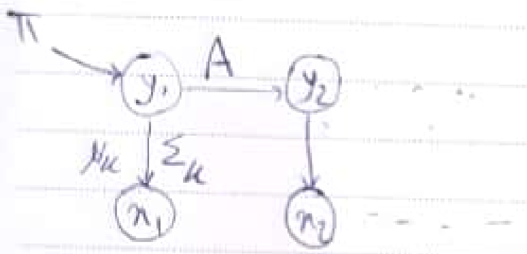
سوال 1)

a)

A: state transition π initial state

$\mu_1 \dots \mu_k$ $\Sigma_1 \dots \Sigma_k$



b)

$$\ln p(D|\theta) = \ln \prod_{n=1}^{N} \Big(p(y_1^n|\pi) \prod_{t=2}^{T} p(y_t^n|y_{t-1}^n, A) \cdot \prod_{t=1}^{T} p(x_t^n|y_t^n) \Big)$$

$\theta: A, \pi, \mu_k, \Sigma_k$

$$\hat{\mu}_k^{mle} = \arg\max_{\mu_k} \ln p(D|\mu_k)$$

$$\Rightarrow \frac{\partial \ln p(D|\mu_k)}{\partial \mu_k} = 0 \Rightarrow \frac{\partial}{\partial \mu_k} \Big(\sum_{t=1}^{T}\sum_{n=1}^{N} \ln p(x_t^n|y_t^n) \Big) = 0$$

$$\Rightarrow \frac{\partial}{\partial \mu_k} \Big(\sum_{t=1}^{T}\sum_{n=1}^{N}\sum_{k=1}^{K} I(y_t^n = k) \ln \mathcal{N}(x_t^n|\mu_k, \Sigma_k) \Big) = 0$$

$$\Rightarrow \frac{\partial}{\partial \mu_k} \Big(\sum_{t=1}^{T}\sum_{n=1}^{N} I(y_t^n = k) \Big(\frac{-1}{2} \ln \big(\sqrt{2\pi\Sigma_k}\big) - \frac{1}{2\Sigma_k} (x_t^n - \mu_k)^2 \Big)\Big) = 0$$

Σ^{-1}

مشتق می‌گیریم

$$\Rightarrow \sum_{t=1}^{T}\sum_{n=1}^{N} I(y_t^n = k) \cdot \frac{1}{\Sigma_k} (x^n - \mu_k) = 0$$

$$\sum_{t=1}^{T}\sum_{n=1}^{N} I(y_t^n = k)\, x^n = \sum_{t=1}^{T}\sum_{n=1}^{N} I(y_t = k)\, \mu_k \Rightarrow \hat{\mu}_k = \frac{\sum_{n=1}^{N}\sum_{t=1}^{T} I(y_t^n = k) \cdot x^n}{\sum_{n=1}^{N}\sum_{t=1}^{T} I(y_t^n = k)}$$

با شهود نیز سازگار است.

Year: Month: Dey : ( ) Subject:

سوال 2)

a) conditional likelihood را می‌نویسیم.

$$\prod_{n=1}^{N} p(y^{(n)} | x^{(n)}, w) = \prod_{n=1}^{N} \frac{1}{z(x^{(n)}, w)} \exp\left\{ \sum_{t=1}^{T} w^T f(x^{(n)}, y_t^{(n)}, y_{t-1}^{(n)}) \right\}$$

$\xrightarrow{\text{ln می‌گیریم}}$ $$= \sum_{n=1}^{N} \left(\sum_{t=1}^{T} \left(w^T f(x^{(n)}, y_t^{(n)}, y_{t-1}^{(n)}) \right) - \ln z(x^{(n)}, w) \right)$$

هدف پیدا کردن w با استفاده از gradient descent و بهینه کردن likelihood

حال است پس نسبت به w از آن مشتق می‌گیریم

$$\frac{\partial}{\partial w} \rightarrow = \sum_{n=1}^{N} \sum_{t=1}^{T} f(x^{(n)}, y_t^{(n)}, y_{t-1}^{(n)}) - \underbrace{\frac{\partial}{\partial w} (\ln z(x^{(n)}, w))}_{1} \quad (1)$$

برای ۱ پس $\frac{\partial z}{\partial w}$ می‌دانیم برای $\log z$ برابر است با امید ریاضی

sufficient statistic. پس داریم:

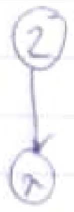
$$\frac{\partial}{\partial w} (\ln z(x^{(n)}, w)) = E_{P(y|x^{(n)}, w)} [f(x^{(n)}, y_t, y_{t-1})] \quad (2)$$

با جایگذاری ② در ① داریم

$$\sum_{n=1}^{N} \sum_{t=1}^{T} f(x^{(n)}, y_t^{(n)}, y_{t-1}^{(n)}) - E_{p(y|x^{(n)}, w)} [f(x^{(n)}, y_t, y_{t-1})]$$

EHSAN

ادامه سوال 2)



b) مدل گرافیکی FA مطابق روبه‌رو x ها را توسط

z که ابعاد کمتری است تولید می‌کنند.

$p(z)=N(z|0,I)$, $p(x|z)=N(x|\mu+Az,\Sigma)$

بنابراین برای کاهش ابعاد پارامترهای Σ و A یاد گرفته شوند. این کار

با استفاده از الگوریتم EM انجام می‌شود.

زمانی که A یادگرفته شد داده‌ها توسط A می‌توانند به ابعاد پایین‌تر (z) انتقال یابند

---

c)

maximum of a multivariate normal is its mean vector.
i-th item of mean vector of multivariate normal distribution is equal to mean of marginal normal distribution for the i-th variable.

سوال 3)

1.3) همان طور که می‌دانیم یادگیری ساختار معادل optimize کردن
$\ln p(D|G,\theta)$ که معادل است با optimize کردن mutual inf
بین نودها و parent های آن ها یعنی
$\sum_i I(X_i, Pa(X_i))$
برای این کار باید ابتدا $I(X_i;X_j)$ بین همه‌ی نودها ساخت
و سپس اجرا کردن الگوریتم MST روی این گراف که درخت با بزرگترین
mutual inf یا بیشترین likelihood را می‌دهد.
برای جهت‌دار کردن درخت می‌توان از BFS استفاده کرد.

2.3)
تنها زمانی $I_p(X,Y)$ مساوی صفر می‌شود که X و Y
از هم مستقل باشند. اما به علت وجود نویز در داده‌های آموزش
$I_p(X,Y)>0$ می‌شود. بنابراین بهینه کردن با توجه به likelihood و
mutual inf منجر به ایجاد یال بین همه‌ی X و Y می‌شود که این موضوع
به علت زیاد کردن تعداد پارامترها و پیچیدگی مدل overfitting ایجاد می‌کند

$$\ln Q(\lambda) = E_{z^{(1)}\dots z^{(n)}}\left[p(x^{(1)}\dots x^{(n)}, z^{(1)}\dots z^{(n)}, \lambda_1\dots\lambda_k)\right]$$ (4 سوال

$$= E_{z^{(1)}\dots z^{(n)}}\left[\ln \prod_{k=1}^{k} p(\lambda_k)\cdot\prod_{n=1}^{N} p(z^{(n)})\cdot p(x^{(n)}|z^{(n)},\lambda_1\dots\lambda_k)\right]$$

$$= \ln\prod_{k=1}^{k} p(\lambda_k) + E_{z^{(1)}\dots z^{(n)}}\left[\sum_{n=1}^{N} \ln p(x^{(n)}|z^{(n)},\lambda_1\dots\lambda_k)\right]$$

$$= \ln\prod_{k=1}^{k} p(\lambda_k) + \sum_{k=1}^{k} E_{z^{(1)}\dots z^{(n)}}\left[z_k^{(n)}\right]\ln \mathrm{Poisson}(x^{(n)}|\lambda_k)$$

$$\boxed{\Rightarrow Q(\lambda_1\dots\lambda_k) = \prod_{k=1}^{k} Q(\lambda_k)}$$

طبق mean field و cordinate ascent داریم:

$$Q(\lambda_k) \propto \exp\left\{\ln p(\lambda_k) + \sum_{n=1}^{N} E[z_k^{(n)}]\ln \mathrm{Poisson}(x^{(n)}|\lambda_k)\right\}$$

$$\propto \exp\left\{(\alpha-1)\ln\lambda - \beta\lambda + \sum_{n=1}^{N} E[z_k^{(n)}]\left(x^{(n)}\ln\lambda - \cancel{x^{(n)}} - \cancel{\ln x^{(n)}!}\right)\right\}$$

cost

$$\propto \exp\left\{\left(\alpha + \sum_{n=1}^{N} E[z_k^{(n)}]\,x^{(n)} - 1\right)\ln\lambda - \beta\lambda\right)$$

$$= \mathrm{Gamma}\left(\alpha + \sum_{n=1}^{N} E[z_k^{(n)}]\,x^{(n)}, \beta\right)$$

$$= Q(\lambda_k)$$

سوال 5)

1.5)

$$p(\theta_i \mid \theta_{-i}, X, \alpha, \beta) = \frac{p(\theta, X, \alpha, \beta)}{\sum_{\theta_i} \text{"}}$$

مخرج مستقل از θ_i است
در صورت و مخرج حذف می‌شود

پس داریم.

$$\propto p(x_i \mid \theta_i) \cdot p(\theta_i)$$

$$= \theta_i^{x_i} (1-\theta_i)^{(1-x_i)} \; \theta_i^{\alpha-1} \cdot (1-\theta_i)^{\beta-1}$$

$$= \theta_i^{(x_i+\alpha-1)} \cdot (1-\theta_i)^{(1-x_i+\beta-1)}$$

$$= Beta(\theta_i \mid \alpha + x_i, \beta + 1 - x_i).$$

$$p(\alpha \mid X, \beta) = \frac{p(\alpha, X, \beta)}{\sum_{\alpha} \text{"}} = \frac{\cancel{p(X \mid \theta)} \cdot p(\theta \mid \alpha, \beta) \cdot p(\alpha, \beta)}{\sum_{\alpha} \cancel{p(X \mid \theta)} \; p(\theta \mid \alpha, \beta) \cdot p(\alpha, \beta)}$$

مستقل از α

$$= \frac{\prod_{i=1}^{D} \left[\frac{\Gamma(\alpha+\beta)}{\Gamma(\beta)}\right] \theta_i^{\alpha} \cancel{\theta_i^{-1}} \cancel{(1-\theta_i)^{\beta-1}} \cdot p(\alpha, \beta)}{\sum_{\alpha} \left[\frac{\Gamma(\alpha+\beta)}{\Gamma(\beta)}\right] \theta_i^{\alpha} \cancel{\theta_i^{-1}} \cancel{(1-\theta_i)^{\beta-1}} \cdot p(\alpha, \beta)}$$

چون مستقل از α اند

ثابت کسر

$$\propto \left[\frac{\Gamma(\alpha+\beta)}{\Gamma(\beta)}\right]^{d} \prod_{i=1}^{d} \theta_i^{\alpha} \cdot p(\alpha, \beta).$$

$$p(\beta|X,\alpha) = \frac{p(\alpha,X,\beta)}{\sum_{\beta} "} = \frac{\cancel{p(x|\theta)} \cdot p(\theta|\alpha,\beta) \cdot p(\alpha,\beta)}{\sum_{\beta} \cancel{p(x|\theta)} \, p(\theta|\alpha,\beta) \cdot p(\alpha,\beta)}$$

$$= \frac{\prod_{i=1}^{D} \left[\frac{\Gamma(\alpha+\beta)}{\Gamma(\beta)}\right] \cancel{\theta_i^{\alpha-1}} \cdot (1-\theta_i)^{\beta} \cdot \cancel{(1-\theta_i)^{-1}} \, p(\alpha,\beta)}{\sum_{\beta} \prod_{i=1}^{D} \left[\frac{\Gamma(\alpha+\beta)}{\Gamma(\beta)}\right] \cancel{\theta^{\alpha-1}} \, (1-\theta_i)^{\beta} \, \cancel{(1-\theta_i)^{-1}} \, p(\alpha,\beta)}$$

مستقل از β

صورت کسر $\propto \left[\frac{\Gamma(\alpha+\beta)}{\Gamma(\beta)}\right]^{d} \prod_{i=1}^{d} (1-\theta_i)^{\beta} \cdot p(\alpha,\beta)$ ✓ ۱۰

5.2) با استفاده از مونت کارلو :

$$p(x_1,\ldots x_d|\alpha,\beta) = \int p(x_1 \ldots x_d|\theta,\alpha,\beta) \, p(\theta|\alpha,\beta) \, d\theta$$

مستقل اند $= \prod_{i=1}^{d} \int p(x_i|\theta_i,\alpha,\beta) \cdot p(\theta_i|\alpha,\beta) \cdot d\theta$

$$= \prod_{i=1}^{d} \left(\frac{1}{S} \sum_{s=1}^{S} p(x_i|\theta_i^{(s)},\alpha,\beta) \right) \qquad \theta_i^{(s)} \sim p(\theta_i|\alpha,\beta)$$ ✓

۱۰