ا- a)

$x \perp y, w \mid z$

$$\Rightarrow p(x,y,w \mid z) = p(x \mid z)\, p(y,w \mid z)$$

می خواهیم ثابت کنیم که :

$$p(x,y \mid z) = p(x \mid z)\, p(y \mid z)$$

$$p(x,y \mid z) = \sum_w p(x,y,w \mid z)$$
$$= \sum_w p(x \mid z)\, p(y,w \mid z)$$
$$= p(x \mid z) \sum_w p(y,w \mid z)$$
$$= p(x \mid z)\, p(y \mid z) \checkmark$$

اثبات شد

b) می دانیم که :

(1) $p(x,y,w \mid z) = p(x \mid z)\, p(y,w \mid z)$

$$p(y,w \mid z) = p(y \mid z)\, p(w \mid z)$$

می خواهیم ثابت کنیم که :

$$p(x,w,y \mid z) = p(x,w \mid z)\, p(y \mid z)$$

$$p(x,w,y \mid z) = p(x \mid z)\, p(y,w \mid z)$$
$$= p(x \mid z)\, p(y \mid z)\, p(w \mid z)$$
$$= p(x \mid z)\, p(w \mid z)\, p(y \mid z)$$
$$= \cancel{\frac{p(x,z)\, p(w,z)}{p(z)^2}\, p(y \mid z)}$$

با توجه به (1) و قسمت a
x و w مستقل اند

$$= p(x,w \mid z)\, p(y \mid z) \checkmark$$

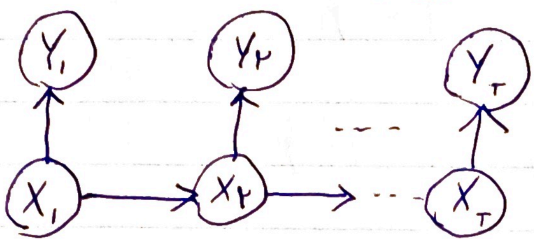
اثبات شد

date : .... subject : ....

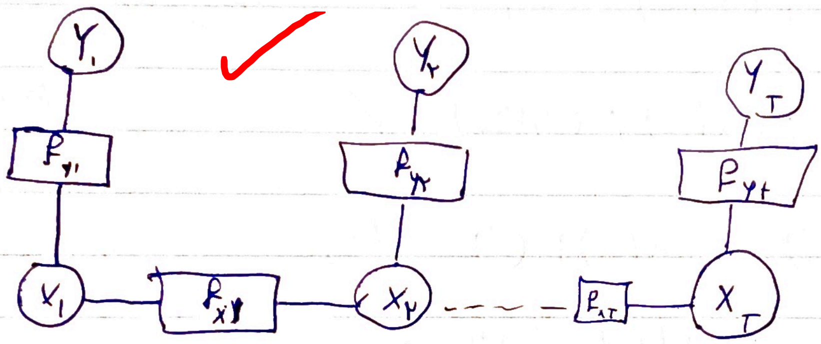
۱. e) درست است چون در مسیر C به E نود B داده شده است و این نود دیگر C است پس مسیر بلاک شده است و $E \perp C \mid B$

۱. d) نادرست است. چون در مسیر A به E روی نود B یک Vstructure قرار دارد لذا این مسیر active شده و نمی‌توان گفت $A \perp E \mid K$

۲. a) شبکه بیزین :



فاکتور گراف



۲-
مثال نقض؟

۱. b) خیر ممکن است چند p-map برای گراف وجود داشته باشد.

۲. a)

$$\phi(x_1) = p(x_1) \quad , \quad \phi(x_{i \neq 1}) = 1$$

$$\phi(x_i , y_i) = p(y_t \mid x_t)$$

$$\phi(x_i , x_{i-1}) = p(x_t \mid x_{t-1})$$

ادامه ۲.b)

$$m_{ji} = \sum_{x_j} \left(\phi(x_j)\, \phi(x_i, x_j) \prod_{k \in N(j) \setminus i} m_{kj}(x_j) \right)$$

$$m_{x_{t+1}, x_t}(x_t) = \sum_{x_{t+1}} \phi(x_{t+1})\, \phi(x_t, x_{t+1})\, m_{y_{t+1}, x_{t+1}}(x_{t+1})\, m_{x_{t+2}, x_{t+1}}(x_{t+1})$$

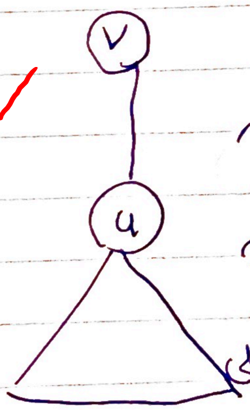
$$m_{y_t x_t}(x_t) = \sum_{y_t} P(y_t \mid x_t)$$

$$\Rightarrow m_{x_{t+1}, x_t} = \sum_{x_{t+1}} P(x_t \mid x_{t+1}) \cdot \sum_{y_t} P(y_t \mid x_t) \cdot m_{x_{t+2}, x_{t+1}}(x_{t+1})$$

اگر $t+2 \leq T$ در غیر این صورت ۱ است

c.2) می دانیم در هر درخت پیامی که از هر نود به نود بالایی آن در الگوریتم sum-product ارسال می شود در واقع حذف آن زیر درخت از گراف است

چون اینجا یک درخت داریم پس نود u یک پدر دارد



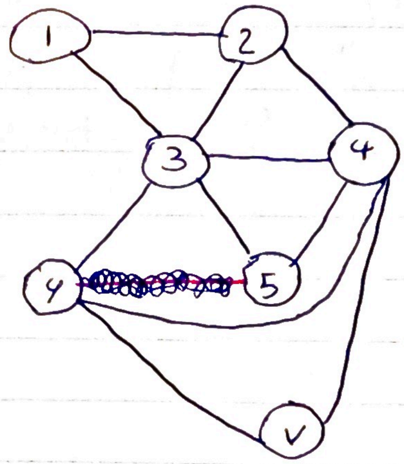
بنابراین اگر m_{uv} را محاسبه کنیم در واقع توزیع توام حذف همه آن متغیرها از درخت را بدست آوردیم و اگر m_{vu} را کامل کنیم در واقع توزیع توام حذف سایر متغیرها از درخت اصلی را بدست آوردیم. بنابراین

$$P(T_u) \propto m_{vu} \prod_{x_i, x_j \in E_{T(u)}} \phi(x_i, x_j) \prod_{x_i \in V_{T(u)}} \phi(x_i)$$

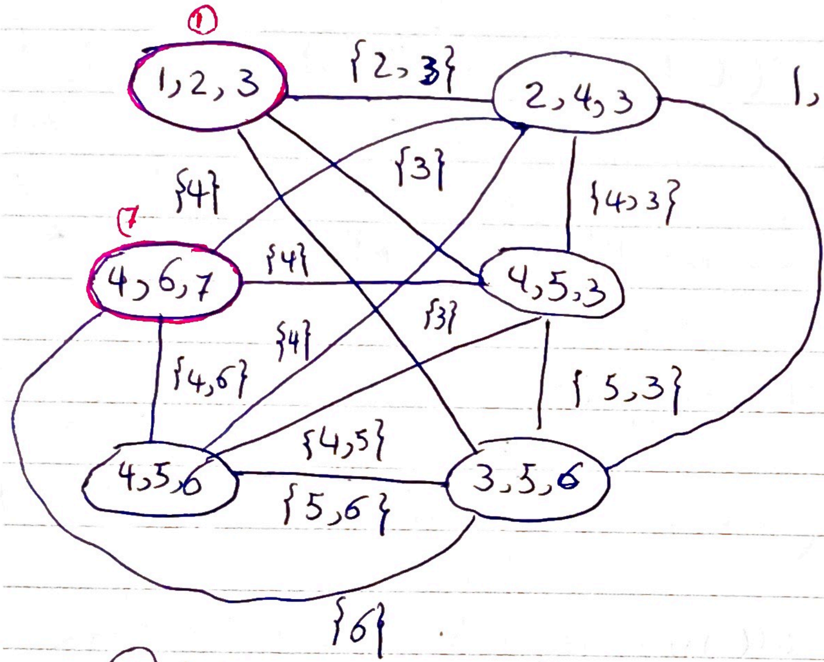
که یکی از نرمالایز کردن روی همه نودهای T_u به جواب می رسیم.

date: Junction Tree subject:

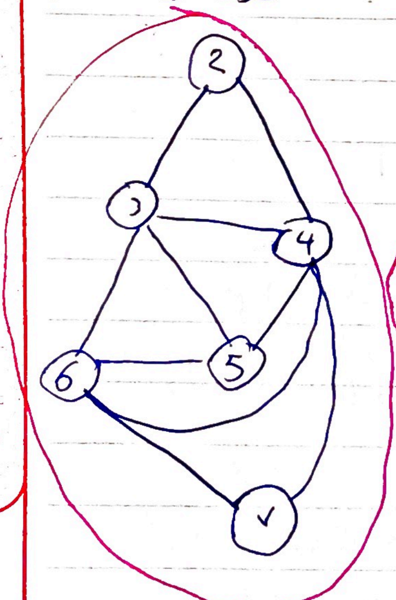
۳-۹) ابتدا گراف moralized شده را تشکیل می دهیم



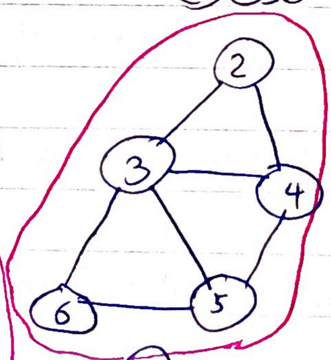
ترتیب حذف: 1, 7, 3, 2, 4, 6, 5



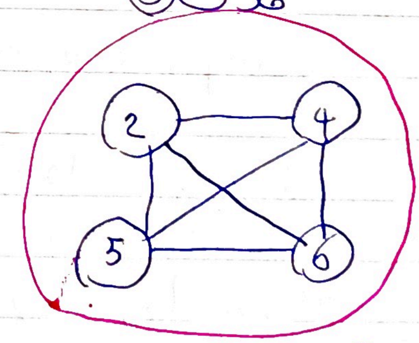
حذف 1



حذف 7



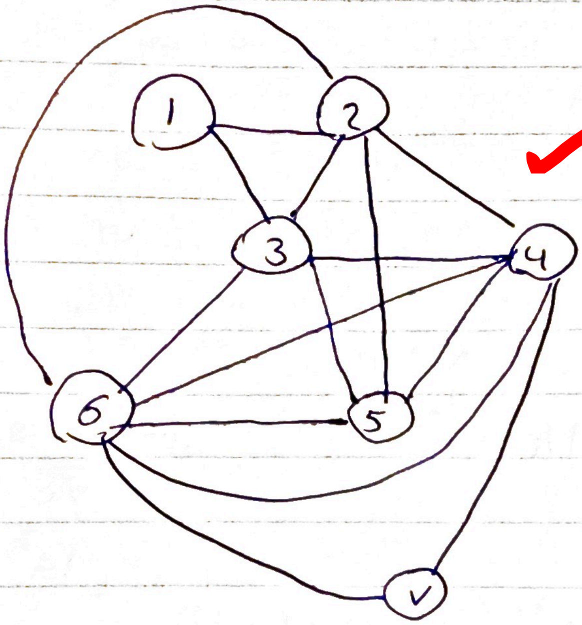
حذف 3



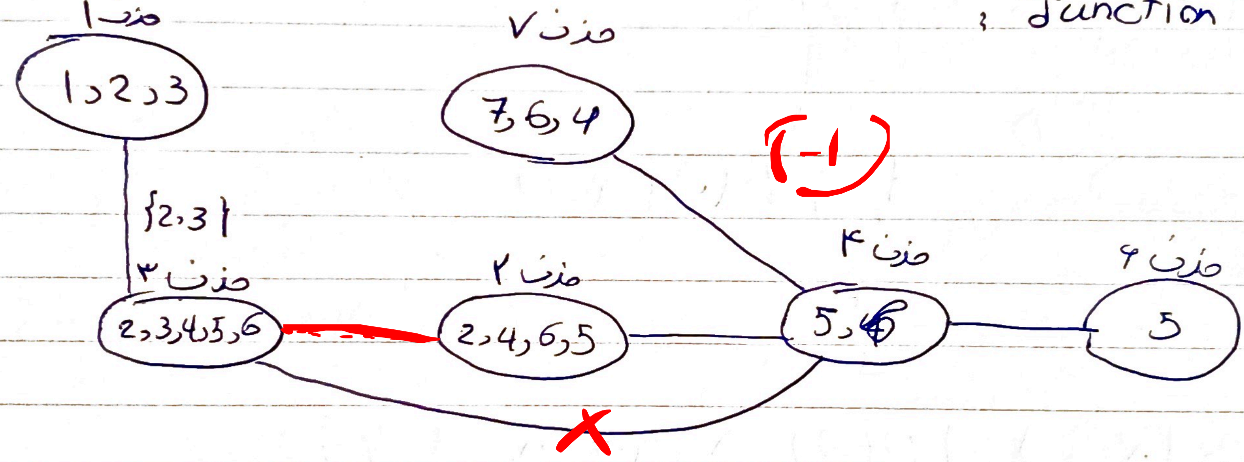
حذف 2 حذف 4 حذف 6

date : .... subject : ....

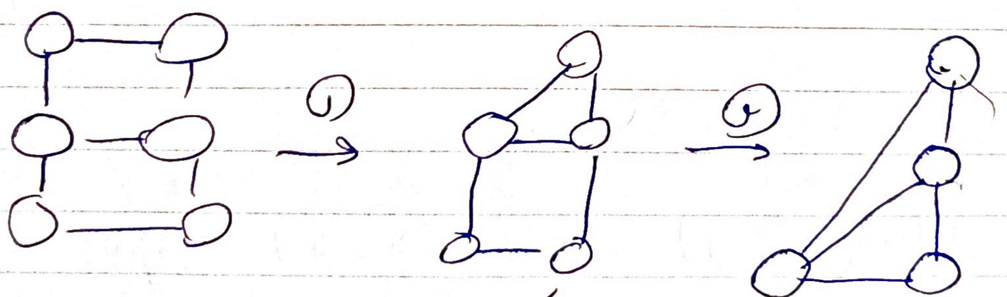
گراف نهایی



Junction Tree :

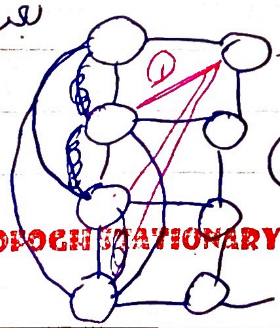


(b



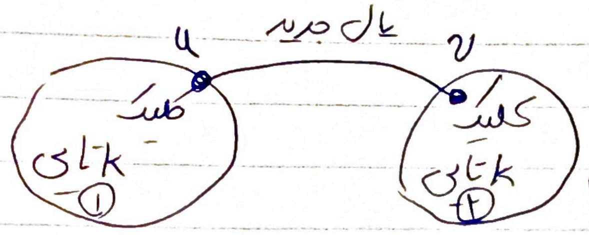
پس با حذف هر ستون یک ~~ستون~~ کلیک m تایی تولید می شود حال ادامه می دهیم تا کلیه ی ستون ها حذف شوند.

بعد حذف ستون ①



حال با حذف یک متغیر از این ستون همه این ستون ها با اولین خانه ستون بعد وصل شده و کلیک می سازند (هنوز اندازه آن m است) پس در نهایت با ادامه این روند سایز بزرگترین کلیک m خواهد بود.

ج)



فرض می‌کنیم بزرگ‌ترین
خوشه درخت فعلی k تایی بوده است
اگر یال جدید بین دو خوشه k تایی اضافه شود در بدترین حالت باعث
آن متغیر u) همه یال های درون کلیک اول به متغیر v وصل می‌شوند
و تشکیل یک ~~خوشه~~ ماکسیمال $2k-1$ تایی می‌دهند.

۴

جواب: $k+1$

۴) a) می‌دانیم که:

$$p(\theta|\alpha)p(\theta) \propto p(\theta|\alpha') \Rightarrow p(\theta|\alpha)p(\theta) = p(\theta|\alpha')\beta$$

می‌خواهیم ثابت کنیم که:

$$\left(\sum_{d=1}^{D} \lambda_d P(\theta|\alpha_d)\right)P(\theta) = \sum_{d=1}^{D} \lambda_d P(\theta|\alpha_d)P(\theta)$$

$$= \sum_{d=1}^{D} \underbrace{\lambda_d \beta_d}_{\lambda_d'} P(\theta|\underbrace{\alpha_d'}_{\alpha_d'})$$

در نتیجه:

$$P(\theta|\alpha_d, \lambda_d)P(\theta) \propto P(\theta|\alpha_d', \lambda_d\beta_d)$$

پس یک conjugate prior برای آن است.

b)

$$P(x) = \frac{1}{z}\exp\sum_{i=1}^{K} \theta_i^T f(x_{c_i})$$

$$P(D|\theta) = \prod_{i=1}^{N} P(x^n|\theta) \rightarrow \ln\prod_{i=1}^{N} P(x^n|\theta) = \sum \ln P(x^n|\theta)$$

$$= \sum^{N} \ln\left(\frac{1}{z}\exp\sum_{i=1}^{K}\theta_i^T f(x_{c_i}^n)\right) =$$

$$\sum_{n=1}^{N}\left(\sum_{i=1}^{K}\theta_i^T f(x_{c_i}^n) - \ln z\right)$$

حال اگر از پارامتر تتای i مشتق بگیریم خواهیم داشت:

$$\nabla_{\theta_i} \ln p(D|\theta) = \sum_{n=1}^{N} \left(f_i(x_{c_i}^{(n)}) - \nabla_{\theta_i} \ln Z(\theta) \right)$$

$$= \sum_{n=1}^{N} f_i(x_{c_i}^{(n)}) - \underbrace{\sum_{n=1}^{N} \nabla_{\theta_i} \ln Z(\theta)}$$

چون نسبت به n ثابت است:

$$= \sum_{n=1}^{N} f_i(x_{c_i}^{(n)}) - N \nabla_{\theta_i} \ln Z(\theta)$$

$$Z^{(n)}(\theta) = \sum_{x} \exp \sum_{i=1}^{K} \theta_i^T f_i(x_{c_i})$$

-۳

$$\ln Z(\theta) = \cancel{\sum_{i=1}^{K}} \theta_i^T f_i(x_{c_i}) =$$

$$\nabla_{\theta_i} \ln Z(\theta) = \cancel{\sum} f_i(x_{c_i}) = E_{p(x|\theta)}[f_i(x_{c_i})]$$

$$\Rightarrow \nabla_{\theta_i} \ln p(D|\theta) = \sum_{n=1}^{N} f_i(x_{c_i}^{(n)}) - N E_{p(x|\theta)}[f_i(x_{c_i})]$$

date: EM

subject:

$$P(x|y) = \left(N(x;\mu_0,1)\right)^{1-y}\left(N(x;\mu_1,1)\right)^{y} (a\ (os$$

با فرض مستقل بودن داده ها

$$\log P(x_1,\dots,x_N,y_1,\dots,y_N) = \log\left(P(x_1,y_1)P(x_2,y_2)\dots P(x_n,y_n)\right)$$

$$= \log\prod_{n=1}^{N} P(x^{(n)},y^{(n)}) = \sum_{n=1}^{N}\log P(x^{(n)},y^{(n)})$$

$$= \sum_{n=1}^{N}\log P(x^n|y^n)P(y^n) \quad \cancel{\sum^{N}\log\left(N(x;\mu_0,1)\right)^{1-y}\left(N(x;\mu_1,1)\right)^{y}}$$

$$\cancel{P(x^n|y^n)\,P(y^n)} \quad \sum_{k=0}^{N_1}\log\theta_k\left(N(x,\mu_0,1)\right)^{1-y}\left(N(x,\mu_1,1)\right)^{y}$$

$$= \sum\sum_{k=0}^{N_1}\log\theta_k + (1-y^{(n)})\log\left(N(x^n,\mu_0,1)\right) + y^{(n)}\log\left(N(x^n,\mu_1,1)\right)$$

$$= \sum\sum_{k=0}^{N_1}\log\theta_k + \log\left(N(x^n,\mu_0,1)\right) + y^{(n)}\left(\log N(x^n,\mu_1,1) - \log N(x^n,\mu_0,1)\right)$$

$$\Rightarrow = \sum_{n=1}^{N} E(y^{(n)})\left(\log\left(N(x^n,\mu_1,1)\right) - \log N(x^n,\mu_0,1)\right)$$

$$= \sum^{N}\log\left((1-\theta)N(x;\mu_0,1)\right)^{1-y}\left(\theta N(x;\mu_1,1)\right)^{y}$$

$$= \sum^{N}(1-y^{(n)})\log\left((1-\theta)N(x^{(n)};\mu_0,1)\right) + y\log\left(\theta N(x^{(n)};\mu_1,1)\right)$$

$$= \sum^{N} y^{(n)}\left(\log\theta N(x^{(n)};\mu_1,1) - \log(1-\theta)N(x^{(n)};\mu_0,1)\right) + C.t$$

عبارت ثابت نسبت به y

$$= \sum^{N} E[y^{(n)}]\left(\ \text{"}\ \right) + C.t$$

پس می توان گفت $E[y^n]$ یک sufficient statistic است. چون با خارج کردن آن دیتای از x^n نمی ماند.

date : .... subject : ....

(b.a

①

$$\underset{P(y_i|x_i)}{E}[y_i] = P(y_i=1|x,\theta^t)$$

$$= \frac{P(x_i|y_i=1,\theta^t)\,P(y_i=1|\theta^t)}{P(x_i|\theta^t)}$$

با جایگذاری هر کدام داریم:

$$= \frac{N(x_i;\mu_1^t,1)\,\theta^t}{\sum_{y_i} P(x_i|y_i,\theta^t)\,P(y_i|\theta^t)}$$

-1

$$= \frac{N(x_i;\mu_1^t,1)\,\theta^t}{\theta^t N(x_i;\mu_1^t,1) + (1-\theta^t)N(x_i;\mu_0^t,1)}$$

دلیل رابطه ①:

$$\underset{P(y_i|x_i)}{E}[y_i] = (y_i=1)P(y_i=1|x_i)$$

چون در صورتی که $y_i=0$ باشد حاصلضرب آن در $P(y_i=0|x_i)$ صفر خواهد شد.

date : .... subject : ....

$$E_{p(y_i|x_i)}[y_i] = \gamma_1^{n}$$ (ج.ه

با توجه به قسمت a داریم:

$$\log\theta + \log N$$

$$= \sum_{i=1}^{N} E(y_i)\left(\log \theta N(x^{(n)}, \mu_1, 1) - \log(1-\theta) N(x^{n}, \mu_0, 1)\right)$$

$$+ \sum_{n=1}^{N} \log\left((1-\theta) N(x^{(n)}, \mu_0, 1)\right)$$

$$\Rightarrow \frac{\partial \log p(D,H|\theta,\mu_0,\mu_1)}{\partial \mu_1} = \frac{\partial \sum_{i=1}^{N} E(y_i) \log N(x_i, \mu_1, 1)}{\partial \mu_1}$$

$$= \frac{\sum_{i=1}^{N} E[y_i]\left(\frac{-1}{2}(x_i-\mu_1)^2 - \log\sqrt{2\pi}\right)}{\partial \mu_1}$$

$$= \sum_{i=1}^{N} E(y_i)(x_i - \mu_1) \Rightarrow \mu_1^{new} = \frac{\sum_{i=1}^{N} E(y_i)\, x_i}{\sum_{i=1}^{N} E(y_i)}$$

$$\frac{\partial \log P(D,H|\theta,\mu_0,\mu_1)}{\partial \mu_0} = \sum_{i=1}^{N} (1 - E(y_i))(x_i - \mu_0)$$

$$\Rightarrow \mu_0^{New} = \frac{\sum_{i=1}^{N} (1 - E(y_i)) x_i}{\sum_{i=1}^{N} (1 - E(y_i))}$$

ادامه ۲.a)

$$\frac{\partial \ell}{\partial \theta} = \sum^{N} E(y_i)\left(\frac{1}{\theta} + \frac{1}{1-\theta}\right) + \sum_{n=1}^{N} \frac{-1}{1-\theta} = 0$$

$$\overset{\theta(1-\theta)}{=} \sum^{N} E(y_i)\left((1-\theta) + \theta\right) \oplus \cancel{\sum_{n=1}} \frac{-N\theta}{\cancel{\quad}} = 0$$

$$\Rightarrow \theta^{new} = \frac{1}{N} \sum_{i=1}^{N} E(y_i)$$

$$\frac{93}{100}$$