EECS E6720: Bayesian Models for Machine Learning
Homework 2

Josh Rutta

October 14, 2018

Problem 1

1. a.) $r_{ij} = \text{sign}(\phi_{ij})$, $\phi_{ij} \sim N(u_i^T v_j, \sigma^2)$; $p(r_{ij}=1, \phi_{ij} | u_i, v_j) = p(r_{ij}=1|\phi_{ij})\, p(\phi_{ij}|u_i, v_j)$

$p(r_{ij}=1, \phi_{ij}|u_i, v_j) = \mathbb{1}(\phi_{ij}>0) \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2\sigma^2}(\phi_{ij} - u_i^T v_j)^2}$; $\int p(r_{ij}=1, \phi_{ij}|u_i, v_j)\, d\phi_{ij} = \int_0^\infty \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2\sigma^2}(\phi_{ij} - u_i^T v_j)^2} d\phi_{ij}$

$\Rightarrow \int_{-\infty}^{\infty} p(r_{ij}=1, \phi_{ij}|u_i, v_j)\, d\phi_{ij} = P(\phi_{ij} > 0)$

we can show this equals $\Phi(u_i^T v_j / \sigma) = \int_{-\infty}^{u_i^T v_j} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}s^2} ds$

we can draw $\phi_{ij} \sim N(u_i^T v_i, \sigma^2)$, $\phi_{ij} = u_i^T v_j + \sigma s$, $s \sim N(0,1)$

$P(\phi_{ij} > 0) = P(u_i^T v_j + \sigma s > 0) = P\left(s > -\frac{u_i^T v_j}{\sigma}\right) = p\left(s \leq \frac{u_i^T v_j}{\sigma}\right)$ (because $N(0,1)$ symmetric)

$\Rightarrow \int p(r_{ij}=1, \phi_{ij}|u_i, v_j)\, d\phi_{ij} = p\left(s \leq \frac{u_i^T v_j}{\sigma}\right) = \Phi(u_i^T v_j / \sigma)$

Similarly: $p(r_{ij}=-1, \phi_{ij}|u_i, v_j) = \mathbb{1}(\phi_{ij} \leq 0) \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2\sigma^2}(\phi_{ij} - u_i^T v_j)^2}$; $\int_{-\infty}^{\infty} p(r_{ij}=-1, \phi_{ij}|u_i, v_j)\, d\phi_{ij} = \int_{-\infty}^{0} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2\pi\sigma^2}(\phi_{ij} - u_i^T v_j)^2} d\phi_{ij}$

$= p(\phi_{ij} < 0)$. $\phi_{ij} = u_i^T v_j + \sigma s$, $s \sim N(0,1)$, $p(u_i^T v_j + \sigma s \leq 0) = p\left(s \leq -\frac{u_i^T v_j}{\sigma}\right) = p\left(s > \frac{u_i^T v_j}{\sigma}\right)$

$\Rightarrow \int p(r_{ij}=-1, \phi_{ij}|u_i, v_j)\, d\phi_{ij} = p\left(s > \frac{u_i^T v_j}{\sigma}\right) = 1 - \Phi\left(\frac{u_i^T v_j}{\sigma}\right)$

initially

$r_{ij} | U,V \overset{ind}{\sim} \text{Bern}(\Phi(u_i^T v_j / \sigma)) \,\Big|\, \forall (i,j) \in \Omega$, $u_i \overset{iid}{\sim} N(0, cI)$, $v_j \overset{iid}{\sim} N(0, cI)$; $i \in \{1, \dots, N\}$, $j \in \{1, \dots, M\}$

$r_{ij} \in \{\pm 1\}$, $u_i, v_j \in \mathbb{R}^d$

1. a) $\ln p(R,U,V) = \underbrace{\int q(\phi) \ln \frac{p(R,U,V,\phi)}{q(\phi)} d\phi}_{\mathcal{L}(U,V)} + \int q(\phi) \ln \frac{q(\phi)}{p(\phi|R,U,V)} d\phi$

$\phi = \{\phi_{ij}\}$ $(i,j) \in \Omega$

$r_{ij} = \text{sign}(\phi_{ij})$

$\phi_{ij} \sim N(u_i^T v_j, \sigma^2)$

$q(\phi) = p(\phi|R,U,V) \propto p(R|\phi)\, p(\phi|U,V)$; R conditionally independent from U,V

$p(\phi|R,U,V) = \frac{p(R|\phi)\, p(\phi|U,V)}{\int\int p(R|\phi)\, p(\phi|U,V)\, d\phi_i\, d\phi_j}$; $p(R|\phi) = \prod_{(i,j)\in\Omega} p(r_{ij}|\phi_{ij})$, $p(\phi|U,V) = \prod_{(i,j)\in\Omega} p(\phi_{ij}|u_i, v_j)$

$p(\phi|R,U,V) = \prod_{(i,j)\in\Omega} \frac{p(r_{ij}|\phi_{ij})\, p(\phi_{ij}|u_i, v_j)}{\int\int p(r_{ij}|\phi_{ij})\, p(\phi_{ij}|u_i, v_j)\, d\phi_{ij}}$

$p(r_{ij}|\phi_{ij}) = \text{sign}(\phi_{ij})$, $p(\phi_{ij}|u_i, v_j) = N(u_i^T v_j, \sigma^2)$

$q(\phi_{ij}) = \frac{p(r_{ij}|\phi_{ij})\, p(\phi_{ij}|u_i, v_j)}{\int\int \text{———}\, d\phi_i\, d\phi_j} = \text{sign}(\phi_{ij}) \frac{1}{\sqrt{2\pi\sigma^2}} e^{\frac{-(\phi_{ij} - u_i^T v_j)^2}{2\sigma^2}}$, $\Rightarrow p(\phi|R,U,V) = \prod_{(i,j)\in\Omega} \text{sign}(\phi_{ij}) \frac{1}{\sqrt{2\pi\sigma^2}} e^{\frac{-(\phi_{ij} - u_i^T v_j)^2}{2\sigma^2}} = q(\phi)$

$= TN_{r_{ij}}(x_i^T w, \sigma^2)$

$\boxed{q(\phi) = p(\phi|R,U,V) = \prod_{(i,j)\in\Omega} TN_{r_{ij}}(\phi_{ij})}$

b.) $\mathcal{L}(U,V) = \int q(\phi) \ln \frac{p(R,U,V,\phi)}{q(\phi)} d\phi = \int q(\phi) \ln p(R,U,V,\phi)\, d\phi - \int q(\phi) \ln q(\phi)\, d\phi$

$p(R,U,V,\phi) = p(R,\phi|U,V)\, p(U,V) = p(R|\phi)\, p(\phi|U,V)\, p(U)\, p(V)$

$p(R|\phi)\, p(\phi|U,V) = \prod_{(i,j)\in\Omega} p(r_{ij}|u_i, v_j)\, p(\phi_{ij}|u_i, v_j)$

$\mathcal{L}(U,V) = \int q(\phi) \ln p(R|\phi)\, p(\phi|U,V)\, p(U)\, p(V)\, d\phi - \int q(\phi) \ln q(\phi)\, d\phi$

$= \int q(\phi)(\ln p(U) + \ln p(V))\, d\phi + \int q(\phi)(\ln p(R|\phi) + \ln p(\phi|U,V))\, d\phi - \int q(\phi) \ln q(\phi)\, d\phi$

$= (\ln p(U) + \ln p(V)) \underbrace{\int q(\phi)\, d\phi}_{1} + \mathbb{E}_q\left[\ln\left(\prod_{(i,j)\in\Omega} \mathbb{1}[\text{sign}(\phi_{ij})]\right) + \ln\left(\prod_{(i,j)\in\Omega} N(u_i^T v_j, \sigma^2)\right)\right] - \int q(\phi) \ln q(\phi)\, d\phi$

$= \ln\left(\prod_{i=1}^{N} N(0, cI)\right) + \ln\left(\prod_{j=1}^{M} N(0, cI)\right) + \mathbb{E}_q\left[\sum_{i=1}^{N}\sum_{j=1}^{M}\left(\ln \mathbb{1}[\text{sign}(\phi_{ij}) = r_{ij}] + \left(-\frac{1}{2\sigma^2}(\phi_{ij} - u_i^T v_j)^2\right)\right)\right] - \int q(\phi) \ln q(\phi)\, d\phi$

$= \sum_{i=1}^{N}\left(-\frac{u_i^T u_i}{2c}\right) + \sum_{j=1}^{M}\left(-\frac{v_j^T v_j}{2c}\right) + \sum_{i=1}^{N}\sum_{j=1}^{M}\left(\mathbb{E}_q(\ln \mathbb{1}[\text{sign}(\phi_{ij}) = r_{ij}]) - \mathbb{E}_q\left[\frac{1}{2\sigma^2}(\phi_{ij} - u_i^T v_j)^2\right]\right) - \int q(\phi) \ln q(\phi)\, d\phi + \text{constant}$

$q(\phi_{ij}) = TN_{r_{ij}}(x_i^T w, \sigma^2) \Rightarrow \mathbb{E}_q(\ln \mathbb{1}[\text{sign}(\phi_{ij}) = r_{ij}]) = 0 \Rightarrow \mathcal{L}(U,V) = \sum_{i=1}^{N}\left(-\frac{u_i^T u_i}{2c}\right) + \sum_{j=1}^{M}\left(-\frac{v_j^T v_j}{2c}\right) + \sum_{i=1}^{N}\sum_{j=1}^{M} \mathbb{E}_q\left[\frac{1}{2\sigma^2}(\phi_{ij} - u_i^T v_j)^2\right] + \text{constant}$

$\mathcal{L}(U,V) = \sum_{i=1}^{N}\left(-\frac{u_i^T u_i}{2c}\right) + \sum_{j=1}^{M}\left(-\frac{v_j^T v_j}{2c}\right) - \sum_{i=1}^{N}\sum_{j=1}^{M} \frac{1}{2\sigma^2} \mathbb{E}_q\left[\phi_{ij}^2 + \underbrace{\phi_{ij}^T u_i^T v_j}_{} + v_j^T u_i \phi_{ij} + (v_j^T u_i u_i v_j)\right] + \text{constant}$

$$\mathcal{L}(U,V) = \sum_{i=1}^{N}\left(-\frac{u_i^T u_i}{2c}\right) + \sum_{j=1}^{M}\left(-\frac{v_j^T v_j}{2c}\right) - \sum_{i=1}^{N}\sum_{j=1}^{M}\frac{1}{2\sigma^2}\left(-2u_i^T v_j \mathbb{E}_q[\phi_{ij}] + \underbrace{v_j^T u_i u_i^T v_j}_{= u_i^T v_j v_j^T u_i}\right) + \text{constant}$$

1. c.) $$\frac{d\mathcal{L}}{du_i} = \frac{-u_i}{c} - \sum_{j=1}^{M}\left(-\frac{1}{\sigma^2}v_j \mathbb{E}_q[\phi_{ij}] + \frac{1}{\sigma^2}v_j v_j^T u_i\right) = 0$$

$$\frac{1}{\sigma^2}\sum_{j=1}^{M}\left(v_j \mathbb{E}_q[\phi_{ij}] - v_j v_j^T u_i\right) - \frac{u_i}{c} = 0, \quad +\left(\frac{1}{\sigma^2}\sum_{j=1}^{M}\left(v_j v_j^T\right) + \frac{I}{c}\right)u_i = +\frac{1}{\sigma^2}\sum_{j=1}^{M} v_j \mathbb{E}_q[\phi_{ij}]$$

$$u_i = \left(\sum_{j=1}^{M}\frac{v_j v_j^T}{\sigma^2} + \frac{I}{c}\right)^{-1}\left(\sum_{j=1}^{M}\frac{v_j \mathbb{E}_q[\phi_{ij}]}{\sigma^2}\right)$$

$$\frac{d\mathcal{L}}{dv_j} = \frac{-v_j}{c} - \sum_{i=1}^{N}\left(-\frac{1}{\sigma^2}u_i \mathbb{E}_q[\phi_{ij}] + \frac{1}{\sigma^2}u_i u_i^T v_j\right) = 0, \quad \frac{1}{\sigma^2}\sum_{i=1}^{N}\left(u_i u_i^T + \frac{I}{c}\right)v_j = \frac{1}{\sigma^2}\sum_{i=1}^{N} u_i \mathbb{E}_q[\phi_{ij}]$$

$$v_j = \left(\sum_{i=1}^{N}\frac{u_i u_i^T}{\sigma^2} + \frac{I}{c}\right)^{-1}\left(\sum_{i=1}^{N}\frac{u_i \mathbb{E}_q[\phi_{ij}]}{\sigma^2}\right).$$

d) 1. Initialize U and V, with $u_i \sim N(0,cI)$, $v_j \sim N(0,cI)$, $u_i, v_j \in \mathbb{R}^d$

2. for iteration $t=1,\dots,T$:

(a) E-step: Calculate $$\mathbb{E}_{q_t}[\phi] = \begin{bmatrix} \mathbb{E}_{q_t}[\phi_{11}] & \cdots & \mathbb{E}_{q_t}[\phi_{1M}] \\ \vdots & & \vdots \\ \mathbb{E}_{q_t}[\phi_{N1}] & & \mathbb{E}_{q_t}[\phi_{NM}] \end{bmatrix}$$

$$\mathbb{E}_{q_t}[\phi_{ij}] = \begin{cases} u_{i_{t-1}}^T v_{j_{t-1}} + \sigma \times \dfrac{\Phi'(-u_{i_{t-1}}^T v_{j_{t-1}}/\sigma)}{1-\Phi(-u_{i_{t-1}}^T v_{j_{t-1}}/\sigma)} & \text{if } r_{ij}=1 \\[2ex] u_{i_{t-1}}^T v_{j_{t-1}} + \sigma \times \dfrac{-\Phi'(-u_{i_{t-1}}^T v_{j_{t-1}}/\sigma)}{\Phi(-u_{i_{t-1}}^T v_{j_{t-1}}/\sigma)} & \text{if } r_{ij}=-1 \end{cases}$$

(b) M-Step: update vectors u_i and v_j using the expectations above in the following equation

$$\forall i \in \{1,\dots,N\}: u_{i_t} = \left(\sum_{j=1}^{M}\frac{v_{j_{t-1}} v_{j_{t-1}}^T}{\sigma^2} + \frac{I}{c}\right)^{-1}\left(\sum_{j=1}^{M}\frac{v_{j_{t-1}} \mathbb{E}_q[\phi_{ij}]}{\sigma^2}\right)$$

$$\forall j \in \{1,\dots,M\}: v_{j_t} = \left(\sum_{i=1}^{N}\frac{u_{i_t} u_{i_t}^T}{\sigma^2} + \frac{I}{c}\right)^{-1}\left(\sum_{j=1}^{N}\frac{u_{i_t} \mathbb{E}_q[\phi_{ij}]}{\sigma^2}\right)$$

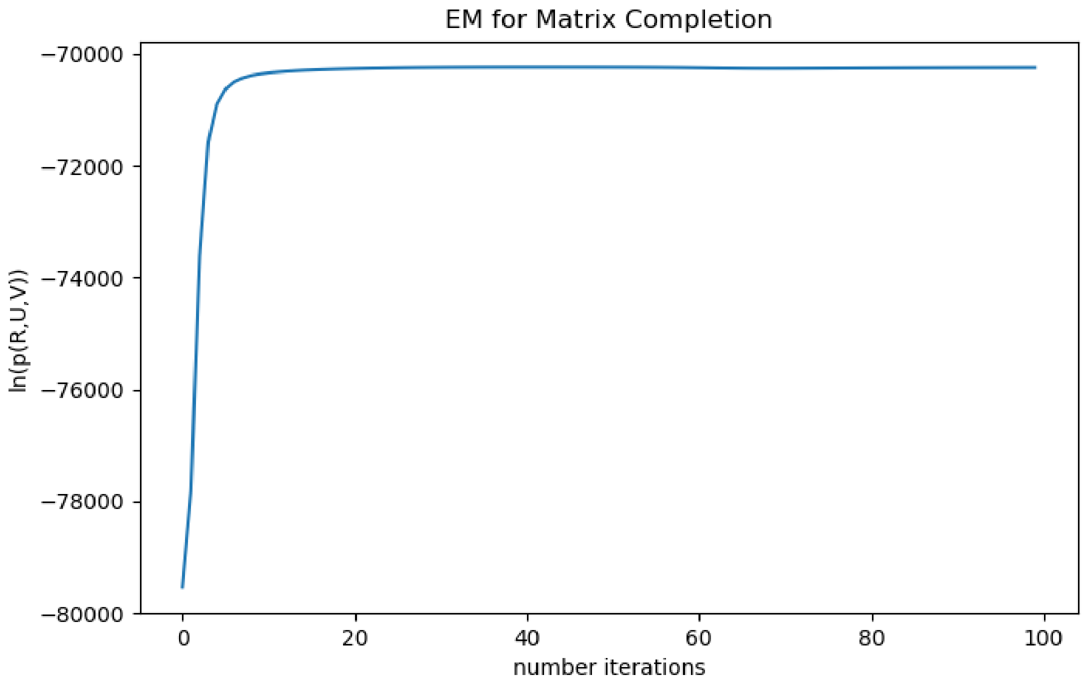
(c) Calculate $\ln p(R,U,V)$

$$\ln p(R,U,V) = \ln\left(p(R|U,V)p(U)p(V)\right) = \ln p(R|U,V) + \ln p(U) + \ln p(V)$$

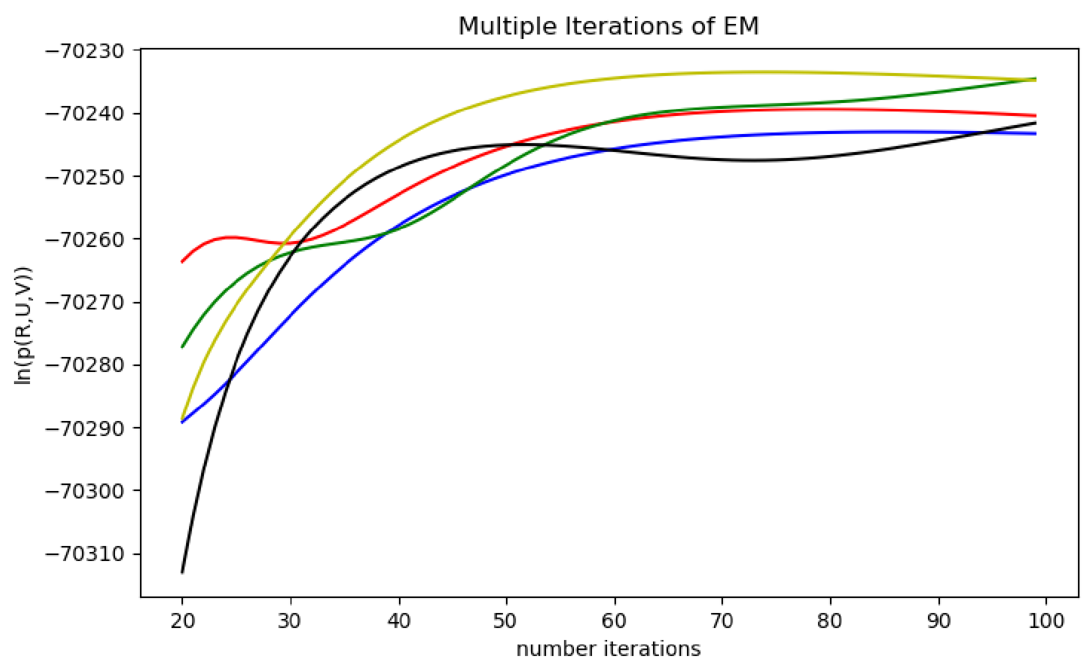
$$= \begin{cases} \sum_{i=1}^{N}\sum_{j=1}^{M} \ln\Phi(u_{i_t}^T v_{j_t}/\sigma) + \sum_{i=1}^{N}\left(-\frac{d}{2}\ln(2\pi c) - \frac{u_{i_t}^T u_{i_t}}{2c}\right) + \sum_{j=1}\left(-\frac{d}{2}\ln(2\pi c) - \frac{v_{j_t}^T v_{j_t}}{2c}\right) & \text{for } r_{ij}=1 \\[2ex] \sum_{i=1}^{N}\sum_{j=1}^{M}\left(\ln\left(1-\Phi\left(\frac{u_{i_t}^T v_{j_t}}{\sigma}\right)\right)\right) + \sum_{i=1}^{N}\left(-\frac{d}{2}\ln(2\pi c) - \frac{u_{i_t}^T u_{i_t}}{2c}\right) + \sum_{j=1}^{M}\left(-\frac{d}{2}\ln(2\pi c) - \frac{v_{j_t}^T v_{j_t}}{2c}\right) & \text{for } r_{ij}=-1 \end{cases}$$

Problem 2

2a)



2b)



2c)

	true -1	true 1
predicted -1	1266	600
predicted 1	1000	2134