# Sparse Estimation of Movie Preferences via Constrained Optimization

Alexander Anemogiannis, Ajay Mandlekar, Matthew Tsao
(anemogia, amandlek, mwtsao)@stanford.edu

## Introduction

- Traditional matrix factorization and collaborative filtering (CF) techniques learn latent user and movie representations using observed user-movie ratings. These can be used to estimate a user's preference for a movie that he hasn't rated.
- Although these models work well for recommendation, the latent factors are uninterpretable, and the model cannot make recommendations for a movie not in the database.
- We develop methods for estimating *interpretable* user preferences that characterize users in an intuitive manner.
- These methods can be used to recommend out-of-sample movies and gain a detailed understanding of the user base.

## Data

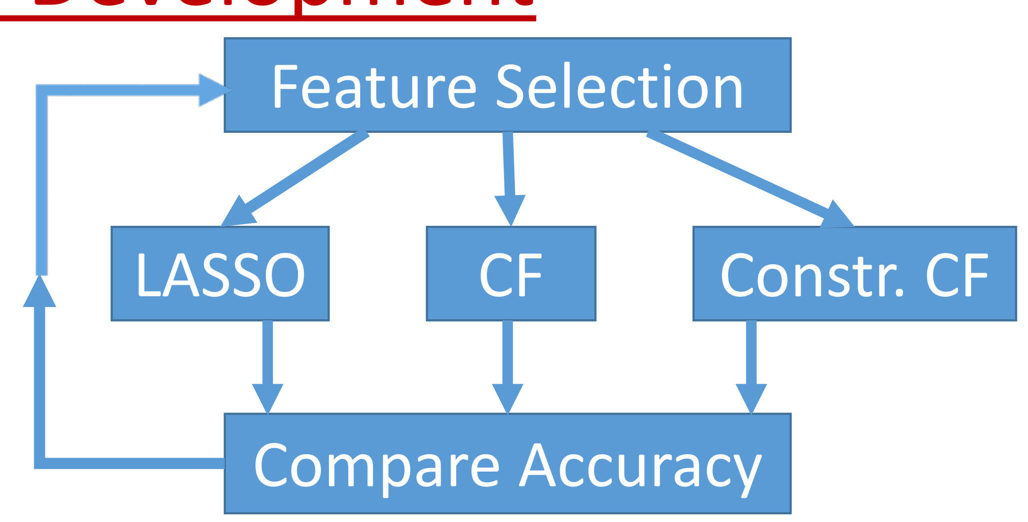
- MovieLens 20m dataset - 20 million ratings of 27,000 movies by 138,000 users
- Each rating is between 0 and 5

## Project Overview

### Feature Selection

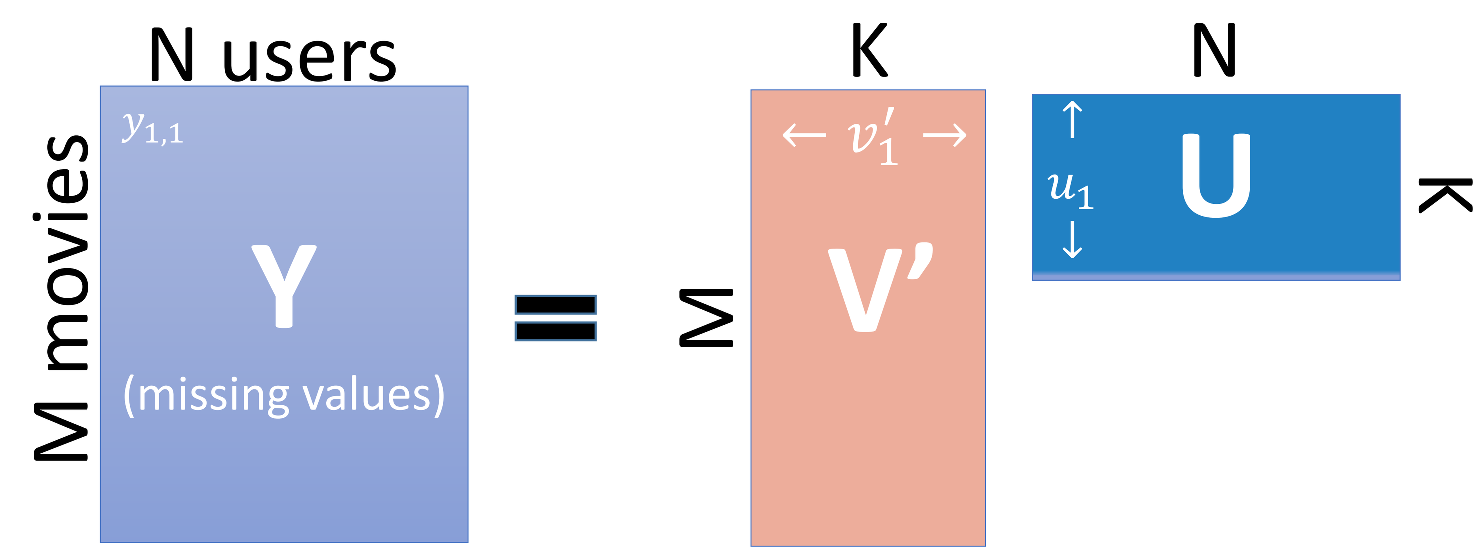
- IMDB provides genre, runtime, IMDB score, lead actors, directors, and year of release.
- The feature mapping $\varphi : \mathbb{N} \to \mathbb{R}^k$ takes as argument a movie ID and outputs a "indicator" feature vector where each entry corresponds to a specific property
- For example, the "action" feature would be 1 if the movie falls in the action genre and 0 otherwise.

### Model Development




## Problem Setup

### Matrix Completion




$$\underset{U,V,a,b}{\arg\min}\ \frac{\lambda}{2}\left(\|U\|_F^2 + \|V\|_F^2\right) + \sum_{(i,j)\in S} \left(Y_{i,j} - (u_i' v_j + a_i + b_j)\right)$$

### LASSO

$$\min_{u \in \mathbb{R}^K} \|y - Vu\|_{\ell 2}^2 + \lambda \|u\|_{\ell 1}$$

## Numerical Results

### Interpretability

- Below is part of the user preference vector belonging to the user with the highest average rating for movies tagged with the "sports" genre
- Notice how CF spreads weight across many features which obscures the user's preference

| | Action | Comedy | History | **Sports** | Western |
|---|---|---|---|---|---|
| LASSO | 0 | 0 | 0 | **1.0** | 0 |
| Constr. CF | 0.116 | 0.128 | 0.092 | **0.171** | 0.022 |

### Rating Holdout (75%/25%)

| | SVD (8-NN) | Constr. SVD | LASSO | Mean |
|---|---|---|---|---|
| Train RMSE | **0.697** | 0.762 | 0.823 | 0.960 |
| Test RMSE | **0.774** | 0.839 | 0.942 | 0.968 |

### Movie Holdout (95%/5%)

| | SVD (8-NN) | Constr. SVD | LASSO | Mean |
|---|---|---|---|---|
| Train RMSE | **0.695** | 0.768 | 0.833 | 1.044 |
| Test RMSE | 0.995 | 0.960 | **0.931** | 1.052 |

## Results (Discussion)

- We expected our RMSE to be larger in the constrained SVD and LASSO methods since they restrict the degrees of freedom in the model.
- For reference, other works were able to achieve test errors of 0.8152 [2] and 0.8748 [3] on the 1 million rating MovieLens dataset, which are comparable to our results for SVD
- As expected, restricting our models to interpretable features in the constrained SVD approach and both interpretable features & sparsity in the LASSO approach gave higher test RMSE when holding out user-movie rating examples.
- In the movie holdout scenario, LASSO outperforms the other approaches. We believe the sparsity discourages overfitting (as observed in it's higher train RSME), which helps it make better recommendations for out-of-sample-movies.
- LASSO is also very interpretable due to the sparsity of learned user preferences

## Future Work

Future work in in this area could focus on the following ideas.

- Refinement of the feature space, which includes adding features, transforming features, and compressing features, can improve the performance of our models.
- Finding a richer data set would also help speed up the testing and exploration process.
- Learning a non-linear mapping that can transform latent factors into interpretable factors can also improve the recommendation accuracy.

## References

[1] Y. Yue, "EE/CS 155 Lecture 13: Latent Factor Models & Non-Negative Matrix Factorization", California Institute of Technology, 2015-2016

[2] Nathan N. Liu , Evan W. Xiang , Min Zhao , Qiang Yang, Unifying explicit and implicit feedback for collaborative filtering, Proceedings of the 19th ACM international conference on Information and knowledge management, October 26-30, 2010, Toronto, ON, Canada [doi>10.1145/1871437.1871643]

[3] N. D. Lawrence R. Urtasun "Non-linear matrix factorization with Gaussian processes" Proc. Int. Conf. Mach. Learn. pp. 601-608 2009.

[4] B. Marlin, "Collaborative Filtering: A Machine Learning Perspective" M.S. thesis, Dept. of CS, University of Toronto, Toronto, CA, 2004.