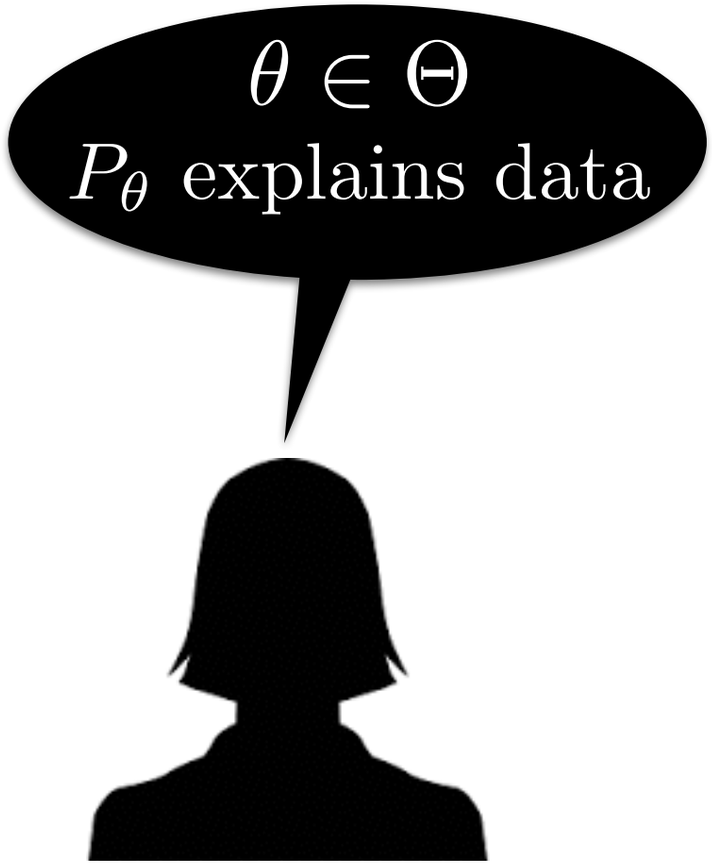
Machine Learning for Data Science (CS4786)
Lecture 17

Graphical Models

Course Webpage :

http://www.cs.cornell.edu/Courses/cs4786/2016fa/

Probabilistic Model



Data

Probabilistic Model

$x_1, \ldots, x_N$

Latent Variables

Observed variables

$\theta \in \Theta$
P_θ explains data

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.

Draw a picture for the generative story
that explains what generates what.

Conditional and Marginal Independence

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

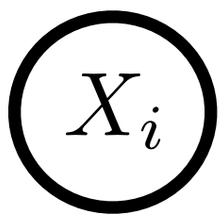
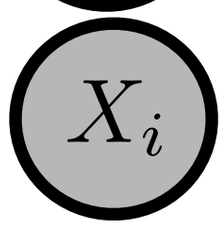
$$X_i \perp X_j | A \Leftrightarrow P_\theta(X_i, X_j | A) = P_\theta(X_i | A) \times P_\theta(X_j | A)$$
$$\Leftrightarrow P_\theta(X_i | X_j, A) = P_\theta(X_i | A)$$

- Marginal independence:

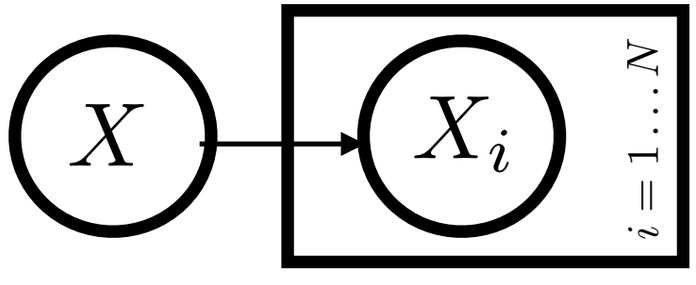
$$X_i \perp X_j | \varnothing \Leftrightarrow P_\theta(X_i, X_j) = P_\theta(X_i) P_\theta(X_j)$$

Example: CI and MI

Graphical Models

- Variables X_i is written as  if X_i is observed
- Variables X_i is written as  if X_i is latent
- Parameters are often left out (its understood and not explicitly written out). If present they dont have bounding objects

- An directed edge $\longrightarrow$ is drawn connecting every parent to its child (from parent to child)



$X_1 \ldots X_N$ drawn repeatedly from $P(Y|X)$

Bayesian Networks

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

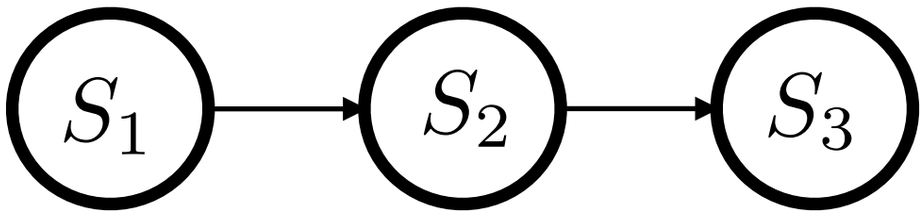
$$P_\theta(X_1, \ldots, X_n) = \prod_{i=1}^{N} P_\theta(X_i | \text{Parent}(X_i))$$

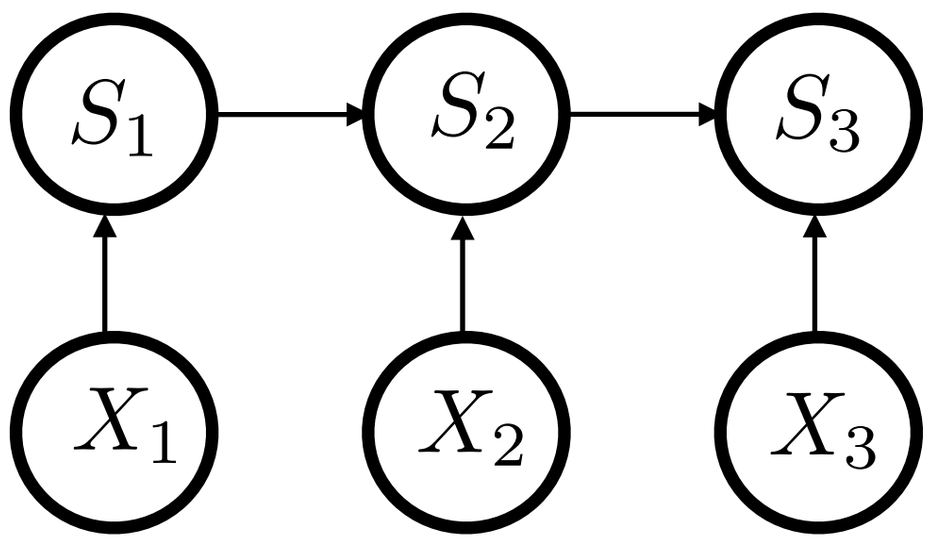
- Hence Bayesian Networks are specified by G along with CPD's over the variables (given their parents)

Bayesian Networks

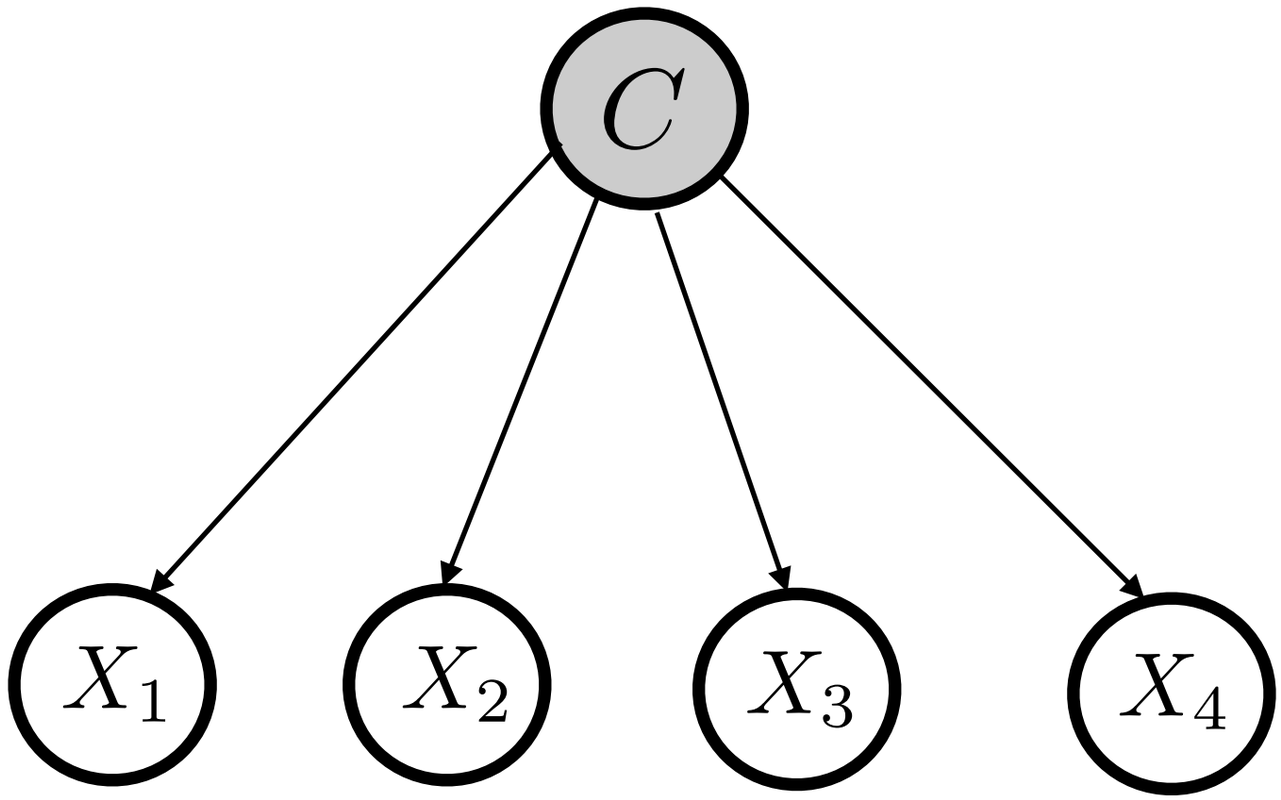
- Provide directed acyclic graph
  - why DAG?
- Provide conditional probability or density of variable given its parents
- At a high level variables can be discrete (like cluster assignments) or continuous like points in d-dimensional space

Example: Sum of Coin Flips



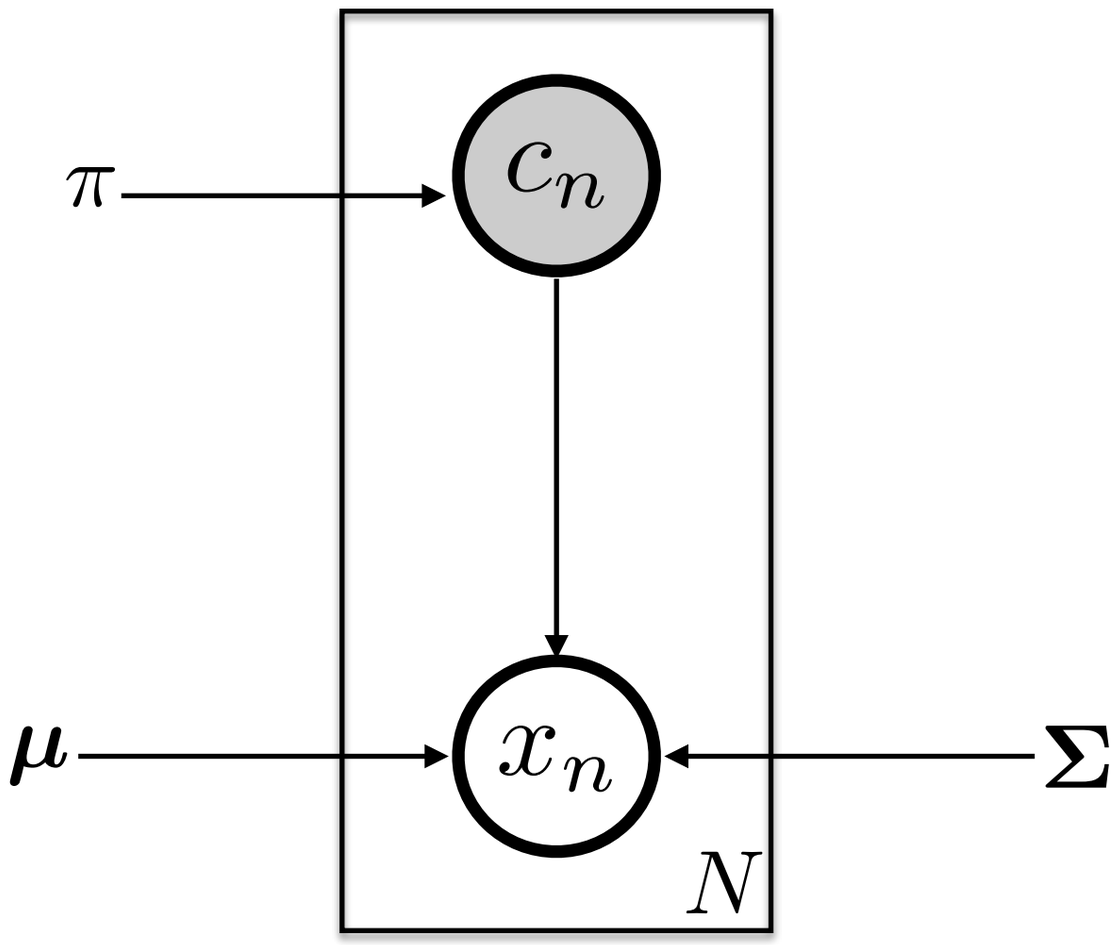


Example: Naive Bayes Classifier



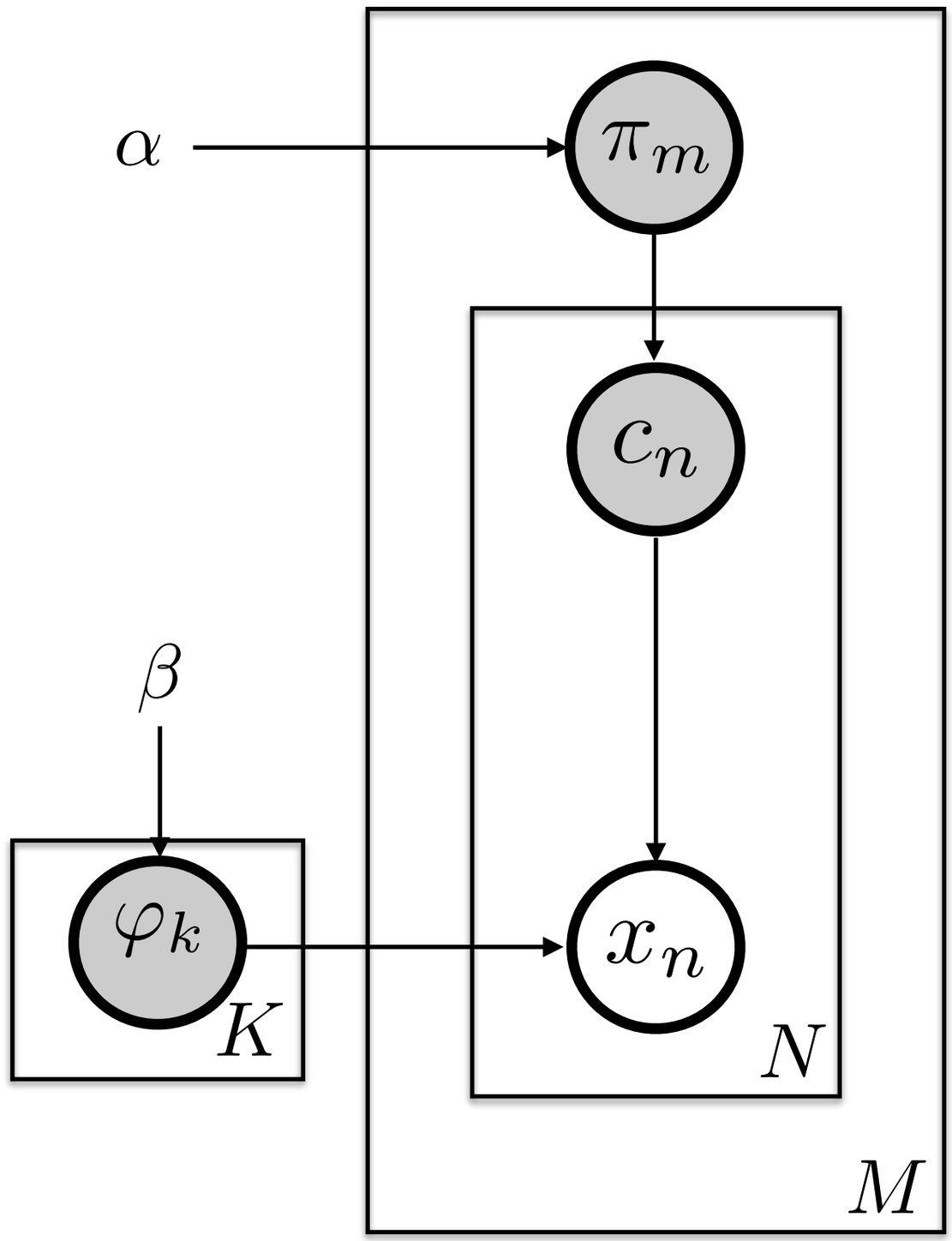
Eg. Spam classification

Example: Mixture Models



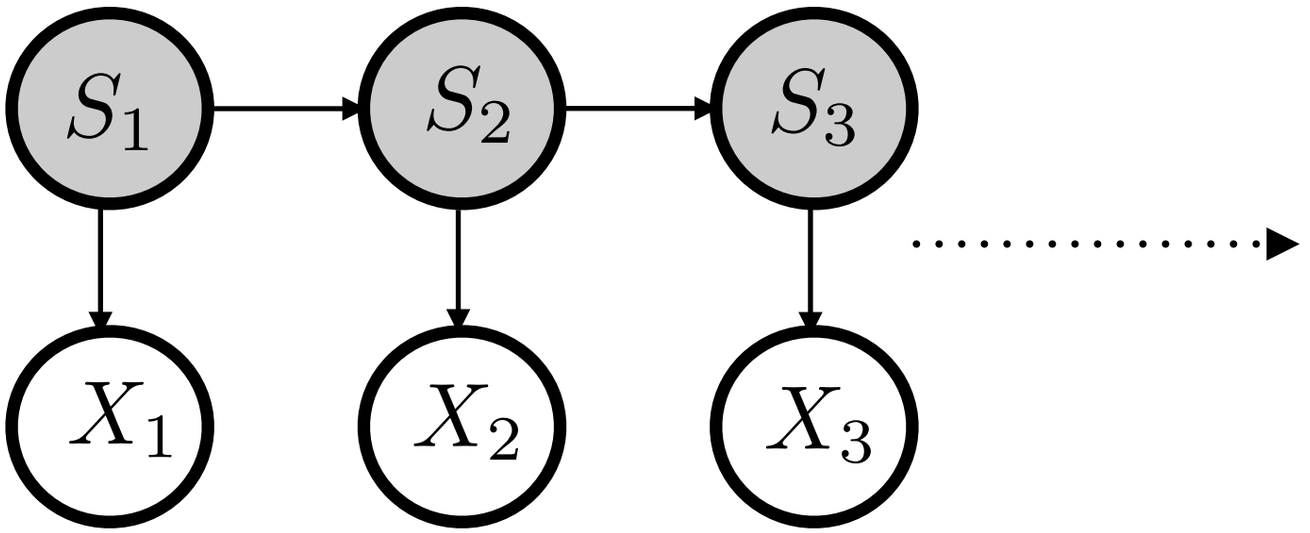
Eg. Clustering

Example: Latent Dirichlet Allocation



Eg. Topic modelling

Example: Hidden Markov Model



Eg. Speech recognition

Representational Power

- Not all joint distributions can be represented by Bayesian Networks
- Eg. $X_1 \perp X_4 \mid X_3, X_2$ and $X_3 \perp X_2 \mid X_1, X_4$

  This dependence can never be captured by a bayesian network, Why?

Which distributions can be represented by Bayesian networks?

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

Why?

Factorizing Joint Probability

For every directed edge from X to Y,

X comes before Y in the sorted order.

- (DAG Factoids) Assume nodes are arranged according to some topological sort
- For any distribution we have:

$$P_\theta(X_1, \ldots, X_N) = \prod_{i=1}^{N} P_\theta(X_i | X_1, \ldots, X_{i-1})$$

...

Two main questions

- Learning/estimation: Given observations, can we learn the parameters for the graphical model ?

- Inference: Given model parameters, can we answer queries about variables in the model

  - Eg. what is the most likely value of a latent variable given observations

Two main questions

- Learning/estimation: Given observations, can we learn the parameters for the graphical model ?

- Inference: Given model parameters, can we answer queries about variables in the model

  - Eg. what is the most likely value of a latent variable given observations

  - Eg. What is the distribution of a particular variable conditioned on others

Inference in Graphical Models

Given parameters of a graphical model, we can answer any questions about distributions of variables in the model

Example queries:

1. What is the probability of a given assignment for a subset of variables (marginal)?
2. What is the probability of a particular assignment of a subset of variables given observed values (evidence) of some subset of the variables (conditional)?

3. Given observed values (evidence) of some subset of variables what is the most likely assignment for a given subset of variables?

Suffices to calculate marginals.

Why?

Bayes rule: for any two sets of variables A and B,

$$P(A|B) = P(A, B)/P(B)$$

Next class

- Start with example of Hidden Markov Model (HMM)