Project Report 3
Empirical Research on Matrix Factorization using Collaborative Filtering

Tingyuan Liang

May 8, 2018

Date Performed:	May 4, 2018
Instructor:	Professor Dit-Yan Yeung
TA:	Zhihan Gao

1 Objectives

The objective of this project is threefold:

1. To acquire a better understanding of matrix factorization and collaborative filtering by implementing a version of probabilistic matrix factorization (PMF).

2. To investigate how the performance of PMF is affected by varying some hyperparameters of the model and the degree of sparsity of the data.

3. To learn how to implement compatible modules in scikit-learn in accordance with its application programming interface (API).

2 Major Tasks

The project consists of the following tasks:
1. To implement PMF with maximum a posteriori (MAP) estimation according to the estimator interface of scikit-learn.
2. To conduct empirical study to compare the performance of PMF by varying the number of latent factors K and the regularization hyperparameters λ_u and λ_v .
3. To conduct empirical study to compare the performance of PMF on dense and sparse data.
4. To write up a project report.

3 PMF Model and Related Derivation

In this project, we utilize PMF model to implement a recommadation system. Suppose we have M movies, N users, and integer rating values from min_rating to max_rating. Let R_{ij} represent the rating of user i for movie j , $U \in R^{K \times N}$ and $V \in R^{K \times M}$ be latent user and movie feature matrices, with column vectors U_i and V_j representing user-specific and movie-specific latent feature vectors respectively. According to [1], maximizing the log-posterior over movie and user features with hyperparameters (i.e. the observation noise variance and prior variances) kept fixed is equivalent to minimizing the sum-of-squared-errors objective function with quadratic regularization terms:

$$E = \frac{1}{2} \sum_{i=1}^{N} \sum_{j=1}^{M} I_{ij} (R_{ij} - U_i^T V_j)^2 + \frac{\lambda_U}{2} \sum_{i=1}^{N} \|U_i\|^2 + \frac{\lambda_V}{2} \sum_{j=1}^{M} \|V_j\|^2 \tag{1}$$

Accorrding the definition of E in (1), we have:

$$\nabla_{\mathbf{U}_i} E = - \sum_{I_{ij}=1} (R_{ij} - \mathbf{U}_i^T \mathbf{V}_j) \mathbf{V}_j + \lambda_U \mathbf{U}_i \tag{2}$$

$$\nabla_{\mathbf{V}_j} E = - \sum_{I_{ij}=1} (R_{ij} - \mathbf{U}_i^T \mathbf{V}_j) \mathbf{U}_j + \lambda_V \mathbf{V}_i \tag{3}$$

4 Task 1: Empirical Experiments on Dense Data

4.1 The Experiment Settings

According to 5-fold cross-validation, the major parameter settings of PMF Estimator based on conventional stochastic gradient descent (SGD) algorithm are presented in Table 1.

Names	Parameter Settings
Estimator	PMF based on SGD
Batch Size	1500
Learning Rate η	0.003
λ_U	0.1
λ_V	0.1
Number of Factors K	3
Epoch	250
MinRating	1
MaxRating	5
Number of Movies	1682
Number of Users	943

Table 1: Experiment Settings of the Empirical Experiments on PMF with Dense Data

4.2 Hyperparameter Tuning Result on λ_U and λ_V

RMSEs of models with different λ_U and λ_V in 5-fold cross-validation are presented in Table 2 and Figure 1. For those result, the number of factors K is set to 2. With 20 processes which handle the CV in parallel, the average time for one of 4x4x5 training-test procedures is 9.65 seconds.

λ_U \ λ_V	0.1	1	10	100
0.1	**0.91375522**	0.97325516	1.20697876	1.26565407
1	0.97134466	1.21901564	1.26653038	1.26653723
10	1.23833647	1.26654346	1.26660413	1.26660622
100	1.26572489	1.26654418	1.26660612	1.26660635

Table 2: RMSE of Models with Different Parameters in 5-Fold Cross-Validation

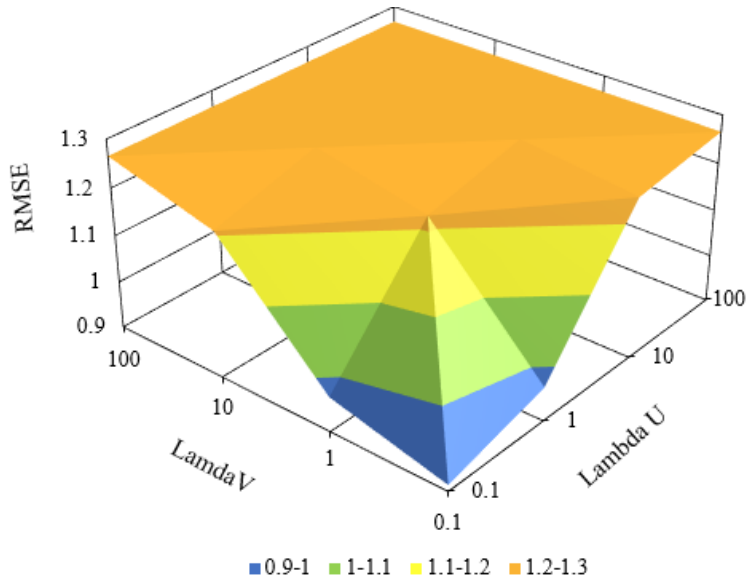


Figure 1: RMSE of Models with Different Parameters in 5-Fold Cross-Validation

4.3 Hyperparameter Tuning Result on the Number of Factors K

Accuracies of models with different K in 5-fold cross-validation are presented in Table 3 and Figure 2. For those result, λ_U and λ_V are set according to the result presented in Section 4.2 to obtain low RMSE. With 20 processes which handle the CV in parallel, the average time for one of 5x5 training-test procedures is 28.574 seconds.

K	1	2	3	4	5
RMSE	1.03115588	0.91307273	**0.88339853**	0.89030351	0.908087

Table 3: RMSE of Models with Different Parameters in 5-Fold Cross-Validation

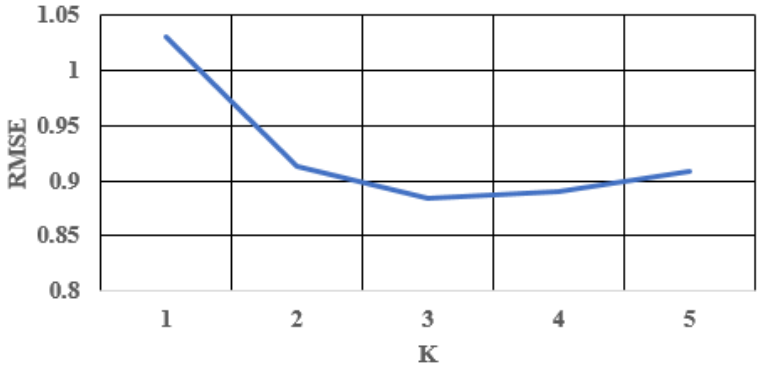


Figure 2: RMSE of Models with Different Parameters in 5-Fold Cross-Validation

4.4 Performance for Dense Training Dataset

Based on the tuned hyperparameters, the RMSE based on training set of the PMF model is 0.8590750967846111 while the RMSE based on test set of the PMF model is 0.9291254281078227. The learning curve is shown in Figure 3. The training time is 78.8635 seconds and the test time is 0.00136 seconds.

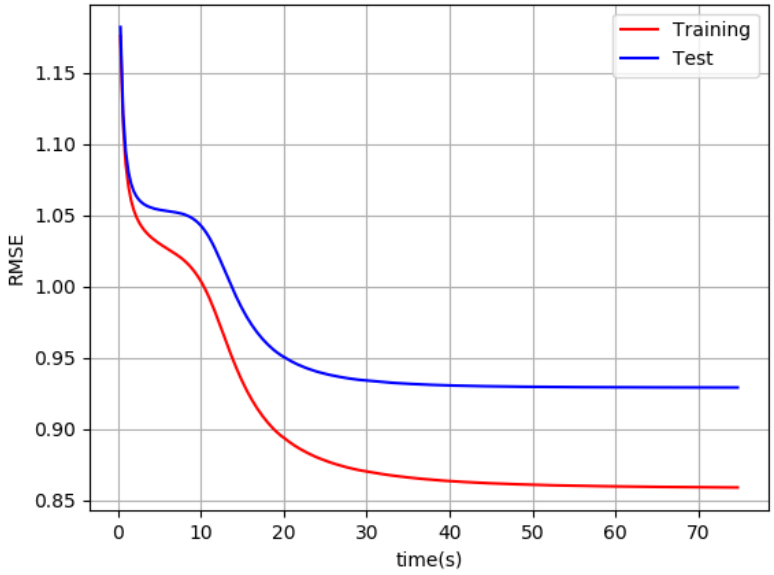


Figure 3: Learning Curve based on Dense Dataset

5 Task 2: Empirical Experiments on Sparse Data

5.1 The Experiment Settings

According to 5-fold cross-validation, the major parameter settings of PMF Estimator based on conventional stochastic gradient descent (SGD) algorithm are presented in Table 4.

Names	Parameter Settings
Estimator	PMF based on SGD
Batch Size	1500
Learning Rate η	0.003
λ_U	0.1
λ_V	1
Number of Factors K	5
Epoch	250
MinRating	1
MaxRating	5
Number of Movies	1682
Number of Users	943

Table 4: Experiment Settings of the Empirical Experiments on PMF with Sparse Data

5.2 Hyperparameter Tuning Result on λ_U and λ_V

RMSEs of models with different λ_U and λ_V in 5-fold cross-validation are presented in Table 5 and Figure 4. For those result, the number of factors K is set to 2. With 20 processes which handle the CV in parallel, the average time for one of 4x4x5 training-test procedures is 2.720 seconds.

λ_U \ λ_V	0.1	1	10	100
0.1	1.3142898	**1.19501144**	1.22051736	1.2685728
1	1.19904226	1.20864976	1.27134222	1.27140254
10	1.22887815	1.2714252	1.27162628	1.27163203
100	1.26823171	1.27138417	1.27162583	1.27164272

Table 5: RMSE of Models with Different Parameters in 5-Fold Cross-Validation

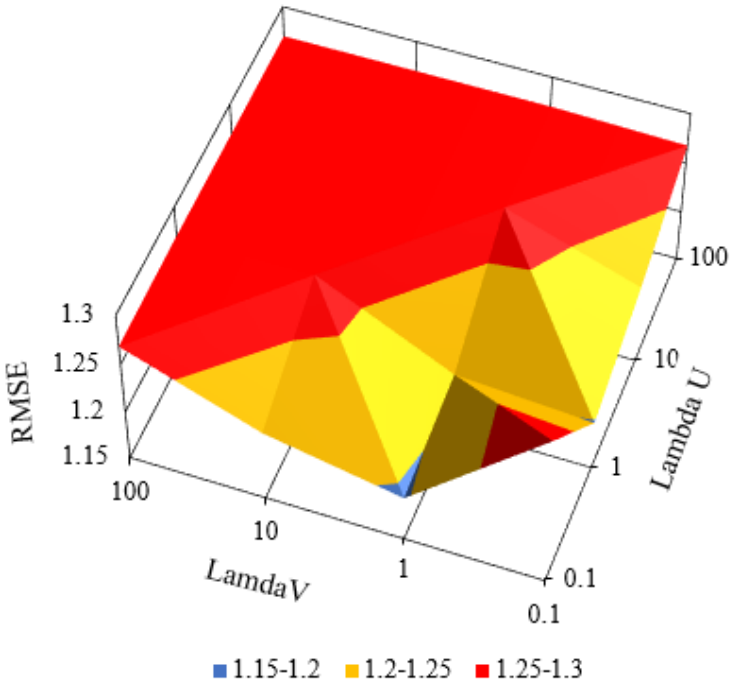


Figure 4: RMSE of Models with Different Parameters in 5-Fold Cross-Validation

5.3 Hyperparameter Tuning Result on the Number of Factors K

Accuracies of models with different K in 5-fold cross-validation are presented in Table 6 and Figure 5. For those result, λ_U and λ_V are set according to the result presented in Section 5.2 to obtain low RMSE. With 20 processes which handle the CV in parallel, the average time for one of 5x5 training-test procedures is 7.1422 seconds.

K	1	2	3	4	5
RMSE	1.16580834	1.19523411	1.18913226	1.21640039	**1.15330528**

Table 6: RMSE of Models with Different Parameters in 5-Fold Cross-Validation

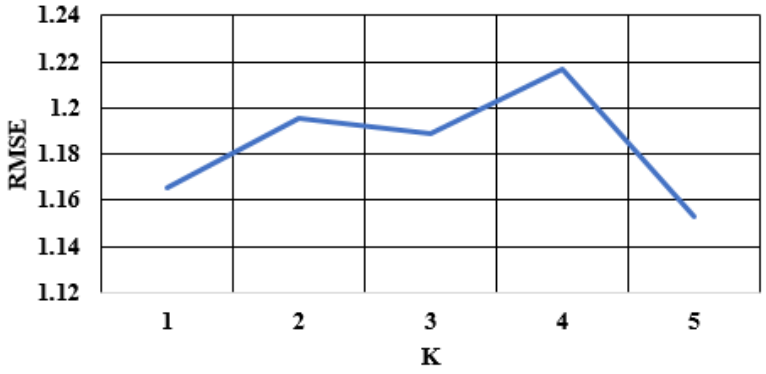


Figure 5: RMSE of Models with Different Parameters in 5-Fold Cross-Validation

5.4 Performance for Sparse Training Dataset

Based on the tuned hyperparameters, the RMSE based on training set of the PMF model is 0.828886747207722 while the RMSE based on test set of the PMF model is 1.0442788465750548. The learning curve is shown in Figure 6. The training time is 29.0133 seconds and the test time is 0.00800 seconds.

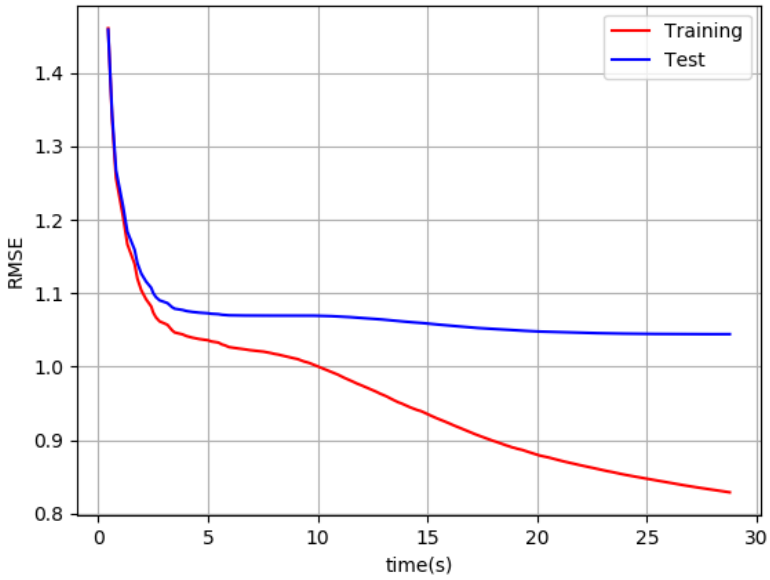


Figure 6: Learning Curve based on Sparse Dataset

6 Task 3: Comparoson of the Performance on the Dense Data and on the Sparse Data

From the perspective of resultant RMSE, PMF can achieve a lower RMSE of test based on dense data than the one based on sparse data.

For dense data, the training RMSE could be higher than the one for sparse data but based on more samples, the test RMSE will be lower. In contrast, for sparse data, the training RMSE can reach a lower level but due to the lack of samples, the test RMSE will be higher. Moreover, the training time for sparse data is much less than the one for dense data, due to less training data. In contrast, the test time for sparse data is more than than the one for dense data, since there are more data left for test when fewer data for training. Finally, due to different charateristics of dataset, the hyperparameters for sparse data and dense data are different.

7 (Extension)Task 4: Experiments on Large Dataset (MovieLens_1M) [2]

7.1 The Experiment Settings

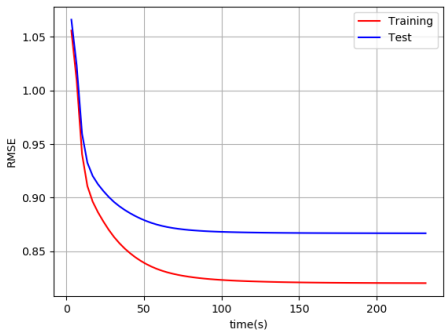
For this project, large dataset have been tried and, for this extensive part, the analysis focuses on the impart of the number of factors K. According to 5-fold cross-validation, the major parameter settings of PMF Estimator based on conventional stochastic gradient descent (SGD) algorithm are presented in Table 7.

Names	Parameter Settings
Estimator	PMF based on SGD
Batch Size	15000
Learning Rate η	0.003
λ_U	0.1
λ_V	0.1
Number of Factors K	8
Epoch	250
MinRating	1
MaxRating	5
Number of Movies	1682
Number of Users	943

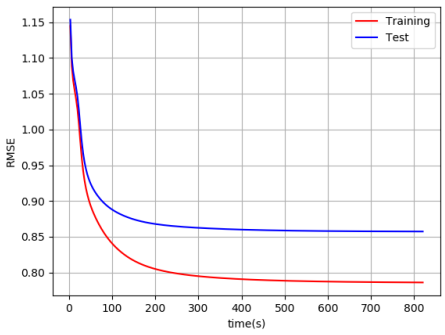
Table 7: Experiment Settings of the Empirical Experiments on PMF with Large Dataset

7.2 Impact of the Number of Factors K

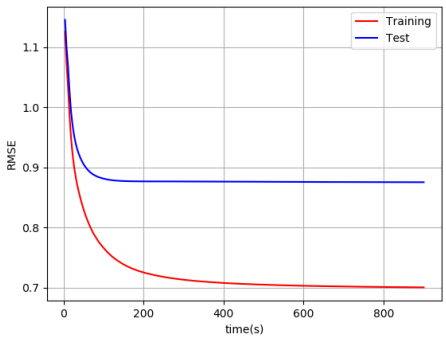
According to cross-validation, results of which are shown below, $K = 8$ can achieve the lowest RMSE. When $K < 8$, the PMF tends to underfit the data while when $K > 8$, the model result into overfitting. It can be noticed that the training time increases as the value of K is raised.



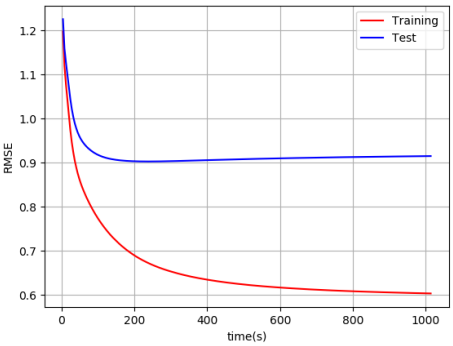
(a) K=5



(b) K=8



(c) K=20



(d) K=40

8 Reference Links

PMF model:`https://github.com/chyikwei/recommend/blob/master/recommend/pmf.py`
Custom Estimator in Sklearn:`http://scikit-learn.org/dev/developers/contributing.html#rolling-your-own-esti`

References

[1] A. Mnih and R. R. Salakhutdinov, "Probabilistic matrix factorization," in *Advances in neural information processing systems*, 2008, pp. 1257–1264.

[2] F. M. Harper and J. A. Konstan, "The movielens datasets: History and context," *ACM Transactions on Interactive Intelligent Systems (TiiS)*, vol. 5, no. 4, p. 19, 2016.