10-708 (L3)

Readings

- Key equations; some intuition on ~~the~~ points

Koller (2009) / Jordan (2003) (need to shorten + hence select + concise)

- BN cannot model a distri that satisfies $(A \perp C \mid \{B, D\})$ and $(B \perp D \mid \{A, C\})$ only.
- UGM $\rightarrow$ undirected edges (prob. interaction)
- similar to BN; parametrisation of MN < local interactions ↓ combine. global model

- product of local factors, normalised
- normalising constant $\rightarrow$ partition function (MRF in statistical physics)
- MN - connection between factorisation and independence properties

D.4.3 (Gibbs distri)

- An undirected graphical model represents a distri $P(X_1, \ldots, X_n)$ defined by an undirected graph H, and a set of (positive) potential functions ψ_c associated with cliques of H s.t.

$$P(X_1, \ldots, X_n) = \frac{1}{Z} \prod_{c \in C} \psi_c(\underline{x}_c) \qquad Z = \sum_{x_1, \ldots, x_n} \prod_{c \in C} \psi_c(\underline{x}_c)$$

D4.4 (MN factoris.)

- we say a distri $P(X_1, \ldots, X_n)$ with $\underline{\psi} = \{\psi_1(\underline{x}_1), \ldots, \psi_C(\underline{x}_C)\}$ factorises over a MN H if each $\underline{x}_c$ $(c = 1, \ldots, C)$ is a complete subgraph (*) of H
- Factors that parametrise MN are clique potentials
- (*) - informally, subs. with idea of a maximal clique
- (*) wlog, parametrisation using maximal cliques obscures structure in original set of factors

- informally: (Jordan)
  - clique (of a graph) $\rightarrow$ fully connected subset of nodes
  - maximal cliques $\rightarrow$ cliques that cannot be extended to include additional nodes without losing property of being fully connected

- Formally :-
- For $G = \{E, V\}$, a complete subgraph (clique) is a subgraph $G' = \{V' \subseteq V, E' \subseteq E\}$ such that nodes V' are fully interconnected

· A maximal clique is a complete subgraph s.t. any superset $V'' \supset V'$ is not complete

· A sub-clique is not necessarily a maximal clique

! sub-cliques - can be edges, singletons

Koller (2009): see Box 4.B. - MN for CV

4.3. MN independencies

- Similar to Kollerian pres. → flows of prob. influence/active trails

D 4.8

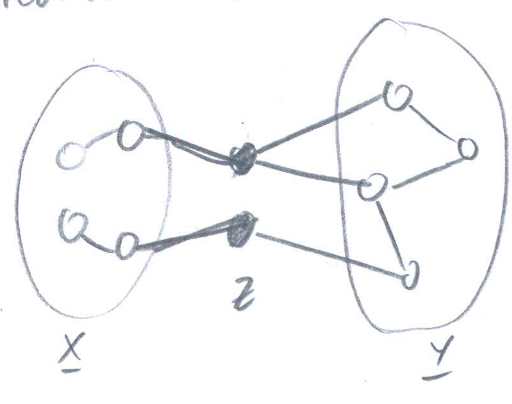
- Let H be an MN structure
- Let $X_1 - \ldots - X_k$ be a path in H.
- Let $\underline{Z} \subseteq \mathcal{X}$ be a set of observed variables
- The path $X_1 - \ldots - X_k$ is active given $\underline{Z}$ if none of the X_is $k=1,\ldots,k$ is in $\underline{Z}$

Allows a definition of separation

D 4.9 (separation/global independencies)

- We say a set of nodes $\underline{Z}$ separates $\underline{X}$ and $\underline{Y}$ in H, denoted $sep_H(\underline{X} : \underline{Y} | \underline{Z})$, if there is no active path between any node $X \in \underline{X}$ and $Y \in \underline{Y}$ given $\underline{Z}$.
- We define the global independencies associated with H to be:-

$$I(H) = \{(\underline{X} \perp \underline{Y} | \underline{Z}) : sep_H(\underline{X}; \underline{Y} | \underline{Z})\}$$



Remark:

Independencies $I(H)$ are precisely those guaranteed to hold for every distri P over H.

Sep. criterion sound for detecting indep. properties over H.

(\\) BN → connection with independence prop implied by MN structure

+

factorising a distri over the graph

(\\) of both representation

(*)

theorems for BN, equivalence of:- Gibbs fact. of distri P over a graph H ⟺ H is an I-map for P (that is P satisfies Markov ass. $I(H)$)

4.3.1.1. (soundness) ⊛ ⊛

(ii) to T.3.3. i.e. Gibbs distri satisfies independencies associated with a graph.

i.e. soundness of separation — (factorisation accord. G $\underset{4.2/3.1}{\overset{4.1/3.2}{\rightleftharpoons}}$ C.I.)

T.4.1.

- Let P be a distri over X, and H an MN structure over X
- If P is a Gibbs distri that factorises over H, then H is an I-map for P

- other direction i.e. C.I. of distri → factorisation: Hammersley-Clifford Theorem
- unlike for BN, HCT does not hold in general.
- Require additional assumption that P is a positive distri

T.4.2

- Let P be a positive distri over X, and H a Markov network graph over X.
- If H is an I-map for P, then P is a Gibbs distri that factorises over H.

(*). For positive distris, the global independencies imply that distri factorises according to the network structure.

(*) For this class of distributions, we have that a distribution P factorises over a Markov Network if and only if H is an I-map of P.

⊛⊛ lecture def ⊛ An UG H is an I-map for a distri P if $I(H) \subseteq I(P)$ i.e. P entails $I(H)$

⊛ If P is a Gibbs distri over H, then H is an I-map of P.

- soundness of separation as criterion for detecting independencies in MN.
- any distri that factorises over G satisfies the independence ass. implied by separation
- completeness - strong version of completeness does not hold
- It is NOT the case that every pair of nodes X and Y that are not separated in H are dependent in every distribution P which factorises over H (use weak)

T.4.3

- Let H be a MN structure.
- If X and Y are not separated given Z in H, then X and Y are dependent given Z in some distri P that factorises over H.

- Same arguments as 4.3.5 to conclude:-

(*) For almost all distri P that factorise over H (all distri. except for a set of measure 0 in space of factor param.), we have $I(P) = I(H)$.

- Our defn of $I(H)$ is maximal one.
- For any independence assertion that is not a consequence of separation in H, we can always find a counterexample distri P that factorises over H.

4.3.2. Indep. revisited

BN: local indep. (each node is indep. of nondesc. given parents)
global indep (induced by d-sep).

- showed that these are equiv, in the sense of one implies the other.

Q: (\\) to BN; can we provide local indep. induced by MN, analogously to local indep of BN

(*): 3 diff poss. defin. of independencies associated with network structure
two local, one global in def 4.9.

4.3.2.1. - Local Markov ass.

D.4.10 → intuitively, two variables directly connected; potential for direct correlation in an unmediated way.
- conversely, two vars not directly linked, some way of rendering c.i.
- X and Y indep given all other nodes.

D.4.10 (Pairwise independencies)

Let H be a MN
We define pairwise independencies associated with H to be:-

$I_p(H) = \{(X \perp Y \mid \mathcal{X} - \{X,Y\} : \{X,Y\} \notin \underset{H}{\mathcal{E}}\}$

D.4.11 → analogue to local indep. associated with B.N.

D.4.11. (Markov blanket)

For a given graph H, we define the Markov blanket of X in H, $MB_H(X)$ to be the neighbours of X in H.

We define the local independencies associated with H to be:-

$I_\ell(H) = \{(X \perp \mathcal{X} - \{X\} - MB_H\{X\} \mid MB_H(X) : X \in \mathcal{X})\}$

i.e. local independencies state that X is indep. of nodes in graph given immed. neighbours.

- We will show that these local indep. ass. hold for any distri that factorises over H, so that X's Markov blanket in H truly does sep. it from all other variables.

4.3.2.2. - Relationships between Markov properties

- 3 sets of indep. assertions assoc. with network structure H.
- For general distri $\mathcal{I}_p(H)$ is weaker than $\mathcal{I}_\ell(H)$ is weaker than $\mathcal{I}(H)$.
- ~~All 3 are eq~~ (*) All 3 are equivalent for <u>positive distri</u>

Prop 4.3

- For any MN H, and any distri P, we have that if $P \models \mathcal{I}_\ell(H)$ then $P \models \mathcal{I}_p(H)$

Prop 4.4

- For any MN H, and any distri P, we have that if $P \models \mathcal{I}(H)$ then $P \models \mathcal{I}_\ell(H)$

T4.4

- Let P be a positive distri. If P satisfies $\mathcal{I}_p(H)$, then P satisfies $\mathcal{I}(H)$.

Corollary 4.1

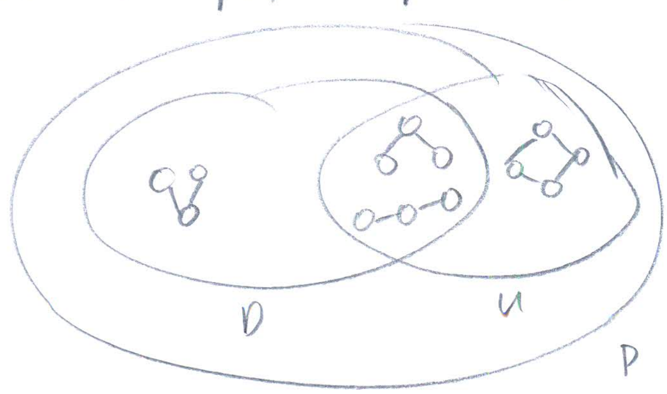
- The following statements are equivalent for a positive distri P.

1. $P \models \mathcal{I}_\ell(H)$
2. $P \models \mathcal{I}_p(H)$
3. $P \models \mathcal{I}(H)$

- ~~Pt~~ <u>P-maps</u>
- An MN H is a perfect-map for P if for any $\underline{X}, \underline{Y}, \underline{Z}$, we have that

$$sep_H(\underline{X} : \underline{Y} \mid \underline{Z}) \iff P \models (\underline{X} \perp \underline{Y} \mid \underline{Z})$$

- <u>Theorem:</u> Not every distri has a perfect map as UGM



From hereon, Jordan (2003).

- Exponential form

- constraining clique potentials to be +ve, inconvenient possibly

(Q)(L): what effect does this have on equivalence of local and global Markov properties?

- represent a clique potential as:- (*) $\psi_c(x_c) = \exp\{-\phi_c(x_c)\}$

- $\phi_c(x_c)$ - a 'potential'

- Additive structure: $p(x) = \frac{1}{Z} \prod_{c \in C} \exp\{-\phi_c(x_c)\}$

$$= \frac{1}{Z} \exp\{-\sum_{c \in C} \phi_c(x_c)\} = \underbrace{\frac{1}{Z} \exp\{-H(x)\}}_{\text{'Boltzmann distri'}}$$

- $H(x) := \sum_{c \in C} \phi_c(x_c)$

- $H(x)$ - free energy

- some notable forms / graph topologies:-

Boltzmann Machines

- Fully connected graph, pairwise edge pot., binary valued nodes (-1, 1)

$p(x_1, x_2, x_3, x_4) = \frac{1}{Z} \exp\{\sum_{i,j} \phi_{ij}(x_i, x_j)\} = \frac{1}{Z} \exp\{\sum_{i,j} \theta_{ij} x_i x_j + \sum_i \alpha x_i + C\}$

- overall energy fn:-

$H(x) = \sum_{ij} (x_i - \mu) \theta_{ij} (x_j - \mu) = (x - \mu)^T \underline{\theta} (x - \mu)$

- Low technique to learn this model, recover graph str. from data.

Ising Models

- Model energy of a physical system (atom interaction)

- Regular grid topology

- Multi-state → Potts.

$p(X) = \frac{1}{Z} \exp\{\sum_{i,j \in N_i} \theta_{ij} X_i X_j + \sum_i \theta_{i0} X_i\}$

Restricted Boltzmann Machines

$$p(\underline{x},\underline{h} \mid \theta) = \exp\left\{ \sum_i \theta_i \phi_i(x_i) + \sum_j \theta_j \phi_j(h_j) \sum_{i,j} \theta_{i,j} \phi_{i,j}(x_i, h_j) - \underline{A}(\underline{\theta}) \right\}$$

- see paper

CRFs

- undirected graph rep. ; encode cond. distri $p(Y \mid X)$ Y_i - target $\underline{X}$ - obs.

$$p_\theta(y \mid x) = \frac{1}{Z(\theta, x) \exp\left\{ \sum_c \theta_c f_c(x, y_c) \right\}}$$