Announcements
- Office hours
- Homework 9 due tomorrow @ 11pm

Outline
- SETs
- Natural language processing
- Sentiment analysis

Natural language processing
- Using a computer to analyze natural (i.e., human) language
- Example use cases
  - Question answering --- given a human-language question (e.g., “Where is Colgate University located?”), determine its answer
  - Natural language generation --- convert data (e.g., map data) into a natural language representation
  - Machine translation --- translating text from one human language to another (e.g., “I love shopping at Ikea.” in Swedish is “Jag älskar att shoppa på Ikea.”)
  - Automatic summarization --- produce a readable summary of a chunk of text
- Usually requires breaking text into its constituent parts and labeling those parts
  - Assume you are given the following snippet of text:
    Donovan’s Pub is located in James C. Colgate Hall. It is open Saturday but not Sunday.
  - Break the text into sentences
    Donovan’s Pub is located in James C. Colgate Hall
    It is open Saturday but not Sunday.
  - Describe an algorithm you could use to break a chunk of text into sentences.
    ■ Split by punctuation marks (e.g., period, question mark, exclamation mark, semicolon)
    ■ Does not work if there are abbreviations!
  - Label each word with its part of speech (noun, pronoun, adjective, verb, adverb, preposition).
    Donovan’s Pub is located in James C. Colgate Hall
    It is open Saturday but not Sunday.
  - Describe an algorithm you could use to label each word in a sentence with its part of speech.
    ■ Lookup a word in a dictionary to determine its possible parts of speech
    ■ Use the parts of speech of the other words in the sentence to choose one of the options
    ■ Construct a statistical model, called a hidden Markov model (HMM), using a large corpus of hand-labeled text; the model encodes statistical trends such as: an article (e.g., the) is followed by a noun 40% of the time, an adjective 40%, and a number 20%
- Recall: human language is inherently ambiguous => natural language processing is hard!
Sentiment analysis

- Determine the attitude conveyed in a piece of text
- Example use cases
  - Restaurant reviews --- positive or negative?
    - “Definitely the best pizza in town. And the people who work there are kind, patient, and helpful.” (5 stars)
    - “Pizza is bland. The only upside is they stay open till 2 am while Colgate is in [...]” (1 star)
    - “Slices (The real name) is a decent pizza spot in Hamilton.” (4 stars)
    - “Tasty pizza and best in town.” (4 stars)
  - Product reviews --- what do people think about a product?
  - Public sentiment --- how are people feeling about something?
    - [https://www.csc2.ncsu.edu/faculty/healey/tweet_viz/tweet_app/](https://www.csc2.ncsu.edu/faculty/healey/tweet_viz/tweet_app/)
  - Politics --- what do people think about a candidate or issue?
- Reviews often include a star rating, so sentiment analysis may be less necessary
  - However, there are always reviews that don’t agree with the star rating…
    - 5-stars: I gave this as a gift. I hope they like it.
    - 1-star: Fantastic product!
    - 5-stars: Product arrived quickly.