

Advanced Topics in Control 2020: Large-Scale Convex Optimization

Exercise 11: Applications

Goran Banjac, Mathias Hudoba de Badyn, Andrea Iannelli,
Angeliki Kamoutsis, Ilmura Usmanova

May 27, 2020

Date due: June 7, 2020 at 23:59.

Please submit your written solutions via Moodle as a PDF with filename `Ex11_Surname.pdf`, replacing `Surname` with your surname. Include any code as separate files (not inside the PDF itself), and please don't zip anything.

1 Problem 1

Thanks to your incredible performance in de-noising the distress signals Mathias sent out, the ATIC team has managed to inverse-Laplace transform him back into normal space. Unfortunately, while he was still transcending space and time, he had acausally watched all English-language media ever made. In order for him to survive the quarantine, he has to branch out into other media. Using your optimization skills, you will now have to recommend him the best possible Japanese anime.

You will find two datasets: `anime_upload_psy_scifi_new_index.csv` contains 2189 different animes, the anime ID, the anime title, and a list of genres that the anime belongs to. All the animes belong to at least one of the Sci-Fi or Psychological genres. The second dataset, `rating_upload_psy_scifi_new_index.csv`, contains data from 200 users of a popular anime streaming website. A user had the option of watching an anime, and then rating it from 1–10. Each row of the dataset contains the user ID, the anime ID, and the rating.

We will model the task of constructing recommendations as a rank-minimization problem. Let X be a matrix whose rows correspond to users, and whose columns correspond to animes. The entry X_{ij} is the i th user's opinion on the anime j . We will be using the nuclear norm as a proxy for the rank of X . Lastly, let $\Omega = \{(i, j) \mid \text{anime } j \text{ is rated by user } i\}$ be the set of ratings that we know. For example, M_{ij} is the i th user's known rating of anime j out of 10. Our optimization problem is thus

$$\begin{aligned} \min_X \quad & \|X\|_* \\ \text{s.t.} \quad & X_{ij} = M_{ij}, \forall (i, j) \in \Omega. \end{aligned} \tag{P1}$$

1. Let M_{ij} be the numerical rating out of 10 that user i assigns to anime j , and solve Problem (P1). What is the resulting recommendation for user 25 for the anime *Pokemon Advanced Generation: Rekkuu no Houmonsha Deoxys* (ID 765)?

Anime	Rating	ID (full dataset)	ID (Sci-Fi dataset)
Steins;Gate	10	2	2
Steins;Gate 0	8	2154	2006
Shinsekai yori	5	29	23
Serial Experiments Lain	10	137	99
Paprika	6	95	70

Table 1: Mathias' anime ratings

- Mathias has watched only a few animes in his life. His ratings for them are in Table 1. Add a new row to X corresponding to Mathias. Given his ratings in Table 1, what are the top 5 animes he should watch that aren't already in Table 1?
- Mathias' otaku friend Sara forced him to watch *Mahou Shoujo Madoka★Magica* (ID 31) with her, and he liked it so much he gave it a 9. Add this rating to your solution to Part (2) (the case with all the data). Do the recommendations for the top 5 animes from part (2) he should watch change? Is this a user-friendly or desired feature? In one or two sentences, discuss any consequences to a anime-streaming business trying to make a recommendation algorithm for their users.
- Notice that all of the animes in Table 1 are of the Sci-Fi genre. Solve problem (P1) using data *only* from animes that are tagged as being Sci-Fi. I've set up the data sets `anime_upload_scifi_new_index.csv` and `rating_upload_scifi_new_index.csv` for this purpose. Note that the animes have new ID's, so that they are numbered sequentially. What are the top 5 Sci-Fi animes Mathias should watch that are not in Table 1? Print out a list of all animes (names, not anime IDs) that are expected for Mathias to rate higher than a 6, and save it as `Surname_recommendations.txt`. Upload this with your solution.
- What happens if you try a binary recommendation strategy, i.e., if a user i rated an anime j 6 or more, then set $M_{ij} = 1$ and zero otherwise. What happens to the recommendations for Mathias compared to part (2)? List one pro and one con of this method. **Hint:** Think about how subjective ratings can be, and how the scales differ from person to person, and also the loss of information by doing this.