Recommender Systems User and Item Biases

Professor Robin Burke

Spring 2019

Fixed bias

- Users may have biases independent of the latent factors
  - We discussed high raters vs low raters
- Items may have biases independent of the latent factors
  - Some items are more popular / better than others
- Remember Resnick's algorithm

Global user bias

$$\mu_u + \frac{\sum_{v \in P_u(j)} \mathrm{Sim}(u,v) \cdot (r_{vj} - \mu_v)}{\sum_{v \in P_u(j)} |\mathrm{Sim}(u,v)|}$$

Biased matrix factorization

User bias term

- Rating prediction

$$\hat{r}_{ij} = o_i + p_j + \sum_{s=1}^{k} u_{is} \cdot v_{js}$$

Item bias term

- Prediction error

$$e_{ij} = r_{ij} - \hat{r}_{ij} = r_{ij} - o_i - p_j - \sum_{s=1}^{k} u_{is} \cdot v_{js}$$

Regularization for bias terms

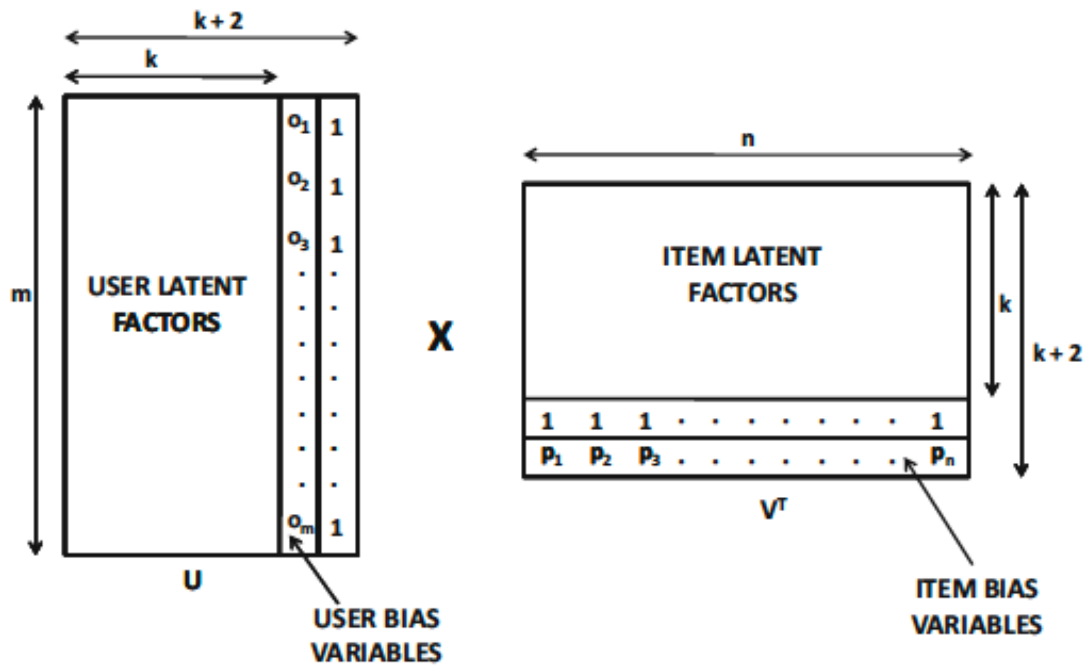
- Objective function

$$J = \frac{1}{2} \sum_{(i,j)\in S} e_{ij}^2 + \frac{\lambda}{2} \sum_{i=1}^{m} \sum_{s=1}^{k} u_{is}^2 + \frac{\lambda}{2} \sum_{j=1}^{n} \sum_{s=1}^{k} v_{js}^2 + \frac{\lambda}{2} \sum_{i=1}^{m} o_i^2 + \frac{\lambda}{2} \sum_{j=1}^{n} p_j^2$$

$$= \frac{1}{2} \sum_{(i,j)\in S} \left(r_{ij} - o_i - p_j - \sum_{s=1}^{k} u_{is} \cdot v_{js} \right)^2 + \frac{\lambda}{2} \left(\sum_{i=1}^{m} \sum_{s=1}^{k} u_{is}^2 + \sum_{j=1}^{n} \sum_{s=1}^{k} v_{js}^2 + \sum_{i=1}^{m} o_i^2 + \sum_{j=1}^{n} p_j^2 \right)$$

But

- This is equivalent to augmenting the factor matrices
  - Adding the constraint that one column of each matrix must always contain 1s
  - This is not how Surprise implements it, though.



LKPy implementation

- Called FunkSVD after Simon Funk's contributions to the Netflix competition
- Problem: it isn't really SVD! More about this later

Why is this a good idea?

- Introduces bias in the technical sense
  - The set of possible models is more constrained
- But in practice this is a useful constraint
  - Esp. for cold-start items and users
- With a small number of ratings
  - Global biases can take over
- Can implement a recommender with just global biases
  - Learned with gradient descent
  - Works surprisingly well in many cases

$$\hat{r}_{ij} = o_i + p_j$$