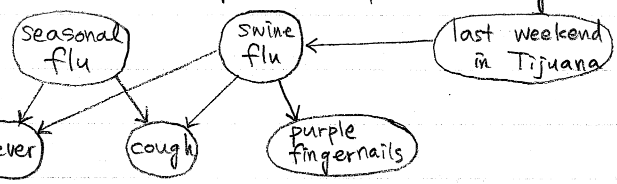
9/23 $\boxed{\text{Probabilistic reasoning}}$

Ex: medical diagnosis

* knowledge representation: diseases cause symptoms.
* Modeling uncertainty: some diseases, some symptoms more likely than others.
* Reasoning: infer diseases from symptoms.

Probability: quantitative, self-consistent framework that captures commonsense patterns of reasoning

$\boxed{\text{graphical model}}$:



How do graphs represent correlation, causation, statistical independence?

Marriage of probability and graph theory.

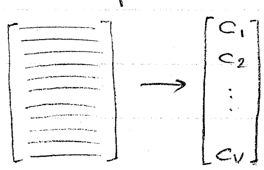
$\boxed{\text{Prediction}}$

Ex: spam filter

input: email message

output: {spam, not spam}

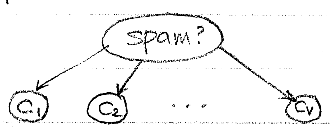
How to represent input? Convert text to vector of word counts:



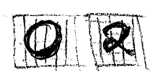
c_i = # times that i^{th} word in dictionary appears in message

V = # entries in dictionary

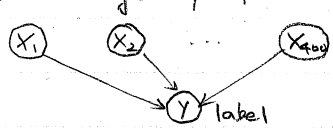
* graphical model



Ex: handwriting recognition

* inputs: grayscale images (say, 20x20) 
* outputs: labels {0,1,2, ..., 9}

Represent image by feature vector $\vec{x} \in \mathbb{R}^{400}$ with one element per pixel.

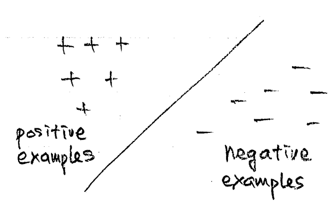


* More generally: classification

inputs: $\{\vec{x}_1, \vec{x}_2, \cdots, \vec{x}_N\}$, $\vec{x}_i \in \mathbb{R}^D$

outputs: $\{y_1, y_2, \cdots, y_N\}$, $y_i \in \{1, 2, \cdots, C\}$

C ↑ #classes



$D=2$

$C=2$

Pattern analysis and discovery

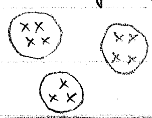
Ex : topic modeling

how to organize large collection of documents?

* more generally, clustering

inputs $\{\vec{x}_1, \vec{x}_2, \cdots, \vec{x}_N\}$, $\vec{x}_i \in \mathbb{R}^D$

How to group inputs when no labels are provided?



: Map inputs $\vec{x} \in \mathbb{R}^D$ to discrete label $y \in \{1, 2, \cdots, C\}$.

Sequential modeling

* How to model systems whose "state" changes over time (or has a similarly extended representation)

Ex : text (written language)

"states" = words

Which sentence is more likely?

1) Mary had a little lamb.

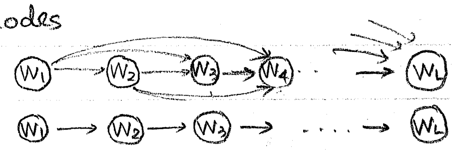
2) Colorless green ideas sleep furiously.

⇒ Markov models for statistical language processing.

Graphical modes

$\boxed{\text{model A}}$

$\boxed{\text{model B}}$



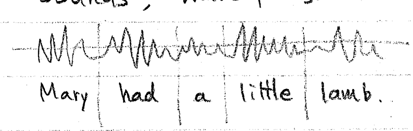
Model (A) is richer but harder to estimate.

Model (B) is wrong but easier to estimate.

Ex : Speech (spoken language)

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

observation = sounds, waveforms.



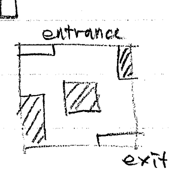
How do we infer words from waveforms?

⇒ hidden Markov models for speech recognition.

Planning and decision-making

* Ex : robot navigation

2d grid world



"States" = cells on 2d grid

actions = attempts to move N, S, E, W.

* noisy dynamics
* rewards = feedback from environment
  - delayed vs. immediate
  - evaluative vs. instructive

More generally: how can autonomous agents learn from experience?

⇒ Markov decision processes, reinforcement learning.

Other "embodied" agents: elevators, helicopters

Other "embedded" agents: game-playing system, spoken dialog system.

Themes of class

1) Probabilistic models of uncertainty
2) Learning as optimizations
3) Power vs. tractable: how to develop compact representation of complex world?
4) Principles vs. heuristics: optimizations, calculations } vs. rules-of-thumb
5) Synergies of AI: inference and learning, perception and action, theory and pratice.

9/28 Motivation

* Modeling of uncertainty
  - Inherent randomness (e.g., radioactive decay)
  - Gross statistical description of complex deterministic world. (e.g., coin toss)
* Probability acts as guardian of commonsense reasoning.
* Many empirical successes: robotics, language, speech, bioinformatics.

Review

* Discrete random variable X

  Domain of possible values $\{x_1, x_2, \cdots, x_m\}$

  Ex: month M $\{m_1 = \text{Jan}, m_2 = \text{Feb}, \cdots, m_{12} = \text{Dec}\}$
* Unconditional (prior) probability $P(X = x_i)$
* Basic axioms: (i) $P(X = x_i) \geq 0$

  (ii) $\sum_i P(X = x_i) = 1$

  (iii) $P(X = x_i \text{ or } X = x_j) = P(X = x_i) + P(X = x_j)$ if $x_i \neq x_j$

  Probs add for union of mutually exclusive events.