Machine Learning for Data Science (CS4786)
Lecture 18

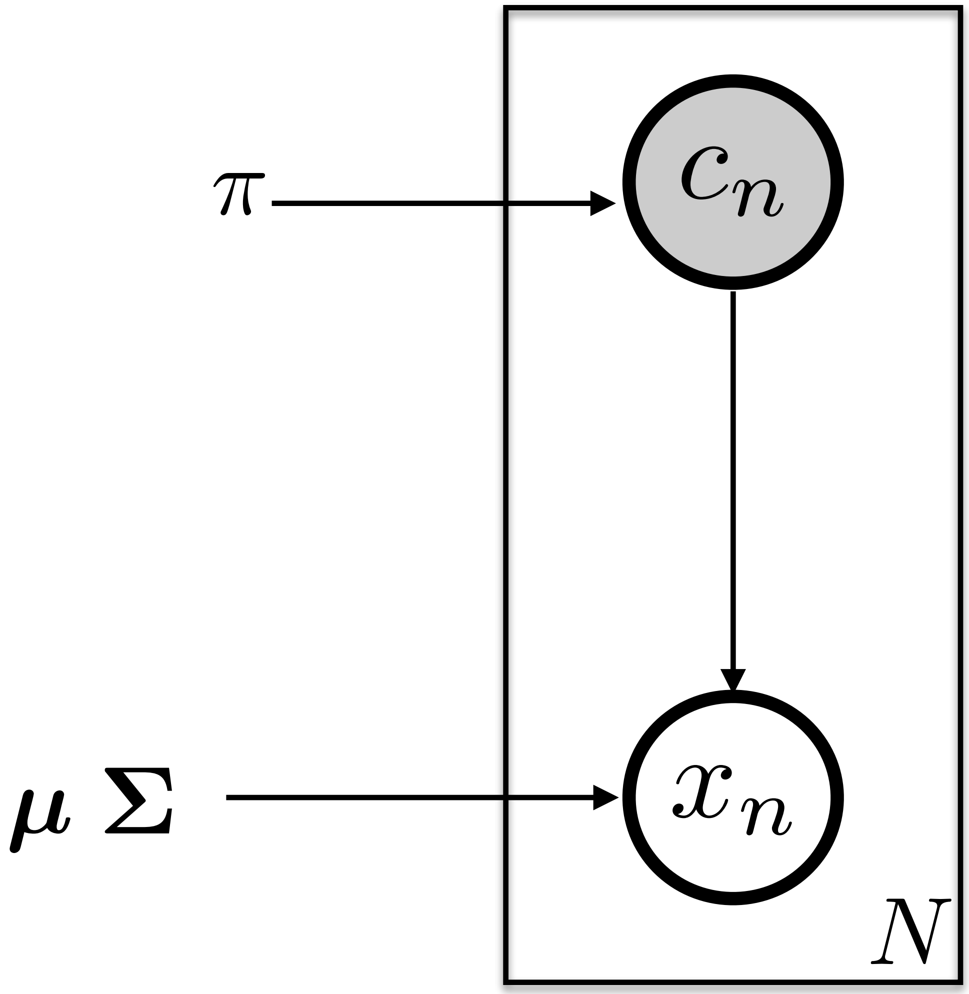
Graphical Models

Course Webpage :
http://www.cs.cornell.edu/Courses/cs4786/2016sp/

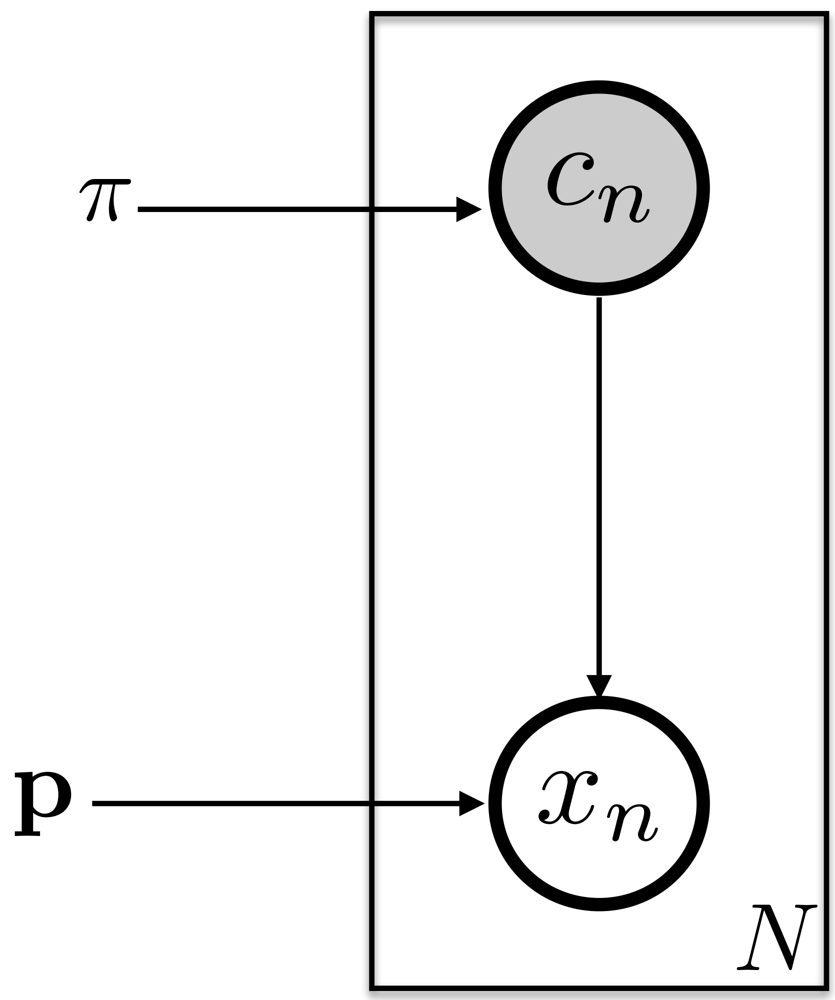
PROBABILISTIC MODELS

- We have a bunch of observed variables
- A bunch of Hidden or Latent variables
- Set Θ consists of parameters s.t. P_θ is the distribution over the random variables by each $\theta \in \Theta$
- Data is generated by one of the $\theta \in \Theta$
- Learning: Estimate value or distribution for $\theta^* \in \Theta$ given data
- Inference: Given parameters and observation infer distribution over variables

Gaussian Mixture Model



Mixture of Multinomials



Graphical Models

- Abstract away the parameterization specifics
- Focus on relationship between random variables

Relationship Between Variables

Let $X = (X_1, \ldots, X_N)$ be the random variables of our model (both latent and observed)

- Joint probability distribution over variable can be complex esp. if we have many complexly related variables
- Can we represent relation between variables in conceptually simpler fashion?
- We often have prior knowledge about the dependencies (or conditional (in)dependencies) between variables

Graphical Models

- A graph whose nodes are variables $X_1, \ldots, X_N$
- Graphs are an intuitive way of representing relationships between large number of variables
- Allows us to abstract out the parametric form that depends on θ and the basic relationship between the random variables.

Conditional and Marginal Independence

- Conditional independence
  - X_i is conditionally independent of X_j given $A \subset \{X_1, \ldots, X_N\}$:

$$\begin{aligned} X_i \perp X_j | A &\Leftrightarrow P_\theta(X_i, X_j | A) = P_\theta(X_i|A) \times P_\theta(X_j|A) \\ &\Leftrightarrow P_\theta(X_i | X_j, A) = P_\theta(X_i|A) \end{aligned}$$

- Marginal independence:

$$X_i \perp X_j | \varnothing \Leftrightarrow P_\theta(X_i, X_j) = P_\theta(X_i) P_\theta(X_j)$$

Example: CI and MI

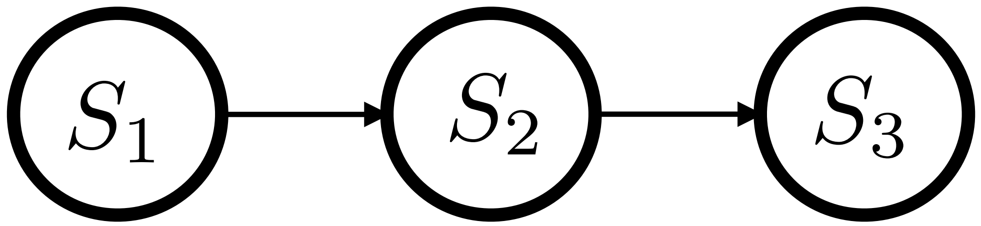
Bayesian Networks

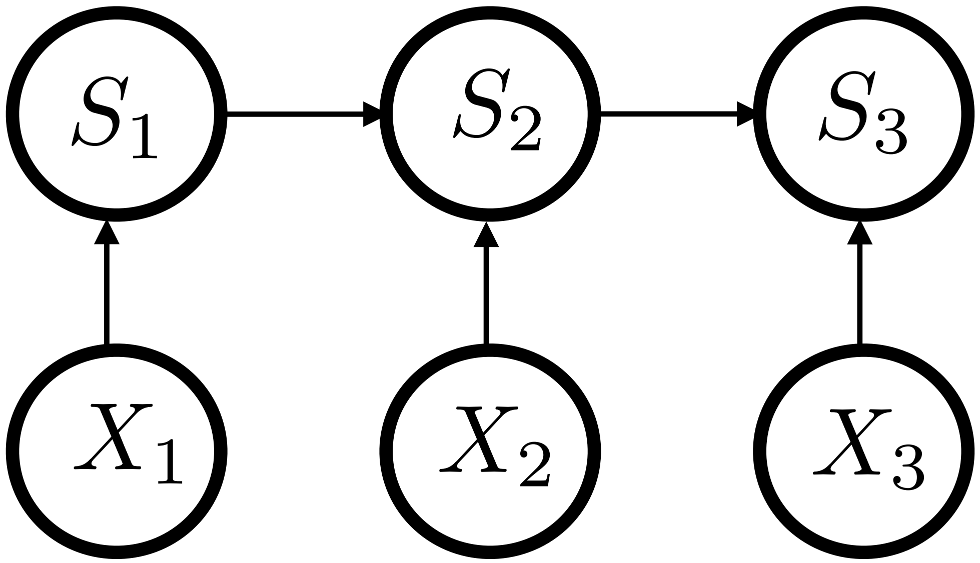
- Directed acyclic graph (DAG): $G = (V, E)$
- Joint distribution P_θ over $X_1, \ldots, X_n$ that factorizes over G:

$$P_\theta(X_1, \ldots, X_n) = \prod_{i=1}^{N} P_\theta(X_i | \text{Parent}(X_i))$$

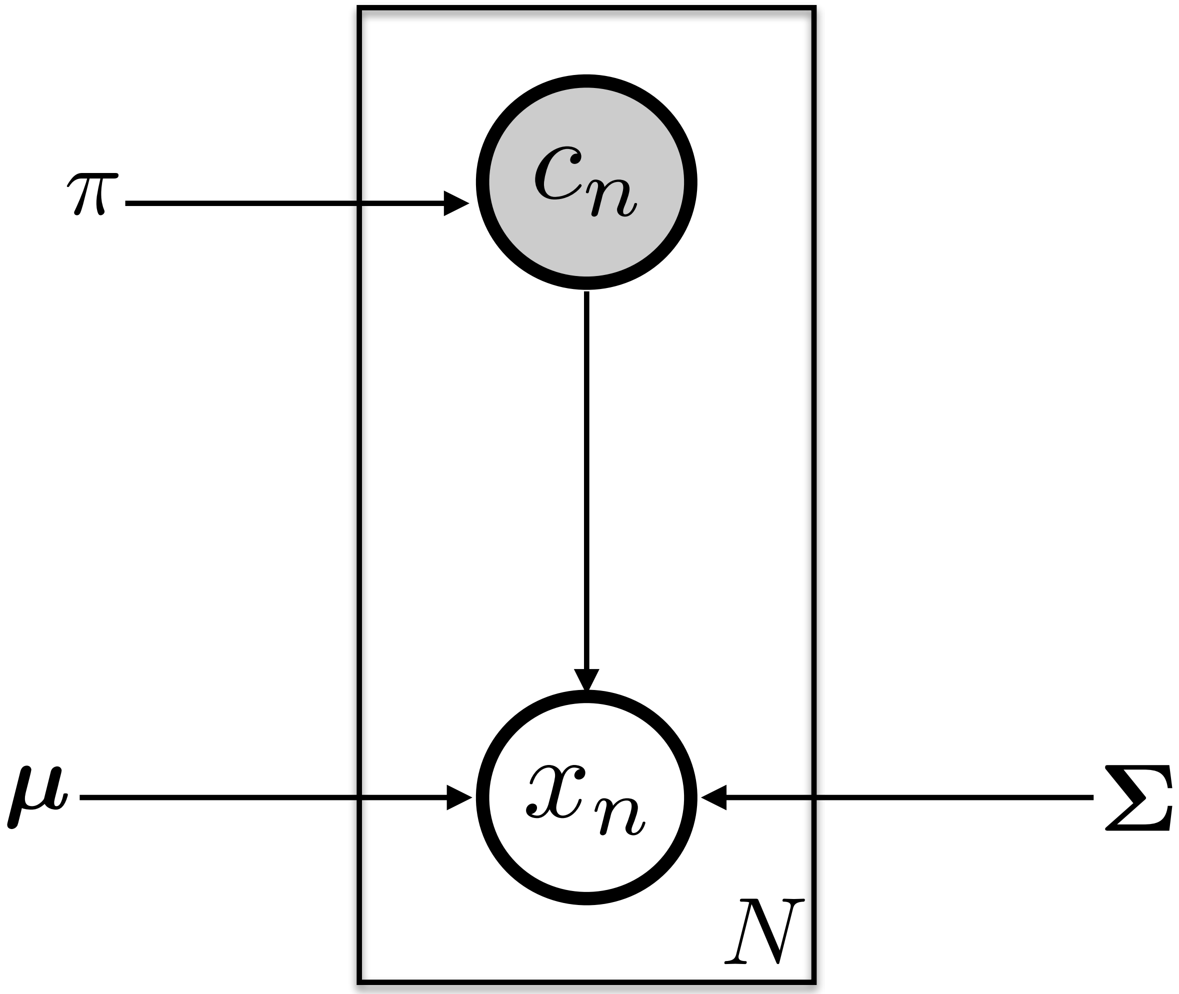
- Hence Bayesian Networks are specified by G along with CPD's over the variables (given their parents)

Example: Sum of Coin Flips

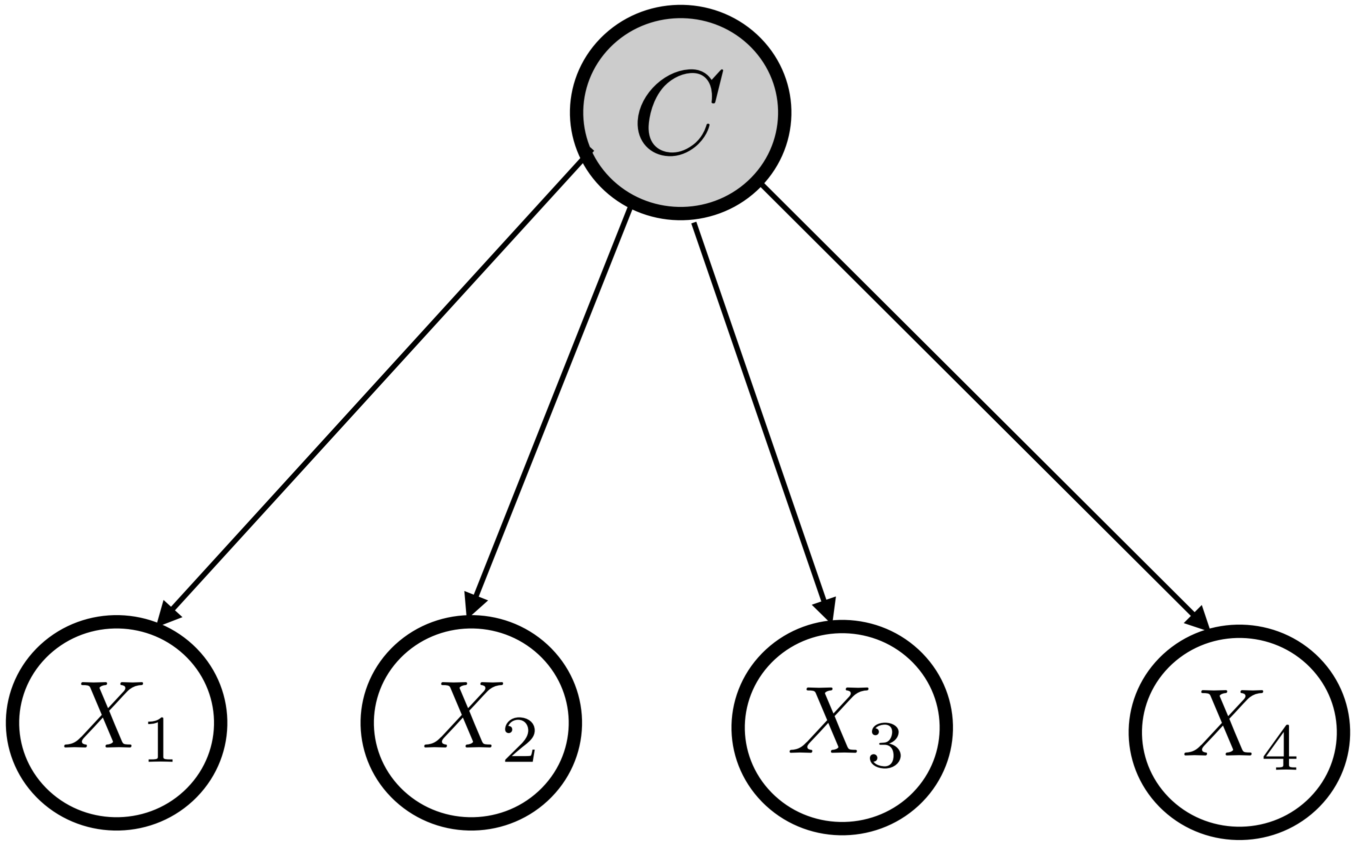




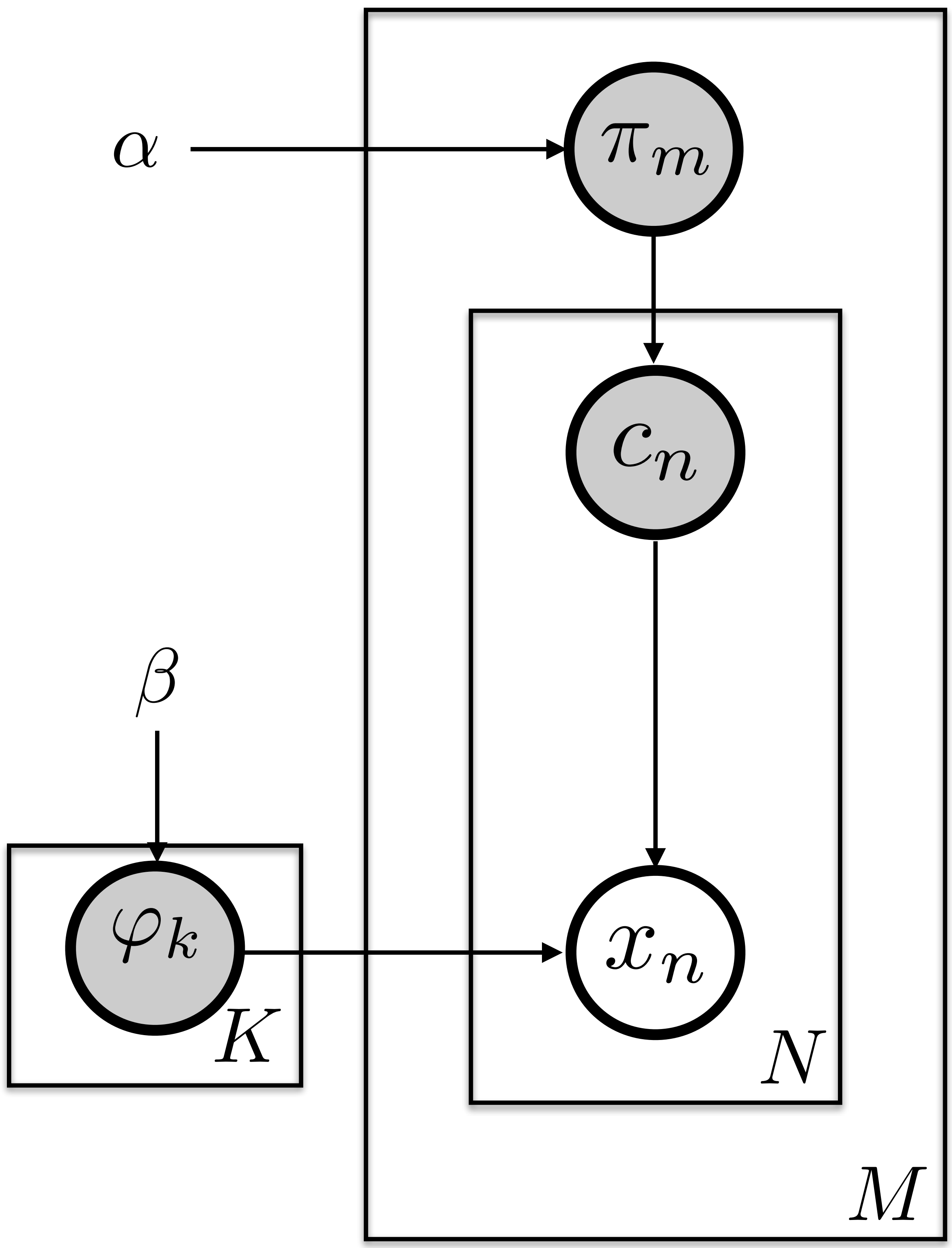
Example: Mixture Models



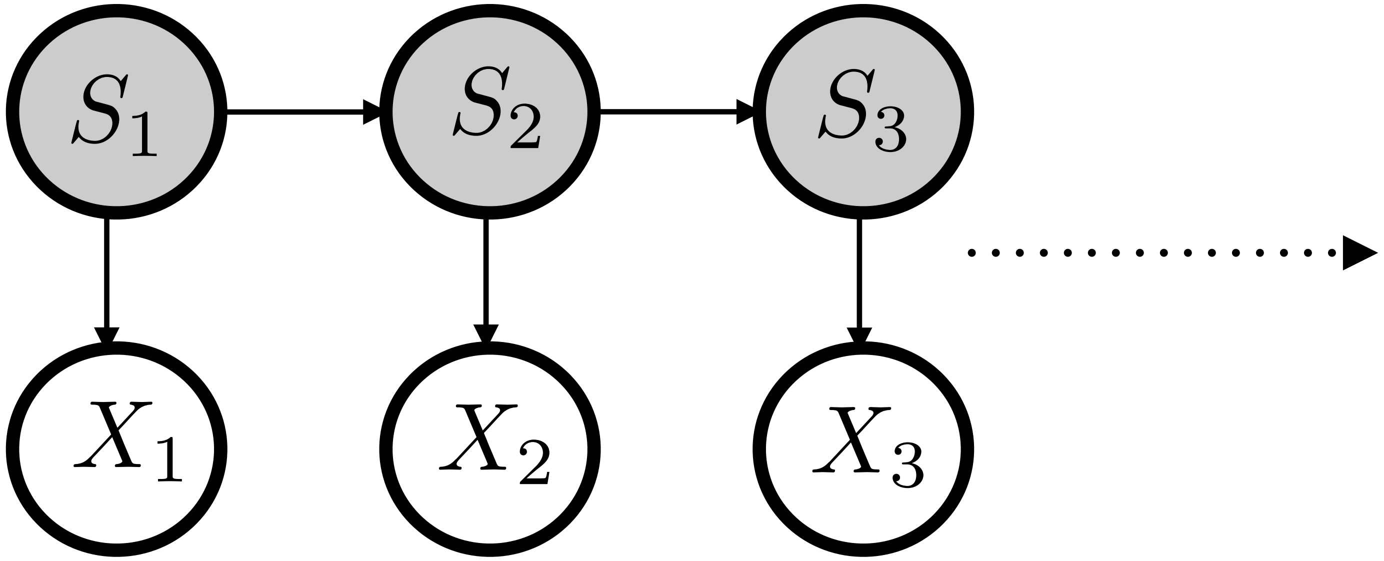
EXAMPLE: NAIVE BAYES CLASSIFIER



Example: Latent Dirichlet Allocation



Example: Hidden Markov Model



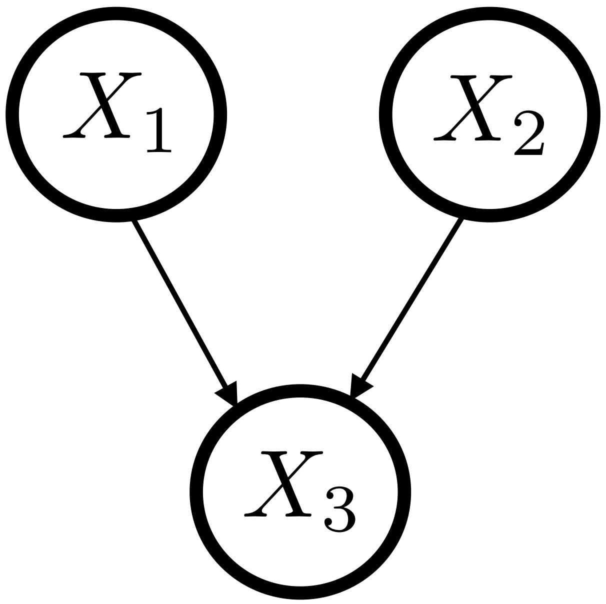
Local Markov Property

- Each variable is conditionally independent of its non-descendants given its parents
- Any joint distribution satisfying the local markov property w.r.t. graph factorizes over the graph

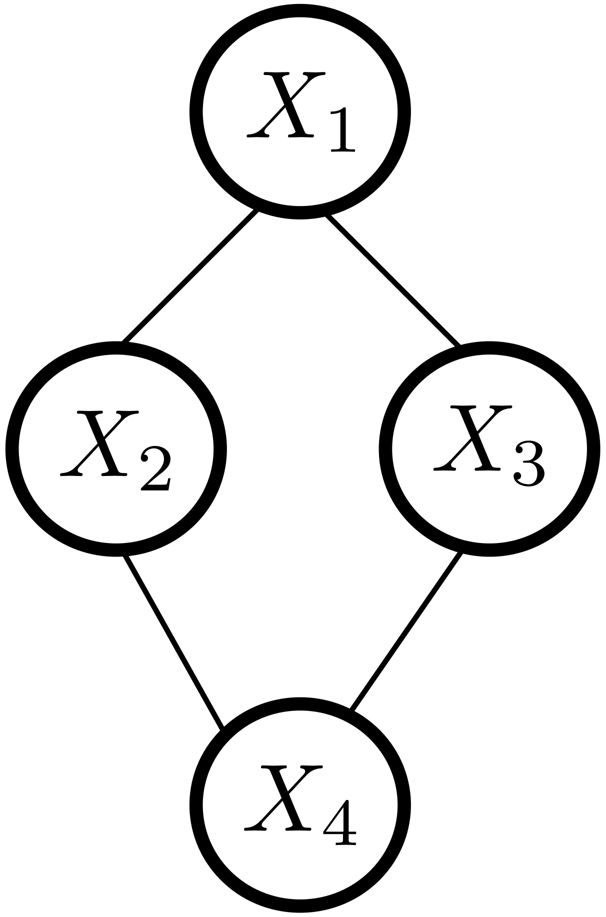
Markov Networks

- Not all distributions can be represented by Bayesian networks
- We also have undirected graphical models.
- Undirected graph $G = (V, E)$ and a set of RV's $X_1, \ldots, X_N$ form a markov network if
  - Any two non adjacent variables are conditionally independent given all other variables
  - Given its neighbors a variable is conditionally independent of all other variables
  - Any two sets of variables are conditionally independent given a separating set

Representational Power: BN Vs MN



No undirected graph can capture the above dependence



No directed graph can capture the above dependence