Inference

בידינו מודל גרפי. נרצה לבצע חישובים בהינתן עדות $E=e$:

1. $P(E=e)$ נראות likelihood
2. $P(Y|E=e)$ posterior
3. $Y=\arg\max P(Y|E=e)$ MAP

$$P(Y|E=e)=\frac{P(Y,e)}{P(e)}=\frac{\sum_{X\setminus E\cup Y} P(X)}{\sum_{X\setminus E} P(X)}$$

מדוע זה קשה?

במקרה הכללי לא ניתן לנצל את המבנה:

1. בעיית ה- Inference היא #P קשה.
2. " " המקורבת היא גם קשה.

אבל עדיין ניתן לנצל את המבנה, והרבה שיטות יעילות בפרקטיקה:

א. שיטות מבוססות הודעות message passing

ב. שיטות מבוססות דגימה

ג. שיטות מבוססות קירובים וריאציוניים variational

⋮

6

Variable Elimination

דוגמה:

$$A \to B \to C \to D \to E$$

חישוב נאיבי: $O(k^n)$

מספר ערכים למשתנה

מספר משתנים.

$$P(D) = \sum_{a,b,c,e} P(A)P(B|A)P(C|B)P(D|C)P(E|D)$$

$$= \sum_{a,b,c,e} \phi(A)\,\phi(A,B)\,\phi(B,C)\,\phi(C,D)\,\phi(D,E)$$

$$m_{AB\to BC} = \sum_{b,c,e} \phi(B,C)\,\phi(C,D)\,\phi(D,E) \underbrace{\sum_{a} \phi(A)\,\phi(A,B)}_{\phi_{AB}(B) \equiv B \text{ פקטור על }}$$

$$m_{BC\to C,D} = \sum_{c,e} \phi(C,D)\,\phi(D,E) \underbrace{\sum_{b} \phi(B,C)\,\phi_{AB}(B)}_{\phi_{ABC}(C)}$$

$$m_{C,D\to D} = \sum_{e} \phi(D,E) \sum_{c} \phi(C,D)\,\phi_{ABC}(C)$$

$$m_{D,E\to D} = \sum_{e} \overbrace{\phi(D,E)}^{1}\,\phi_{ABCD}(D) = \phi_{ABCDE}(D)$$

③

מה הרווחנו?

$$P(X_{i+1}) = \sum_{X_i} \underbrace{\underbrace{\phi(X_i)\, P(X_{i+1} \mid X_i)}_{k^2 \text{ ערכים ומכפלות}}}_{k(k+1) \text{ סכומים}} \Rightarrow O(nk^2)$$

$$\wedge$$

$$O(k^n)$$

מה זה תצפיות?

קלט: (1) אוסף פקטורים Φ

(2) משתנים שרוצים לסכום Z

(3) סדר על Z

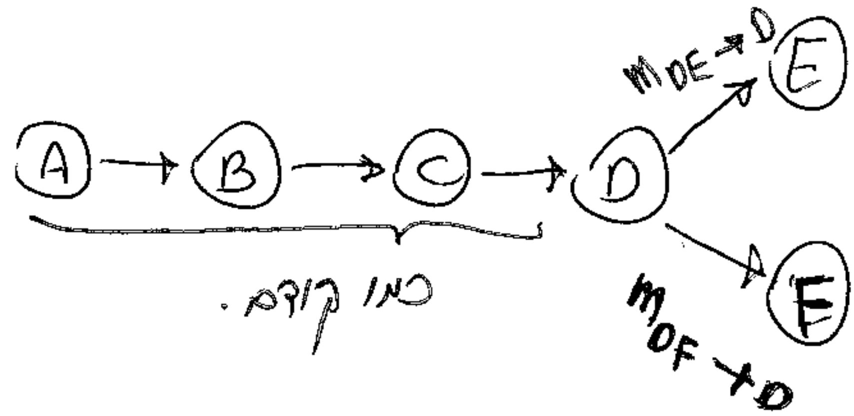
for $z_1, \dots, z_k$

$\Phi_z = \{\phi \in \Phi : z_i \in scope(\phi)\}$ פקטורים בהם מופיע z_i

$\phi_z = \prod_{\phi \in \Phi_{z_i}} \phi$ מכפלה את כל אלו

$\phi = \sum_{z_i} \phi_{z_i}$ אלימינציה (סכימה) של z_i

$\Phi = \Phi_{\neg z_i} \cup \phi$ מה שנשאר



$P(D) = ?$

האם זה תמיד עובד? מה סדר המשתנים?

נראה שהאלגוריתם עובד גם פה טוב.

Clique Tree

1. עץ
2. כל צומת מכיל כמה משתנים.
3. Family Perserving: כל פקטור מקורי מוכל לפחות באחת מהצמתים.
4. RIP: אם $X \in C_i$ וגם $X \in C_j$ אז X מופיע במסלול בין הצמתים.

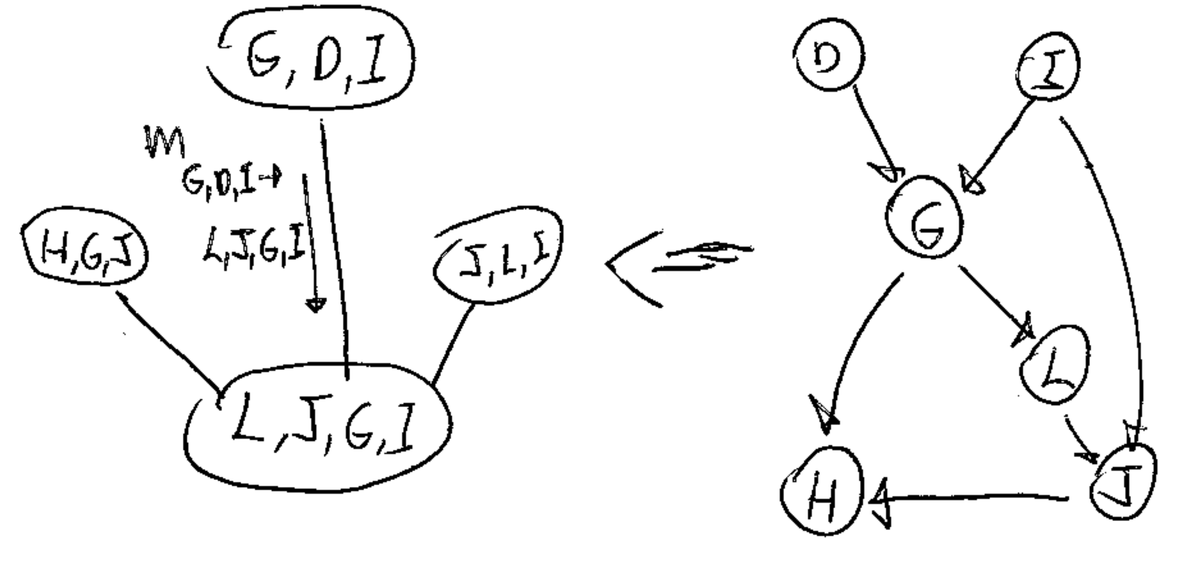
נגדיר אלגוריתם BP על CT:

הודעה מקליקה i לקליקה j

$$m_{i \to j} = \sum_{C_i \setminus S_{ij}} \underbrace{\psi_i}_{\text{פקטורת הצומת}} \prod_{k \in Ne(i)/j} m_{k \to i}$$

S_{ij} – חיתוך בין C_i ל-C_j

דוגמא:



חישוב הסתברויות (לאחר התכנסות)

$$b(C_i) = \varphi_i \prod_{k\in Ne(i)} m_{k\to i}$$

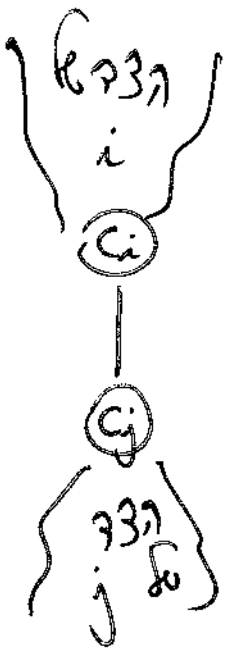
מה סדר הפעולות הרצוי?

אם זה עובד, מה הרווח?

איפה הקושי ב- inference?

<u>משפט:</u>

$$b(C_i) = P(\cancel{X}_{C_i}) = \sum_{X\setminus C_i} \prod_j \varphi_j$$



<u>טענה:</u> כאשר x_i נסכם החוצה בהודעה בין C_i ל-C_j אזי הוא לא מופיע יותר בצד של C_j

<u>הגדרה:</u> $F_{<(i\to j)}$ אוסף הפקטורים בקליקות שבצד של i

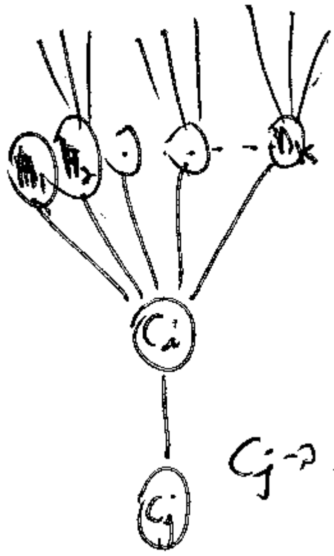
$V_{<(i\to j)}$ אוסף המשתנים באותו צד <u>מלבד</u> S_{ij}

<u>משפט עזר:</u> $m_{i\to j}$ "מסכם" את המידע בצד של i לגבי S_{ij}:

$$m_{i\to j}(S_{ij}) = \sum_{V_{<(i,j)}} \prod_{\varphi\in F_{<(i,j)}} \varphi$$

<u>הוכחה באינדוקציה:</u>

<u>בסיס:</u> נכון עבור עלים מהגדרת $m_{i\to j}$



<u>צעד:</u> $V_{<(i,j)} = \left(\bigcup_{n_k \in Ne(i)\setminus j} V_{<(n_k\to i)}\right) \cup \{\underbrace{C_i \setminus C_j}_{\text{שמבצעים } Y_i \text{ תוך אלימינציה}}\}$

<u>במילים:</u> $V_{<(i,j)}$ כולל את כל המשתנים בצד של i ולא ב-C_i <u>וגם</u> את מה שב-C_i ולא ב-C_j

$$\underbrace{\sum_{Y_i}\sum_{V_{<(n_1\to i)}}\cdots\sum_{V_{<(n_k\to i)}}}_{\text{סכום על כל } V_{<(i\to j)}} \underbrace{\left(\prod_{\varphi\in F_{<(n_1\to i)}}\varphi\right)\cdots\left(\prod_{\varphi\in F_{<(n_k\to i)}}\varphi\right)\overset{\text{מסין:}}{\left(\prod_{\varphi\in F_i}\varphi\right)}}_{\text{כל הפקטורים ב- } F_{<(i\to j)}}$$

$$\overset{\text{הכנסת סכומים}}{=} \sum_{Y_i}\left(\prod_{\varphi\in F_i}\varphi\right)\prod_{k}\sum_{V_{<(n_k\to i)}}\left(\prod_{\varphi\in F_{<(n_k\to i)}}\varphi\right)$$

$$= \sum_{Y_i}\psi_i\, m_{n_1\to i}\cdots m_{n_k\to i} \equiv m_{i\to j}$$

∎

הוכחת משפט מרכזי:

$$B(C_i) = \psi_i \prod_k m_{k\to i} = \psi_i \prod_k \sum_{V<(k\to i)} \prod_{F<(k\to i)} \psi$$

$$= \sum_{X\setminus C_i} \underbrace{\psi_i \prod_{\psi\neq\psi_i} \psi}_{\text{כל הפקטורים.}} = \sum_{X\setminus C_i} P(X) = P(X_{C_i})$$

תוצאות: (1)

$$\underbrace{m_{i\to j}\, m_{j\to i}}_{\mu_{ij}} = \sum_{C_i\setminus S_{ij}} B_i = \sum_{j\setminus S_{ij}} B_i = P(S_{ij})$$

קליברציה

(2)

$$P(X) = \frac{\prod_i B_i(C_i)}{\prod_{i,j} \mu_{ij}(S_{ij})}$$

איך בונים CT?