CF Tutorial

Some Notes on Collaborative Filtering

Ralph Schlosser

https://github.com/bwv988

February 2017

Overview

- Introduction
- Model-based Collaborative Filtering
- Alternating Least Squares: Intuition

Introduction

- Want to predict user ratings for movies which they haven't rated.
- This can be achieved with Collaborative Filtering.
- CF is not *one* algorithm, it's a broad set of different techniques.
- Here: Model-based approach using Alternating Least Squares.

Users	Movie 1	Movie 2	Movie 3	...
User 1	?	5.9	2.6	...
User 2	1.4	5.8	?	...
User 3	1.5	5.8	?	...
...	...	...	...	...

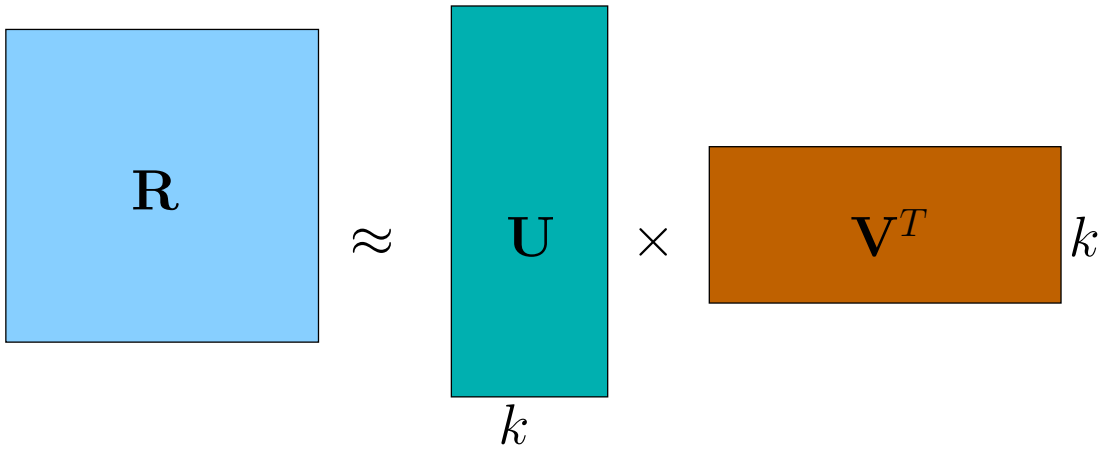
Model-based Collaborative Filtering

- Have a *sparse* and usually large matrix $\mathbf{R} \in \mathbb{R}^{u \times v}$ of user ratings.
- u - Number of users (rows)
- v - Number of movies (columns)
- R_{ij} - Rating of user i for movie j.
- Example:

$$\mathbf{R} = \begin{pmatrix} 2.2 & 2.7 & 7.7 & 2.9 & 3.3 \\ ? & ? & 2.6 & 1.2 & 8.9 \\ 7.0 & ? & 3.5 & 0.7 & 2.1 \\ 9.1 & 0.6 & ? & 1.8 & ? \\ ? & ? & 7.4 & 3.1 & 5.9 \end{pmatrix}$$

Alternating Least Squares: Intuition

- Model-based: Try to uncover latent factors that model the data.
- Can be achieved through approximate matrix decomposition: $\mathbf{R} \approx \mathbf{U} \times \mathbf{V}^T$



- The k columns in $\mathbf{U}$ and $\mathbf{V}^T$ correspond to the latent, i.e. *unobserved* factors.
- ALS: Approximate $\mathbf{U}$ and $\mathbf{V}$ through linear regression.

Links

FIXME: Add links