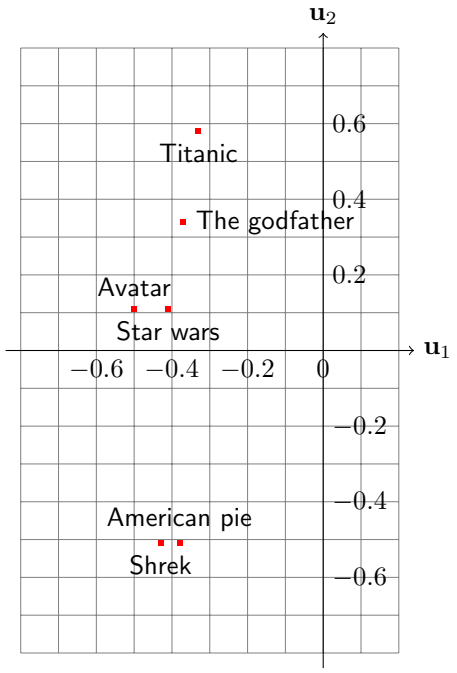
Exercises
Computational Intelligence Lab
SS 2019

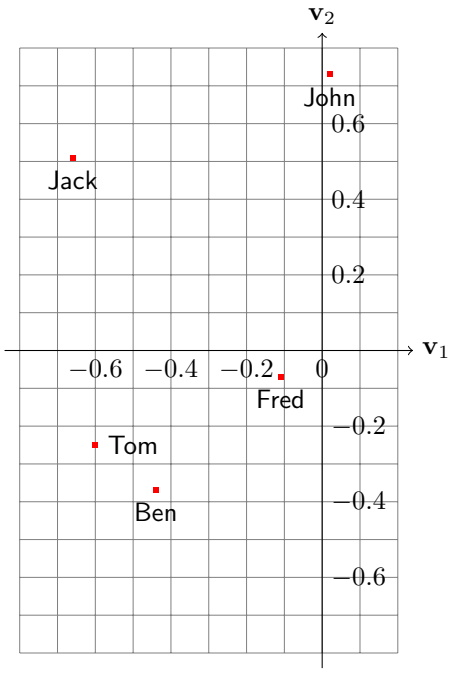
Machine Learning Institute
Dept. of Computer Science, ETH Zürich
Prof. Dr. Thomas Hofmann
Web http://cil.inf.ethz.ch/

Series 2, Feb 27-28, 2020 (SVD)

Solution 1 (SVD Theory):

1. $\mathbf{K} \in \mathbb{N}^{6,6}$. $\mathbf{K}_{ij}$ represents the (dot-product) similarity between different movies based on their received ratings. When normalized, the cross-products amount to the cosine similarity.
2. $\mathbf{L} \in \mathbb{N}^{5,5}$. $\mathbf{L}_{ij}$ represents the (dot-product) similarity between different users based on their given movie ratings. When normalized, the cross-products amount to the cosine similarity.
3. $\mathbf{U}$ is 6 by 6, $\mathbf{D}$ is 6 by 5 and $\mathbf{V}$ is 5 by 5. Note that $\mathbf{U}$ and $\mathbf{V}$ form orthonormal bases of the column and row space of $\mathbf{A}$ respectively.
4. SVD is used to obtain a low-rank approximation of the original matrix. This is used to predict missing ratings from the original matrix as shown in the lecture.
5. One way to do this is to plot the singular values over their index and look for a "knee" where the values suddenly drop. Then all of the small values are discarded. The motivation for this is that the reconstruction errors depends directly on the magnitude of the discarded singular values.
6. The Matrix $\mathbf{U}$ maps users to concepts. At the same time, since $\mathbf{AV} = \mathbf{UD}$, $\mathbf{U}$ contains the movie embeddings (in rows, omitting the weighting by $\mathbf{D}$).
7. The Matrix $\mathbf{V}$ maps movies to concepts. At the same time, since $\mathbf{U}^\top\mathbf{A} = \mathbf{DV}^\top$, $\mathbf{V}$ contains the user embeddings (in rows, omitting the weighting by $\mathbf{D}$).
8. Movie embeddings plotted in the figure. Similar movies (e.g. comedies) get cluster together in the latent embedding space.
9. User embeddings plotted in the figure. Users that prefer similar types of movies (e.g. Tom and Ben) get clustered together.





10. Matrix $\mathbf{D}$ shows the frequency of each movie genre in the input data.

11. Left as exercise. Similar to what was done at points 8) and 9).

12. If two singular values are kept (2-rank approximation), then the approximation error is:

    - Euclidean norm: 7.09
    - Frobenius norm: $7.09^2 + 2.75^2 + 0.67^2$

13. Bob's ratings columns needs to be added to the original matrix $\mathbf{A}$ and the SVD computation needs to be re-run from scratch. Avatar will receive the highest predicted score for user Bob. Alternatively, using an SGD approach where the optimization is done only on known ratings, the additional computation would be performed only on the embedding of the user Bob.