

# Compsci 753 Assignment 4

## Recommender Systems

Latent Factor Modeling with and without bias terms on user-business review data.

Chase Robertson

UPI: crob873

ID: 686249907

## Tasks

### Task 1

Estimate global bias  $b_g$ , user specific bias  $b_i$ , and item specific bias  $b_j$  on training data. Report global bias, and user/item specific bias estimates for one example user and business.

- First, all unique user and business ID's are collected from all datasets.
- Then, global, per-user, and per-business biases are estimated by rating means.

```
Global bias: 3.83  
Bias of user "b4aIMeX0x4cn3bjtdIOo6Q": -0.46  
Bias of business "7VQYoXk3Tc8EZeKuXeixeg": -0.05
```

- The user chosen tends to review below the global mean of reviews by half a star. The selected business is quite close to having an average mean review.

### Task 2

Train a Latent Factor Model without bias, with  $k=8$  factors, for 10 epochs, with learning rate 0.01 and regularisation 0.3, reporting RMSE for each epoch.

- User and business  $k$ -length factor vectors are initialised from the standard normal distribution.
- For each epoch, a rating prediction is calculated with the dot product of each training data point's user and business factor vectors. Squared error is then calculated and used to update factors by the error gradient.

```
Epoch 0 RMSE: 3.3828  
Epoch 1 RMSE: 1.6362  
Epoch 2 RMSE: 1.2519  
Epoch 3 RMSE: 1.1775  
Epoch 4 RMSE: 1.1530  
Epoch 5 RMSE: 1.1413  
Epoch 6 RMSE: 1.1347  
Epoch 7 RMSE: 1.1303  
Epoch 8 RMSE: 1.1273  
Epoch 9 RMSE: 1.1252
```

- There is a nicely exponential decline in RMSE over epochs, with 10 actually seeming an ideal stopping point, by chance.

### Task 3

Report RMSE on validation set of LFM trained with each  $k$  in  $\{4,8,16\}$ . Choose the model with best RMSE and report RMSE on test set.

```
LFM k=4 validation RMSE: 1.66596
LFM k=8 validation RMSE: 1.69591
LFM k=16 validation RMSE: 1.67299

Best LFM (k=4) test RMSE: 1.75655
```

- The validation performance is quite similar across models. The test RMSE is also not very good - predictions are nearly 2 stars away from their true value.

## Task 4

Add bias terms to LFM, initialising with the estimated bias from Task 1. Train a model like that in Task 2, reporting RMSE of each epoch and specific bias of a single user and business.

- Initial bias estimates, addition of bias to rating prediction calculation, and user/business bias gradient updates are added to the previous LFM training code.

```
Epoch 0 RMSE: 1.5420
Epoch 1 RMSE: 1.1150
Epoch 2 RMSE: 1.0829
Epoch 3 RMSE: 1.0720
Epoch 4 RMSE: 1.0667
Epoch 5 RMSE: 1.0635
Epoch 6 RMSE: 1.0615
Epoch 7 RMSE: 1.0600
Epoch 8 RMSE: 1.0589
Epoch 9 RMSE: 1.0581
```

- RMSE started at a much lower value now that bias is included - it seems that starting with reasonable bias estimates makes quite a big impact upfront.

```
Global bias: 3.83
Bias of user "b4aIMeX0x4cn3bjtdIOo6Q": -0.30
Bias of business "7VQYoXk3Tc8EZeKuXeixeg": 0.07
```

- The selected user and business bias estimates are not too far off from their naive estimates after training, but have changed somewhat.

## Task 5

Report RMSE on validation set of LFM trained with bias on each  $k$  in  $\{4,8,16\}$ . Choose the model with best RMSE and report RMSE on test set.

- I've chosen to default to a user/business bias of zero, if the user/business does not already have a bias estimate.

```
LFM with bias k=4 validation RMSE: 1.25018
LFM with bias k=8 validation RMSE: 1.26316
LFM with bias k=16 validation RMSE: 1.28440

Best LFM with bias (k=4) test RMSE: 1.2335
```

- The test RMSE of the best model with bias terms included is much better than the best without bias from Task 3. Adding bias led to a nearly 50% improvement, an impressive gain in performance for a relatively minor change. The model with  $k=4$  seems best now, though not by much.