Project 1:
Collaborative Filtering

Alejandro Jorba, Luka Lafaye de Micheaux, Newman Chen

Problem: Collaborative Filtering

A recommender system technique to predict users' ratings of objects (in our case, movies), with the help of the known ratings of other users and items. We predict the missing values in a large, sparse, $\mathbb{R}^{mxn}$ dimension matrix, R, which when completed will fully describe the ratings of all users for all objects. This will enable personalized recommendations of new items the users have not yet rated.

Our methods

- Matrix factorization with gradient descent
  - L2 Normalization
  - Bias Terms
  - Momentum
  - Naive rounding + fitting output to input distribution
- Graph method

We also considered using hyperdimensional computing

Evaluated with RMSE and accuracy

Matrix Factorization

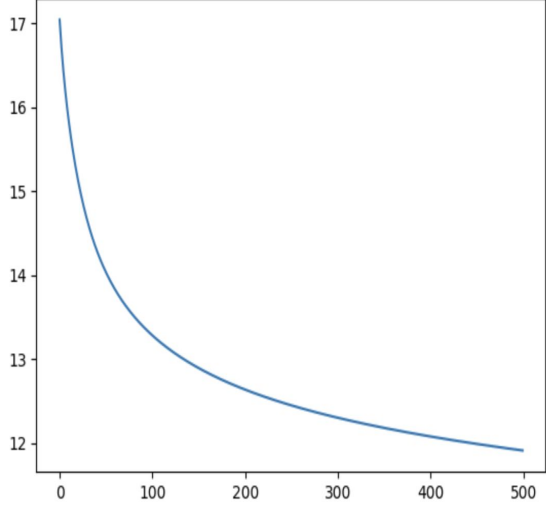
Decomposition into UV^T with L2 normalization and bias terms for U and V

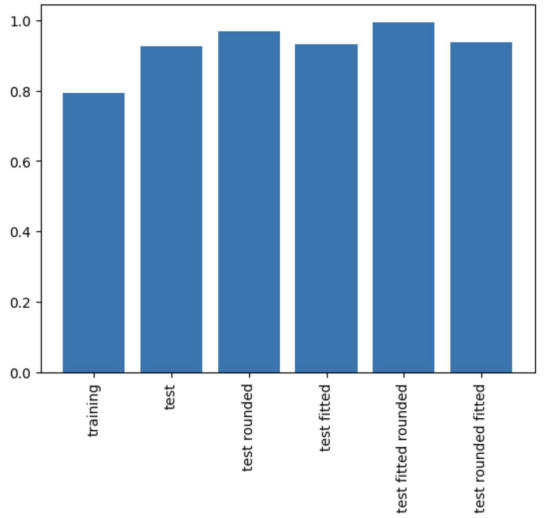
Loss function:

$$J(\boldsymbol{U},\boldsymbol{V}) = 1/2(tr(\boldsymbol{R^T R})-2\ tr(\boldsymbol{R^T U V^T}) + tr(\boldsymbol{V U^T U V^T}) + (tr(\boldsymbol{U^T U}) + tr(\boldsymbol{V^T V})))$$

Update functions:

$$U \leftarrow U\ (1- \alpha \cdot \boldsymbol{\lambda}) + EV \qquad V \leftarrow V(1 - \alpha \cdot \boldsymbol{\lambda}) + E^T U$$





Momentum

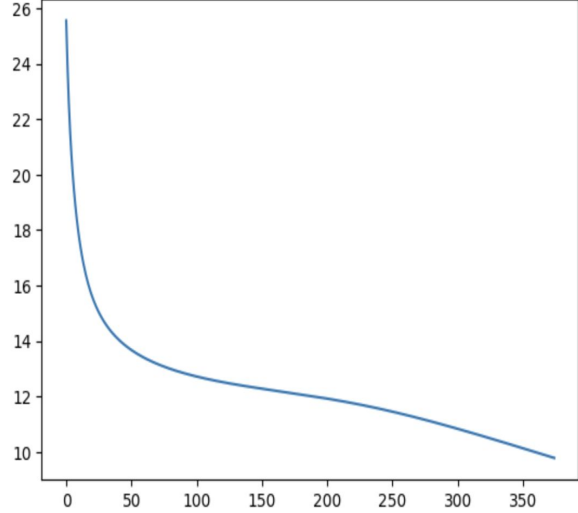
Basic momentum to smooth out gradient and improve speed

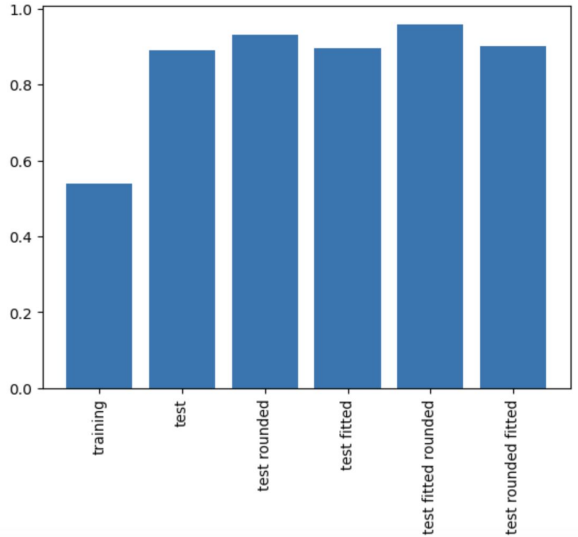
Loss function remains the same

Update functions:

$vU \leftarrow m * vU + \alpha(EV - \boldsymbol{\lambda}U)$ $\quad vV \leftarrow m * vV + \alpha(E^TU - \boldsymbol{\lambda}V)$

$U \leftarrow U + vU$ $\quad V \leftarrow V + vV$

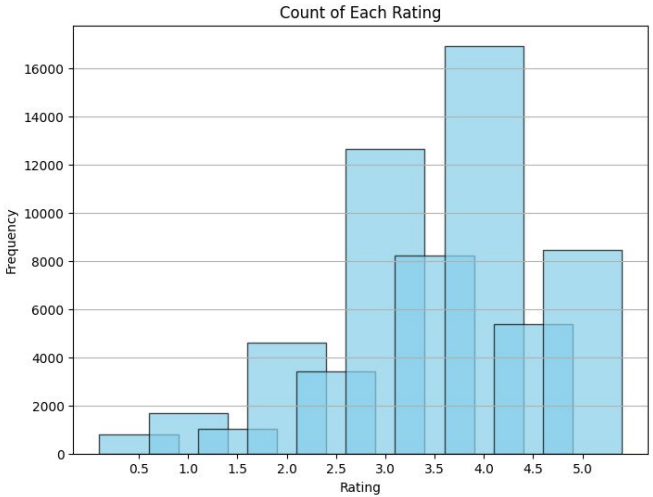


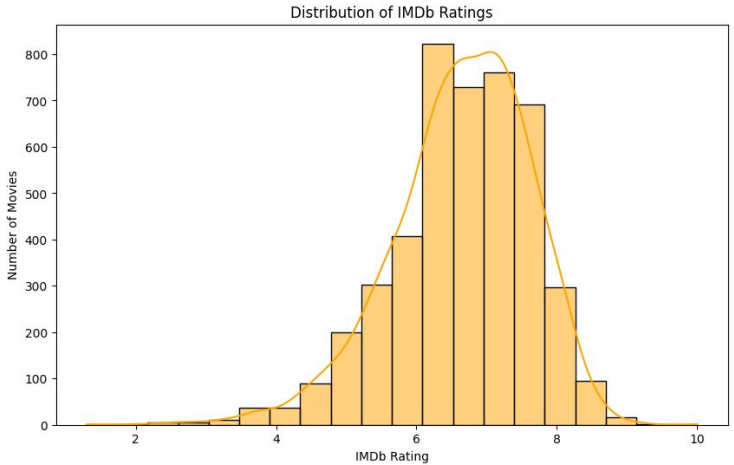


Fitting + Rounding

Shifted and scaled output distribution by mean and standard deviation of input distribution

Rounded to the nearest half integer



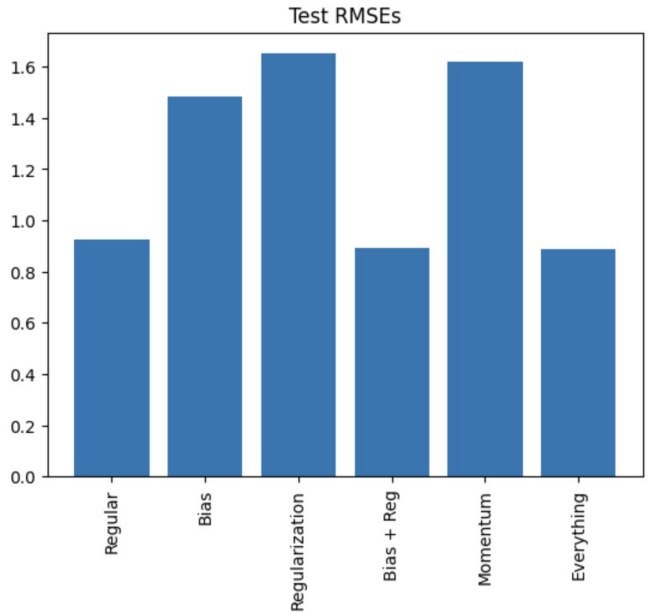


Experimental Results

Our best performing set of hyperparameters for matrix factorization came out to be:

- k=160
- alpha=0.0008
- momentum=0.1
- L2 mu = L2 lambda=0.01
- epochs=375

Which gave us an RMSE of 0.888



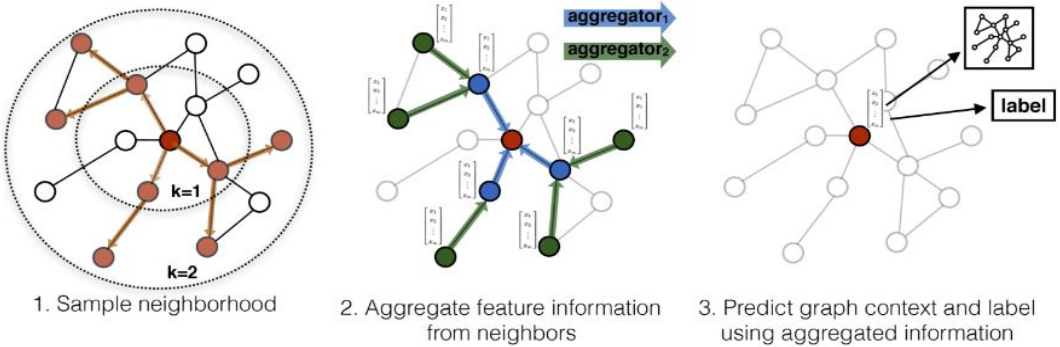
Graph Method

Heterogeneous graph with nodes are **users** and **movies,** and edges as **interactions** (ratings) between users and movies

Initialize nodes as feature vectors

- For users, **one-hot encoding of size 610** (for 610 users in total)
- For movies, **embeddings with movie metadata** from IMDB (recovered using OMDB API)

To actually predict the ratings, we perform an edge prediction task and predict the edge label, the rating a user would give to a movie



1. Sample neighborhood

2. Aggregate feature information from neighbors

3. Predict graph context and label using aggregated information

Training the model

Encoding

Algorithm 1: GraphSAGE embedding generation (i.e., forward propagation) algorithm

Input : Graph $\mathcal{G}(\mathcal{V}, \mathcal{E})$; input features $\{\mathbf{x}_v, \forall v \in \mathcal{V}\}$; depth K; weight matrices $\mathbf{W}^k, \forall k \in \{1, ..., K\}$; non-linearity σ; differentiable aggregator functions $\text{AGGREGATE}_k, \forall k \in \{1, ..., K\}$; neighborhood function $\mathcal{N} : v \rightarrow 2^{\mathcal{V}}$

Output : Vector representations $\mathbf{z}_v$ for all $v \in \mathcal{V}$

$\mathbf{h}_v^0 \leftarrow \mathbf{x}_v, \forall v \in \mathcal{V}$;
for $k = 1...K$ **do**
  for $v \in \mathcal{V}$ **do**
    $\mathbf{h}_{\mathcal{N}(v)}^k \leftarrow \text{AGGREGATE}_k(\{\mathbf{h}_u^{k-1}, \forall u \in \mathcal{N}(v)\})$;
    $\mathbf{h}_v^k \leftarrow \sigma\left(\mathbf{W}^k \cdot \text{CONCAT}(\mathbf{h}_v^{k-1}, \mathbf{h}_{\mathcal{N}(v)}^k)\right)$
  end
  $\mathbf{h}_v^k \leftarrow \mathbf{h}_v^k / \|\mathbf{h}_v^k\|_2, \forall v \in \mathcal{V}$
end
$\mathbf{z}_v \leftarrow \mathbf{h}_v^K, \forall v \in \mathcal{V}$

Decoding

Z = CONCAT(z_user , z_movie)

R' = LeakyReLU(Wd x Z)

z_user and z_movie are the embeddings of the user and movie nodes

Wd is the weight matrix of the linear layer of the decoder

R is the predicted rating of the edge that joins the user and movie

To train, we minimize MSE throughout continual forward/back passes

Experimental Results