Graphical Models

April 9, 2015

Course Webpage:
http://www.cs.cornell.edu/Courses/cs4786/2015sp/
Its on!

- Competition dataset and instructions are out
- Due date: April 22nd, 11:59pm
- Consists of two challenges
- One of the challenges is posted on Kaggle
- Group size: 1-4
- Report/writeup 5-15 pages
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
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 $\theta$ and the basic relationship between the random variables.
Conditional and Marginal Independence

- **Conditional independence**
  - $X_i$ is conditionally independent of $X_j$ given $A \subset \{X_1, \ldots, X_N\}$:
    
    \[
    X_i \perp X_j | A \iff P_{\theta}(X_i, X_j | A) = P_{\theta}(X_i | A) \times P_{\theta}(X_j | A)
    \]
    
    \[
    \iff P_{\theta}(X_i | X_j, A) = P_{\theta}(X_i | A)
    \]

- More generally for $C, B \subset \{X_1, \ldots, X_N\}$,
  
  \[
  B \perp C | A \iff \{X_i \perp X_j | A, \forall X_i \in B, X_j \in C\}
  \]

- **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) $G = (V, E)$
  - Nodes represent variables $X_1, \ldots, X_n$
  - Edges indicate causality structure or structure of generative model
- We say that a joint distribution $P_\theta$ factorizes over $G$ if:

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

In other words, the distribution which we can model using a given Bayesian Networks are the distributions that factor over the network
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.
(DAG Factoids) Assume nodes are arranged according to some topological sort
For any distribution we have:

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

\[ \ldots \]
EXAMPLES

- Gaussian Mixture Models
- Mixtures of Multinomials
- Latent Dirichlet Allocation
- Naive Bayes
- Hidden Markov models and Kalman filters
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 $X$ 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.