



华南理工大学

South China University of Technology

The Experiment Report of *Machine Learning*

SCHOOL: SCHOOL OF SOFTWARE ENGINEERING

SUBJECT: SOFTWARE ENGINEERING

Author:

Wei Wen
Renxin Zhuang
Ziwen Zhao

Supervisor:

Mingkui Tan

Student ID:

201530613009
201530613979
201530613764

Grade:

Undergraduate

January 7, 2018

Logistic Regression, Linear Classification and Stochastic Gradient Descent

Abstract—In this paper, I do the Recommender System Based on Matrix Decomposition experiment on MovieLens-100k dataset using stochastic gradient descent (SGD). We use different factorization parameter k . And we found that the factorization parameter k has a great impact on experiment result.

I. INTRODUCTION

In recent years, Recommender System is widely used in many areas. Matrix Factorization is the most widely used algorithm of Recommender System.

Stochastic gradient descent (SGD) is a good optimization algorithm to get the parameters of model. We need to use SGD to optimize our matrixes.

In order to explore the construction of recommended system, understanding the principle of matrix decomposition, being familiar to the use of gradient descent, cultivating our engineering ability, we conduct the following experiment.

In this experiment, we implement the Recommender System Based on Matrix Decomposition on MovieLens-100k dataset using stochastic gradient descent (SGD).

II. METHODS AND THEORY

Recommender System applies statistical and knowledge discovery techniques to the problem of making product recommendations.

The Matrix Factorization is a model based collaborative filtering algorithm. It is based on a rating matrix, with sparse ratings from m users to n items.

It assume rating matrix R can be factorized into the multiplication of two low-rank feature matrices P and Q . We use SGD to get the proper P and Q in order to fill the matrix.

SGD is to minimize the following objective function:

$$\mathcal{L} = \sum_{u,i \in \Omega} (r_{u,i} - \mathbf{p}_u^\top \mathbf{q}_i)^2 + \lambda_p \|\mathbf{p}_u\|^2 + \lambda_q \|\mathbf{q}_i\|^2$$

$$\mathbf{P} = [\mathbf{p}_1, \mathbf{p}_2, \dots, \mathbf{p}_m]^\top \in \mathbb{R}^{m \times k}.$$

$$\mathbf{Q} = [\mathbf{q}_1, \mathbf{q}_2, \dots, \mathbf{q}_n] \in \mathbb{R}^{k \times n}.$$

$r_{u,i}$ denotes the **actual rating** of user u for item i .

Ω denotes the set of **observed samples** from rating matrix R .

λ_p and λ_q are **regularization parameters** to avoid overfitting.

In order to apply SGD to obtain the proper P and Q , we calculate the gradient of P and Q , and use the gradient to update P and Q :

Randomly select an observed sample $r_{u,i}$.

Calculate the **prediction error**:

$$E_{u,i} = r_{u,i} - \mathbf{p}_u^\top \mathbf{q}_i$$

Calculate the **gradient**:

$$\frac{\partial \mathcal{L}}{\partial \mathbf{p}_u} = E_{u,i}(-\mathbf{q}_i) + \lambda_p \mathbf{p}_u$$

$$\frac{\partial \mathcal{L}}{\partial \mathbf{q}_i} = E_{u,i}(-\mathbf{p}_u) + \lambda_q \mathbf{q}_i$$

Update the feature matrices P and Q with **learning rate** α :

$$\mathbf{p}_u = \mathbf{p}_u + \alpha(E_{u,i}\mathbf{q}_i - \lambda_p \mathbf{p}_u)$$

$$\mathbf{q}_i = \mathbf{q}_i + \alpha(E_{u,i}\mathbf{p}_u - \lambda_q \mathbf{q}_i)$$

III. EXPERIMENTS

A. Dataset

We use the MovieLens-100k dataset, which consists 10000 comments from 943 users out of 1682 movies. We use the `u1.data` and `u1.test` for the splitting of training set and test set.

B. Implementation

We realized Recommender System Based on Matrix Decomposition in MovieLens-100k dataset. To optimize the parameter, we used stochastic gradient descent.

We try different factorization parameter k using the same training sample in the same order during the training phase.

We set the iteration time to 6000, and the learn rate to 0.1. The experimental result is in figure 1, figure 2, and figure 3.

According to the experiment result, we can see that over the iteration, the loss value drops significantly. And we found the value of k has an impact on the loss curve.

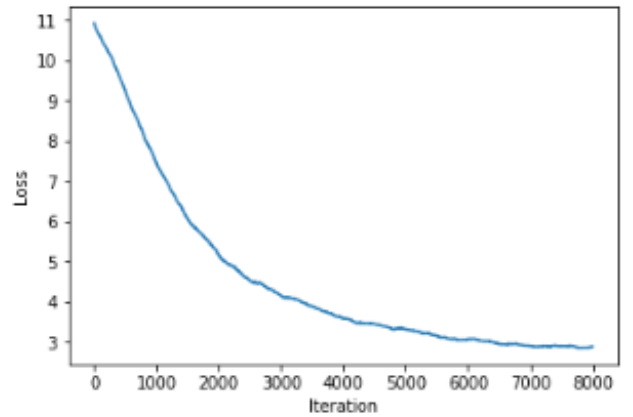


Fig. 1: loss curve for $k = 2$.

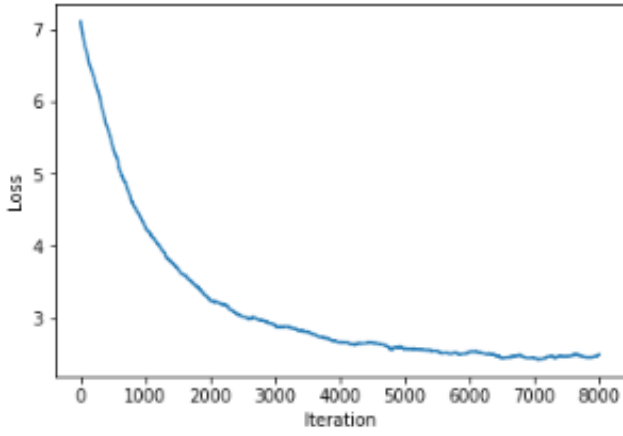


Fig. 2: loss curve for $k = 5$.

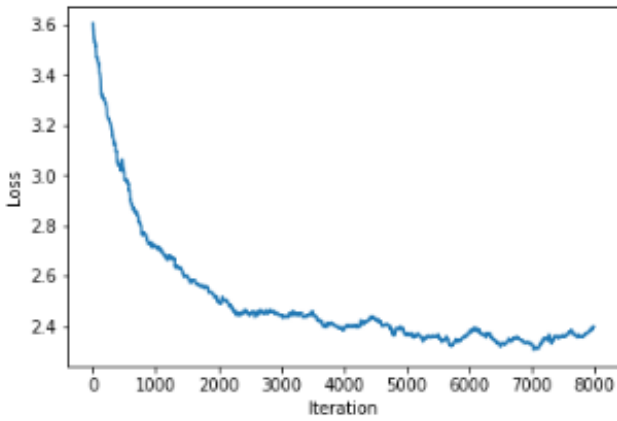


Fig. 3: loss curve for $k = 10$.

IV. CONCLUSION

Through this experiment, we found that Matrix Factorization is a good algorithm to make Recommender System. We have a deeper understanding for the principle of matrix decomposition. And we are more familiar to the use of gradient descent. This experiment cultivates our engineering ability. More importantly, we are more interested in machine learning. We got a big harvest. And I will continue to learn more knowledge about machine learning in the future.