3/10/14

MN/MRF INFERENCE

(1) Setup [Compare to BN Inference Notes]

→ Random Vars $\vec{X} = \{X_1, ---, X_n\}$ [all categorical for now]

→ Given MRF induced graph $G = (V, E)$ undirected edges

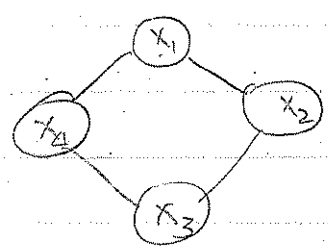
→ Given a set of Factors F defined on a set of clique $\vec{C}$

$$F = \{ \Psi_i(C_i) : \forall C_i \in \vec{C} \}$$

each clique $C_i \in \vec{C}$ is a subset of variable-indices

e.g. $\vec{X} = \{X_1, X_2, X_3, X_4\}$

$G \equiv$



$$\vec{C} = \{C_1, C_2, C_3, C_4\}$$
$$C_1 = \{X_1, X_2\}$$
$$C_2 = \{X_2, X_3\}$$
$$C_3 = \{X_3, X_4\}$$
$$C_4 = \{X_4, X_1\}$$

$$F = \{ \Psi_1(C_1), \underbrace{\Psi_2(C_2)}_{\substack{\parallel\parallel \\ \Psi_2(X_2, X_3)}}, \Psi_3(C_3), \Psi_4(C_4) \}$$

→ [Optionally] Given some "evidence" variables $\vec{E} \subseteq \vec{X}$
and their states $\vec{E} = \vec{e}$ ie $E_1 = e_1$ $E_2 = e_2$ . . . etc

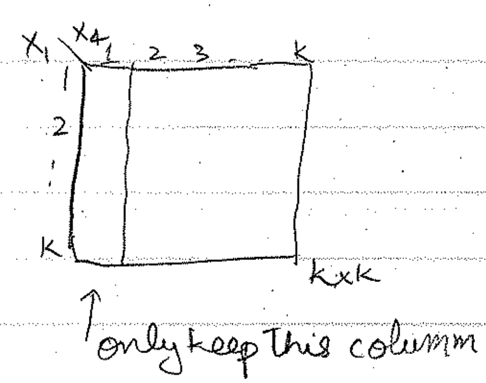
We say "Optionally" because:

(a) Even without evidence variables, MRF inference querries are non-trivial & hard. (Unlike BNs)

(b) Typically evidence variables can just be "instantiated" & factors simply "reduced" to construct a new MRF with no evidence variables.

e.g in previous 4-cycle example
say $\vec{E} = \{X_4\}$ $\vec{e} = \{1\}$
So we know $X_4 = 1$
All we have to do

$$\underbrace{\tilde{\varphi}_4(X_1)}_{\text{New factor}} = \varphi_4(X_1, X_4 = 1)$$



&

$$\tilde{\varphi}_3(X_3) = \varphi_3(X_3, X_4 = 1)$$

Now we have an evidence instantiated MRF on $\{X_1, X_2, X_3\}$

→ [Optionally] Given some Query variable
$\vec{Y} \subseteq \vec{X}$
$\{Y_1, Y_2 \ldots\}$

Types of MRF Inference Queries:

→ CONDITIONAL PROB / MARGINAL Inference
Find $P(\vec{Y} = \vec{y} \mid \vec{E} = \vec{e})$

→ MAP Inference
Find argmax $P(\vec{X} \setminus \vec{E} \mid \vec{E} = \vec{e})$

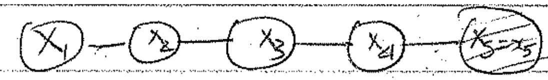
[We will not study this in class]
→ Marginal MAP Inference
Find $\underset{\vec{y}}{\operatorname{argmax}}\ P(\vec{Y} = \vec{y} \mid \vec{E} = \vec{e})$

$$= \underset{\vec{y}}{\operatorname{argmax}} \sum_{\vec{x}} P(\vec{Y} = \vec{y}, \vec{X} \setminus \{\vec{Y} \cup \vec{E}\} = \vec{x} \mid \vec{E} = \vec{e})$$

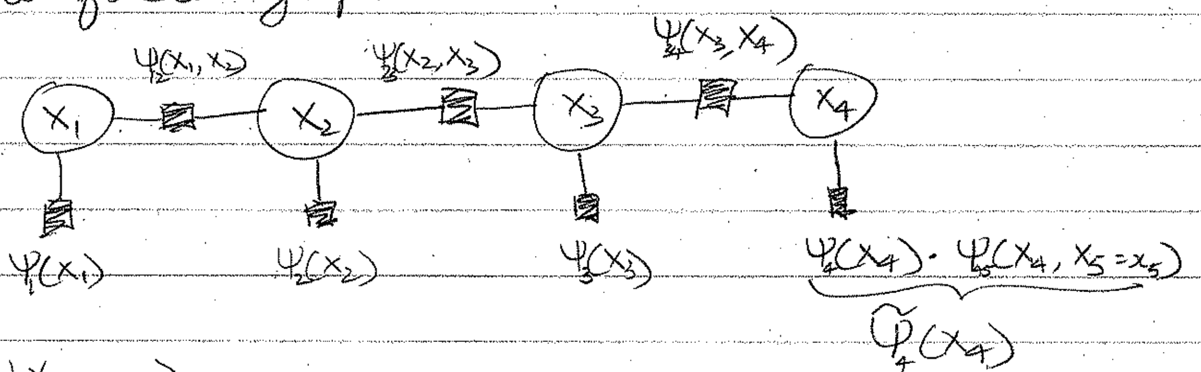
(2) Algorithms
→ Exact
→ VE
→ BP on Junction Trees
→ Graph-Cuts [for MAP only & special cases]
→ Approximate
→ BP
→ Variational Inference: Mean Field
→ LP relaxations [for MAP only]

(3) Exact Inference: Variable Elimination on MRFs / Factor Graphs

→ Basically algorithm unchanged from BNs
  → Pick elimination ordering [Min-fill heuristic]
  → For var X_i to be eliminated
    → multiply factors that depend on X_i
    → $\underbrace{\text{(sum out)}}_{\text{Marginal}}$ or $\underbrace{\text{(max out)}}_{\text{MAP}}$ X_i to create new factor

Example: Chain MRF 

Instantiate evidence
& write factor graph:



Find: $P(X_1 \mid X_5 = x_5)$
Elim. Ordering: X_4, X_3, X_2

Eliminating	All Factors	Elimination	New Factor
X_4	$\psi_i(X_i)\ \psi_{ij}(X_i, X_j)$	$\sum_{x_4} \tilde{\psi}_4(X_4)\psi_{34}(X_3, X_4)$	$g_{4\to3}(X_3)$
X_3	$\psi_1, \psi_2, \psi_3, \psi_{12}, \psi_{23}, g_{4\to3}$	$\sum_{x_3} \psi_3(X_3)\psi_{23}(X_2, X_3) g_{4\to3}(X_3)$	$g_{3\to2}(X_2)$
X_2	$\psi_1, \psi_2, \psi_{12}, g_{3\to2}$	$\sum_{x_2} \psi_2(X_2)\psi_{12}(X_1, X_2) g_{3\to2}(X_2)$	$g_{2\to1}(X_1)$

Finally,

$$P(X_1 | X_5 = x_5) = \frac{P(X_1, X_5 = x_5)}{P(X_5 = x_5)} \propto \psi_1(X_1)\, g_{2\to 1}(X_1)$$

$$= \frac{\psi_1(X_1)\, g_{2\to 1}(X_1)}{\sum_{x_1} \psi(X_1 = x_1)\, g_{2\to 1}(X_1 = x_1)}$$

Question: Complexity of VE?

If n-vars Each takes k-states

$O(k^2)$-time foreach elimination

$(n-1)$ eliminations

total: $O(nk^2)$

What if I want to compute ALL marginal $P(X_i)$

Naively, $O(n^2 k^2)$ [Run VE n-times]

Can we do better? Say $O(nk^2)$?

Yes!

with BP. [Next]

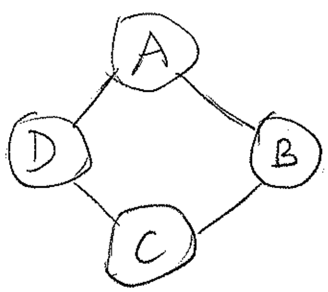
④ Belief Propagation

→ Intuitively, passes "messages" between Factors / Variables each communication their local beliefs.

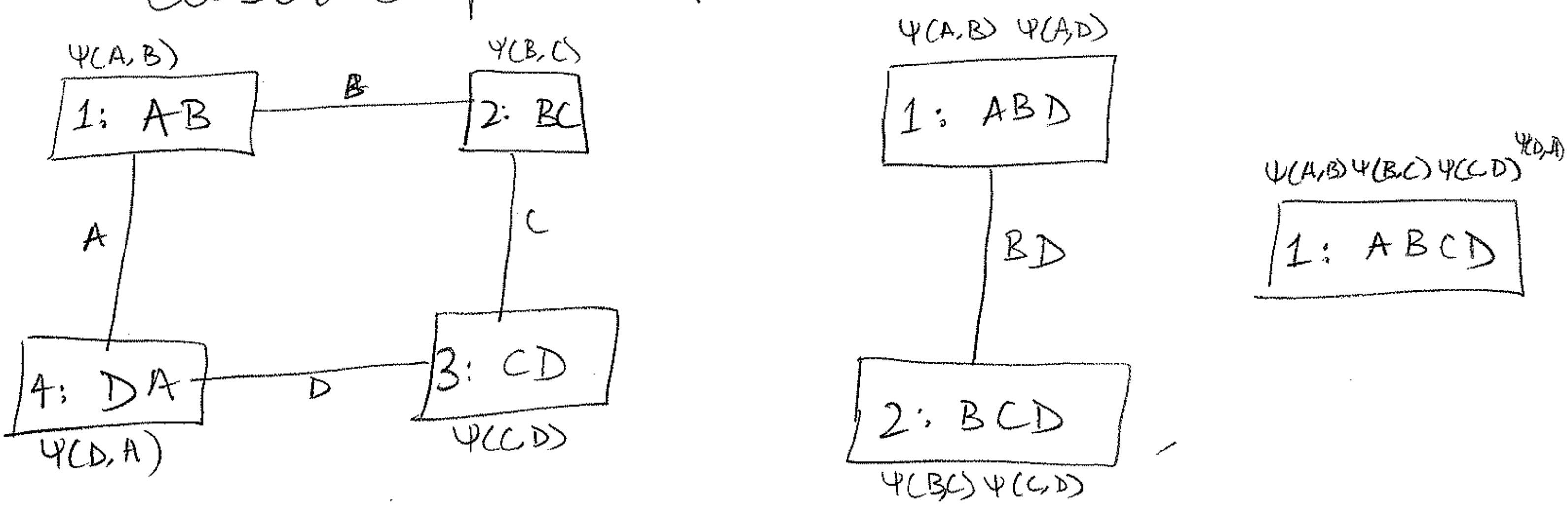
→ [Def:] Cluster Graph: An undirected graph where each node is a subset of Vars $\{C_1, C_2 \dots\}$ each edge $C_i - C_j$ is associated with a "separator set" of variables

$$S_{ij} \subseteq C_i \cap C_j$$

Example: MRF



Cluster Graphs examples



→ Given Factors $F = \{\psi_i(\cdot)\}$, we assign or "place them" to/at clusters where they can be placed, ie

$$\text{Scope}[\psi_i] \subseteq C_i$$

Examples shown above.

Generalized BP

→ On each undirected $C_i - C_j$ edge in Cluster graph, we will pass "messages"

$$\underbrace{\delta_{i\to j}(S_{ij})}_{\text{Message from } i \text{ to } j \text{ about } S_{ij}} = \sum_{C_i \setminus S_{ij}} \underbrace{\prod \psi_i(\cdot)}_{\text{Factors placed at Cluster } i} \cdot \underbrace{\prod_{k\in N(i)\setminus j} \delta_{k\to i}(S_{ij})}_{\text{Product of incoming messages at } i \text{ EXCEPT the one coming from } j}$$

[Some papers use $\delta_{ij}(S_{ij})$]

Message from i to j about S_{ij}

Factors placed at Cluster i

Product of incoming messages at i EXCEPT the one coming from j (to which we are sending this message)

Thus, Generalized BP algorithm

→ Make Cluster Graph (How? Next)

→ Assign Factors $F = \{\psi_i(\cdot)\}$ to clusters [How? As long as scope is satisfied, doesn't matter]

→ Initialize Messages $\delta_{i\to j}(S_{ij}) = \vec{1}$

→ While Not Converged [Defined next]

→ For loop on edges in some schedule

→ Update $\delta_{i\to j}$

Final
→ Belief or Pseudo-marginals $\mu_i(C_i) = \prod_i \psi_i(\cdot) \prod_{k\in N(i)} \delta_{k\to i}(S_{ij})$

→ Properties of Cluster Graphs

→ Family Preserving: $\forall \varphi_j \in F$, $\exists i$ s.t $Scope[\varphi_j] \subseteq C_i$

Basically, I can place all factors somewhere.

→ RIP (Running Intersection Prop)

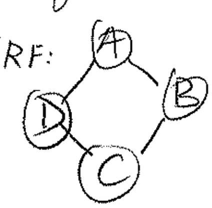
If $X \in C_i$ & $X \in C_j$, then $\exists$ a unique $C_i - C_j$ path with $X_i \in S_{uv}$ $\forall$ u,v edges on the path.

* Basically, communication about X should be possible (no disjoint communities who know something about X) AND no cycles in X $\underbrace{\text{(can't re-enforce your own beliefs)}}_{\text{although this problem is still possible. How?}}$

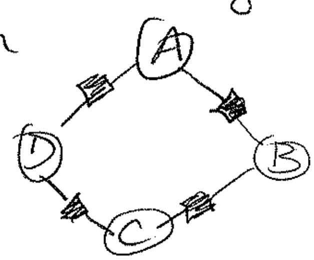
→ Most Common Cluster Graph: Bethe Cluster Graph
[Almost looks like a factor graph but isn't]

→ Make each factor it's own cluster + Unary clusters

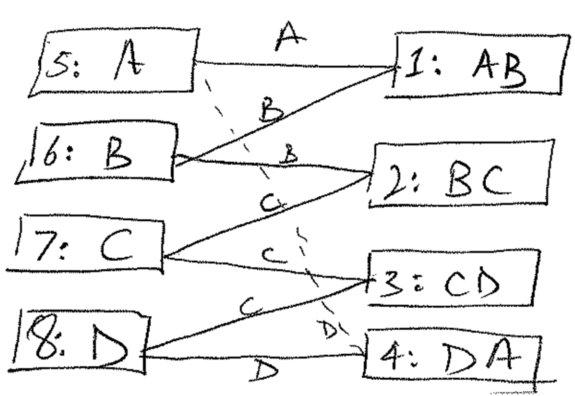
→ Example: MRF:



Factor Graph



Bethe Cluster Graph:



Now, 2 types of Messages

→ Variable to Factor

$$\delta_{5\to 1}(A) = \prod_{\substack{k \text{ is a factor} \\ \text{involving } A \\ \text{Except } 1}} \delta_{k\to 1}(A) = \delta_{4\to 5}(A)$$

Nothing to marginalize $C_5 = S_{51} = \{A\}$

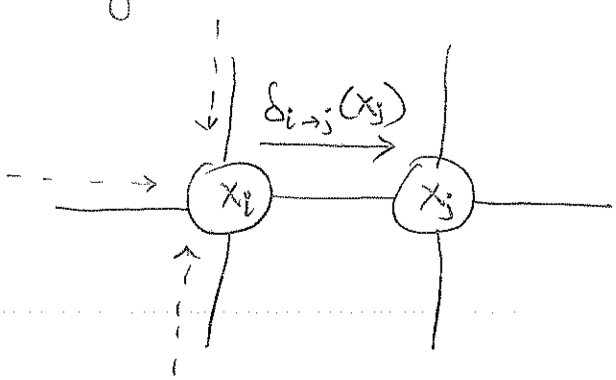
→ Factor to Variables

$$\delta_{1\to 5}(A) = \sum_{B} \psi(A,B) \cdot \underbrace{\delta_{6\to 1}(B)}_{\text{All incoming messages except from 5}}$$

LOOPY BP on PAIRWISE MRFs
[Simplest Instantiation of BP]

→ When all factors are only pairwise [+ optionally unary factors]

BP can be run VERY VERY easily directly on the MRF Graph. (Snippet below)



$$\delta_{i\to j}(X_j) = \sum_{X_i} \psi_i(X_i)\, \psi_{ij}(X_i, X_j) \cdot \prod_{k \in N(i) \setminus j} \delta_{k\to i}(X_i)$$