STAT 37710 Homework 5

Yuhui Ni.

1. (a)

$$l(\theta) = \log p(x)$$

$$= \log \prod_{i=1}^{n} [\theta_i]_{x_i, x_1^{(i)}, \dots x_{p_i}^{(i)}}$$

$$= \sum_{i=1}^{n} \log [\theta_i]_{x_i, x_1^{(i)}, \dots x_{p_i}^{(i)}}$$

$$l(\theta | x_1, \dots x_m) = \log \sum_{j=1}^{m} \sum_{i=1}^{n} [\theta_i]_{x_i^j, x_1^{j(i)}, \dots x_{p_i}^{j(i)}}$$

$$= \log \prod_{i=1}^{n} \prod_{x_1^{(i)}} \cdots \prod_{x_{p_i}^{(i)}} C_i \cdot [\theta_i]_{1, x_1^{(i)}, \dots x_{p_i}^{(i)}}^{N_1, x_1^{(i)} \dots x_{p_i}^{(i)}} \cdots [\theta_i]_{k, x_1^{(i)} \dots x_{p_i}^{(i)}}^{N_k, x_1^{(i)} \dots x_{p_i}^{(i)}}$$

$$= \sum_{i=1}^{n} \sum_{x_1^{(i)}} \cdots \sum_{x_{p_i}^{(i)}} \sum_{j=1}^{k} [N_i]_{j, x_1^{(i)} \dots x_{p_i}^{(i)}} \log [\theta_i]_{j, x_1^{(i)}, \dots x_{p_i}^{(i)}} + C$$

(b) we have the constraints $\sum_{j=1}^{k} [\theta_i]_{j, x_1^{(i)}, \dots x_{p_i}^{(i)}} = 1$

for $[\theta_i]_{1, x_1^{(i)}, \dots x_{p_i}^{(i)}}, \dots [\theta_i]_{k, x_1^{(i)}, \dots, x_{p_i}^{(i)}}$,

we have. $g = \sum_{j=1}^{k} [N_i]_{j, x_1^{(i)}, \dots, x_{p_i}^{(i)}} \log [\theta_i]_{j, x_1^{(i)}, \dots x_{p_i}^{(i)}} + \lambda \left(\sum_{j=1}^{k} [\theta_i]_{j, x_1^{(i)}, \dots x_{p_i}^{(i)}} - 1 \right)$

$$\frac{\partial g}{\partial [\theta_i]_{j, x_1^{(i)}, \dots x_{p_i}^{(i)}}} = 0, \quad j=1, 2, \dots k, \qquad \frac{\partial g}{\partial \lambda} = 0$$

$$\Rightarrow \quad [\hat{\theta}_i]_{j, x_1^{(i)}, \dots x_{p_i}^{(i)}} = \frac{[N_i]_{j, x_1^{(i)}, \dots x_{p_i}^{(i)}}}{\sum_{j=1}^{k} [N_i]_{j, x_1^{(i)} \dots x_{p_i}^{(i)}}}$$ for specific, $j, x_1^{(i)} \dots x_{p_i}^{(i)}$

if $[N_i]_{j, x_1^{(i)} \dots x_{p_i}^{(i)}} = 0$, then we get $[\hat{\theta}_i]_{j, x_1^{(i)}, \dots x_{p_i}^{(i)}} = 0$

~~The likelihood will be zero, it's meaningless and we can not compute~~ ~~log-likelihood,~~ it's meaningless to compute $[\hat{\theta}_i]_{j, x_1^{(i)}, \dots x_{p_i}^{(i)}}^{N_j, x_1^{(i)}, \dots x_{p_i}^{(i)}} = 0^0$

(c) $p([\theta_i]_{1,b}, \cdots [\theta_i]_{k,b} \mid x_1, \cdots x_m) \propto p(x_1 \cdots x_m \mid [\theta_i]_{1,b}, \cdots [\theta_i]_{k,b}) \cdot p([\theta_i]_{1,b}, \cdots [\theta_i]_{k,b})$

$$\propto [\theta_i]_{1,b}^{N_{1,b}} \cdots [\theta_i]_{k,b}^{N_{k,b}} \cdot [\theta_i]_{1,b}^{a-1} \cdots [\theta_i]_{k,b}^{a-1}$$

$$\propto \prod_{j=1}^{k} [\theta_i]_{j,b}^{N_{j,b}+a-1}$$

Therefore, for any value b,

$$p([\theta_i]_{1,b}, \cdots [\theta_i]_{k,b} \mid x_1, \cdots x_m) = Dir([N_i]_{1,b}+a, [N_i]_{2,b}+a, \cdots [N_i]_{k,b}+a)$$

when i has two parents, for any value b and c

$p([\theta_i]_{1,b,c} \cdots [\theta_i]_{k,b,c} \mid x_1, \cdots x_m) \propto p(x_1, \cdots x_m \mid [\theta_i]_{1,b,c} \cdots, [\theta_i]_{k,b,c}) \cdot p([\theta_i]_{1,b,c}, \cdots [\theta_i]_{k,b,c})$

$$\propto [\theta_i]_{1,b,c}^{N_{1,b,c}} \cdots [\theta_i]_{k,b,c}^{N_{k,b,c}} \cdot [\theta_i]_{1,b,c}^{a-1} \cdots [\theta_i]_{k,b,c}^{a-1}$$

$$\propto \prod_{j=1}^{k} [\theta_i]_{j,b,c}^{N_{j,b,c}+a-1}$$

Therefore, for any value b and c, we have

$$p([\theta_i]_{1,b,c} \cdots [\theta_i]_{k,b,c} \mid x_1, \cdots x_m) = Dir([N_i]_{1,b,c}+a, \cdots, [N_i]_{k,b,c}+a)$$

(d) when we use bayesian strategy, we use posterior mean to estimate parameter. That's to say, we get

$$[\hat{\theta}_i]_{j, x_1^{(i)} \cdots x_{p_i}^{(i)}} = \frac{[N_i]_{j, x_1^{(i)} \cdots x_{p_i}^{(i)}} + a}{\sum_{j=1}^{k} [N_i]_{j, x_1^{(i)} \cdots x_{p_i}^{(i)}} + a}$$

even if some $[N_i]_{j, x_1^{(i)} \cdots x_{p_i}^{(i)}} = 0$, $[N_i]_{j, x_1^{(i)} \cdots x_{p_i}^{(i)}} + a > 0$ (since $a > 0$)

our estimated parameters are always > 0, hence ~~the likelihood will never be zero.~~ $[\hat{\theta}_i]_{j, x_1^{(i)} \cdots x_{p_i}^{(i)}}^{N_{j, x_1^{(i)} \cdots x_{p_i}^{(i)}}}$ is meaningful.

2. (a) $p(x_0, \dots x_T, y_0, \dots y_{m-1}, y_{m+1}, \dots y_T)$

$$= \sum_{y_m} p(x_0) \cdot \prod_{t=1}^{T} p(x_t|x_{t-1}) \prod_{t=0}^{T} p(y_t|x_t)$$

$$= p(x_0) \cdot \prod_{t=1}^{T} p(x_t|x_{t-1}) \cdot \prod_{t=0}^{m-1} p(y_t|x_t) \cdot \prod_{t=m+1}^{T} p(y_t|x_t)$$

① when ~~$t<m-1$~~ $t=0$,

$$a_0(x_0) = p(x_0, y_0) = p(y_0|x_0)\, p(x_0) = \pi_0 W_{x_0,y_0}.$$

when $0 < t < m-1$.

$$a_{t+1}(x_{t+1}) = p(x_{t+1}, y_0, \dots y_{t+1}) = \sum_{x_t} p(x_{t+1}, x_t, y_0, \dots y_{t+1})$$

$$= \sum_{x_t} p(x_{t+1}|x_t)\, p(y_{t+1}|x_{t+1})\, p(x_t, y_0, \dots y_t)$$

$$= \sum_{x_t} \theta_{x_t, x_{t+1}}, W_{x_{t+1}, y_{t+1}}\, a_t(x_t)$$

when $t = m-1$:

$$a_m(x_m) = p(x_m, y_0, \dots y_{m-1}) = \sum_{x_{m-1}} p(x_m, x_{m-1}, y_0, \dots y_{m-1})$$

$$= \sum_{x_{m-1}} p(x_m|x_{m-1}) \cdot p(x_{m-1}, y_0, \dots y_{m-1})$$

$$= \sum_{x_{m-1}} W_{x_{m-1}, x_m}\, a_{m-1}(x_{m-1})$$

when $t > m-1$

~~when~~ we have $a_{t+1}(x_{t+1}) = \sum_{x_t} W_{x_{t+1}, y_{t+1}}\, \theta_{x_t, x_{t+1}}\, a_t(x_t)$ similarly.

Thus,

$$a_t(x_t) = \begin{cases} \pi_0 W_{x_0,y_0} & t=0 \\ \sum_{x_{t-1}} \theta_{x_{t-1}, x_t} W_{x_t, y_t}\, a_{t-1}(x_{t-1}), & 0<t<m,\ t>m \le T \\ \sum_{x_{t-1}} W_{x_{t-1}, x_t}\, a_{t-1}(x_{t-1}) & t=m \end{cases}$$

③ Let $\beta_T(x_T)=1$.

when $m < t+1 \leq T$

$$\beta_t(x_t) = p(y_{t+1}, \ldots y_T \mid x_t) = \sum_{x_{t+1}} p(x_{t+1}, y_{t+1}, \ldots y_T \mid x_t)$$

$$= \sum_{x_{t+1}} p(y_{t+2}, \ldots y_T \mid x_{t+1}) \cdot P(x_{t+1} \mid x_t) \cdot p(y_{t+1} \mid x_{t+1})$$

$$= \sum_{x_{t+1}} \beta_{t+1}(x_{t+1}) \cdot w_{x_{t+1}, y_{t+1}} \theta_{x_t, x_{t+1}}$$

when $t+1 = m$,

$$\beta_{m-1}(x_{m-1}) = \sum_{x_m} p(x_m, y_{m+1}, \ldots y_T \mid x_{m-1})$$

$$= \sum_{x_m} p(y_{m+1}, \ldots y_T \mid x_m) \, P(x_m \mid x_{m-1})$$

$$= \sum_{x_m} \beta_m(x_m) \, \theta_{x_{m-1}, x_m}$$

When $0 < t+1 < m$, we similarly have

$$\beta_t(x_t) = \sum_{x_{t+1}} \beta_{t+1}(x_{t+1}) \, w_{x_{t+1}, y_{t+1}} \theta_{x_t, x_{t+1}}$$

Thus,

$$\beta_t(x_t) = \begin{cases} 1 & t=T \\ \sum_{x_{t+1}} \beta_{t+1}(x_{t+1}) \, w_{x_{t+1}, y_{t+1}} \theta_{x_t, x_{t+1}} & m < t+1 < T, \ 1 \leq t+1 < m \\ \sum_{x_m} \beta_{t+1}(x_{t+1}) \, \theta_{x_{m-1}, x_m} & , \ t = m-1 \end{cases}$$

$$p(y_m \mid y_0, \ldots y_{m-1}, y_{m+1}, \ldots y_T) = \sum_{x_m} p(y_m, x_m \mid y_0, \ldots y_{m-1}, y_{m+1}, \ldots y_T)$$

$$= \sum_{x_m} p(y_m \mid x_m) \cdot P(x_m \mid y_0, \ldots y_{m-1}, y_{m+1}, \ldots, y_T)$$

$$= \sum_{x_m} w_{x_m, y_m} \frac{\alpha_m(x_m) \, \beta_m(x_m)}{\sum_{x_m'} \alpha_m(x_m') \cdot \beta_m(x_m')}$$

we find filling values by $\arg\max_{y_m} p(y_m \mid y_1, \ldots y_{m-1}, y_{m+1}, \ldots y_T)$

2.(b) $$p(x_t, x_{t+1} \mid y_0, \dots, y_T) = \frac{p(x_t, x_{t+1}, y_0, \dots, y_T)}{p(y_0, \dots, y_T)}$$

$$p(x_t, x_{t+1}, y_0, \dots, y_T) = p(y_0, \dots y_T, x_{t+1} \mid x_t) \cdot p(x_t)$$
$$= p(y_0 \dots y_t \mid x_t) \cdot p(y_{t+1}, x_{t+1} \mid x_t) \cdot p(x_t) \cdot \beta_{t+1}(x_{t+1})$$
$$= p(y_0, \dots y_t \mid x_t) \cdot p(x_t)\ p(y_{t+1} \mid x_{t+1})\ p(x_{t+1} \mid x_t)\ \beta_{t+1}(x_{t+1})$$
$$= p(y_0, \dots y_t, x_t) \cdot p(y_{t+1} \mid x_{t+1})\ p(x_{t+1} \mid x_t) \cdot \beta_{t+1}(x_{t+1})$$
$$= \alpha_t(x_t)\ p(y_{t+1} \mid x_{t+1}) \cdot p(x_{t+1} \mid x_t) \cdot \beta_{t+1}(x_{t+1})$$

$$p(y_0, \dots y_T) = \sum_{x_t} \alpha_t(x_t) \cdot \beta_t(x_t)$$

Thus, $$p(x_t, x_{t+1} \mid y_0, \dots, y_T) = \frac{\alpha_t(x_t) \cdot \beta_t(x_t)}{\left[\sum_{x_t} \alpha_t(x_t)\beta_t(x_t)\right] \cdot \beta_t(x_t)}\ p(y_{t+1} \mid x_{t+1}) \cdot p(x_{t+1} \mid x_t)\ \beta_{t+1}(x_{t+1})$$
$$= \frac{\gamma_t(x_t) \cdot \beta_{t+1}(x_{t+1})}{\beta_t(x_t)} \cdot p(x_{t+1} \mid x_t) \cdot p(y_{t+1} \mid x_{t+1})$$
$$=: \xi_t(x_t, x_{t+1})$$

(c) $$A = \sum_{x_0} \sum_{x_1} \cdots \sum_{x_T} \log \pi_{x_0}\, p(x_0, \dots x_T \mid y_0, \dots y_T)$$
$$= \sum_{x_0} \log \pi_0\, p(x_0 \mid y_0, \dots y_T)$$

Let $f = \sum_{x_0} \log \pi_{x_0}\, p(x_0 \mid y_0, \dots y_T) - \lambda\left(\sum_i \pi_i - 1\right)$

$$\frac{\partial f}{\partial \pi_i} = \frac{1}{\pi_i}\gamma_0(i) - \lambda = 0 \qquad \frac{\partial f}{\partial \lambda} = \sum_i \pi_i - 1 = 0$$

$$\Rightarrow \quad \pi_i^{new} = \gamma_0(i)$$

for each $\theta_{i,j}$, we have

$$\cancel{B = \sum_{x_0}\sum_{x_1}\cdots\sum_{x_{T-2}}\sum_{x_{T-1}}\sum_{x_T}\left(\sum_{t=1}^{T}\log\theta_{x_{t-1},x_t}\right)}$$

$$h = \sum_{t=1}^{T} \log\theta_{i,j}\, P(x_{t-1}=i, x_t=j \mid y_0, \dots, y_T)$$

$$= \sum_{t=1}^{T} \log\theta_{i,j}\, \xi_t(i,j)$$

At the same time, we have the constraint, $\sum_j \theta_{i,j} = 0$

$$\cancel{\frac{\partial h}{\partial \theta_{ij}}} \qquad H = \sum_{j=1}^{J}\sum_{t=0}^{T-1} \log\theta_{i,j}\, \xi_t(i,j) - \lambda\left(\sum_j \theta_{i,j} - 1\right)$$

$$\frac{\partial H}{\partial \theta_{ij}} = \frac{\sum_{t=0}^{T-1}\xi_t(i,j)}{\theta_{i,j}} - \lambda = 0 \qquad \frac{\partial H}{\partial \lambda} = \sum_j \theta_{i,j} - 1 = 0$$

$$\Rightarrow \quad \theta_{i,j}^{new} = \frac{\sum_{t=0}^{T-1}\xi_t(i,j)}{\sum_j \sum_{t=0}^{T-1}\xi_t(i,j)} = \frac{\sum_{t=0}^{T-1}\xi_t(i,j)}{\sum_{t=0}^{T-1}\gamma_t(i)}$$

For each $w_{i,j}$, we have

$$C = \sum_{t: y_t = j} \log w_{i,j}\, P(x_t = i \mid y_0, \dots, y_T) = \sum_{t: y_t = j} \log w_{i,j}\, \gamma_t(i)$$

then $$F = \sum_j \sum_{t: y_t = j} \log w_{ij}\, \gamma_t(i) - \lambda\left(\sum_j w_{ij} - 1\right)$$

$$\frac{\partial F}{\partial w_{i,j}} = 0 \qquad \frac{\partial F}{\partial \lambda} = 0$$

$$\Rightarrow \quad w_{i,j}^{new} = \frac{\sum_{t: y_t = j} \gamma_t(i)}{\sum_{t: y_t = j}\sum_j \gamma_t(i)} = \frac{\sum_{t: y_t = j} \gamma_t(i)}{\sum_{t=0}^{T-1} \gamma_t(i)}$$

π_i^{new} represents the posterior probability of $x_0 = i \mid y_0, \dots y_T$, i.e. $P(x_0 = i \mid y_0, \dots, y_T)$

$\theta_{i,j}^{new}$ represents the ~~proportation~~ proportion of posterior probability $\sum_{t=0}^{T-1}\xi_t(i,j)$ in.

all posterior probability $\sum_{t=0}^{T-1} P(X_t=i \mid y_0, \ldots, y_T)$

$w_{i,j}^{new}$ represents the proportion of posterior probability $\sum_{t:y_t=j} \gamma_t(i)$ in all posterior probability

$\sum_{j} \sum_{t:y_t=j} \gamma_t(i)$