CSE250A HW8

Jiping Lin A15058075

1. (a) See Code

(b) $P(\{R_j=r_j^{(t)}\}_{j\in\Omega_t}) = \sum_{i=1}^{k} P(Z=i, \{R_j=r_j^{(t)}\}_{j\in\Omega_t})$ 「Marg.」

$= \sum_{i=1}^{k} P(Z=i)\, P(\{R_j=r_j^{(t)}\}_{j\in\Omega_t} \mid Z=i)$ 「PR」

$= \sum_{i=1}^{k} P(Z=i) \prod_{j\in\Omega_t} P(R_j=r_j^{(t)} \mid Z=i)$ 「CI」

(c) $P(Z=i \mid \{R_j=r_j^{(t)}\}_{j\in\Omega_t}) = \dfrac{P(Z=i, \{R_j=r_j^{(t)}\}_{j\in\Omega_t})}{\sum_{i'=1}^{k} P(Z=i', \{R_j=r_j^{(t)}\}_{j\in\Omega_t})}$ 「By Baye's Rule and denom. is Marg. $(= P(\{R_j=r_j^{(t)}\}_{j\in\Omega_t})$」

$= \dfrac{P(Z=i) \prod_{j\in\Omega_t} P(R_j=r_j^{(t)} \mid Z=i)}{\sum_{i'=1}^{k} P(Z=i') \prod_{j\in\Omega_t} P(R_j=r_j^{(t)} \mid Z=i')}$ 「pt. (b) w/o Σ」

(d) Root: $P(Z=i) = \frac{1}{T}\sum_{t=1}^{T} P(Z=i \mid V_t=v_t)$ ↳ visible

$= \frac{1}{T}\sum_{t=1}^{T} P(Z=i \mid \{R_j=r_j^{(t)}\}_{j\in\Omega_t})$

$= \frac{1}{T}\sum_{t=1}^{T} \rho_{it}$

Other: $P(R_j=1 \mid Z=i) = \dfrac{\sum_{t=1}^{T} P(R_j=1, Z=i \mid \{R_j=r_j^{(t)}\}_{j\in\Omega_t})}{\sum_{t=1}^{T} P(Z=i \mid \{R_j=r_j^{(t)}\}_{j\in\Omega_t}}$

$= \dfrac{\sum_{\{t \mid j\in\Omega_t\}} I(R_j^{(t)}, 1)\, P(Z=i \mid \{R_j=r_j^{(t)}\}_{j\in\Omega_t} + \sum_{\{t \mid j\notin\Omega_t\}} P(Z=i \mid \{R_j=r_j^{(t)}\}_{j\in\Omega_t}) \cdot P(R_j=1 \mid Z=i, \{R_j=r_j^{(t)}\}_{j\in\Omega_t})}{\sum_{t=1}^{T} \rho_{it}}$

「Split t into t, s.t. $j\in\Omega_t$ so we can use indicator & t, s.t. $j\notin\Omega_t$ where we use PR」

$= \dfrac{\sum_{\{t \mid j\in\Omega_t\}} I(R_j^{(t)}, 1)\, \rho_{it} + \sum_{\{t \mid j\notin\Omega_t\}} \rho_{it}\, P(R_j=1 \mid Z=i)}{\sum_{t=1}^{T} \rho_{it}}$ 「CI, since in this sum, j is not in Ω_t」

(e) likelihood increases, the form is

iter	$\mathcal{L}$
0	-27.0358
1	-17.5604
2	-16.0024
4	-15.0606
8	-14.5016
16	-14.2638
32	-14.1802
64	-14.1701
128	-14.1640
256	-14.1637

(f) See code

The rec. seems fit for my taste for first few movies.

2.(a) $P(y=1|\vec{x}) = \dfrac{P(\vec{x}|y=1)P(y=1)}{\sum_j P(\vec{x}, y=j)}$ ⌜BR & Marg.⌟

$= \dfrac{P(\vec{x}|y=1)P(y=1)}{P(\vec{x}|y=0)P(y=0)+P(\vec{x}|y=1)P(y=1)}$

$= \dfrac{(2\pi)^{-d/2}|\Sigma_1|^{-1/2}\exp(-\frac{1}{2}(\vec{x}-\vec{\mu}_1)^T\Sigma_1^{-1}(\vec{x}-\vec{\mu}_1))\pi_1}{(2\pi)^{-d/2}|\Sigma_0|^{-1/2}\exp(-\frac{1}{2}(\vec{x}-\vec{\mu}_0)^T\Sigma_0^{-1}(\vec{x}-\vec{\mu}_0))\pi_0 + \text{num.}}$

(b) In denom. of part (a), 2 parts now share the term $(2\pi)^{-d/2}|\Sigma|^{-1/2}$, so this cancels out w/ the same part in num., so we have

$P(y=1|\vec{x}) = \dfrac{\exp(-\frac{1}{2}(\vec{x}-\vec{\mu}_1)^T\Sigma^{-1}(\vec{x}-\vec{\mu}_1))\pi_1}{\exp(-\frac{1}{2}(\vec{x}-\vec{\mu}_0)^T\Sigma^{-1}(\vec{x}-\vec{\mu}_0))\pi_0+\exp(-\frac{1}{2}(\vec{x}-\vec{\mu}_1)^T\Sigma^{-1}(\vec{x}-\vec{\mu}_1))\pi_1}$

so $\dfrac{1}{P(y=1|\vec{x})} = 1 + \dfrac{\exp(-\frac{1}{2}(\vec{x}-\vec{\mu}_0)^T\Sigma^{-1}(\vec{x}-\vec{\mu}_0))\pi_0}{\exp(-\frac{1}{2}(\vec{x}-\vec{\mu}_1)^T\Sigma^{-1}(\vec{x}-\vec{\mu}_1))\pi_1}$

$= 1+\exp[-\frac{1}{2}((\vec{x}-\vec{\mu}_0)^T\Sigma^{-1}(\vec{x}-\vec{\mu}_0)-(\vec{x}-\vec{\mu}_1)^T\Sigma^{-1}(\vec{x}-\vec{\mu}_1))]\cdot\exp(\log\frac{\pi_0}{\pi_1})$ ⌜$\frac{\pi_0}{\pi_1}=e^{\log\frac{\pi_0}{\pi_1}}$⌟

let $-z = -\frac{1}{2}((\vec{x}-\vec{\mu}_0)^T\Sigma^{-1}(\vec{x}-\vec{\mu}_0)-(\vec{x}-\vec{\mu}_1)^T\Sigma^{-1}(\vec{x}-\vec{\mu}_1))+\log\frac{\pi_0}{\pi_1}$

$= -\frac{1}{2}((\vec{x}^T\Sigma^{-1}\vec{x}-\vec{x}^T\Sigma^{-1}\vec{\mu}_0-\vec{\mu}_0^T\Sigma^{-1}\vec{x}+\vec{\mu}_0^T\Sigma^{-1}\vec{\mu}_0)$
$\quad -(\vec{x}^T\Sigma^{-1}\vec{x}-\vec{x}^T\Sigma^{-1}\vec{\mu}_1-\vec{\mu}_1^T\Sigma^{-1}\vec{x}+\vec{\mu}_1^T\Sigma^{-1}\vec{\mu}_1))+\log\frac{\pi_0}{\pi_1}$

$= -\frac{1}{2}((\underbrace{\vec{x}^T\Sigma^{-1}\vec{x}}-2\vec{x}^T\Sigma^{-1}\vec{\mu}_0+\vec{\mu}_0^T\Sigma^{-1}\vec{\mu}_0)$
$\quad -(\underbrace{\vec{x}^T\Sigma^{-1}\vec{x}}_{\text{cancel}}-2\vec{x}^T\Sigma^{-1}\vec{\mu}_1+\vec{\mu}_1^T\Sigma^{-1}\vec{\mu}_1))+\log\frac{\pi_0}{\pi_1}$

⌜Since $\vec{\mu}_i^T\Sigma^{-1}\vec{x}$ is scalar and Σ^{-1} is positive definite, i.e. $(\Sigma^{-1})^T=\Sigma^{-1}$, then $(\vec{\mu}_i^T\Sigma^{-1}\vec{x})^T=\vec{\mu}_i^T\Sigma^{-1}\vec{x}=\vec{x}^T\Sigma^{-1}\vec{\mu}_i$⌟

$= -\frac{1}{2}(\vec{x}^T(2\Sigma^{-1}(\vec{\mu}_1-\vec{\mu}_0))+\vec{\mu}_0^T\Sigma^{-1}\vec{\mu}_0-\vec{\mu}_1^T\Sigma^{-1}\vec{\mu}_1)+\log\frac{\pi_0}{\pi_1}$

$= \vec{x}^T\Sigma^{-1}(\vec{\mu}_0-\vec{\mu}_1)-\frac{1}{2}(\vec{\mu}_0^T\Sigma^{-1}\vec{\mu}_0-\vec{\mu}_1^T\Sigma^{-1}\vec{\mu}_1)+\log\frac{\pi_0}{\pi_1}$

Since $\Sigma^{-1}(\vec{\mu}_0-\vec{\mu}_1)$ is col. vector
then $\vec{x}^T\Sigma^{-1}(\vec{\mu}_0-\vec{\mu}_1)$ is equivalent to $\vec{x}\cdot(\Sigma^{-1}(\vec{\mu}_0-\vec{\mu}_1))$

then writing z as $\vec{w}\cdot\vec{x}+b$, we have ($-z$ before, flip sign)

$\vec{w} = \Sigma^{-1}(\vec{\mu}_1-\vec{\mu}_0)$, $b = -\log\frac{\pi_0}{\pi_1}+\frac{1}{2}(\vec{\mu}_0^T\Sigma^{-1}\vec{\mu}_0-\vec{\mu}_1^T\Sigma^{-1}\vec{\mu}_1)$

(c) $\dfrac{P(y=1|\vec{x})}{P(y=0|\vec{x})} = \dfrac{P(y=1|\vec{x})}{1-P(y=1|\vec{x})} = \dfrac{1/(1+e^{-z})}{1-1/(1+e^{-z})} = \dfrac{1/(1+e^{-z})}{e^{-z}/(1+e^{-z})} = \dfrac{1}{e^{-z}}$

then if this fraction is constant, then $\frac{1}{e^{-z}}=k$, which means z must be constant, then $\vec{w}\cdot\vec{x}+b=z=-\log\frac{1}{k}$, which is indeed constant. if $k=1$, then $z=-\log 1=0 \Rightarrow \vec{w}\cdot\vec{x}+b=0$.

3. (a) $\mathcal{L}(\bar{v}) = \sum_{t=1}^{T} \log P(y_t|\bar{x}_t)$

$$= \sum_{t=1}^{T} y_t \log P(y_t=1|\bar{x}_t) + (1-y_t)\log P(y_t=0|\bar{x}_t)$$

$$= \sum_{t=1}^{T} y_t \log(1-e^{-\bar{v}\cdot\bar{x}_t}) + (1-y_t)\log e^{-\bar{v}\cdot\bar{x}_t}$$

$$= \sum_{t=1}^{T} y_t \log(1-e^{-\bar{v}\cdot\bar{x}_t}) - (1-y_t)(\bar{v}\cdot\bar{x}_t)$$

(b) $$\frac{\partial \mathcal{L}}{\partial v} = \sum_{t=1}^{T} y_t \frac{-(-\bar{x}_t e^{-\bar{v}\cdot\bar{x}_t})}{1-e^{-\bar{v}\cdot\bar{x}_t}} - (1-y_t)$$

$$= \sum_{t=1}^{T} y_t \bar{x}_t \frac{e^{-\bar{v}\cdot\bar{x}_t}}{1-e^{-\bar{v}\cdot\bar{x}_t}} - (1-y_t)\bar{x}_t$$

$$= \sum_{t=1}^{T} \left(\frac{y_t e^{-\bar{v}\cdot\bar{x}_t}}{1-e^{-\bar{v}\cdot\bar{x}_t}} + y_t - 1\right)\bar{x}_t$$

$$= \sum_{t=1}^{T} \left(\frac{y_t e^{-\bar{v}\cdot\bar{x}_t} + y_t - y_t e^{-\bar{v}\cdot\bar{x}_t} - 1 + e^{-\bar{v}\cdot\bar{x}_t}}{1-e^{-\bar{v}\cdot\bar{x}_t}}\right)\bar{x}_t$$

$$= \sum_{t=1}^{T} \left(\frac{y_t - (1-e^{-\bar{v}\cdot\bar{x}_t})}{1-e^{-\bar{v}\cdot\bar{x}_t}}\right)\bar{x}_t = \sum_{t=1}^{T}\left(\frac{y_t-\rho_t}{\rho_t}\right)\bar{x}_t, \quad \rho_t = 1-e^{-\bar{v}\cdot\bar{x}_t}$$

(ii) $$= P(y_t=1|\bar{x}_t)$$

(c) $P(y=1|\bar{x}) = 1-e^{-\bar{v}\cdot\bar{x}}$

$$= 1-e^{-\sum v_i x_i}$$

$$= 1-e^{\sum \log(1-p_i)^{x_i}}$$

$$= 1-\prod e^{\log(1-p_i)^{x_i}}$$

$$= 1-\prod(1-p_i)^{x_i},$$

$P(y=0|\bar{x}) = 1-P(y=1|\bar{x})$

$$= \prod(1-p_i)^{x_i}$$

So this is equivalent to noisy-OR

(d) $$\frac{\partial \mathcal{L}}{\partial p_i} = \frac{\partial \mathcal{L}}{\partial v_i}\cdot\frac{\partial v_i}{\partial p_i} = \left(\frac{1}{1-p_i}\right)\frac{\partial \mathcal{L}}{\partial v_i}$$

「Chain rule $v=f(p)$, $\mathcal{L}=g(v)$,

$\Rightarrow \mathcal{L}'(p) = \mathcal{L}'(f(p))\,v'(p)$

$= \mathcal{L}'(v)\,v'(p)$」

(e) GA: $$p_i + \eta\left(\frac{\partial \mathcal{L}}{\partial p_i}\right) = p_i + \frac{p_i(1-p_i)}{T_i}\left(\frac{1}{1-p_i}\right)\frac{\partial \mathcal{L}}{\partial v_i}$$

$$= p_i + \frac{p_i}{T_i}\sum_{t=1}^{T}\left(\frac{y_t-\rho_t}{\rho_t}x_{it}\right)$$, since it is ith term of $\bar{x}_t$

$$= p_i + \frac{p_i}{T_i}\sum_{t=1}^{T}\left(\frac{y_t x_{it}}{\rho_t} - x_{it}\right) = p_i + \frac{p_i}{T_i}\left(\sum_{t=1}^{T}\frac{y_t x_{it}}{\rho_t} - \sum_{t=1}^{T} x_{it}\right) = p_i + \frac{p_i}{T_i}\left(\sum_{t=1}^{T}\frac{y_t x_{it}}{\rho_t} - T_i\right)$$

$$= \frac{p_i}{T_i}\sum_{t=1}^{T}\frac{y_t x_{it}}{\rho_t}$$

4. (a) $P(y=1, y'=1 \mid \bar{x}, \bar{x}', s=1)$

$$= \frac{\overbrace{P(s=1 \mid y=1, y'=1, \bar{x}, \bar{x}')}^{(=1 \text{ since } y=y')} P(y=1, y'=1 \mid \bar{x}, \bar{x}')}{P(s=1 \mid \bar{x}, \bar{x}')} \quad \ulcorner BR \lrcorner$$

$$= \frac{P(y=1 \mid \bar{x}, \bar{x}') P(y'=1 \mid y=1, \bar{x}, \bar{x}')}{P(s=1 \mid \bar{x}, \bar{x}')} \quad \ulcorner PR \lrcorner$$

$$= \frac{P(y=1 \mid \bar{x}) P(y'=1 \mid \bar{x}')}{P(s=1 \mid \bar{x}, \bar{x}')} \quad \ulcorner d.sep(3) \lrcorner$$

$$= \frac{\sigma(\bar{w} \cdot \bar{x}) \sigma(\bar{w} \cdot \bar{x}')}{\sum_{y,y'} P(s=1, y, y' \mid \bar{x}, \bar{x}')}$$

$$= \frac{\sigma(\bar{w} \cdot \bar{x}) \sigma(\bar{w} \cdot \bar{x}')}{\sum_{y,y'} \underbrace{P(s=1 \mid y, y', \bar{x}, \bar{x}')}_{(=1 \text{ iff } y=y') \Rightarrow y=y'=0 \ \& \ y=y'=1} P(y \mid \bar{x}) P(y' \mid \bar{x}')} \rightarrow \text{same d.sep(3) as last step}$$

$$= \frac{\sigma(\bar{w} \cdot \bar{x}) \sigma(\bar{w} \cdot \bar{x}')}{P(y=0 \mid \bar{x}) P(y'=0 \mid \bar{x}) + P(y=1 \mid \bar{x}) P(y'=1 \mid \bar{x})}$$

$$= \frac{\sigma(\bar{w} \cdot \bar{x}) \sigma(\bar{w} \cdot \bar{x})'}{(1-\sigma(\bar{w} \cdot \bar{x}))(1-\sigma(\bar{w} \cdot \bar{x}')) + \sigma(\bar{w} \cdot \bar{x}) \sigma(\bar{w} \cdot \bar{x}')}$$

(b) Similarly,

$P(y=1, y'=0 \mid \bar{x}, \bar{x}', s=0)$

$$= \frac{P(y=1 \mid \bar{x}) P(y'=0 \mid \bar{x}')}{\sum_{y,y'} P(s=0, y, y' \mid \bar{x}, \bar{x}')}$$

$\hookrightarrow$ need $y=0, y'=1$ & $y=1, y'=0$

$$= \frac{\sigma(\bar{w} \cdot \bar{x})(1-\sigma(\bar{w} \cdot \bar{x}'))}{P(y=0 \mid \bar{x}) P(y'=1 \mid \bar{x}) + P(y=1 \mid \bar{x}) P(y'=0 \mid \bar{x})}$$

$$= \frac{\sigma(\bar{w} \cdot \bar{x})(1-\sigma(\bar{w} \cdot \bar{x}'))}{(1-\sigma(\bar{w} \cdot \bar{x}))\sigma(\bar{w} \cdot \bar{x}') + \sigma(\bar{w} \cdot \bar{x})(1-\sigma(\bar{w} \cdot \bar{x}'))}$$

(c) BBAC

$s=1 \Rightarrow y=y'$ $\hookrightarrow s=0 \Rightarrow y \neq y'$

(d) $\mathcal{L}(\bar{w}) = \sum_t \log P(s_t \mid \bar{x}_t, \bar{x}_t')$

$$= \sum_t s_t \log P(s_t=1 \mid \bar{x}_t, \bar{x}_t') + (1-s_t) \log P(s_t=0 \mid \bar{x}_t, \bar{x}_t')$$

$$= \sum_t s_t \log\left[(1-\sigma(\bar{w} \cdot \bar{x}_t))(1-\sigma(\bar{w} \cdot \bar{x}_t')) + \sigma(\bar{w} \cdot \bar{x}_t)\sigma(\bar{w} \cdot \bar{x}_t')\right] \quad \ulcorner \text{from denom. in (a), (b)} \lrcorner$$

$$+ (1-s_t) \log\left[(1-\sigma(\bar{w} \cdot \bar{x}_t))\sigma(\bar{w} \cdot \bar{x}_t') + \sigma(\bar{w} \cdot \bar{x}_t)(1-\sigma(\bar{w} \cdot \bar{x}_t'))\right]$$

(e) $$\frac{\partial \mathcal{L}}{\partial \vec{w}} = \sum_t \Big[\bar{y}_t \frac{\sigma'(\vec{w}\cdot\bar{x}_t)}{\sigma(\vec{w}\cdot\bar{x}_t)} \sigma(-\vec{w}\cdot\bar{x}_t)\bar{x}_t + (1-\bar{y}_t)\frac{\sigma(-\bar{w}\cdot\bar{x}_t)}{\sigma(-\bar{w}\cdot\bar{x}_t)}\sigma(\bar{w}\cdot\vec{x}_t)\bar{x}_t$$
$$+ \bar{y}_t' \frac{\sigma(\bar{w}\cdot\bar{x}_t')}{\sigma(\bar{w}\cdot\bar{x}_t')}\sigma(-\bar{w}\cdot\bar{x}_t')\bar{x}_t' + (1-\bar{y}_t')\frac{\sigma(-\bar{w}\cdot\bar{x}_t')}{\sigma(-\bar{w}\cdot\bar{x}_t')}\sigma(\bar{w}\cdot\bar{x}_t')\bar{x}_t' \Big]$$
$$= \sum_t \left[(\bar{y}_t - \sigma(\bar{w}\cdot\bar{x}_t))\bar{x}_t + (\bar{y}_t' - \sigma(\bar{w}\cdot\bar{x}_t'))\bar{x}_t' \right]$$
and the expression in [] goes into the bracket.

5. (a) $$\alpha_{it} = P(Y_t = i \mid y_0, \bar{x}_1, \cdots, \bar{x}_t)$$
$$= \sum_{j\in\{0,1\}} P(Y_t = i, Y_{t-1} = j \mid y_0, \bar{x}_1, \cdots, \bar{x}_t)$$
$$= \sum_j P(Y_{t-1} = j \mid y_0, \bar{x}_1, \cdots, \bar{x}_t)\, P(Y_t = i \mid y_0, \bar{x}_1, \cdots, \bar{x}_t, Y_{t-1} = j)$$
$$= \sum_j \underbrace{P(Y_{t-1} = j \mid y_0, \bar{x}_1, \cdots, \bar{x}_{t-1})}_{d\text{-sep}(3)}\, \underbrace{P(Y_t = i \mid \bar{x}_t, Y_{t-1} = j)}_{d\text{-sep}(1)}$$
$$= \begin{cases} \alpha_{0(t-1)}\sigma(\bar{w}_0\cdot\bar{x}_t) + \alpha_{1(t-1)}\sigma(\bar{w}_1\cdot\bar{x}_t), & \text{if } i=1 \\ \alpha_{0(t-1)}(1-\sigma(\bar{w}_0\cdot\bar{x}_t)) + \alpha_{1(t-1)}(1-\sigma(\bar{w}_1\cdot\bar{x}_t)), & \text{if } i=0. \end{cases}$$

(b) $$\ell^*_{it} = \max_{y_1,\cdots,y_{t-1}} \left[\log P(y_1, \cdots, y_t = i \mid y_0, \bar{x}_1, \cdots, \bar{x}_t) \right]$$
$$= \max_{y_1,\cdots,y_{t-1}} \left[\log P(y_1, \cdots, y_{t-1} \mid y_0, \bar{x}_1, \cdots, \bar{x}_t) + \log P(y_t = i \mid y_0, \bar{x}_1, \cdots, \bar{x}_t, y_1, \cdots, y_{t-1}) \right]$$
PR
$$= \max_{y_1,\cdots,y_t} \underbrace{\log P(y_1, \cdots, y_{t-1} \mid y_0, \bar{x}_1, \cdots, \bar{x}_{t-1})}_{d\text{-sep}(3)} + \max_{y_1,\cdots,y_t} \underbrace{\log P(y_t = i \mid \bar{x}_t, y_{t-1})}_{d\text{-sep}(1)}$$
$$= \ell^*_{i(t-1)} + \max(\log\sigma(\bar{w}_0, \bar{x}_t), \log\sigma(\bar{w}_1, \bar{x}_t), \log(1-\sigma(\bar{w}_0, \bar{x}_t)), \log(1-\sigma(\bar{w}_1, \bar{x}_t))).$$

(c) $$\alpha_{it} = \sum_{i\in\{0,1\}} P(Y_t = i, Y_{t+1} = j \mid y_0, \bar{x}_1, \cdots, \bar{x}_t)$$
$$= \sum_{i\in\{0,1\}} \underbrace{P(Y_t = i \mid y_0, \bar{x}_1, \cdots, \bar{x}_{t+1})\, P(Y_{t+1} = j \mid Y_t = i, \bar{x}_{t+1})}_{(*)}$$
Take log we have $(*) = \ell^*_{it} + \log P(Y_{t+1} = j \mid Y_t = i, \bar{x}_{t+1})$.
this is analogous to the viterbi algo. for HMM,
so $\Phi_{t+1}(j) = \operatorname{argmax}_{i\in\{0,1\}} \left[\ell^*_{it} + \log P(Y_{t+1} = j \mid Y_t = i, \bar{x}_{t+1}) \right]$
and $y_t^* = \Phi_{t+1}(y_{t+1}^*)$.