Baseline Predictors

Baseline Predictors

- Good movies will have higher than average global ratings
- Certain users are more critical than others
- An item above or below average is biased
- A baseline predictor is the sum of the global average plus the bias of the item plus the bias of the user

$$b_{ui} = \mu + b_u + b_i$$

Finding Bias using least squares

- Find **bs that make the below equation as small as possible**

$$\min_{b} \sum_{(u,i) \in K} (r_{(u,i)} - \mu - b_u - b_i)^2$$

Funk SVD

- Regularized SVD
- Gradient Descent is used to reduce the RMSE of the Baseline predictors

$$\begin{bmatrix} 5 & 3 & 0 & 2 & 2 & 2 \\ 4 & 3 & 4 & 0 & 3 & 3 \\ 5 & 2 & 5 & 2 & 1 & 1 \\ 3 & 5 & 3 & 0 & 1 & 1 \\ 3 & 3 & 3 & 2 & 4 & 5 \\ 2 & 3 & 2 & 3 & 5 & 5 \end{bmatrix} = \begin{bmatrix} u_{1,1} & u_{1,2} \\ u_{2,1} & u_{2,2} \\ u_{3,1} & u_{3,2} \\ u_{4,1} & u_{4,2} \\ u_{5,1} & u_{5,2} \\ u_{6,1} & u_{6,2} \end{bmatrix} \begin{bmatrix} v_{1,1} & v_{1,2} & v_{1,3} & v_{1,4} & v_{1,5} & v_{1,6} \\ v_{2,1} & v_{2,2} & v_{2,3} & v_{2,4} & v_{2,5} & v_{2,6} \end{bmatrix}$$

$$\min_{u,v} \sum_{(u,i) \in known} (r_{ui} - u_u v_i)$$

$$f(u_1, \dots, u_N, v_1, \dots, v_M) = \sqrt{\frac{1}{|known|} \sum_{(u,i) \in known} (r_{ui} - u_u v_i)^2}$$

Find all the values of U and V which minimises the RMSE

Minimization with regularization

$$\min_{uv} \sum_{(u,i) \in known} (r_{ui} - u_u v_i) + \lambda(\|u\|^2 + \|v\|^2)$$

Getting Recommendations with Funk SVD

1. *Item factor matrix—Where each column represents a content item described by the latent factors that you calculated.*

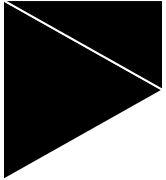
2. *User factor matrix—Where each column represents a user described by the latent factors.*

3. *Item bias—Where certain items are generally considered better—or worse—than others. The bias describes the difference between the global mean and the item's mean.*

4. *User bias—Encompasses different rating scales for different users.*

With these four things, you can calculate a predicted rating for any item for any user using the formula we discussed earlier and shown here:

$$\widehat{r_{u,i}} = \mu + b_u + b_i + q_i p_u$$



Getting Recommendations with Funk SVD

Brute force recommendation calculation - Calculate ratings for each user for each item, sort all the predictions and return the top N items.Compute heavy and takes time

Neighborhoods recommendation calculation - instead of using the original rating data, you can use the factors that you calculated. This means that you're calculating similarities where things are closer and in a smaller dimension, which makes it easier

Faster Implementation

- If you have millions of products and users, Funk SVD will be slower
- Alternating Least Squares (ALS) which isn't as precise as gradient descent, but should work fairly well

Summary

- SVD is a way to do matrix factorization. Imputation of empty cells are required
- Baseline predictors can be used to do the imputation
- Baseline predictors are used to understand the user and item biases
- Funk SVD can use sparse matrix
- We can use Funk SVD along with Baseline predictors
- Gradient descent is used to optimize Funk SVD