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 $u \rightarrow v$

$$\Pi(S) \equiv \text{parents of node } S$$

$$P_s(X_s | X_{\pi(S)}) \equiv \text{denotes the graphical model of the DAG. }$$

$$P(X) \equiv P(X_1, X_2, \ldots, X_n) = \prod S P_S(X_S | X_{\pi(S)})$$

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

$$P(\vec{X}) = P(X_1) P(X_2 | X_1) P(X_3 | X_1)$$

$$\int_{X_1} \int_{X_3} \int_{X_2} P(X_1) P(X_2 | X_1) P(X_3 | X_1) dX_2 dX_3 dX_1 = \int_{X_1} \int_{X_3} P(X_1) P(X_3 | X_1) dX_3 dX_1 = 1$$

**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.
is called Maximal Clique.

\[ C_G \equiv \text{set of all possible cliques in graph } G \]
\[ c \in C_G, \Psi_c(X_c) \equiv \text{compatibility function} \]
\[ P(X) = \frac{1}{Z} \prod_{c \in C_G} \Psi_c(X_c) \text{ where } Z \text{ is the normalising constant} \]

**Factor Graph Model** : Here the graph is considered to constitute of a product of factors \( \Psi \) 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 \((X_3, X_4, X_5, X_6)\)

**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 \( X_4 \).

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 $\overrightarrow{X} = x_1, x_2, \ldots, x_p$ 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. $\arg\max \overrightarrow{x} P(\overrightarrow{x})$. Finding maximum a posteriori (MAP)

**Applications:**

1. Constraint Satisfaction:
   $X_1, X_2, \ldots, X_p$ 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 $X$ corre-
   sponding to the maximum probability.

![Figure 4: Signal decoding with parity check](image)

3
\[ X_1, X_2, \ldots, X_p \quad Y_1, Y_2, \ldots, Y_p \]

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

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

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

\[
P(X) = \frac{1}{Z} \prod_s \psi_s(X_s) \psi_f(X_f)
\]

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

![Figure 5: Hidden Markov Model](image-url)