φ

Markov Networks

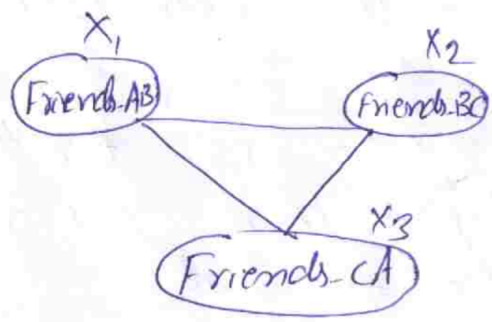
CSL 865

Oct 3, 4

Markov Networks:-

Undirected Graphical Models of data :-

Motivation:- May need to represent undirected dependences:-



Transitive relationship between $x_1, x_2 \wedge x_3$

UnDirected graph ~~&~~ represents these dependences well. Directed graph :- issue with cycles.

Equivalent of a CPT in Markov network is a potential functions. A potential function is defined over ~~the~~ the cliques in the Markov network graph.

Example:-

$\phi(x_1, x_2, x_3)$

X_1	X_2	X_3	ϕ	
0	0	0	10	ϕ_{000}
0	0	1	1	ϕ_{001}
0	0	0	1	ϕ_{010}
0	1	1	1	ϕ_{011}
1	0	0	1	ϕ_{100}
1	0	1	1	ϕ_{101}
1	1	0	10	ϕ_{110}
				$\phi_{11..}$

$\phi(X_1, X_2, X_3)$: Potential function

ϕ is allowed to be any non-negative function of the variables in the clique over which it is defined.

The parameters represent the odds associated with different states.

i.e. $\dfrac{\phi_{000}}{\phi_{100}} = \dfrac{10}{1} = 10$ mean that

state ~~θ~~ $X_1=0, X_2=0, X_3=0$ is ten times more likely to happen than $X_1=1, X_2=0, X_3=0$ given everything else ~~as~~ same.

$\phi: Dom(X_1) \times Dom(X_2) \times Dom(X_3) \rightarrow R^+ \cup \{0\}$

Markov network :- A ~~m~~ Markov network is an undirected graph $G = (V, E)$ s.t. for every maximal clique C_j in G, there is an associated potential function $\phi_c(X_c)$ ~~define~~ (X_{C_j} denotes subset of variable in C_j).
The distribution defined by a Markov network is given by:-

$$P(X=x) = \frac{1}{Z} \prod_{c_j \in C} \phi_{c_j}(X_{c_j} = x_{c_j})$$

Normalization constant

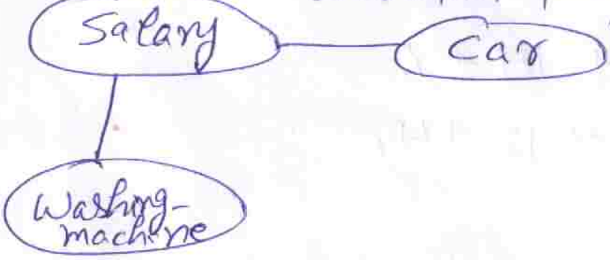
C is set of cliques on G.

x_{c_j} is x restricted to variables in X_{c_j}.

$$\boxed{Z = \sum_{x'} \prod_{c_j \in C} \phi_{c_j}(X_{c_j} = x'_{c_j})}$$

Example :-

Note:- Distribution is invariant of to a constant factor scaling potential values i.e. each term in a potential function can be multiplied by $k > 0$ without changing distribution



$\phi_{S,C}(S,C)$

S	C	$\phi_{S,W}$
0	0	8
0	1	2
1	0	4
1	1	6

$\phi_{S,W}(S,W)$

S	W	$\phi_{S,W}$
0	0	7
0	1	3
1	0	2
1	1	8

$$P(S=0, C=0, W=0) = \frac{1}{Z} \phi_{S,C}(0,0) \cdot \phi_{S,W}(0,0)$$

$$= \frac{1}{Z} \cdot 8 \times 7$$

$$\cancel{Z = \phi_{S,C}(0,0)\phi_{S,W}(0,0) + \phi_{S,C}(0,1)\phi_{S,W}(0,0) + \phi_{S,C}(0,0)\,\phi_{S,C}(0,1) + \phi_{S,C}(0,1)\,\phi_{S,W}(0,1)}$$

$$Z = \sum_{S,C,W \in \{0,1\}} \phi_{S,C}(S,C)\ \phi_{S,W}(S,W)$$

$$= \sum_{S \in \{0,1\}} \sum_{C,W \in \{0,1\}} \phi_{S,C}(S,C) \cdot \phi_{S,W}(S,W)$$

$$= \sum_{S \in \{0,1\}} \phi(S,0) \cdot \phi(S,0) + \phi(S,0) \cdot \phi(S,1) + \phi(S,1)\ \phi(S,0) + \phi(S,1) \cdot \phi(S,1)$$

$$= 56 + 24 + 14 + 6 + 8 + 32 + 12 + 48$$

$$= 100 + 100 = 200$$

$\cancel{P(0,0}$ $P(S=0, C=0, W=0) = \frac{56}{200}$

$\phi(S, C, W)$

S,	C,	W	ϕ
0	0	0	56
0	0	1	24
0	1	0	14
0	1	1	6
1	0	0	8
1	0	1	32
1	1	0	12
1	1	1	48

~~Variable elimin~~

Can answer arbitrary questions of the form:-

$$P(Y \mid X)$$

$P(Y$ ⇓ Query, $X)$ → Evidence

$$= \frac{\sum_{Z} P(X,Y,Z)}{\sum_{Y,Z} P(X,Y,Z)}$$

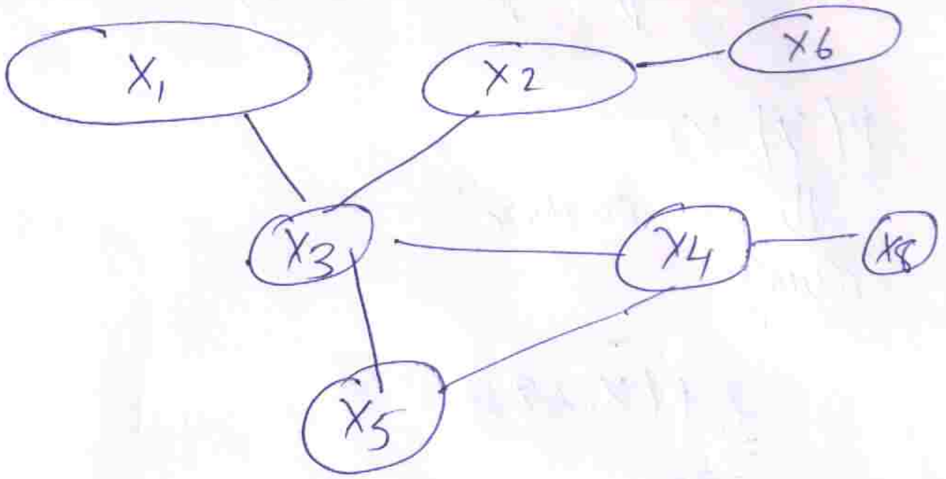
Variable elimination can be applied

What kind of conditional independences does a Markov network entail?

Let $Nbr(X_i)$ denote ~~se~~ the set of neighbors of $\underline{X_i}$.

Given a set of nodes X, $MB(X)$ be defined as the set of nodes s.t. $Y_k \in MB(X)$ if $Y_k \notin X$ & ~~$Y_k \in$~~ $Y_k \in Nbr(X_i)$ for some $X_i \in X$.

i.e. $MB(X)$ is set of all the neighbors of nodes in X (excluding the nodes in X).



$MB(X_5) = \{X_3, X_4\}$

$MB(X_3) = \{X_1, X_2, X_4, X_5\}$

$MB(\{X_1, X_2\}) = \{X_3, X_6\}$

$MB(X)$ disconnect X from the rest of the network

In a Markov network (distribution defined by a Markov network)

$X \perp Z \mid MB(X)$

where $Z \cap X = \phi$, $Z \cap MB(X) = \phi$

i.e a node (or set of nodes) is independent of everything else in the graph given its Markov blanket.

Hammersley Clifford Theorem:-

Consider a set of variables $\{X_1 \ldots X_n\}$ with joint distribution defined over them as $P(X_1 \ldots X_n)$. Let $G = (X, E)$ be s.t. $P(X_1 \ldots X_n) > 0$. ~~or~~ a Markov network defined over these variables. ~~with potential functions ϕ_{c_j} if $\phi_{c_j} > 0 \ \forall j$. then (i.e. potential functions are strictly positive)~~

Set of conditional independences imposed by network structure hold

i.e. $X \perp Z \mid MB(X)$

$Z \notin X, Z \notin MB(X)$

$\Updownarrow$

Joint distribution $P(X_1 \ldots X_n)$ can be factored as a product of potential functions defined over maximal cliques on graph G.

Learning the parameters :-

No closed form expression

$$P(X) = \frac{1}{Z} \prod_{c_j \in C} \phi(x_{c_j})$$

Parameters → potential function values

K cliques ⇒ 2^k parameters.

(some redundancy because of scaling factor)

$$\boxed{\begin{aligned} & P(\{x^{(i)}\}_{i=1,m}) \\ &= \prod_{i=1}^{m} \frac{1}{Z} \prod_{c_j \in C} \phi(X_{c_j} = x^{(i)}_{c_j}) \\ &= \frac{1}{Z} \prod_{i=1}^{m} \prod_{c_j \in C} \phi(x_{c_j} = x^{(i)}_{c_j}) \end{aligned}}$$

Use Gradient Descent.

Maximum likelihood estimator

Can learn the structure as well.

$\boxed{\text{Widely Used in Vision}}$:→ $\boxed{\text{Pairwise Networks}}$