רשתות בייסיאניות

עד עכשיו עבדנו עם עולם דטרמיניסטי, אבל במציאות העולם יהיה סטוכסטי - דברים יקרו בהסתברות גם עבור אותה פעולה המופעלת באותו מצב. בשביל לתכנן אלגוריתמים שידעו להתמודד עם זה, צריך קוד דרך לייצג את זה.

הגדרות
- Factored Representation זה שבירה לגורמים של העולם, כאשר חלק מהדברים חסרים. כלומר - לא כל הקשרים בין הדברים שאנחנו יודעים לייצג ידועים לנו. למשל - אם יש מתג ויש מנורה, אבל לא יודעים שהמתג מדליק את המנוקה.
- Relational Representation זה ייצוג של הקשרים בין הגורמים הללו.

רשת בייסיאנית היא גרף מכוון (DAG), שבו קשת מX לY אומרת שX משפיע על Y. כל צומת מחזיק CPD (Conditional Probability Distribution):

$$P\left(X|\text{parents}\left(X\right)\right)$$

בשביל לחשב הסתברות של מאורע, משתמשים בנוסחה:

$$P\left(X_1,\ldots,X_n\right)=\prod_{i=1}^{n}P\left(X_i|\text{parents}\left(X_i\right)\right)$$

כאשר עבור $\text{parents}\left(X_i\right)$ משתמשים בערכים הידועים האחרים.
חשוב לבחור סדר טוב לייצג את התלויות, ולא לשים תלויות מיותרות שיתנו לנו קשתות מיותרות שיגדילו את הגרף וייקרו את החישוב.

The Markov Blanket

מי הם הקודקודים שהם בלתי תלויים יחד עם X בהינתן קבוצה $MB\left(X\right)$ הכוללת:

- ההורים של X
- הילדים של X
- הורים אחרים של הילדים של X (נקראים spouses)

הסקה הסתברותית

בהינתן ראיה e, רוצים למצוא את הסתברוויות הפוסטריור של Q

$$P\left(Q|e\right)$$

בד"כ מנרמלים באמצעות $\mu=\frac{1}{P(e)}$:

$$P\left(Q|e\right)=\mu P\left(Q,e\right)$$

באופן נאיבי, אפשר לעבור עם DFS על כל האפשרויות לכל המשתנים. אבל זה יותר מדי זיכרון. עדיין - נוכל לבחור טבלאות מסויימות (factors) שאנחנו יודעים שנצטרך.

ייצוג CPD באמצעות עץ החלטה

כל צומת בעץ מייצגת משתנה תלות. העץ לא חייב להיות מאוזן - יכול להיות משתנה שערך מסויים שלו מייתר הרבה משתנים אחרים, ואז זה חוסך ומקטין את העץ (בניגוד לטבלה שם עדיין היינו צריכים שורה לכל אפשרות)

מודל Noisy-OR

נניח שיש לנו k משתנים, שדורשים 2^k שורות. במקום זה ניצור $k+1$ משתנים - k משתנים בוליאנים, ועוד משתנה שנקרא "leak".

$$P\left(\text{Effect}=1\right)=1-P\left(\text{no cause triggers the effect}\right)=1-\text{leak}\cdot\prod_{\text{cause}}P\left(\text{cause}\right)$$

Dynamic Bayesian Network

מדובר ברשת שחוזרת על עצמה - עם שינוי. משתנה אחד S_t או מספר משתנים $X_t^1, X_t^2, \ldots, X_t^n$. חצים יכולים להיות בין משתנים מזמנים שונים - אבל רק קדימה.

שרשראות מרקוביות

מקרה פרטי של DBN. יש לנו רק משתנה אחד - S_t - עם שתי הנחות:

- אי תלות בעבר - התלות היא רק צעד אחד אחורה: $P\left(S_{t+1}|S_0,\ldots S_t\right)=P\left(S_{t+1}|S_t\right)$.
- אנחנו לא תלויים בזמן - $P\left(S_{t+1}|S_t\right)$ אינו תלוי בt.

Hidden Markov Models (HMM)

כאשר כל time slice מכיל:

- S_t - המצב החבוי
- O_t - המצב הנצפה

O_t נותן מידע עם רעש על S_t. הS_t הם מרקוביים - S_{t+1} תלוי בS_t. אבל O_t תלוי רק בS_t.