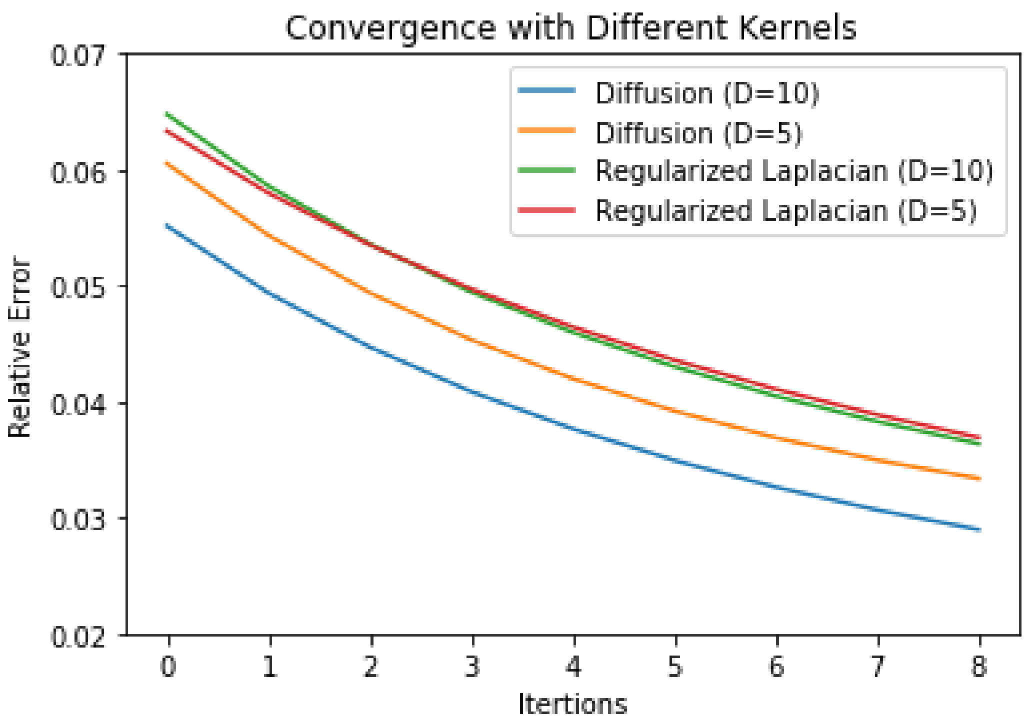


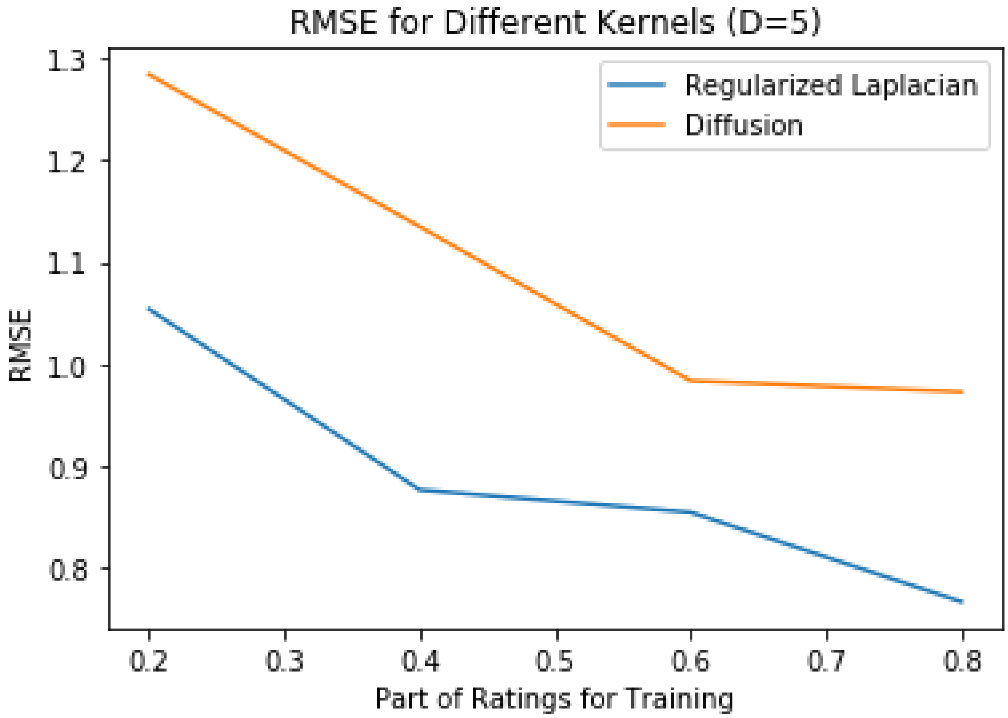
Convergence with Diffusion (D=5)
0.10
0.08
0.06
0.04
0.02
0.00
Relative Error
0
1
2
3
4
5
6
7
8
Itertions
0.8
0.6
0.4
0.2

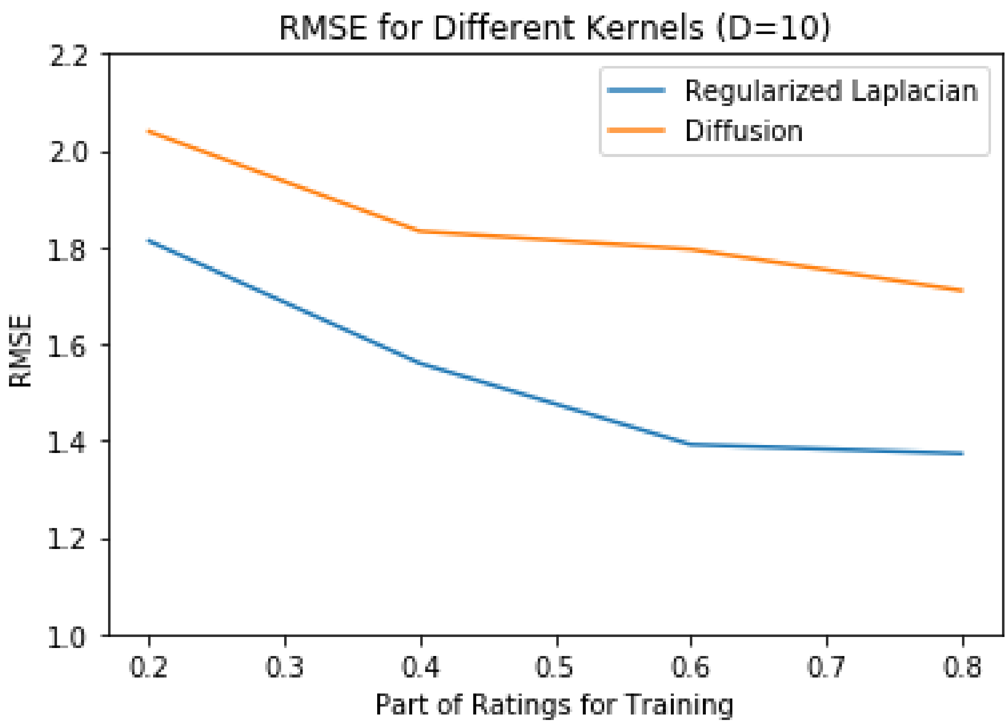


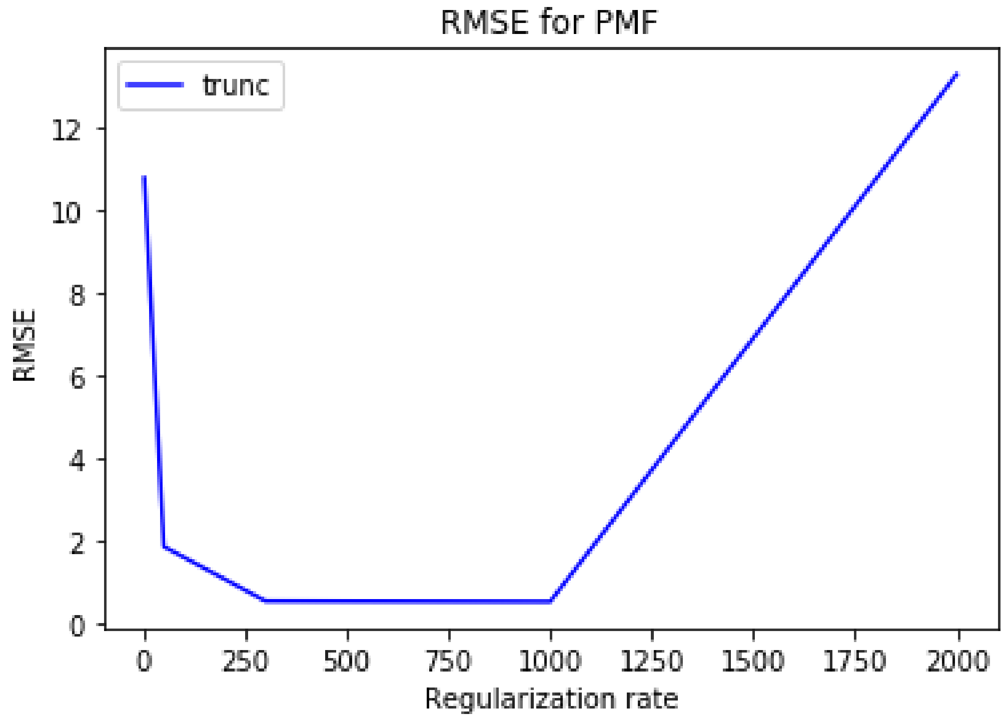
Convergence with Regularized Laplacian (D=10)
0.10
0.08
0.06
0.04
0.02
0.00
Relative Error
0
1
2
3
4
5
6
7
8
Itertions
0.8
0.6
0.4
0.2

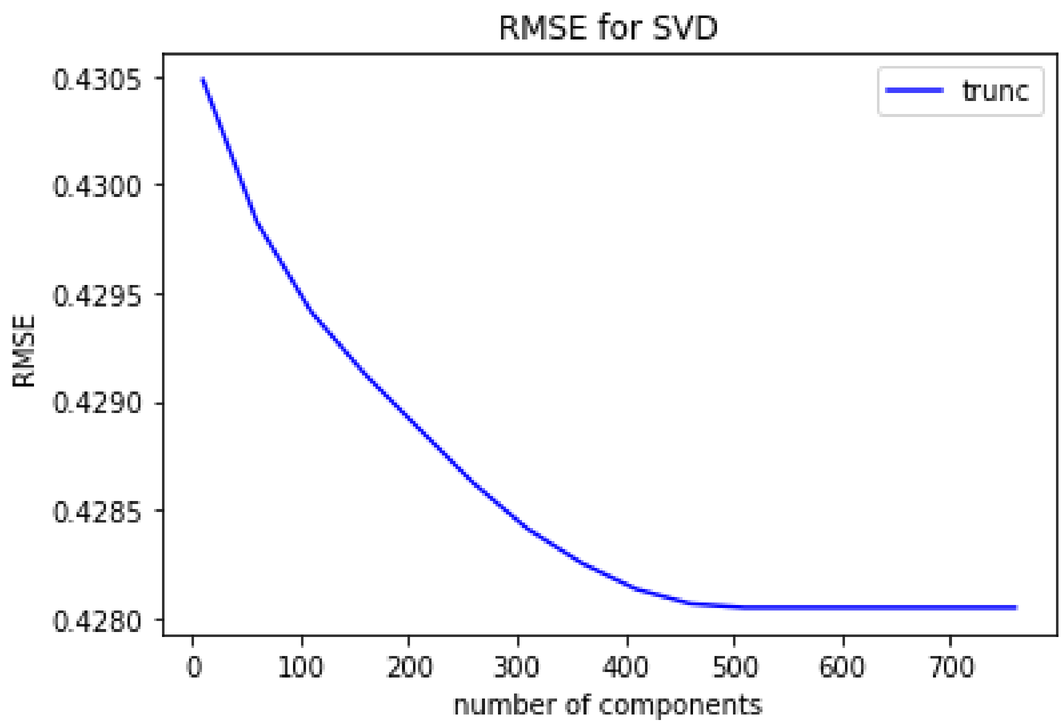

Convergence with Regularized Laplacian (D=5)
0.8
0.6
0.4
0.2
Relative Error
Itertions
0.00
0.02
0.04
0.06
0.08
0.10
0
1
2
3
4
5
6
7
8


Convergence with Different Kernels
Diffusion (D=10)
Diffusion (D=5)
Regularized Laplacian (D=10)
Regularized Laplacian (D=5)
Relative Error
Itertions
0.02
0.03
0.04
0.05
0.06
0.07
0
1
2
3
4
5
6
7
8


RMSE for Different Kernels (D=5)
Regularized Laplacian
Diffusion
RMSE
1.3
1.2
1.1
1.0
0.9
0.8
0.2
0.3
0.4
0.5
0.6
0.7
0.8
Part of Ratings for Training


RMSE for Different Kernels (D=10)
Regularized Laplacian
Diffusion
RMSE
2.2
2.0
1.8
1.6
1.4
1.2
1.0
0.2
0.3
0.4
0.5
0.6
0.7
0.8
Part of Ratings for Training





Method	RMSE
KPMF (RL Kernel)	0.763 (!needed more iterations)
SVD	0.431
PMF	0.423

References

[1] Tinghui Zhou, Hanhuai Shan, Arindam Banerjee, Guillermo Sapiro. *Kernelized Probabilistic Matrix Factorization: Exploiting Graphs and Side Information*, Proceedings of the 2012 SIAM International Conference on Data Mining, 2012.

[2] R. Salakhutdinov and A. Mnih. *Probabilistic matrix factorization*, in NIPS, 2007.

[3] R. Salakhutdinov and A. Mnih, *Bayesian probabilistic matrix factoriza- tion using Markov chain Monte Carlo*, in ICML, 2008.

[4] H. Ma, H. Yang, and I. K. M. Lyu, *Social recommendation using probabilistic matrix factorization*, in CIKM, 2008.

[5] H. Shan and A. Banerjee, *Generalized probabilistic matrix factorizations for collaborative filtering*, in ICDM, 2010.