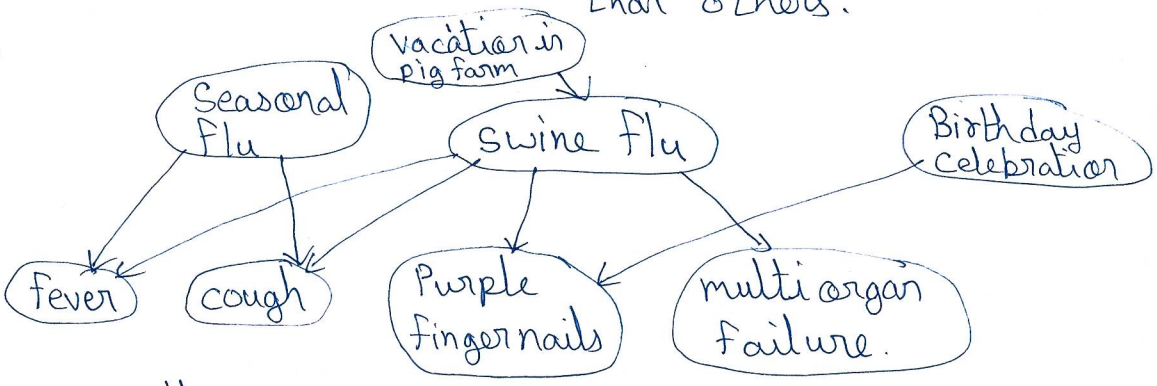
CSE 250A 26-Sept-2019

Lecture 0

Probabilistic Reasoning

Ex. * Medical diagnosis
* Knowledge representation: disease, cure, symptoms
* Modeling uncertainty: some diseases, symptoms more likely than others.



More generally
* How to update beliefs in light of new evidence?
* How do graphs represent correlation, causation, interdependence

Graphical models ⟷ of graph theory and probability theory.

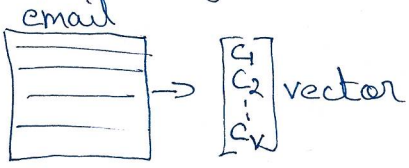
Prediction

Ex: Spam filter

I/P: email messages O/P: {spam, not spam}

* How to represent input?

Simplest way: convert text to fixed length vector of word counts.

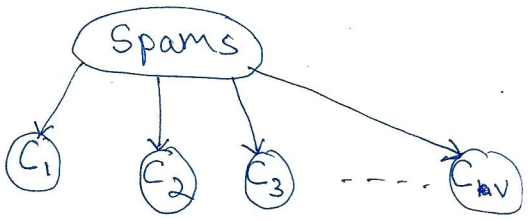


V: vocabulary size

~~a. cow~~

c_i: count of times that i^{th} word in dictionary appear

* Graphical model.



Certain words are more likely to appear in spam.
- How to quantify

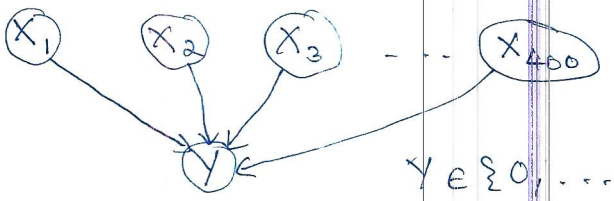
Ex: Character recognition.

I/P: Gray scale image of 20×20

O/P: Label $\{0,1,\ldots,9\}$

* Represent image as $X \in R^{400}$ with one element per pixel.

Graphical model:



Pattern analysis and discovery

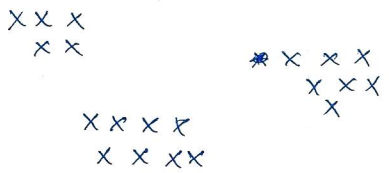
Ex: Topic modelling.

- 100 K documents & organize this very large collection of (unlabelled) documents.

→ More generally, this is a problem of clustering.

How to cluster inputs $\{x_1, x_2, \ldots x_n\}$ when no labels are provided?

Collection of objects. How to group them into 3 groups.

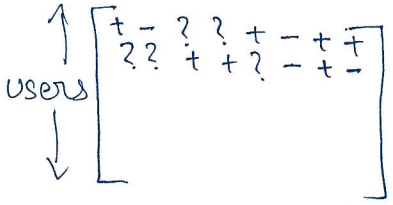


Map inputs to some discrete label $y \in \{0,1,2\}$

Ex: collaborative filtering

- How to build a movie recommendation system?

Collect data set (user item matrix) of movie ratings



* more generally,

how to complete partially observed matrix?
Given some elements, how to infer the rest?

Sequential modeling

* How do we model systems whose 'state' changes overtime or have some altered representation?

Ex: Text (written language)

- 'states' = words.

which sentence is more likely?

- Mary had a little lamb.
- Colorless green ideas sleep furiously.

We can use Markov models for statistical language processing.

Model A

Fully connected to all previous words.

Model B

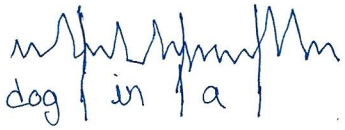
Model A is a richer model but harder to estimate
Model B is wrong (too simple) but easier to work with.

Goal: Find simplest model that does the job.

Ex: Speech (spoken language)

states - words (or syllables or smaller units of speech)

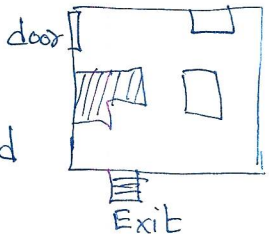
observations: sounds, waveforms



* How do we infer words from waveforms?
  ⇒ Hidden Markov Models for speech recognition.

Planning and decision making

Ex: Robot navigation
2d grid world



states: cells on 2d grid world

actions: Try to move
N S E W

More general question: How can an autonomous agent learn from the experiences from the world?

- Stochastic environment: actions change agent's states, but not deterministically.
- rewards: feed back from environment.
  - delayed rather than immediate rewards.
  - evaluative vs instructive feedback

} Markov decision process, Reinforcement learning

Other embodied agents: self driving cars, drones

Other embedded agents: game playing AIs (eg. Alpha Go)

Theme of class

1) Probability: model of ~~un~~ uncertainty
2) Principle
   Inference as calculations.
   Learning as optimizations.
3) Power vs tractability: How to develop compact representations of complex worlds?