Ludwig-Maximilians-Universität München
Institut für Informatik
Prof. Dr. Matthias Schubert
Evgeniy Faerman
Daniyal Kazempour

Munich, 10.01.2019

Big Data Management and Analytics

WS 2018/19

Tutorial 10: High Dimensionality Data

Assignment 10-1 *Game Recommendation with SVD*

Download the Steam Video Game Dataset which is originally from Kaggle: https://www.kaggle.com/tamber/steam-video-games from our website. The dataset contains game purchases of users and how long the users have played their purchased games. We want to implement a simple collaborative filtering based recommender system to recommend games to users. To achieve this, perform the following steps:

1. Load the dataset as a $pandas$ dataframe and get an overview.
2. We will interpret the number of hours a user played a certain game as a rating. Delete the rows corresponding to purchase actions.
3. To reduce sparsity, delete users who have played less than 5 games.
4. Create a rating matrix from your dataframe, where each row corresponds to a user, each column corresponds to a game, and the entries of the matrix are the hours a user played a certain game. You can use the function $pandas.pivot\_table$.
5. Compute the SVD of your rating matrix. You can use the function $numpy.linalg.svd$.
6. Write a function $get\_topk\_similar\_games(game\_name, n\_components, topk)$, which returns a list of the top k similar games for a given game, using the top $n\_components$ components of the SVD. Use the cosine similarity measure. You can use the function $scipy.spatial.distance.cdist$ with parameter $metric =' cosine'$.
7. If you played the game *Fallout 3*, what are the top 20 recommendations for you? Play around with different parameter settings.

Assignment 10-2 *CUR Decomposition*

Given the matrix

	Matrix	Alien	Star Wars	Casablanca	Titanic
Joe	1	1	1	0	0
Jim	3	3	3	0	0
John	4	4	4	0	0
Jack	5	5	5	0	0
Jill	0	0	0	4	4
Jenny	0	0	0	5	5
Jane	0	0	0	2	2

Find the CUR-decomposition of the matrix, when we pick **two** ”random” rows and columns. The columns we pick are *Alien* and *Star Wars* and the rows are the ones of *Jack* and *Jill*.