CS395T: Learning Theory; Fall 2011

Lecture 03 ; Date: 09/05/11

Lecturer: Pradeep Ravikumar ; Scribe: Deepjyoti Deka

Keywords: Graphical Models, Inference

Graphical Models: Introduction

Graphical Models can be of two types: Directed (Bayes Network) and Undirected (Markov Random fields).
Directed Acyclic Graph(DAG) : This graph has directed edges. u is the ancestor of v iff there is a directed path from u to v depicted by $u \longrightarrow v$

$$\Pi(S) \equiv \text{parents of node } S$$
$$P_s(X_s \mid X_{\pi(S)}) \equiv \text{denotes the graphical model of the DAG.}$$
$$P(X) \equiv P(X_1, X_2, ...X_n) = \prod_S P_S(X_S \mid X_{\pi(S)})$$

This can reduce the number of expressions needed to represent $P(\overrightarrow{X})$. For example,in the following DAG, $X_2 \longleftarrow X_1 \longrightarrow X_3$

$$P(\overrightarrow{X}) = P(X_1)P(X_2 \mid X_1)P(X_3 \mid X_1)$$
$$\int_{X_1}\int_{X_3}\int_{X_2} P(X_1)P(X_2 \mid X_1)P(X_3 \mid X_1)dX_2dX_3dX_1 = \int_{X_1}\int_{X_3} P(X_1)P(X_3 \mid X_1)dX_3dX_1$$
$$= 1$$

Undirected Acyclic Graph : Here there are no directed edges and hence no notion of ancestor as shown in Fig.1.

Cliques : These refer to fully connected sub-graphs of a undirected graph. The clique which cannot be grown any further,i.e, has maximum possible vertices in it

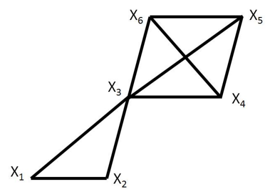

Figure 1: undirected graphical model

is called *Maximal Clique*.

$$\mathbb{C}_{\mathbb{G}} \equiv \text{set of all possible cliques in graph G}$$

$$c \in \mathbb{C}_{\mathbb{G}}, \Psi_c(X_c) \equiv \text{compatibility function}$$

$$P(X) = \frac{1}{Z} \prod_{c \in \mathbb{C}_{\mathbb{G}}} \Psi_c(X_c) \text{ where Z is the normalising constant}$$

Factor Graph Model : Here the graph is considered to constitute of a product of factors Ψ with each factor being contributed by variables/vertices within a clique. For example, in the Fig. 2, $P(X) = \frac{1}{Z} \prod_f \Psi_f(X_f)$ where X_f can be (X_1, X_2, X_3) or (X_3, X_4, X_5, X_6)

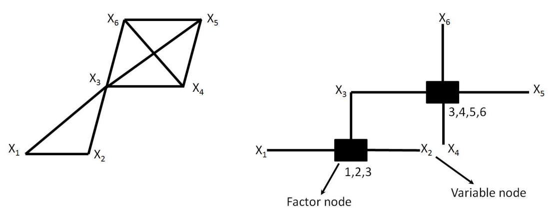

Figure 2: Factor Graph

Conditional Independence Assumption : Given a particular connected graph, two sub-graphs are considered conditionally independent given the set of nodes which separates them . The separating nodes form the *Separator Set*. Here, A and B are conditionally independent given the separator set consisting of X_3 and X_4 .

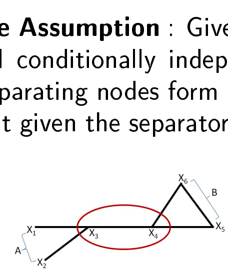

Figure 3: Separator Set separating A and B

The **Hammersley-Clifford Theorem** states that a probability distribution function satisfies pairwise Markov property with respect to an undirected graphical model if the distribution function can be factorized according to the graph.

Graphical Model Inference: There can be several kinds of inference desired from a given graphical model for a set of variables $\overrightarrow{X} = x_1, x_2, ...x_p$ like

1. finding $P(\overrightarrow{X})$. This is hard as finding the normalizing constant Z is hard as it requires a summation over all possible configurations of $\overrightarrow{X}$.
2. $A \subseteq V$ (vertex set). Finding $P_A(X_A)$
3. $A, B \subseteq V$. Finding $P(X_A \mid X_B)$
4. $argmax_{\overrightarrow{x}} P(\overrightarrow{x})$. Finding maximum a posteriori (MAP)

Applications:

1. Constraint Satisfaction:

$$X_1, X_2, ..X_p \text{ where } X_k \in \{0, 1\}$$

$$\psi_{1,2,3}(X_1, X_2, X_3) = \begin{cases} 0 & (X_1, X_2, X_3) = 001, \\ 1 & \text{otherwise} \end{cases}$$

2. Signal Decoding: As seen in Fig. 4, we find the configuration of X corresponding to the maximum probability.

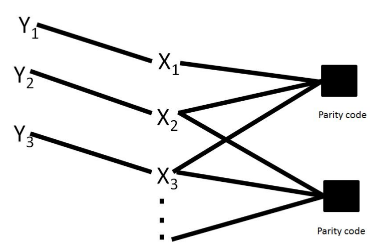

Figure 4: Signal decoding with parity check

$$X_1, X_2, ..X_p \qquad\qquad\qquad Y_1, Y_2, ...Y_p$$

$$\psi_{1,2,3}(X_1, X_2, X_3) = \begin{cases} 1 & \text{if check(parity code) is satisfied,} \\ 0 & \text{otherwise} \end{cases}$$

$$\psi_1(X_1) = P(Y_1 \mid X_1) \in \{0, 1\} \text{ noise probability}$$

$$\psi_f(X_f) = \{0 \text{ or} 1\} \text{ parity code configuration}$$

$$P(X) = \frac{1}{Z} \prod_s \Psi_s(X_s) \Psi_f(X_f)$$

3. Hidden Markov Model: This has application in vision and speech recognition and represented by the following graphical model.

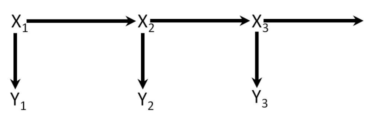

Figure 5: Hidden Markov Model