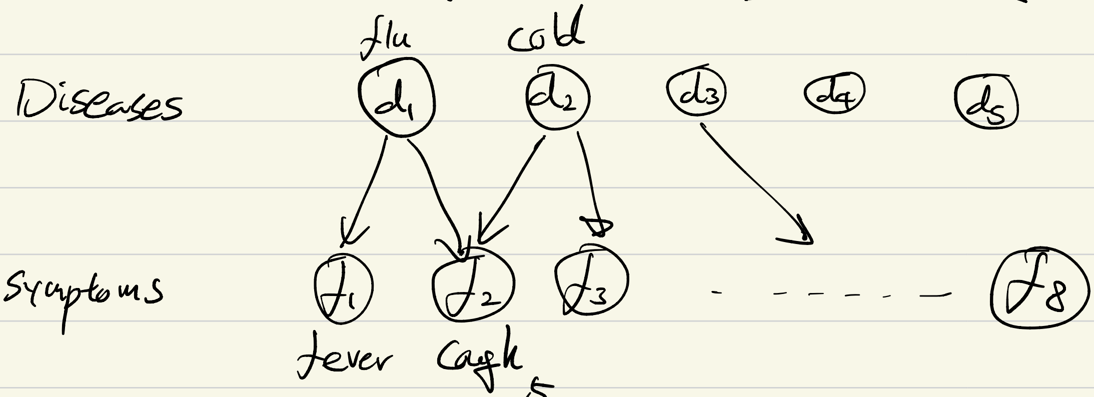
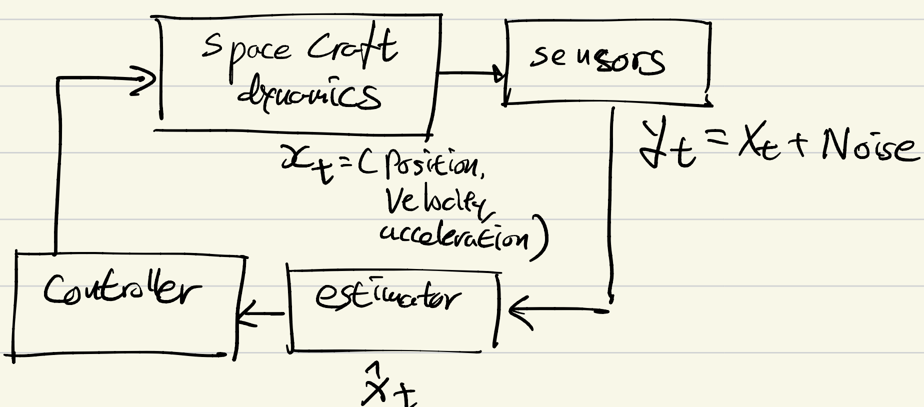
* Probabilistic Graphical Models (PGM) provide a unifying framework to solve inference tasks + efficient solvers

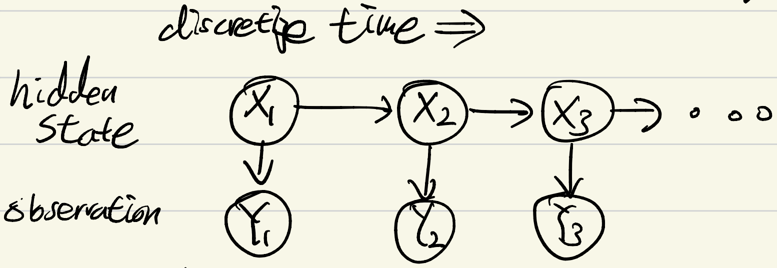
Example 1. medical diagnosis



- $d = [d_1, \dots, d_5] \in \{0,1\}^5$ is the status of which disease the patient
- $f = [f_1 \dots, f_8] \in \{0,1\}^8$
- know the graphical model from historical data
  - the graph $G = ($ ~~set of~~ nodes V, set of edges $E)$
  - their relations $\mathbb{P}(f_1 = 1 \mid d = [1,0,0,0,0])$
- Goal: given a patient with symptoms $f = [1,1,0,0,0,0,0,0]$ what is most likely diagnosis $\arg\max_d \mathbb{P}(d \mid f)$

Example 2. Hidden Markov model for navigation (moon landing)

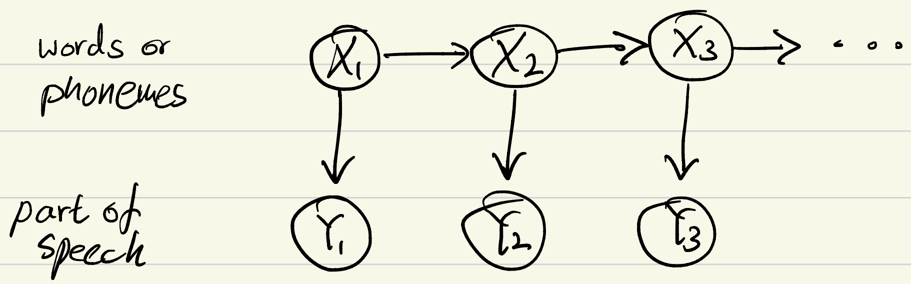




Graph $G(V,E)$ is obvious
relations learned from data

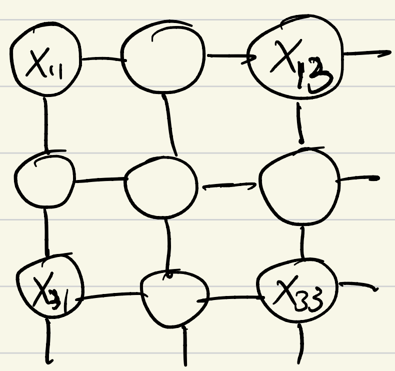
- we know the physical law governing the dynamics & noise statistics
- Goal: given noisy observations of past states, estimate current state. compute $\mathbb{P}(X_3 \mid Y_1, Y_2, Y_3, X_1)$

Example 3. Hidden Markov models for speech recognition



- Baum-Welch algorithm: for learning the graphical model
- Viterbi algorithm: for inferring the most likely sequence

Example 4. Markov Random Field model for super-resolution

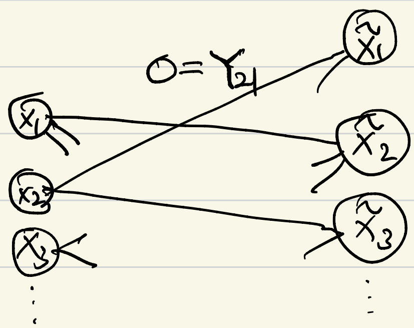


- Goal: given a low-resolution image, infer high-resolution image.
- loopy belief-propagation

Example 5. Peer grading

		$\tilde{X}_1$ student 1	$\tilde{X}_2$ student 2	$\tilde{X}_3$ student 3	← Reliability ...
X_1	student 1 HW0 Problem 1		2		
X_2	student 2 HW0 Problem 1	0		1	
	⋮				

↑ hidden true score



Goal: infer the most likely true scores.

General theme of the course

(we know) Probability distribution over $X = [X_1, X_2, \cdots, X_n]$

given observations $Y = [Y_1, Y_2, \cdots, Y_m]$

$$P_Y(x) \triangleq P(X_1, X_2, \cdots, X_n \mid Y_1, \cdots, Y_m)$$

$$X_i \in \mathcal{X} \text{ with } |\mathcal{X}| < \infty$$

(we want) inference

- Find the most probable realization

$$\hat{x} \in \arg\max_{X} P_Y(X)$$

- Calculate the marginals

$$P_Y^{(1)}(x_1) = \sum_{x_2, \cdots, x_n \in \mathcal{X}^{n-1}} P_Y(x)$$

- Sampling X from $P_Y(X)$

* Major challenge is $n \gg 1$, and brute force inference takes $O(|\mathcal{X}|^n)$

inference can be dramatically efficient by exploiting <u>structure</u>

claim: suppose variables are conditionaly independent.

$$\Leftrightarrow \quad P_Y(x) = P_Y^{(1)}(x_1)\, P_Y^{(2)}(x_2) \cdots P_Y^{(n)}(x_n)$$

then inference takes $O(|\mathcal{X}| \cdot n)$ time.

- Finding most probable realization

$$\hat{x}_1 \in \arg\max_{x_1} P_Y^{(1)}(x_1), \quad \hat{x}_2 \in \arg\max_{x_2} P_Y^{(2)}(x_2)$$

$$O(|\mathcal{X}|)$$

- Calculate Marginals

$$\left.P_Y^{(1)}(x_1) \propto P_Y(x_1, 0, 0, \dots 0)\right\} O(|\mathcal{X}|)$$

- sampling

$$x_1 \sim P_Y^{(1)} \text{ is } O(|\mathcal{X}|).$$

$$\boxed{\text{* When the probability of distribution factorizes we can solve inference efficiently.}}$$

more structure

Complete Conditional Independance

$$P_Y(x) = \prod_{i=1}^{n} P_Y^{(i)}(x_i)$$

Independance ?

Factorization ?

Complete dependence

$$P_Y(x)$$

- Graphical models
  - Markov Random Fields = Undirected Graphical Models
  - Factor Graphs
  - Bayesian Networks = Directed Graphical Models
- Inference on Graphical Models
  - Sumproduct algorithm = Belief Propagation
  - max-product algorithm
- Learning Graphical Models
  - Parameter estimation
  - structure learning
  - Learning with hidden variables.