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:

a. "Generate" new examples.
b. "In-Painting"
c. Applied Sciences useful in identifying relations.

## PROBABILITY:

### INDEPENDENCE:

$(X, Y)$ random variables (joint distribution)

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

$$\iff \Pr[Y = y \mid X = x] = \Pr[Y = y]$$

eg: weather today is independent of stock market.

| | Heart Problems | Wake Up Time | Age |
|---|---|---|---|
| P1 | NO | 11 AM | 19 |
| P2 | YES | 6:30 AM | 45 |
| P3 | YES | 7 AM | 40 |
| P4 | NO | 12 PM | 24 |
| P5 | NO | 10 AM | 25 |
| ⋮ | ⋮ | ⋮ | ⋮ |
| P10 | YES | 6 AM | 59 |

Comparing features directly:

Waking up early is bad for heart.

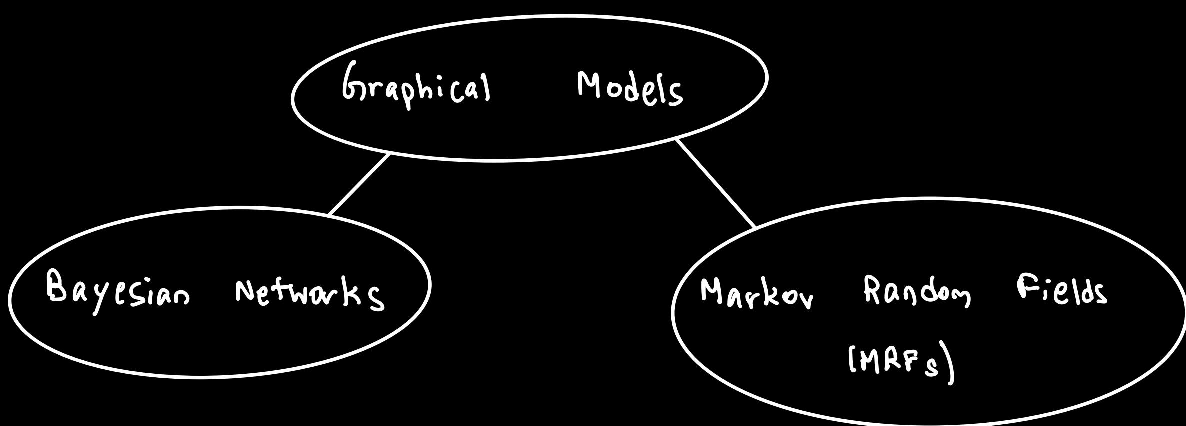
**CONDITIONAL INDEPENDENCE:**

$X, Y, Z$ 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 BayesNet with graph $G$.

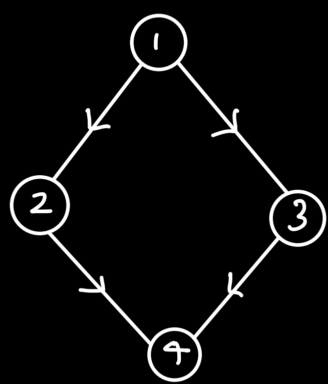
$$P(X = x_1, x_2, \ldots, x_d) = \prod_{i=1}^{d} P[X = x_i \mid X_{pa(i)} = x_{pa(i)}]$$

$Pa(1) = \phi$

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

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

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



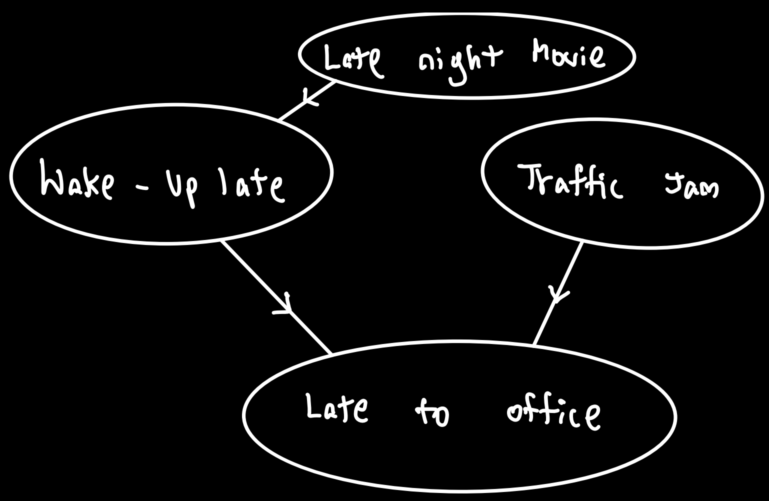

$pa(i)$ = the parent of $i$.

Example

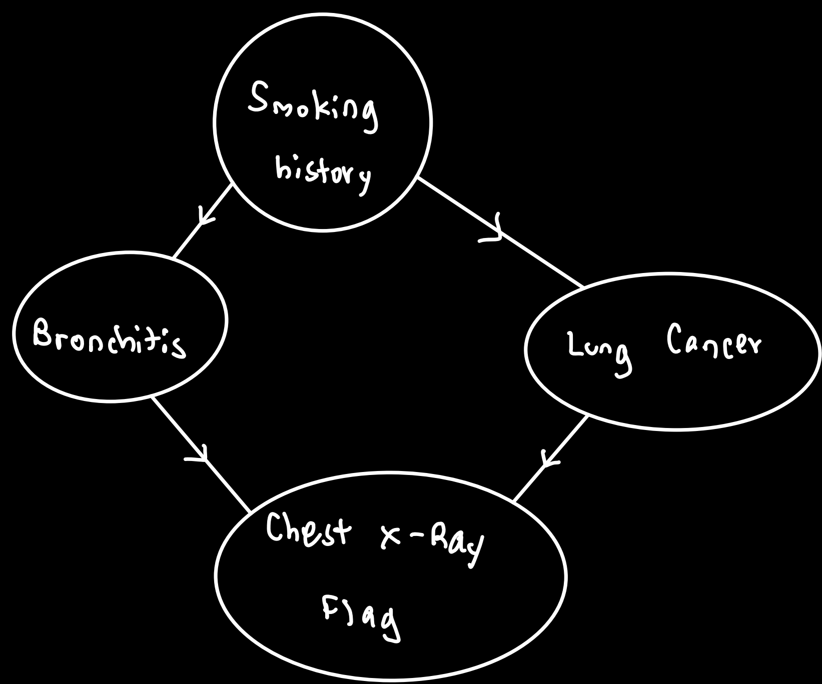
$$P(X = x_1, x_2, x_3, x_4) = Pr[X_1 = x_1] \cdot Pr[X_2 = x_2 \mid X_1 = x_1] \cdot$$

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

EXAMPLE:




EXAMPLE:

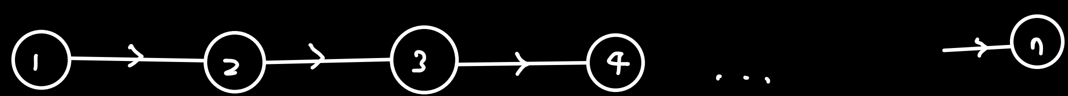


## LEARN THE BAYES NET FROM SAMPLES :

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

→ Learn the underlying directed dependency graph.

Example:



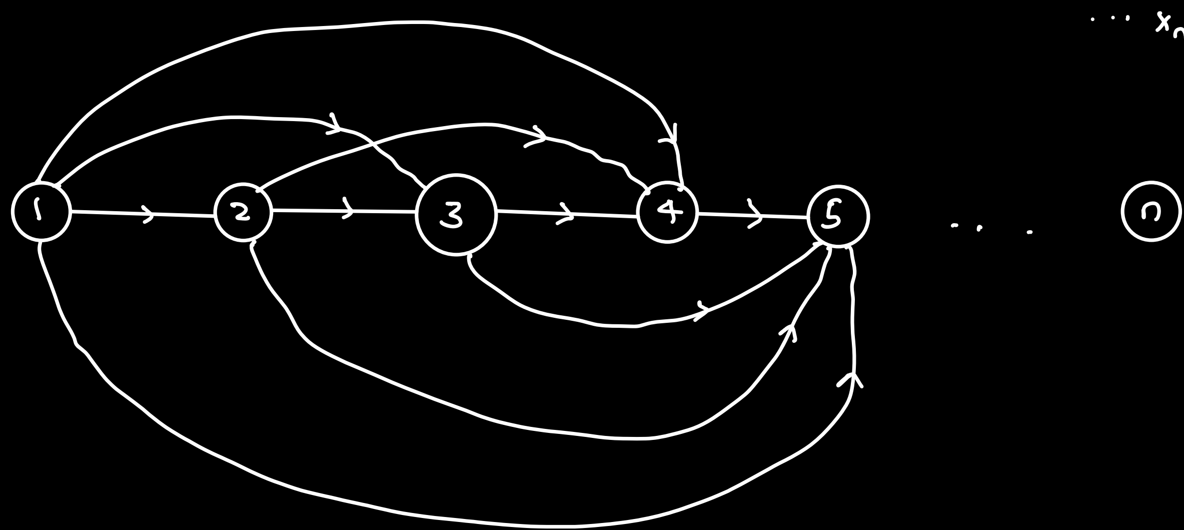

"Markov Chain"

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

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

$$\Pr[X = x_1, x_2, \ldots, x_n] = \Pr[X_1 = x_1] \cdot \Pr[X_2 = x_2 \mid X_1 = x_1]$$

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



Make an assumption about the true "graph".

(being "simple" in some sense).

eg: - Degree is bounded?

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

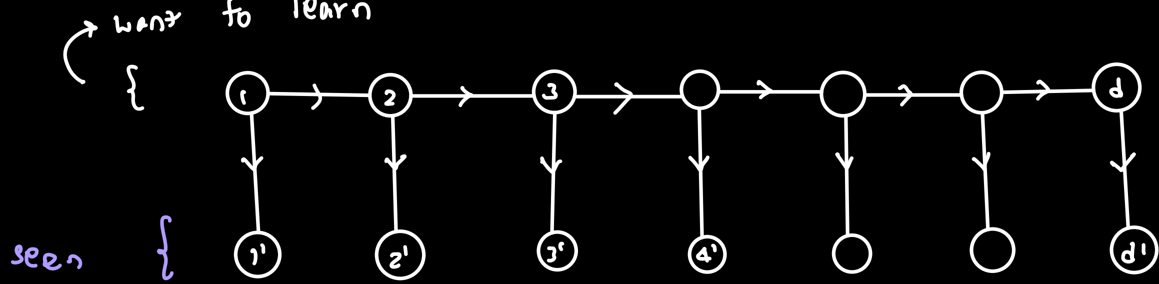
→ Learn the structure graph under assumptions on the graph.

→ learn the distribution assuming you even know the graph.

→ Some features are missing?

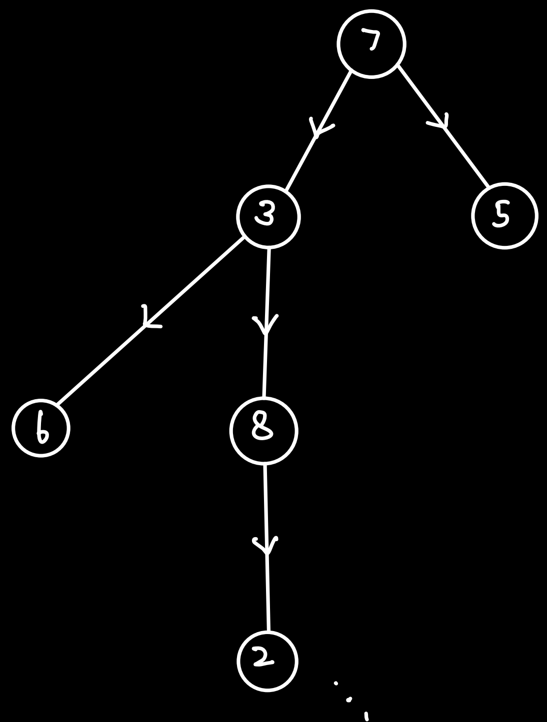
(Hidden Markov Models)

want to learn



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.