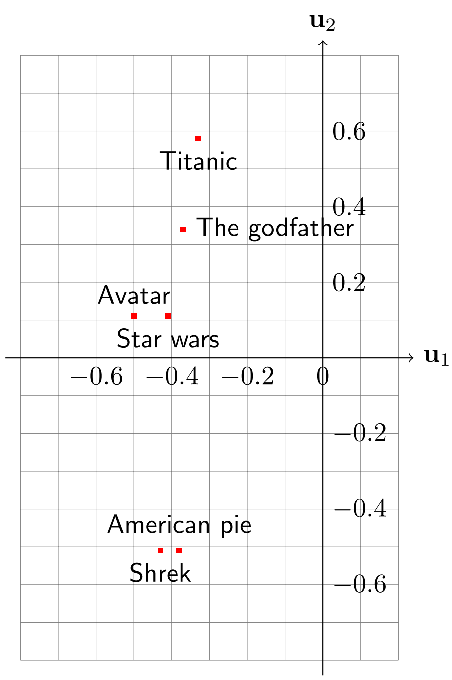
Exercises
Computational Intelligence Lab
SS 2017

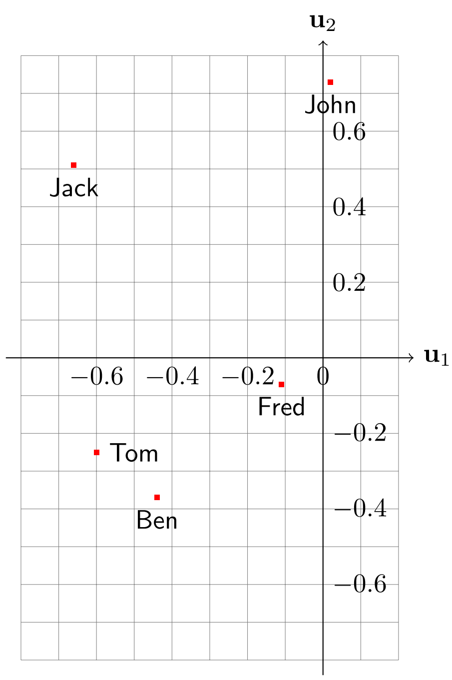
Machine Learning Institute
Dept. of Computer Science, ETH Zürich
Prof. Dr. Thomas Hofmann
Web http://cil.inf.ethz.ch/

Series 3, March 16-17, 2017 (SVD)

Problem 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.
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.
3. Sizes shown in the figure.
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. Plot the singular values similar to the way it was done for PCA. Choose to keep top singular values based on the knee of this graph.
6. Matrix $\mathbf{U}$ contains movie embeddings.
7. Matrix $\mathbf{V}$ contains user embeddings.
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.