* Probabilistic Graphical Models (PGM)

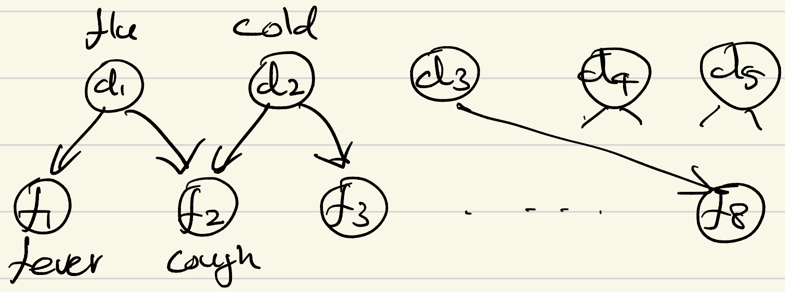
<u>inference tasks</u> + <u>efficient solvers</u>

Example 1.

Medical Diagnosis

Diseases

symptoms



$d = [d_1, d_2, d_3, d_4, d_5] \in \{0,1\}^5$

$f = [f_1 \cdots \quad f_8] \in \{0,1\}^8$

Know the <u>graphical Model</u> from historical data.

- graph $G = ($ nodes V, edges $E)$
- their relations $\mathbb{P}(f_1 = 1 \mid d = [1,0,0,0,0])$

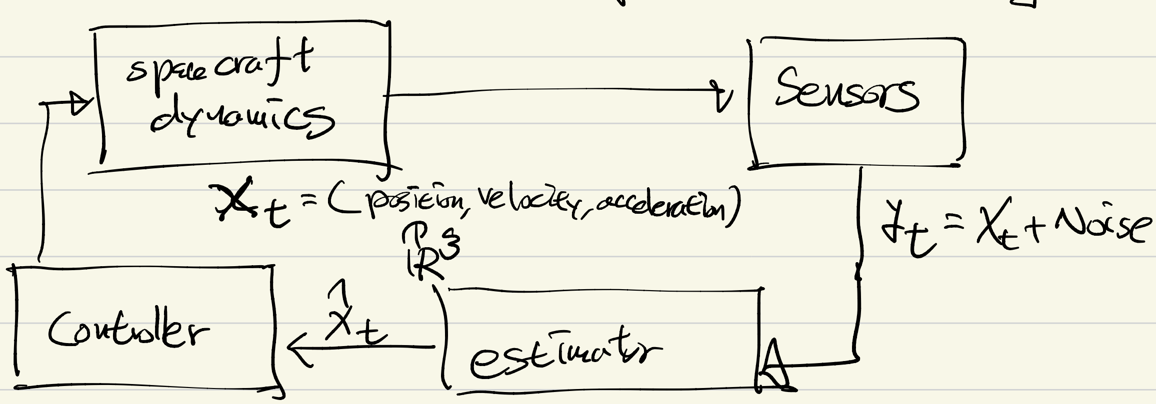
. Goal: given a patient with symptoms $f = [1,1,1,0,0,0,0,0]$

what is the most probable diagnosis: $\arg\max_{d \in \{0,1\}^5} \mathbb{P}(d \mid f)$

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

$x_t \in \mathcal{X}$

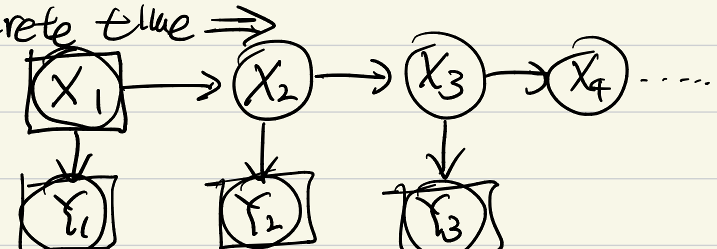
$|\mathcal{X}| = \infty$



discrete time ⇒

hidden state

observation



- $G(V,E)$ is given.
- relation learned from data

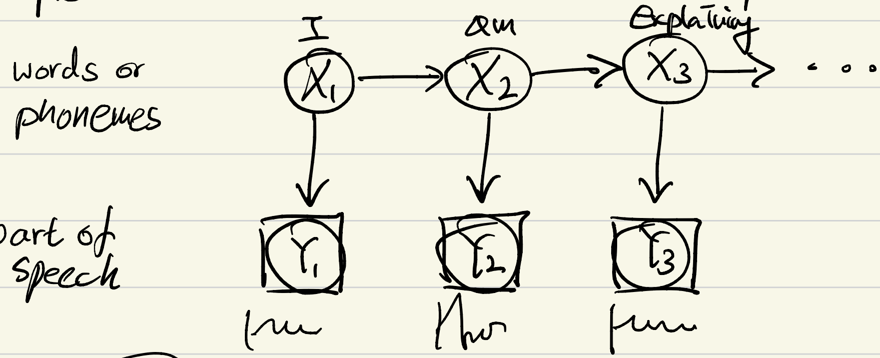
$Y_1^t \triangleq (Y_1, Y_2, \cdots, Y_t)$

. Goal: given noisy observations Y_1^t, estimate current state

compute $\mathbb{P}(X_t \mid \underbrace{X_1, Y_1 Y_2 \cdots Y_t}_{\text{known}})$

Hidden Markov Model

Example 3. Hidden Markov models for speech recognition



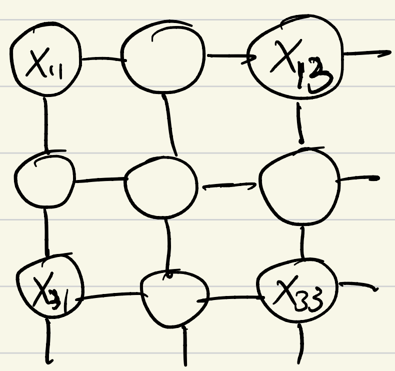
$x_i \in \mathcal{X}$, $|\mathcal{X}| < \infty$, $\mathcal{X}$ = Dictionary
400,000

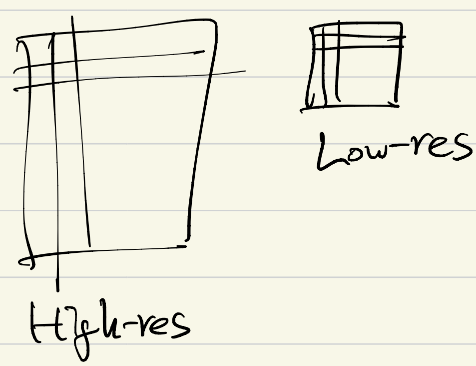
- Baum-Welch algorithm: for learning the graphical model

- Viterbi algorithm: for inferring the most likely sequence

$$\arg\max_{x_1, \dots, x_t \in S_{words}} \mathbb{P}(X_1, \dots, X_t \mid Y_1 \dots Y_t)$$

Example 4. Markov Random Field model for super-resolution





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

- loopy belief-propagation

$\tilde{X}_1$ $\tilde{X}_2$ $\tilde{X}_3$ ← Hidden Reliability

Example 5. Peer grading

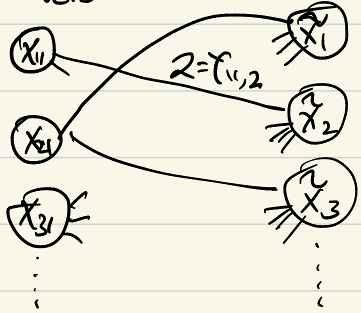
student 1   student 2   student 3

Hidden true score

X_{11} student 1 problem 1

X_{21} student 2 problem 1

0 ↖ $Y_{21,1}$ 2 ↑ $Y_{11,2}$ 1 ↑ $Y_{21,3}$



we know $\mathbb{P}(Y_{21,1} \mid X_{21}, \tilde{X}_1)$,   ex) $Y_{21,1} = X_{21} + \text{Noise} \times \tilde{X}_1$

inference : $\underset{\{X_{ij}\},\{\tilde{X}_\ell\}}{\arg\max}\ \mathbb{P}(\{X_{ij}\}, \{\tilde{X}_\ell\} \mid \{Y_{ij,\ell}\})$

General theme of the course

(we know) Probabilistic Distribution over $X = [X_1, \cdots, X_n]$

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

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

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

(we want) Inference tasks

- find the most probable realization

$$\hat{x} \in \underset{x \in \mathcal{X}^n}{\arg\max}\ P_Y(x_1, \cdots, x_n)$$

- Calculate the marginal distribution

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

- Sampling X from $P_Y(X)$

$n \gg 1$. $O(|\mathcal{X}|^{n-1})$ run-time is required inference can be dramatically efficient by exploiting <u>structure</u>

claim: suppose all variables are mutually independent X_1^n ← Independence conditioned on γ_1^m. $\Longleftrightarrow$ $P_\gamma(X) = f_1(x_1) \cdot f_2(x_2) \cdots f_n(x_n)$ ← Factorization

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

- $\arg\max_X P_\gamma(X) \Leftarrow \arg\max f_n(X_n)$ } n tasks
- $P_\gamma^{(1)}(x_1) = \frac{f_1(x_1)}{\sum_{x_1} f_1(x_1)}$, $O(|\mathcal{X}|)$, n marginals to compute.
- Sampling $X_1 \sim P_\gamma^{(1)}(X_1)$, X_2 ? . . .
  $O(|\mathcal{X}|)$

* When the distribution factorizes,
we can solve inference efficiently.

$\longleftrightarrow$

Complete dependence	Independence ?	Complete Independence
$P_\gamma(X)$	factorization ?	$P_\gamma(X) = \prod_i P_\gamma^{(i)}(x_i)$
slow	complexity ?	fast

- Graphical Models
  - Directed Graphical Models = Bayesian Networks
  - Undirected G. M. = Markov Random Fields
  - Factor Graphs
- Inference on Graphical Models
  - sum-product algorithm = Belief Propagation
  - max-product algorithm
- Learn Graphical Models
  - Parameter estimation (known G)
  - structure learning
  - learning with hidden variables.