Exercise 4

Bookmark this page

Lecture Sequence due Dec 15, 2022 07:30 +08 Completed

Exercise 4

5/5 points (graded)

The company Internet Movies, Inc. has found wide success in their streaming movie business. After many long and successful years of delivering content, they have decided to use machine learning to make their business even more successful. Luckily, they already possess a huge dataset that has grown over years and years of user activity – but they need your help to make sense of it! Answer the following questions

1. Let's start with a simple case. Assume user Alice is a particularly good member and she makes sure to rate every movie she ever watches on the website. What machine learning approach would be better for the company to use for determining whether she would be interested in a new specific movie?

- (●) Supervised
- ( ) Unsupervised

✔

2. Bob, on the other hand, is not that much into ratings. He does watch a lot of movies, but never takes the time to rate them. For users like Bob, which of the following data can the company use to determine potential interest in a specific movie? Check all that apply.

- [x] Metadata of movies: actors, director, genre, etc.
- [ ] Length of the movie
- [x] Popularity of the movie amongst other users
- [ ] User login patterns

✔

3. What machine learning approach should the company use for cases like Bob?

- ( ) Supervised
- (●) Unsupervised

✔

Now that the company has some idea about how to use the data, it's time to design a classifier. Our classifier will be very simple: given a movie and a user, it will classify the movie as either "Good" or "Bad" for this user.

4. Assume all the users of the company have a very simple rule in their movie taste: they like it if Tom Cruise has the lead role. Any other data is mostly irrelevant. However, no one in the company knows about this fact. Which of the following clustering models might be able to detect this rule? Check all that apply.

- [ ] Supervised (label: rating), with data: Director, language, genre
- [x] Supervised (label: rating), with data: Movie length, lead role, director
- [x] Unsupervised, with data: Lead role, movie length, rating
- [ ] Unsupervised, with data: Lead role, genre, director
- [ ] Unsupervised, with data: Number of ratings, lead role

✔

5. Looking at the options they're given, the board members choose to go with a super[illegible]ead role

Calculator

as data. You be … no watches a 3 hour long mo… …l results. You see, if we remove movie length, ..." What can your colleague (correctly) say to convince you? Check all that apply.

- [ ] "we can reduce inter-cluster dissimilarity."
- [x] "we can reduce intra-cluster dissimilarity."