**Setup**

This project aimed to exploit the user artist interactions and side information contained in the lastfm-2k dataset (Cantador, Brusilovsky and Kuflik, 2011) to implement a recommender system using matrix factorisation. It was hypothesised that methods encoding the user-artist direct interactions as well as the user and artist side information would produce better recommendations than those only using the former.

As proposed in stage 1, pseudo ratings were derived from log transforming the user artist listen interactions. This allowed the matrix factorisation methods to take advantage of the log normally distributed data with a span of 0 to approximately 13.

Accuracy of recommendations can come from two places in matrix factorisation, that is, either directly from the magnitude of errors in the explicit values of the user artist interaction matrix that are being predicted or from the categorical differences in the generated recommendation compared to the implicit artist listens in the held out set. For this project the former was evaluated using RMSE and the latter using Recommender Precision and Recall. As it was relatively computationally expensive generate recommendations for every iteration of hyperparameter tuning, during this step RMSE would be accuracy metric used, while both metrics were to be used when evaluating on test data. For reliability the RMSE between all model variants would be significance tested with an F-test under the assumption of asymptotic normality. Computational Efficiency was measured using system time per fitting procedure.

A very large proportion (60.6%) of the artist cardinality only contains one users listens. For initial pre processing the user artists interaction matrix was filtered to only contain artist with at least two recommendations.

**Approach**

The models tested were Traditional Matrix Factorisation, Collective Matrix Factorisation, Item side collective matrix factorisation, user side collective matrix factoriation and a user bias baseline model. The specific mathematical formula of these methods are found in appendix 1. The ablation test between the inclusion of side information was chosen to provide insight into whether user or artist encoded more useful information for reconstructing the user-artist matrix.

The side information datasets had to be pre-processed into matrices to fit in the recommender reconstruction term. To achieve this, the user side information, in the form of a social adjacency list, was transformed into an adjacency matrix. Similarly, the artist tag data frame was transformed in a matrix using one-hot-encoding on the tags.

Due to the volume of models fitted, grid search was employed over global hyper parameters per model, rather than per parameter. Of chief interest were the regularisation rate over all minimised norms, the number of latent factor parameters and whether to include user and item biases. This allowed for the same grid to be tested across all Matrix Factorisation models.

**Results**

One of the hyper parameters tuned was whether to use dimension reduction on the user side and item side information matrices prior to matrix factorisation. The two matrices were independently standardised using the training data and fitted and transformed using PCA, with proportion of variance explained set to 95%. Without further hyperparameter tuning the collective MF model was fit with and without PCA performed on the two side matrices. The two accuracy measures were worse for the PCA reduced side matrices and training time was significantly longer (364.31 secs vs 8.59 secs). As a result of the considerably longer training time as well as there reduction in accuracy measure, PCA wasn’t implemented for any further hyperparameter tuning.

**Conclusions**

**References**

**Appendix 1**

Model formulations:

1. Traditional Matrix Factorisation (Cortes, 2020):
2. The Collective Matrix Factorisation method (Cortes, 2020):

1. Item side Collective Matrix Factorisation:
2. User side Collective Matrix Factorisation:
3. Baseline Bias Matrix

Where:

, and are column means for the interaction, user side and item side matrices respectively.  
,,,are the factor matrices to learn, the former two learnt across all least square terms.  
,are the item bias vector and user bias vectors respectively and are the penalty norms to regularise