

This homework is due October 31, 2016, at Noon.

1. Homework process and study group

- (a) Who else did you work with on this homework? List names and student ID's. (In case of homework party, you can also just describe the group.)
- (b) How long did you spend working on this homework? How did you approach it?

2. Recommendation system

On Saavan's recommendation, the EE16B TAs hang out all the time outside of work. Every Friday night, we watch movies on Netflix and we have been collecting ratings for all the movies we've watched. A sample of this data set is shown below, and gives star ratings (between 1 and 5 stars) for each of the movies we've watched. These data are saved in the file `data_TAs.csv`. Professors Maharbiz and Arcaak sometimes crash movie night, and when they do we also collect their ratings. These data are saved in `data_arcaak.json` and `data_maharbiz.json`.

	Harry Potter	Groundhog Day	Life of Brian	Indiana Jones	Lawrence of Arabia	Jurassic Park	Rushmore	Grease	The Godfather	Spinal Tap	...
Brian	3	2	3	3	3	3	3	3	4	3	...
Rachel	2	3	4	3	2	3	4	3	3	4	...
Sid	2	5	4	2	3	2	4	4	1	4	...
Reia	3	4	3	3	3	3	3	3	3	3	...
Varun	4	1	1	5	4	4	2	2	4	1	...
Saavan	4	2	2	4	3	3	2	3	4	2	...
John	2	5	5	1	2	2	4	4	2	5	...
Edward	3	4	3	3	3	3	3	3	2	3	...
Deborah	4	1	2	4	3	4	2	2	4	2	...
Ana	3	4	3	3	3	3	3	3	2	3	...

In this problem, we will use the SVD to build a system that will predict ratings for unrated movies based on a small sample of rated movies. This will allow us to make customized movie recommendations for the professors, like Netflix does for its viewers. Before starting, we recommend you review the lecture notes to see how to interpret the U , S , and V matrices of the SVD for this problem. You can use the iPython notebook `Recommender_System.ipynb` as a starting point.

- (a) Load the TAs' ratings from the CSV file `data_TAs.csv`.
- (b) Extract the matrix of ratings, compute the SVD of this matrix and plot the singular values.

- (c) You should see that only 3 singular values contribute most of the information in the data set. Truncate the SVD to build a rank 3 model of the data.
- (d) Load the professors' ratings from the JSON files.
- (e) Since the professors have each only seen a small number of movies, we will need to estimate their preferences based on only those movies. Write a function that takes as input the truncated V matrix from the SVD and a list of the movies an individual has watched and returns a reduced V matrix with only the appropriate columns.
- (f) Now use your function from part (e) to write a function that takes a professor's data and returns a least squares estimate of weights U for that professor's preferences. Use this function to predict the weights for both professors.
- (g) Now write a function that predicts how the professors would rate each of the movies. The function should take as input the weights U estimated in part (f) for a given professor and return a vector that looks like a row of the TA data matrix.
- (h) Now use your functions from parts (f) and (g) to write a function that takes a professor's ratings and returns a movie recommendation for that professor (the movie with the largest predicted rating). What movie would you recommend to Professor Arcak? What movie would you recommend to Professor Maharbiz?

3. Redo Problem 1 on the midterm

- (a)
- (b)

4. Redo Problem 2 on the midterm

- (a)
- (b)
- (c)

5. Redo Problem 3 on the midterm

6. Redo Problem 4 on the midterm

- (a)
- (b)
- (c)
- (d)

7. Redo Problem 5 on the midterm

- (a)
- (b)
- (c)

8. Redo Problem 6 on the midterm

(a)

(b)

Contributors:

- John Maidens.