Graphical Models

Course Webpage: 
http://www.cs.cornell.edu/Courses/cs4786/2016sp/
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 $\theta$ and the basic relationship between the random variables.
Conditional and Marginal Independence

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

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

$$\iff P_\theta(X_i | X_j, A) = P_\theta(X_i | A)$$

- Marginal independence:

$$X_i \perp X_j | \emptyset \iff P_\theta(X_i, X_j) = P_\theta(X_i)P_\theta(X_j)$$
Bayesian Networks

- Directed acyclic graph (DAG): $G = (V, E)$
- Joint distribution $P_\theta$ 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)
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
Example: Sum of Coin Flips
Example: Mixture Models
Example: Naive Bayes Classifier
**Example: Latent Dirichlet Allocation**

\[ \alpha \rightarrow \pi_m \]

\[ \beta \rightarrow \varphi_k \]

\[ \varphi_k \rightarrow x_n \]

\[ c_n \rightarrow x_n \]

\[ M \]

\[ N \]

\[ K \]
Example: Hidden Markov Model
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.
No undirected graph can capture the above dependence

No directed graph can capture the above dependence
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.
Variable Elimination: Examples
VARIABLE ELIMINATION: ORDER MATTERS

\[ X_5 \]
\[ X_1 \quad X_2 \quad X_3 \quad X_4 \]
VARIABLE ELIMINATION: EXAMPLES

\[ S_1 \rightarrow S_2 \rightarrow S_3 \]

\[ X_1 \rightarrow X_2 \rightarrow X_3 \]
Variable Elimination: Bayesian Network

Initialize List with conditional probability distributions

Pick an order of elimination I for remaining variables

For each $X_i \in I$

Find distributions in List containing variable $X_i$ and remove them

Define new distribution as the sum (over values of $X_i$) of the product of these distributions

Place the new distribution on List

End

Return List
Often we need more than one marginal computation

Over variables we need marginals for, there are many common distributions/potentials in the list

Can we exploit structure and compute these intermediate terms that can be reused?