DS2000

11/19-Tues.

Admin

- TA survey (at end of class)

Agenda

1. Recommendation Design
2. Attributes + values
3. Python

} OOP for a data science algorithm

1. Recommendation Design

morning 0 afternoon 1

next: TV recommendations to Carey

promise to watch ≥1 episode of recommendation section w/ best show wins!

data science algorithms:

- sentiment analysis (-1 to +1 on text data)
  bad good
- recommendation system (rec a show, product, account, college, course, website, etc.)
  - ↳ goal ~ make recs based on what we know, like a 

manage expectations not perfect

in ML/OS ~  make recs the way would

rec F1 (for Laney) rec Baking Show (for Laney)

today's test case — for Laney to not on couch and watch over Thanksgiving

recent Agatha All Along — to make rec based on any info

What data to use? shows ppl — assume we have everything! (can't have too much!)

- director
- actor
- writer
- genre
- tone ✓
- user history (shows in common)
- release date ✓
- # seasons ✓
- rating ✓
- search history
- ep length ✓
- age
- cultural identity
- mood

(all are used in real life!)

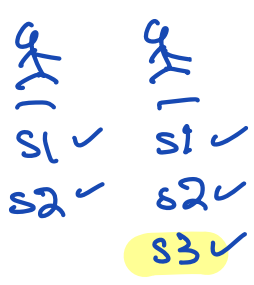
2. Attributes + Values

Data we want → code

Starting point: Laney, Bake Agatha

Show vs.  — 

User history
Shows in common
rec something new one person liked



collaborative filtering

Show details
find similarity between known, new shows

content based filtering

S2 S3 S4 → pick most similar to S1

our implementation: content based!

- stick w/ numeric values

↳ rating
tone (1 sad to 10 happy)
seasons
year
length of ep

bit.ly/sec3_recs