Recommender System

Asmae Tounsi, Lukas Moser, Deepak Karkala
Ecole Polytechnique Federale de Lausanne, Switzerland

***Abstract*—**

I. Introduction

II. Methodology

A. Brief description of data

B. Baseline methods

1) Global Mean:

2) User Mean:

3) Item Mean:

C. Pre-processing

1) Pre processing method 1:

2) Pre processing method 2:

D. Models and Methods

1) Matrix factorisation: Given the items $\mathbf{d = 1, 2 ... D}$ and the users $\mathbf{n = 1, 2 ... N}$, let $\mathbf{X}$ be the $\mathbf{DxN}$ data matrix containing all the rating entries.

Matrix factorization models map both users and items to a joint latent factor space of dimensionality K, such that user-item interactions are modeled as inner products in that space. The aim is to determine the two matrices $\mathbf{W}$, $\mathbf{Z}$ such that the data matrix $\mathbf{X}$ is approximated by

$$X \approx WZ^T \tag{1}$$

where $\mathbf{W}$ and $\mathbf{Z}$ are matrices of dimensions $\mathbf{DxK}$ and $\mathbf{NxK}$ respectively and $\mathbf{K} << \mathbf{D}, \mathbf{N}$. Each row of $\mathbf{W}$ and $\mathbf{Z}$ is the feature representation of a movie and a user respectively.

In general, the matrix $\mathbf{X}$ is very sparse. Let Ω be the indices of the observed ratings of the input matrix $\mathbf{X}$. The objective is to determine the matrices $\mathbf{W}$ and $\mathbf{Z}$ such that the following cost function is minimised.

$$min_{W,Z} L(W, Z) = \frac{1}{2} \sum_{(d,n \in \Omega)} [x_{dn} - (WZ^T_{dn})]^2 \tag{2}$$

Further in order to avoid over-fitting, we can use regularisation and penalise arbitrarily large entries in $\mathbf{W}$ and $\mathbf{Z}$. In this case, we will minimise the following cost function,

$$\frac{1}{2} \sum_{(d,n \in \Omega)} [x_{dn} - (WZ^T_{dn})]^2 + \frac{\lambda_w}{2} \parallel W \parallel^2_{Frob} + \frac{\lambda_z}{2} \parallel Z \parallel^2_{Frob} \tag{3}$$

where λ_w, $\lambda_z > \mathbf{0}$ are scalars.

The above minimisation can be done using multiple methods. In this work, we use Stochastic Gradient Descent, Alternating least squares and the Coordinate descent methods to determine the matrices $\mathbf{W}$ and $\mathbf{Z}$.

- Stochastic Gradient Descent In this method, for each rating in training set, we predict $\mathbf{x_{d,n}}$ and compute the prediction error as

$$e_{dn} = x_{dn} - (WZ^T)_{dn} \tag{4}$$

- Alternating Least Squares
- Coordinate Descent

E. Results

1) Baseline performance:

2) Matrix factorisation with SGD: (Plot cross validation results)

- Number of features: K

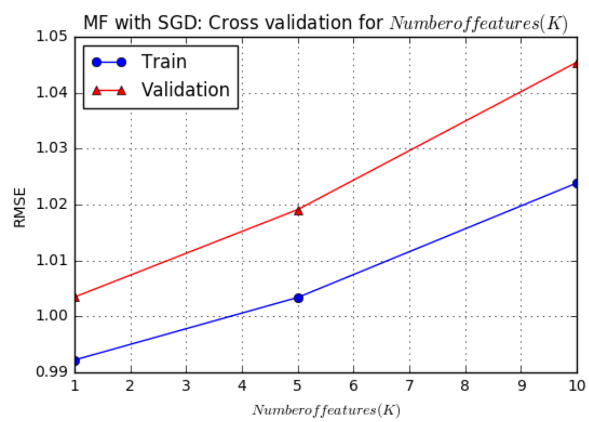

Figure 1. Cross validation: Number of features in Matrix factorisation with SGD.

- Regularisation parameter: λ_{user}
- Regularisation parameter: λ_{item}
- Learning rate γ

3) Matrix factorisation with ALS: (Plot cross validation results)

- Number of features: K
- Regularisation parameter: λ_{user}
- Regularisation parameter: λ_{item}

4) Results with other method:

5) Compare results for SGD, ALS, other method:

III. Discussion

IV. Summary

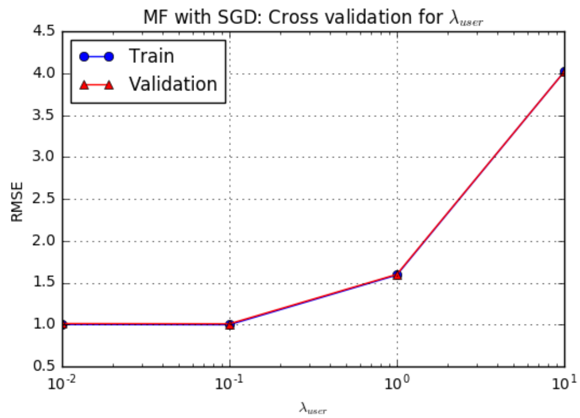

Figure 2. Cross validation: λ_{user} in Matrix factorisation with SGD.

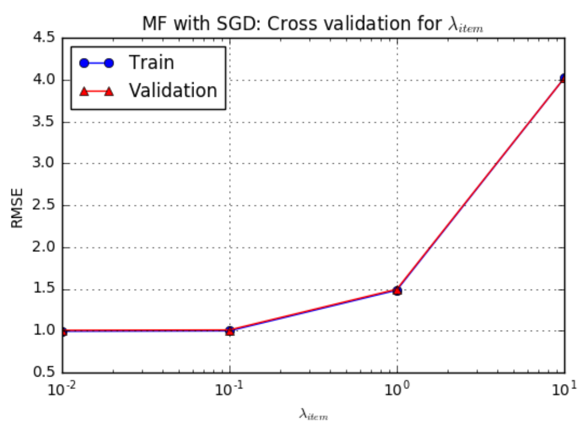

Figure 3. Cross validation: λ_{item} in Matrix factorisation with SGD.