LECTURE 17

Today

→ What are graphical Models?

→ Bayesian Networks : Definition, Examples,
Learning challenges, Chow-Liu Algorithm.

1. The class on 05/25/2022 will be remote on zoom.

2. Final exam:
   - Covers all topics until the exam
   - Easier than assignments.
   - In-person exam ~~as~~ per university schedule
   - MSOL students online.

Unsupervised Learning: Given a dataset, we want to build a "model" for the dataset.

PROBABILISTIC MODELS OF DATA

→ Graphical models are probabilistic models for generating data

→ Precursors for modern "deep" generative models

Core issues : ① Succinct representations of distributions

② Modeling dependencies.

Applications : ⓐ "Generate" new examples.

ⓑ "In-Painting"

ⓒ Applied Sciences useful in identifying relations.

Independence: (X, Y) random variables (joint dist.)

$$X \perp Y \iff \Pr[X=x \wedge Y=y] = \Pr[X=x]\cdot \Pr[Y=y]$$
$$\iff \Pr[Y=y \mid X=x] = \Pr[Y=y].$$

weather today is independent of Stock market.

	Heart Problems	Wake up time	Age
P_1	NO	11 AM	19
P_2	YES	6:30 AM	45
P_3	YES	7: AM	40
P_4	NO	12 PM	24
P_5	NO	10 AM	25
	⋮	⋮	
P_{10}	YES	6: AM	59
P_{11}	YES	6:30	62
		⋮	⋮

Comparing features directly:

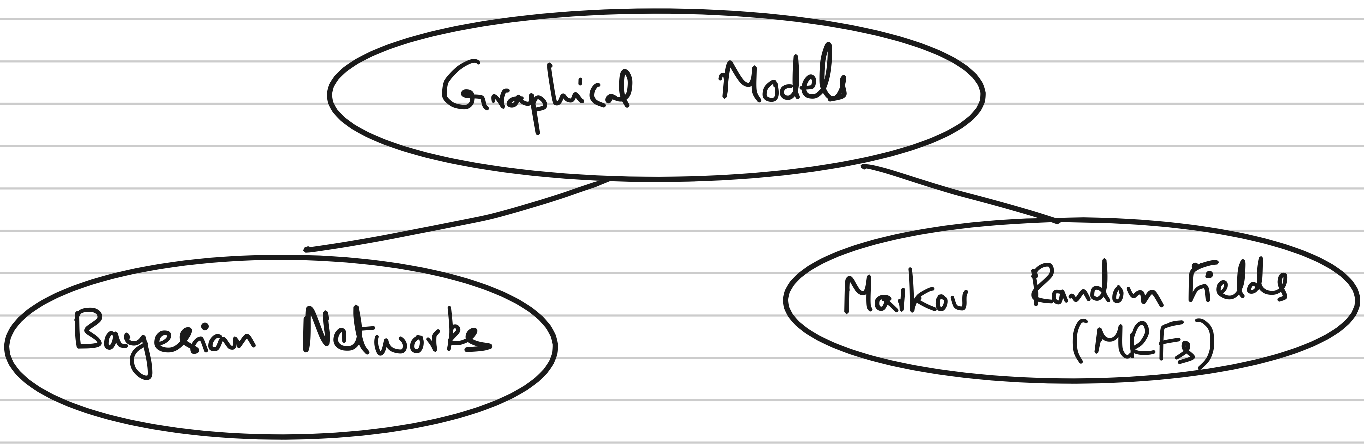
Waking up early is bad for heart

Conditional Independence: X, Y, Z random variables

Conditional independence: X, Y, Z random variables

$$X \perp Y \mid Z \iff \Pr[X=x, Y=y \mid Z=z] = \Pr[X=x \mid Z=z] \cdot \Pr[Y=y \mid Z=z]$$
$$\iff \Pr[Y=y \mid X=x, Z=z] = \Pr[Y=y \mid Z=z].$$

Graphical Models are distributions with "constrained" conditional independence ("CI") relations.



"Directed" Graphical Models

"Undirected" Graphical Models.

Bayesian Networks

$X \in \Sigma^d$ $\qquad (X_1, X_2, \ldots, X_d)$

A Directed Acyclic Graph (DAG) on $[d]$ vertices. G

A distribution D is a Bayes Net with graph G

$P(X_1 = x_1, \ldots, X_d = x_d) = \prod^{d} P[X_i = x_i \mid X_{\ldots} = x_{\ldots}]$

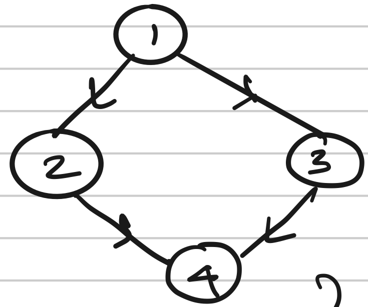
$$P(X = x_1, x_2, \ldots, x_d) = \prod_{i=1} \Pr[X = x_i \mid X_{Pa(i)} = x_{Pa(i)}]$$

$Pa(1) = \phi$

$Pa(2) = \{1\}$

$Pa(3) = \{1\}$

$Pa(4) = \{2, 3\}$

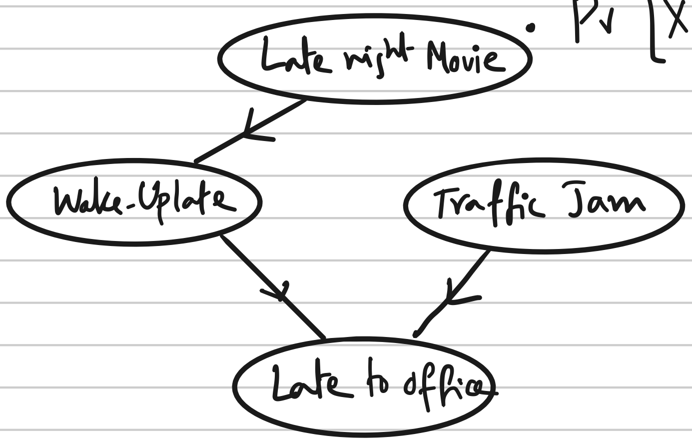


$pa(i)$ = the parents of i.

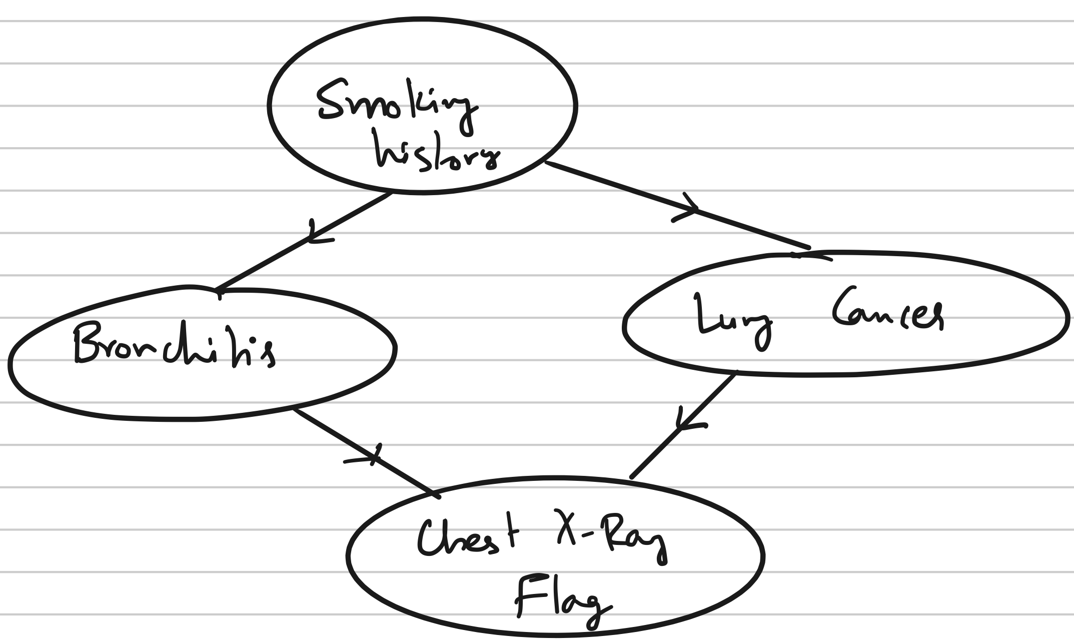
Example: $P(X = x_1, x_2, x_3, x_4)$

$$= \Pr[X_1 = x_1] \cdot \Pr[X_2 = x_2 \mid X_1 = x_1]$$
$$\cdot \Pr[X_3 = x_3 \mid X_1 = x_1]$$
$$\cdot \Pr[X_4 = x_4 \mid X_2 = x_2, X_3 = x_3].$$

Example:



Examples:



Learn the Bayes Net from

Samples

→ Given n samples $x^1, x^2, \ldots, x^n$

→ Learn the underlying directed dependency graph

Ex:

$$1 \to 2 \to 3 \to 4 \; \cdots\cdots \to n$$

"Markov Chain":

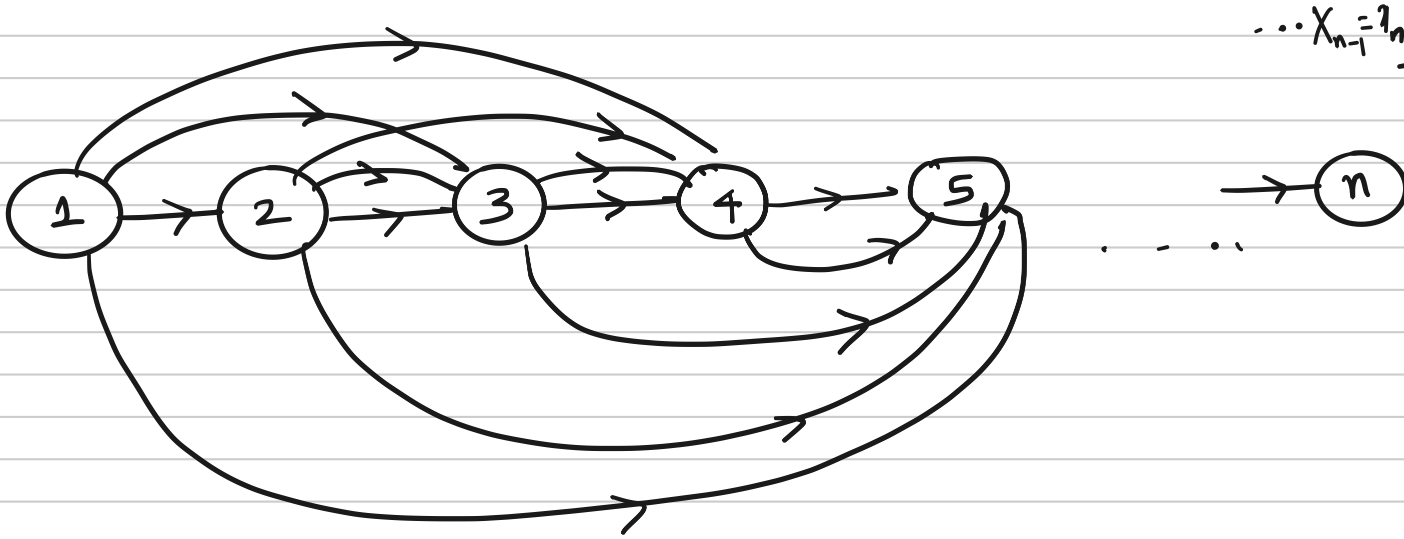
$$X_i \perp X_j \mid X_{i-1} \quad \text{for} \quad j < i-1$$

$$X_{10} \perp X_2 \mid X_9$$

$$\Pr[X = x_1, x_2, \ldots, x_n]$$

$$= \Pr[X_1 = x_1] \cdot \Pr[X_2 = x_2 \mid X_1 = x_1]$$

$$\cdot \Pr[X_3 = x_3 \mid X_1 = x_1, X_2 = x_2] \cdots\cdots \Pr[X_n = x_n \mid X_1 = x_1, \ldots X_{n-1} = x_{n-1}]$$



Make an assumption about the true "graph" (being "simple" in some sense).

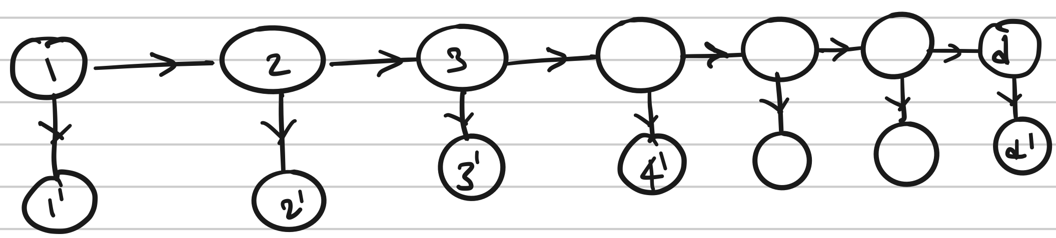
→ Degree is bounded ?

→ The "unknown" graph is a tree / path ...

→ Learn the structure graph under assumptions on the graph

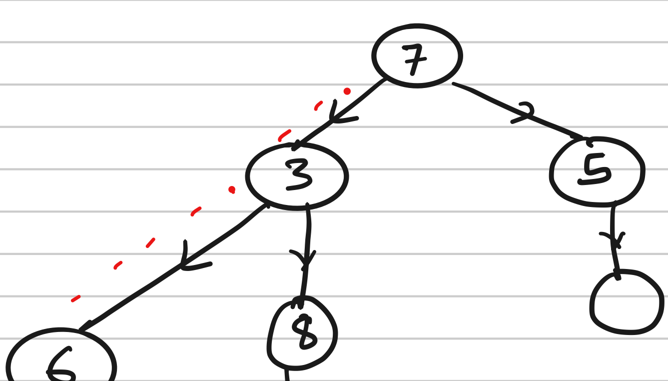
→ Learn the distribution assuming you even know the graph.

→ Some features are missing ?
(Hidden Markov Models)

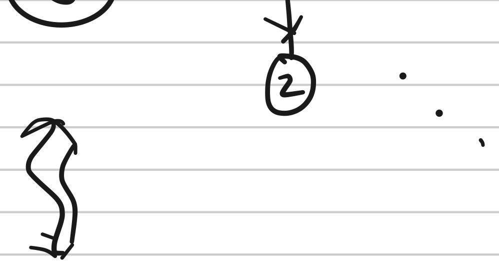


Suppose we get samples from a distribution generated by a tree Bayes network.

→ Can you learn the distribution from samples?



Each node has a single parent.



CHOW-LIU Algorithm 1968:

→ We Can learn tree-shaped Bayesian networks.