

# Machine Learning

4771

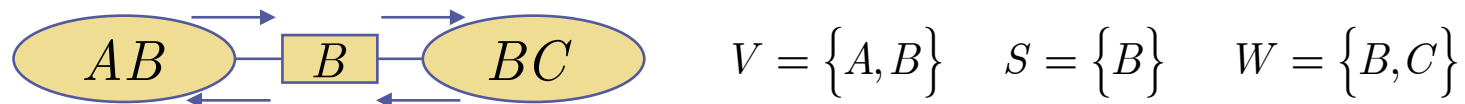
Instructor: Tony Jebara

# Topic 18

- The Junction Tree Algorithm
- Collect & Distribute
- Algorithmic Complexity
- ArgMax Junction Tree Algorithm

# Review: Junction Tree Algorithm

- Send message from each clique *to* its separators of what it thinks the submarginal on the separator is.
- Normalize each clique by incoming message *from* its separators so it agrees with them



**If agree:**  $\sum_{V \setminus S} \psi_V = \phi_S = p(S) = \phi_S = \sum_{W \setminus S} \psi_W$     **...Done!**

**Else: Send message  
From V to W...**

$$\begin{aligned} \phi_S^* &= \sum_{V \setminus S} \psi_V \\ \psi_W^* &= \frac{\phi_S^*}{\phi_S} \psi_W \\ \psi_V^* &= \psi_V \end{aligned}$$

**Send message  
From W to V...**

$$\begin{aligned} \phi_S^{**} &= \sum_{W \setminus S} \psi_W^* \\ \psi_V^{**} &= \frac{\phi_S^{**}}{\phi_S^*} \psi_V^* \\ \psi_W^{**} &= \psi_W^* \end{aligned}$$

**Now they  
Agree...Done!**

$$\begin{aligned} \sum_{V \setminus S} \psi_V^{**} &= \sum_{V \setminus S} \frac{\phi_S^{**}}{\phi_S^*} \psi_V^* \\ &= \frac{\phi_S^{**}}{\phi_S^*} \sum_{V \setminus S} \psi_V^* \\ &= \phi_S^{**} = \sum_{W \setminus S} \psi_W^{**} \end{aligned}$$

# JTA with Evidence

- Example: if *evidence* is observed, say variable  $A=1$

Initialize as before...

$$\psi_{AB} = p(A, B) \quad \psi_{BC} = p(C | B) \quad \phi_B = 1$$

Update with slice...

$$\phi_B^* = \sum_A \psi_{AB} \delta(A = 1) = \sum_A p(A, B) \delta(A = 1) = p(A = 1, B)$$

$$\psi_{BC}^* = \frac{\phi_S^*}{\phi_S} \psi_{BC} = \frac{p(A = 1, B)}{1} p(C | B) = p(A = 1, B, C)$$

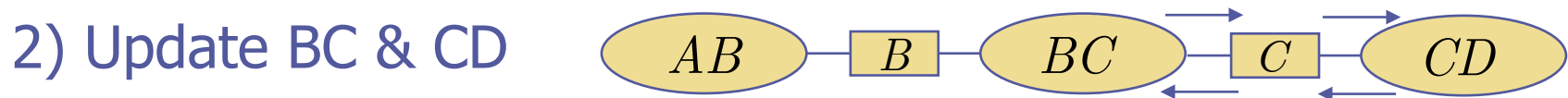
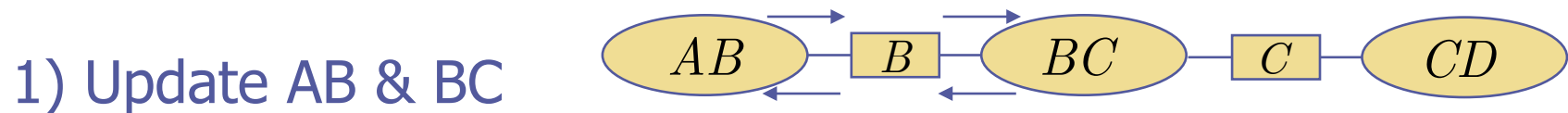
$$\psi_{AB}^* = \psi_{AB} = p(A = 1, B)$$

All  $\psi, \phi$  become marginals *conditioned* on evidence

$$p(B, C | A = 1) = \frac{\psi_{BC}^*}{\sum_{B,C} \psi_{BC}^*}$$

# JTA with many cliques

- Problem: what if we have more than two cliques?



- Problem: AB has not heard about CD!  
After BC updates, it will be inconsistent for AB
- Need to iterate the pairwise updates many times
- This will eventually converge to consistent marginals
- But, inefficient... can we do better?

# JTA: Collect & Distribute

- Trees: recursive, no need to reiterate messages mindlessly!
- Send your message only after hearing from all neighbors...

**initialize(DAG)** { **Pick root**  
                   **Set all variables as:**  $\psi_{C_i} = p(x_i | \pi_i), \phi_S = 1$  }

**collectEvidence(node)** {  
   **for each child of node** {  
     **update(node, collectEvidence(child));** }  
**return(node);** }

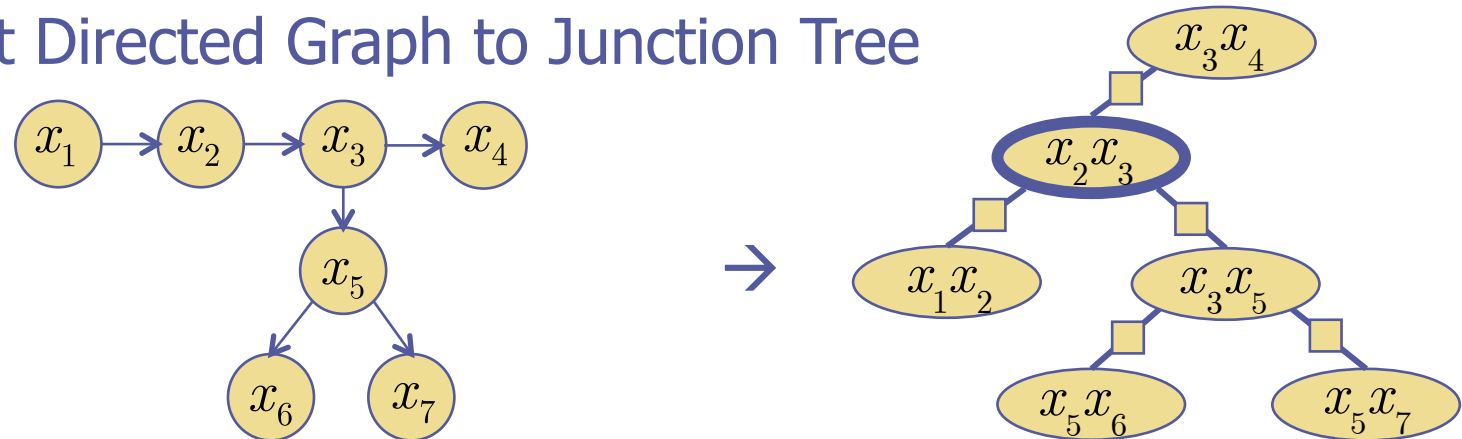
**distributeEvidence(node)** {  
   **for each child of node** {  
     **update(child, node);**  
     **distributeEvidence(child);** } }

**normalize(DAG)** {  $p(X_C) = \frac{\psi_C}{\sum_{X_C} \psi_C}, p(X_S) = \frac{\phi_S}{\sum_{X_S} \phi_S}$  }

**update(node  $\psi$ , evidence  $\phi$ )** {  $\psi_C^* = \frac{\phi_S}{\sum_{C \setminus S} \psi_C} \psi_C$  }

# Junction Tree Algorithm

- Convert Directed Graph to Junction Tree



- *Initialize* separators to 1 (and  $Z=1$ ) and set clique tables to the CPTs in the Directed Graph

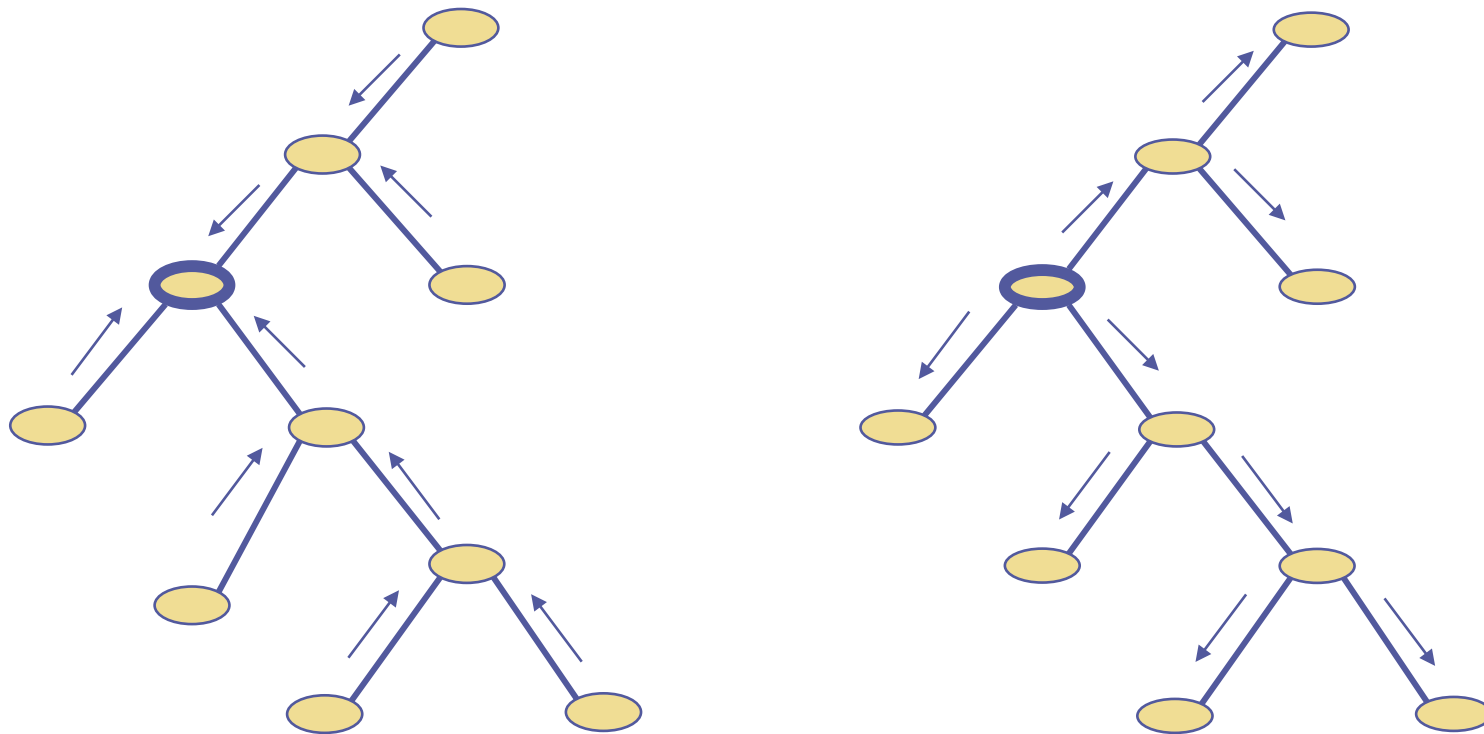
$$p(X) = p(x_1) p(x_2 | x_1) p(x_3 | x_2) p(x_4 | x_3) p(x_5 | x_3) p(x_6 | x_5) p(x_7 | x_5)$$

$$p(X) = \frac{1}{Z} \frac{\prod_C \psi(X_C)}{\prod_S \phi(X_S)}$$

$$= \frac{1}{1} \frac{p(x_1, x_2) p(x_3 | x_2) p(x_4 | x_3) p(x_5 | x_3) p(x_6 | x_5) p(x_7 | x_5)}{1 \times 1 \times 1 \times 1 \times 1}$$

# Junction Tree Algorithm

- JTA: 1)*Initialize* 2)*Collect* 3)*Distribute* 4)*Normalize*



- Note: leaves do not change their  $\psi$  during *collect*
- Note: the first cliques *collect* changes are parents of leaves
- Note: root does not change its  $\psi$  during *distribute*

# Algorithmic Complexity

- The 5 steps of JTA are all efficient:

## OFFLINE

### 1) Moralization

Polynomial in # of nodes

### 2) Introduce Evidence (fixed or constant)

Polynomial in # of nodes (convert pdf to slices)

### 3) Triangulate (Tarjan & Yannakakis 1984)

Suboptimal=Polynomial, Optimal=NP

### 4) Construct Junction Tree (Kruskal)

Polynomial in # of cliques

## ONLINE (for each query, new evidence, etc.)

### 5) Propagate Probabilities (Junction Tree Algorithm)

Polynomial (linear) in # of cliques, *Exponential* in Clique Cardinality

# ArgMax Junction Tree Algorithm

- We can also use JTA for finding the max not the sum over the joint to get argmax of marginals & conditionals
- Say have some evidence:  $p(X_F, \bar{X}_E) = p(x_1, \dots, x_n, \bar{x}_{n+1}, \dots, \bar{x}_N)$
- Most likely (highest p)  $X_F$ ?  $X_F^* = \arg \max_{X_F} p(X_F, \bar{X}_E)$
- What is most likely state of patient with fever & headache?

$$\begin{aligned}
 p_F^* &= \max_{x_2, x_3, x_4, x_5} p(x_1 = 1, x_2, x_3, x_4, x_5, x_6 = 1) \\
 &= \max_{x_2} p(x_2 | x_1 = 1) p(x_1 = 1) \max_{x_3} p(x_3 | x_1 = 1) \\
 &\quad \max_{x_4} p(x_4 | x_2) \max_{x_5} p(x_5 | x_3) p(x_6 = 1 | x_2, x_5)
 \end{aligned}$$

- Solution: update in JTA uses max instead of sum:

$$\phi_S^* = \max_{V \setminus S} \psi_V \quad \psi_W^* = \frac{\phi_S^*}{\phi_S} \psi_W \quad \psi_V^* = \psi_V$$

- Final potentials aren't marginals:  $\psi(X_C) = \max_{U \setminus C} p(X)$
- Highest value in potential is most likely:  $X_C^* = \arg \max_C \psi(X_C)$