The Experiment Report of Machine Learning



December 25, 2017

Grade:

Undergraduate

Student ID：201530612286 and 201530612279 and 201536612525

Supervisor:

Mingkui Tan

Author:

Yingjie Lin Yang Lin Dongcheng Mai

林英杰 林杨 麦栋铖

**SUBJECT:**SOFTWARE ENGINEERING

**SCHOOL:** SCHOOL OF SOFTWARE ENGINEERING

[[1]](#footnote-1)

Face Classification Based on AdaBoost Algorithm

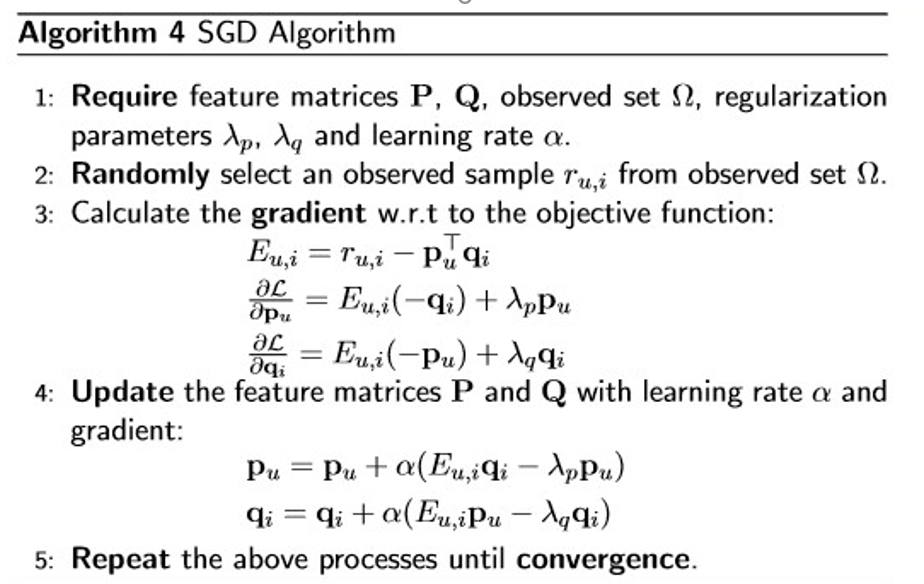
Abstract—This experiment focus on building a Recommender System Based on Matrix Decomposition.

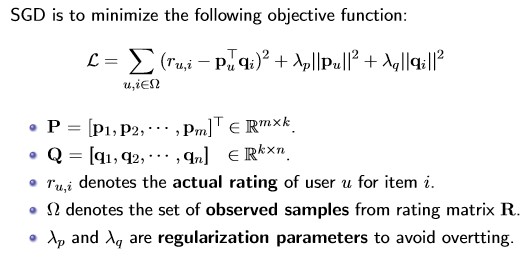
# INTRODUCTION

This experiment is for:

1. Explore the construction of recommended system.
2. Understand the principle of matrix decomposition.
3. Be familiar to the use of gradient descent.
4. Construct a recommendation system under small-scale dataset, cultivate engineering ability.

# METHODS AND THEORY





# Experiment

## Dataset

1. Utilizing MovieLens-100k dataset.
2. u.data -- Consisting 10,000 comments from 943 users out of 1682 movies. At least, each user comment 20 videos. Users and movies are numbered consecutively from number 1 respectively. The data is sorted randomly
3. u1.base / u1.test are train set and validation set respectively, seperated from dataset u.data with proportion of 80% and 20%. It also make sense to train set and validation set from u1.base / u1.test to u5.base / u5.test.
4. You can also construct train set and validation set according to your own evaluation method.

## Experiment Step

Using stochastic gradient descent method(SGD):

1. Read the data set and divide it (or use u1.base / u1.test to u5.base / u5.test directly). Populate the original scoring matrix against the raw data, and fill 0 for null values.
2. Initialize the user factor matrix and the item (movie) factor matrix , where is the number of potential features.
3. Determine the loss function and hyperparameter learning rate and the penalty factor .
4. Use the stochastic gradient descent method to decompose the sparse user score matrix, get the user factor matrix and item (movie) factor matrix:

4.1 Select a sample from scoring matrix randomly;

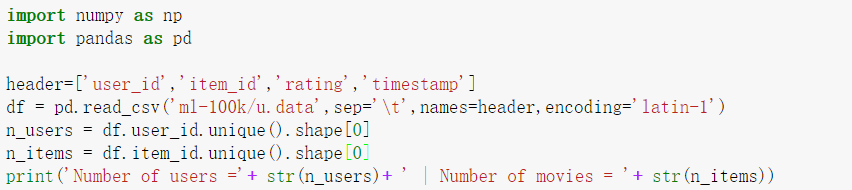
4.2 Calculate this sample's loss gradient of specific row(column) of user factor matrix and item factor matrix; 4.3 Use SGD to update the specific row(column) of and ;

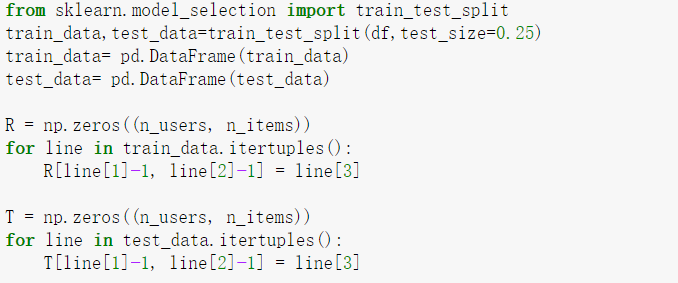
4.4 Calculate the on the validation set, comparing with the of the previous iteration to determine if it has converged.

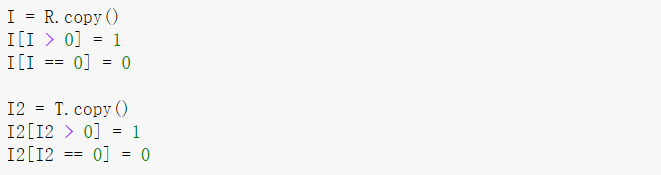
1. Repeat step 4. several times, get a satisfactory user factor matrix and an item factor matrix , **Draw a curve with varying iterations**.
2. The final score prediction matrix is obtained by multiplying the user factor matrix and the transpose of the item factor matrix .

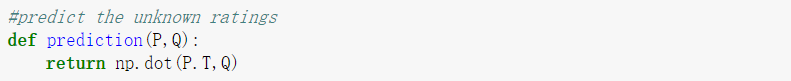
## Implementation

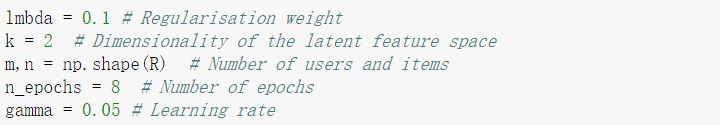
**Source code:**



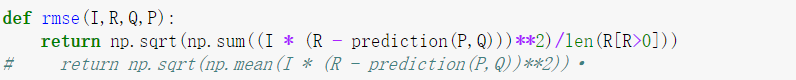


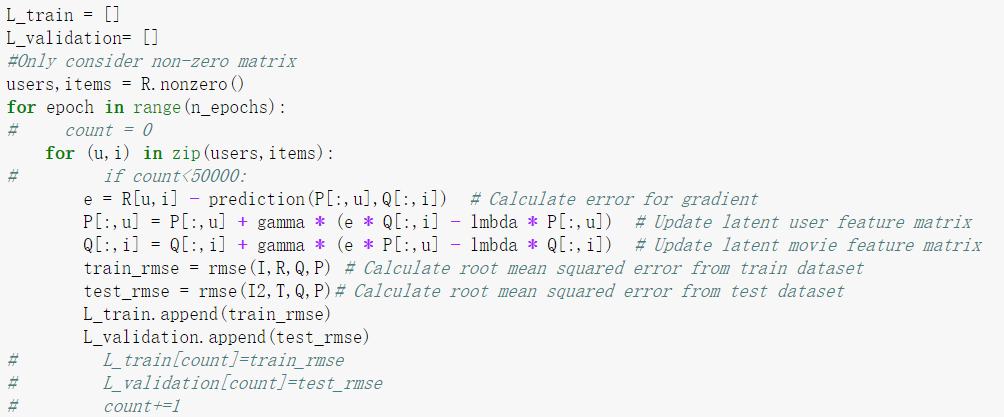


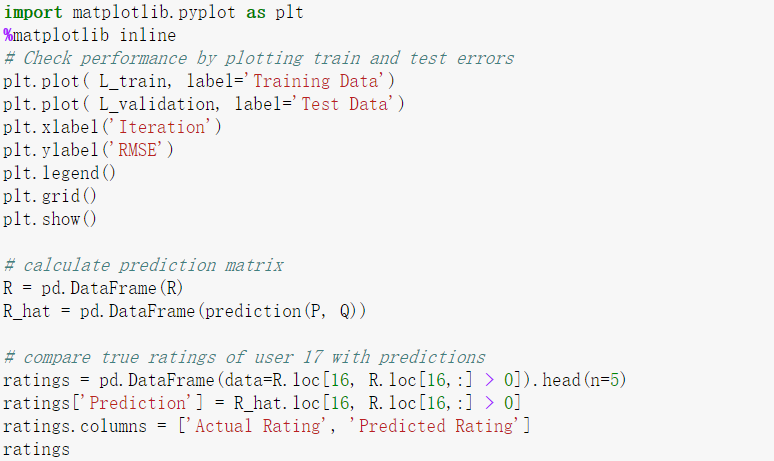


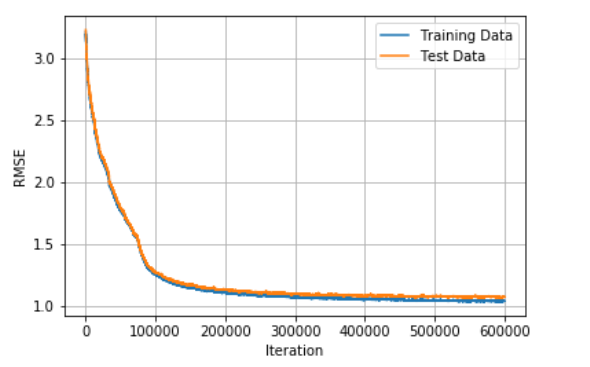


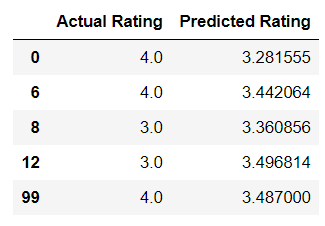












# conclusion

In this experiment we chose to use SGD method to complete the experiment, we encountered some problems, but in the end are discussed and access to relevant information to solve, by setting the parameters again and again, and finally we get a relatively good Convergence and loss.

Through this experiment, we see the SGD algorithm for the versatility of machine learning, but also deeply understand his importance. We understand the way to implement the recommender system, mastered the principle of matrix decomposition, and we gained a lot through this experiment.

1. [↑](#footnote-ref-1)