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 $\Theta$ consists of parameters s.t. $P_\theta$ 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
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
- 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:
  \[
  X_i \perp X_j | \emptyset \iff P_\theta(X_i, X_j) = P_\theta(X_i) P_\theta(X_j)
  \]
EXAMPLE: CI AND MI
**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)
Example: Sum of Coin Flips

$S_1 \rightarrow S_2 \rightarrow S_3$

$X_1 \rightarrow X_2 \rightarrow X_3$
Example: Mixture Models
Example: Naive Bayes Classifier
Example: Latent Dirichlet Allocation

\[ \alpha \rightarrow \pi_m \]

\[ \varphi_k \rightarrow \mathcal{C}_n \]

\[ \beta \rightarrow \mathcal{X}_n \]

\[ K \rightarrow \mathcal{C}_n \]

\[ N \rightarrow \mathcal{X}_n \]
Example: Hidden Markov Model

\[ S_1 \rightarrow S_2 \rightarrow S_3 \]

\[ X_1 \rightarrow X_2 \rightarrow X_3 \]
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.
No undirected graph can capture the above dependence

No directed graph can capture the above dependence

**Representational Power: BN Vs MN**