Midterm Exam

Nan Zhao nz2250

1. (a) $P(\pi|x_1,\cdots,x_n) = \prod_{i=1}^{N} P(\pi|x_i) = \prod_{i=1}^{N}\frac{P(x_i|\pi)\,P(\pi)}{P(x_i)} \propto \prod_{i=1}^{N} P(x_i|\pi)\,P(\pi)$

$P(\pi) = Beta(a,b) = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)}\cdot\pi^{a-1}(1-\pi)^{b-1}$

$P(x|\pi,r) = \binom{x+r-1}{x}\pi^{x}(1-\pi)^{r}$

$\prod_{i=1}^{N} P(x_i|\pi)\cdot P(\pi)$

$= \prod_{i=1}^{N}\binom{x_i+r-1}{x_i}\cdot\pi^{\sum_{i=1}^{N}x_i}(1-\pi)^{r\cdot N}\cdot\frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)}\cdot\pi^{a-1}(1-\pi)^{b-1}$

$\propto \pi^{\sum_{i=1}^{N}x_i+a-1}(1-x)^{N\cdot r+b-1}$

$a' = a+\sum_{i=1}^{N}x_i \qquad b' = N\cdot r+b$

$P(\theta|x_1,\cdots,x_n) \propto Beta(a+\sum_{i=1}^{N}x_i,\ b+Nr)$

(b) $P(x_{n+1}|x_1,x_2\cdots x_n) = \int p(x_{n+1}|\pi,r)\,p(\pi,r|x_1,\cdots,x_n)\,d\pi\,dr$

$= \int p(x_{n+1}|\pi,r)\,p(\pi|x_1,\cdots,x_n)\,d\pi$

$= \int p(x_{n+1}|\pi,r)\cdot Beta(\pi|a+\sum_{i=1}^{N}x_i,\ b+r\cdot N)$

$= \underbrace{\prod_{i=1}^{N+1}\binom{x_i+r-1}{x_i}\cdot\frac{\Gamma(a+b+\sum_{i=1}^{N}x_i+rN)}{\Gamma(a+\sum_{i=1}^{N}x_i)\Gamma(b+r\cdot N)}}_{\Theta}\cdot\int \pi^{\sum_{i=1}^{N+1}x_i+a+\sum_{i=1}^{N}x_i}(1-\pi)^{b+r\cdot N+r\cdot(N+1)}\,d\pi$

$= \Theta\cdot\underbrace{\left\{\frac{\Gamma(a+b+\sum_{i=1}^{N+1}x_i+\sum_{i=1}^{N}x_i+r\cdot N+r\cdot(n+1))}{\Gamma(a+\sum_{i=1}^{N+1}x_i+\sum_{i=1}^{N}x_i)\,\Gamma(b+r\cdot n+r\cdot(n+1))}\right\}^{-1}}_{\alpha}\cdot\underbrace{\int Beta(\pi|a+\sum_{i=1}^{N+1}x_i+\sum_{i=1}^{N}x_i,\ b+r\cdot N+r\cdot(n+1))\,d\pi}_{=1}$

$P(x_{n+1}|x_1,\cdots x_n) = \alpha\cdot\Theta$

$= \prod_{i=1}^{N+1}\binom{x_i+r-1}{x_i}\cdot\frac{\Gamma(a+\sum_{i=1}^{N+1}x_i+\sum_{i=1}^{N}x_i)\,\Gamma(b+2rN+r)\,\Gamma(a+b+\sum_{i=1}^{N}x_i+rN)}{\Gamma(a+\sum_{i=1}^{N}x_i)\,\Gamma(b+r\cdot N)\,\Gamma(a+b+\sum_{i=1}^{N+1}x_i+\sum_{i=1}^{N}x_i+2r\cdot N+r)}$

2.

$\ln(y,\lambda|x) = \int \ln(y,\lambda,w|x)\,dw$ over λ

$$= \underbrace{\int q(w) \ln \frac{p(y,\lambda,w|x)}{q(w)}\,dw}_{\mathcal{L}(\lambda)} + \underbrace{\int q(w) \ln \frac{q(w)}{p(w|y,\lambda,w)}\,dw}_{KL(q\|p)}$$

E step:

set $q(w) = p(w|y,\lambda,x)$

$$p(w|y,\lambda,x) \propto p(y|w,\lambda,x)\,p(w) = \prod_{i=1}^{N} p(y_i|w,\lambda,x_i)\,p(w)$$

$$\propto \exp\left\{-\frac{\alpha}{2}\sum_{i=1}^{N}(y_i - x_i^T w)^2\right] \exp\left(-\frac{\lambda}{2}w^T w\right)$$

$$\propto \exp\left\{-\frac{1}{2}(w-\mu)^T \Sigma (w-\mu)\right)$$

$$\Sigma = \left(\lambda I + \alpha \sum_{i=1}^{N} x_i x_i^T\right)^{-1} \qquad \mu = \Sigma \cdot \left(\alpha \sum_{i=1}^{N} y_i x_i\right)$$

$$p(w|y,\lambda,x) = \text{Normal}(w|\mu,\Sigma) \quad \therefore \quad E_{q_t}[w] = \mu$$

$$E_{q_t}[ww^T] = \mu\mu^T + \Sigma$$

at iteration t, we have $q_t(w) = p(w|y,\lambda_{t-1},x)$

$$\mathcal{L}(\lambda) = E_t\left[\ln \frac{p(y,\lambda,w|X)}{q_t(w)}\right]$$

$$= E_q[\ln p(y|w,\lambda,x)] + E_q[\ln p(w)] + E_q[\ln p(\lambda)] - E_q[\ln q(w)]$$

$$= (a-1)\ln(\lambda) - b\lambda + \frac{1}{2}\ln\lambda - \frac{\lambda}{2}\,tr(\mu\mu^T + \Sigma) + \text{const w.r.t. } \lambda$$

$$= (a+\tfrac{1}{2})\ln(\lambda) - b\lambda - \frac{\lambda}{2}tr(\mu\mu^T + \Sigma) + \text{const}$$

M step:

$$\nabla_\lambda \mathcal{L}(\lambda) = 0$$

$$\frac{a+\frac{1}{2}}{\lambda} - b - tr(\mu\mu^T + \Sigma) = 0$$

$$\lambda = \frac{b + tr(\mu\mu^T + \Sigma)}{a + \frac{1}{2}}$$

Last step

$$\ln(y, \lambda | x) = \mathcal{L}(\lambda_{t-1}) = E_t[\ln p(y, \lambda_{t-1}, w | x) - \ln q_t(w)]$$

$$= E_q[\ln p(y|\lambda_{t-1}, X, w)] + E_q[\ln p(\lambda_{t-1})] + E_q[\ln p(w) - E_q[\ln q(w)]$$

$$= \frac{N}{2}\ln(\lambda_{t-1}) - \frac{\lambda_{t-1}}{2}\sum_{i=1}^{N}\{ y_i^2 + tr[x_i x_i^T(\mu\mu^T + \Sigma)] - 2tr(\mu y_i x_i)\}$$

$$+ a\ln b - \ln\Gamma(a) + (a-1)\ln(\lambda_{t-1}) - b\lambda_{t-1}$$

$$+ \frac{1}{2}\ln\lambda - \frac{\lambda}{2} tr(\mu\mu^T + \Sigma)$$

$$+ \frac{1}{2}\ln\det(2\pi e\Sigma)$$

Algorithim

1. Initilize λ_0 to a vector of zeros
2. For each iteration t:

(a). E step: get $q(w) = p(w|y, \lambda_{t-1}, X) = \prod_{i=1}^{N} p(y_i | w, \lambda_{t-1}, x_i)\, p(w)$.

(b). M step: get $\lambda_t = \arg\max_\lambda \mathcal{L}(\lambda) = \arg\max_\lambda E_t[\ln p(w, \lambda, y | X) - \ln q_t(w)]$

(c). Let $\ln(\vec{y}, \lambda_t | X) = \mathcal{L}_t(\lambda_{t-1})$

3.

From general step, we have

$$q_i(\theta_i|\phi_i) = \frac{1}{Z} \exp[\ln p(y, \theta_1, \cdots |x)]$$

- $q(\alpha)$

$$q(\alpha) \propto \exp\{E_{q(w)} \ln p(y, w, \alpha, \lambda | x)\}$$

$$\propto \exp\{E_{q(w)} \ln[\prod_{i=1}^{N} p(y_i|w, \lambda, \alpha, x_i)\, p(w|\lambda)\, p(\lambda)\, p(\alpha)]\}$$

$$\propto \exp\{E_{q(w)}[\underbrace{-\frac{\alpha}{2}\sum_{i=1}^{N}(y_i - x_i^T w)^2]\lambda^{\frac{d}{2}} \exp(-\frac{\lambda}{2}w^T w)\lambda^{e-1} e^{-\lambda f}\, \alpha^{a-1} e^{-\alpha b}}_{\Theta}]\}$$

Remove terms not related to α

$$\Theta \propto -\frac{1}{2}\sum_{i=1}^{N}(y_i - x_i^T w)^2 \cdot \alpha \cdot \alpha^{a-1} \cdot e^{-\alpha b}$$

$$q(\alpha) = \text{Gamma}(\alpha | a', b') \quad a' = a+1 \quad b' = b$$

- $q(\lambda)$

$$q(\lambda) \propto \exp\{E_{q(w)} \underbrace{\ln p(y, w, \alpha, \lambda | x)}_{\Theta}\}$$

$$\Theta \propto \prod_{i=1}^{N} p(y_i | w, \alpha, \lambda, x_i)\, p(w|\lambda) p(\lambda) p(\alpha)$$

$$\propto \exp[-\frac{\alpha}{2}\sum_{i=1}^{N}(y_i - x_i^T w)^2]\lambda^{\frac{d}{2}} \exp(-\frac{\lambda}{2}w^T w)\lambda^{e-1} e^{-\lambda f}\, \alpha^{a-1} e^{-\alpha b}$$

Remove terms not related to α

$$\Theta \propto \lambda^{\frac{d}{2}+e-1} e^{-\lambda(\frac{w^T w}{2} + f)}$$

$$q(\lambda) \propto \lambda^{\frac{d}{2}+e-1} \exp\{-\lambda(E_{q(w)}[\frac{w^T w}{2}] + f)\}$$

$$q(\lambda) = \text{Gamma}(\lambda | e', f'), \quad e' = \frac{d}{2} + e \qquad f' = E_{q(w)}[\frac{w^T w}{2}] + f.$$

· $q(w)$

$$q(w) \propto \exp[E_{q(\alpha,\lambda)}[\ln[p(y,w,\alpha,\lambda|X)]]\}$$

$$p(y,w,\alpha,\lambda|X) \propto \prod_{i=1}^{N} p(y_i|w,x\ \alpha\ x_i)\ p(w|\lambda))\ p(\lambda))\ p(\alpha)$$

$$\propto \exp[-\frac{\alpha}{2}\sum_{i=1}^{N}(y_i - x_i^T w)^2]\ \lambda^{\frac{d}{2}}\exp(-\frac{\lambda}{2}(w^T w))\ p(\lambda))\ p(\alpha)$$

Remove terms not related to w

$$\propto \exp[-\frac{E_{q(\alpha)}[\alpha]}{2}\sum_{i=1}^{N}(y_i - x_i^T w)^2]\ \exp\{-\frac{E_{q(\lambda)}[\lambda]}{2} w^T w\}$$

$$q(w) = \mathrm{Normal}(w|\mu', \Sigma')$$

$$\Sigma' = (E_{q(\lambda)}[\lambda]I + E_{q(\alpha)}[\alpha]\sum_{i=1}^{N} x_i x_i^T)^{-1},\quad \mu' = \Sigma' \cdot (E_{q(\alpha)}\sum_{i=1}^{N} x_i y_i)$$

$$E_{q(\alpha)}[\alpha] = a'/b'$$

$$E_{q(\lambda)}[\lambda] = e'/f'$$

$$E_{q(w)}[w^T w] = \mu\mu'^T + \Sigma'$$

$$\mathcal{L}_t = E_q[\ln p(y, a'_t, b'_t, e'_t, f'_t, \mu_t, \Sigma_t|X)] - E_{q(a'_t)}[\ln(a'_t)] - E_{q(b'_t)}[\ln(b'_t)]$$
$$- E_{q(e'_t)}[\ln(e'_t)] - E_{q(f'_t)}[\ln(f'_t)] - E_{q(\mu'_t)}[\ln(\mu'_t)] - E_{q(\Sigma'_t)}[\ln(\Sigma'_t)]$$

Algorithm:

1. Initialize $a'_0\ b'_0\ e'_0\ f'_0\ \mu'_0\ \Sigma'_0$
2. For iteration t.

update $q(\alpha)$ .... $a'_t = a+1 \quad b' = b$

update $q(\lambda)$ ... $e'_t = \frac{d}{2} + e - 1 \quad f' = \frac{1}{2}(\mu'_{t-1}\mu'^T_{t-1} + \Sigma'_{t-1}) + b$

update $q(w)$ .... $\Sigma'_t = (\frac{e'_t}{f'_t} I + \frac{a'_t}{b'_t}\sum_{i=1}^{n} x_i x_i^T)^{-1},\ \mu'_t = \Sigma'_t \cdot (\frac{a'_t}{b'_t}\sum_{i=1}^{n}(x_i y_i))$

Get $\mathcal{L}_t$ and compare with $\mathcal{L}_{t-1}$, terminate if $\mathcal{L}_{t-1}$ is small.
Continue iteration until teminate.