Amirreza Velae 400102227

1. $p(Y=1)=p(Y=2)=p(Y=3)=\frac{1}{3}$

$$1\le i\le 3: X|Y=i \sim N(\mu_i,\Sigma_i) \Rightarrow p(X|Y=i)=\frac{1}{2\pi|\Sigma_i|^{\frac{1}{2}}}\exp\left(-\frac{1}{2}(X-\mu_i)^T\Sigma_i(X-\mu_i)\right)$$

$$p(C_i|X)\propto p(X|C_i)\,p(C_i)$$

آ) $X=[50,0.5]^T$

$$\Rightarrow\begin{cases} p(C_1|X)\propto \frac{1}{2\pi\times\sqrt{0.49}}\exp\left(-\frac{1}{2}\overbrace{[50\quad \tfrac{1}{2}]\begin{bmatrix}\frac{7}{10}&0\\0&\frac{7}{10}\end{bmatrix}\begin{bmatrix}50\\ \frac{1}{2}\end{bmatrix}}^{\approx 1750}\right)=\frac{e^{-875}}{1.4\pi}\\ p(C_2|X)\propto \frac{1}{2\pi\times\sqrt{0.07}}\exp\left(-\frac{1}{2}\overbrace{[49\quad -\tfrac{1}{2}]\begin{bmatrix}\frac{8}{10}&\frac{3}{10}\\ \frac{3}{10}&2\end{bmatrix}\begin{bmatrix}49\\ -\frac{1}{2}\end{bmatrix}}^{\approx 1909.8}\right)=\frac{e^{-955}}{0.53\pi}\\ p(C_3|X)\propto \frac{1}{2\pi\times\sqrt{0.52}}\exp\left(-\frac{1}{2}\overbrace{[49\quad -\tfrac{1}{2}]\begin{bmatrix}\frac{7}{10}&\frac{2}{10}\\ \frac{2}{10}&\frac{8}{10}\end{bmatrix}\begin{bmatrix}49\\ -\frac{1}{2}\end{bmatrix}}^{\approx 1674}\right)=\frac{e^{-8355}}{1.99\pi}\end{cases}$$

$$X=\begin{bmatrix}50\\ \frac{1}{2}\end{bmatrix}\Rightarrow i=1$$

ب) $X=[0.5,0.5]^T$

$$\Rightarrow\begin{cases} p(C_1|X)\propto \frac{1}{2\pi\times\sqrt{0.49}}\exp\left(-\frac{1}{2}[\tfrac{1}{2}\quad \tfrac{1}{2}]\begin{bmatrix}\frac{7}{10}&0\\0&\frac{7}{10}\end{bmatrix}\begin{bmatrix}\frac{1}{2}\\ \frac{1}{2}\end{bmatrix}\right)=\frac{e^{-0.175}}{1.4\pi}\\ p(C_2|X)\propto \frac{1}{2\pi\times\sqrt{0.07}}\exp\left(-\frac{1}{2}[-\tfrac{1}{2}\quad -\tfrac{1}{2}]\begin{bmatrix}\frac{8}{10}&\frac{3}{10}\\ \frac{3}{10}&\frac{2}{10}\end{bmatrix}\begin{bmatrix}-\frac{1}{2}\\ -\frac{1}{2}\end{bmatrix}\right)=\frac{e^{-0.2}}{0.53\pi}\\ p(C_3|X)\propto \frac{1}{2\pi\times\sqrt{0.52}}\exp\left(-\frac{1}{2}[-\tfrac{1}{2}\quad -\tfrac{1}{2}]\begin{bmatrix}\frac{7}{10}&\frac{2}{10}\\ \frac{2}{10}&\frac{8}{10}\end{bmatrix}\begin{bmatrix}-\frac{1}{2}\\ -\frac{1}{2}\end{bmatrix}\right)=\frac{e^{-0.2375}}{1.99\pi}\end{cases}$$

$$X=\begin{bmatrix}\frac{1}{2}\\ \frac{1}{2}\end{bmatrix}\Rightarrow i=2$$

2

$$y(x_n,w) = w_0 + \sum_i^D w_i x_{ni} = WX$$

$$E_D(w) = \frac{1}{2}\sum_{n=1}^N [y(x_n,w)-y_n]^2 \Rightarrow E_D(w) = \frac{1}{2}\sum_{n=1}^N [WX_n - y_n]^2$$

$$X' = X+\varepsilon_i \Rightarrow \tilde{E}_D(w) = \frac{1}{2}\sum_{n=1}^N [(w_0 + \sum_i^D w_i x'_{ni}) - y_n]^2 = \frac{1}{2}\sum_{n=1}^N [w_0 + \sum_i^D w_i (x_{ni} + \varepsilon_{ni}) - y_n]^2$$

$$W = [w_0 \; w_1 \; \dots \; w_D]^T_{D\times 1},\; X = [X_1 \; \dots \; X_N]_{D\times N},\; \varepsilon = [\varepsilon_i \; \dots \; \varepsilon_N]_{D\times N},\; Y = [y_1 \cdots y_N]_{1\times N}$$

$$\Rightarrow \tilde{E}_D(w) = \frac{1}{2}\sum_i [W(X_i+\varepsilon_i) - Y_i]^2 = \frac{1}{2}\sum_i \{[W(X_i+\varepsilon_i)]^2 - 2Y_i W(X_i+\varepsilon_i) + Y_i^2\}$$

$$\Rightarrow E_D[\tilde{E}_D(w)] = E_D\left[\frac{1}{2}\sum_i \{[W(X_i+\varepsilon_i)]^2 - 2Y_i W(X_i+\varepsilon_i) + Y_i^2\}\right] \quad , [W(x_i+\varepsilon_i)]^2 = (Wx_i)^2 + 2W\varepsilon_i + (W\varepsilon_i)^2$$

extract

$$= E_D\left[\frac{1}{2}\sum_i \{[WX_i]^2 - 2Y_i WY_i + Y_i^2\}\right] + \frac{1}{2}\sum_i E_D\left[2W\varepsilon_i + (W\varepsilon_i)^2 - 2Y_i W(\varepsilon_i)\right]$$

$$= E_D(w) + \sum_i (W - Y_i)\underbrace{\cancel{E[\varepsilon_i]}}_{=0} + \frac{1}{2}\sum_i W E[\varepsilon_i \varepsilon_i^T] W^T = E_D(w) + \frac{1}{2} N\sigma^2 WW^T$$

3) 1) $M=2$: $LR(x) = \ln\left(\frac{P(\Omega_1|x)}{P(\Omega_2|x)}\right) = \ln\left(\frac{P(\Omega_1)P(x|\Omega_1)}{P(\Omega_2)P(x|\Omega_2)}\right) = w^T x + w_0$

$M>2$: $LR(x) = \ln\left(\frac{P(\Omega_i)P(x|\Omega_i)}{P(\Omega_m)P(x|\Omega_m)}\right) = w_i^T x + w_{i,0}$ $\quad 1<i<M$ $\Rightarrow \begin{cases} p(\Omega_i|x,\theta);\; \theta = \dfrac{e^{(w_i^T x + w_{i,0})}}{1+\sum_{i=1}^{M-1} e^{(w_i^T x + w_{i,0})}} \\ p(\Omega_M|x,\theta);\; \theta = \dfrac{1}{1+\sum_{i=1}^{M-1} e^{(w_i^T x + w_{i,0})}} \end{cases}$

we can also use softmax function: $p(\Omega_i|x) = Ber\left(y_n \middle| \dfrac{e^{(W_i^T x)}}{\sum_i e^{(W_i^T x)}}\right)$

where $w_k = 0$

$$L(w_1,\dots,w_{k-1}) = \sum_{i=1}^n \ln P(Y|x_i) = \sum_{i=1}^n \ln\left(\prod_{j=1}^k Ber(y_j|x_i,w)\right)$$

y_i's are mutually exclusive

assume "softmax $(w_j^T x_j) = \mu_j$" and y is one hot encoded $\Rightarrow L(w_1,\dots,w_{k-1}) = \sum_{i=1}^n \ln\left(\prod_{j=1}^k Ber(y_{ij}|\mu_{ij})\right) = \sum_{i=1}^N \sum_{j=1}^k \ln(Ber(y_{ij}|\mu_{ij}))$

$Ber(y_{ij}|\mu_{ij}) = \mu_{ij}^{y_{ij}} \Rightarrow L(w_1, \dots, w_{k-1}) = \sum_{i=1}^{n} \sum_{j=1}^{c} \ln \mu_{ij}^{y_{ij}} = \sum_{i=1}^{n} \sum_{j=1}^{k} y_{ij} \ln \mu_{ij}$

ب) define $L_i = y_{ic} \ln \mu_{ic}$ and $a_{ij} = e^{(w_j x_i)}$ so $\mu_{ic} = \dfrac{e^{a_{ic}}}{\sum_c e^{a_{ic}}}$

$\nabla_{w_j} L_i = \sum_c \dfrac{\partial L_i}{\partial \mu_{ic}} \dfrac{\partial \mu_{ic}}{\partial a_{ij}} \dfrac{\partial a_{ij}}{\partial w_j} = \sum_c \dfrac{y_{ic}}{\mu_{ic}} \times \mu_{ic}(\delta_{jc} - \mu_{nj}) \times x_n = \sum_c y_{ic}(\delta_{jc} - \mu_{nj}) x_i$

$\underbrace{\sum_c \delta_{ic}}_{=1\{c=i\}} y_{ij} x_i - (\underbrace{\sum_c y_{ic}}_{=1}) \mu_{ij} x_i = y_{ij} x_i - \mu_{ij} x_i = (y_{ij} - \mu_{ij}) x_i$

Thus : $\nabla_{w_j} L = \sum_{i=1}^{n} x_n (\mu_i - y_i)^T$

ج) $\nabla_{w_j} f = \nabla_{w_j} L - \frac{\lambda}{2} \nabla_{w_j} \sum_{j=1}^{k-1} |w_j|_2^2 \rightarrow = \dfrac{\partial}{\partial w_j} w_j^T w_j = 2 w_j$

$\Rightarrow \nabla_w f = \nabla_w L - \lambda w_j$

4) $p(y|x, W) = \prod_{j=1}^{J} \mathcal{N}(y_j | w_j^T x, \sigma_j^2) \Rightarrow RSS(w) = \frac{1}{2} \sum_{n=1}^{N} (y_n - w_j^T x_n)^2$ $(X = x_j)$

$= \frac{1}{2} \| X w_j - Y \|_2 = \frac{1}{2} (X w_j - Y)^T (X w_j - Y) = \frac{1}{2} w_j^T X^T X w_j - 2 w_j^T X^T Y + Y^T Y$

$\nabla_{w_j} RSS(w_j) \left(\dfrac{\partial (X^T A X)}{\partial X} = (A + A^T) X \right) = (X^T X + X X^T) w_j - 2 X^T Y = X^T X w_j - X^T Y = 0$

$\Rightarrow X^T X w_j = X^T Y \Rightarrow w_j = (X^T X)^{-1} X^T Y$ (x_i's are independent) $\Rightarrow w_j = \dfrac{X_j^T Y}{X_j^T X_j}$

ب) $\underbrace{X_i X_j = 0 \quad , i \neq j}_{*}$ $\qquad X_{nm} \; , W_{mn}$

$RSS(W) = \frac{1}{2} \| XW - Y \|_2^2 \Rightarrow XW = Y + \varepsilon \; , \; \varepsilon \in N(X)$

$XW = \sum_i x_i w_i \Rightarrow \sum_i x_i w_i = Y + \varepsilon \quad \times x_j^T : \sum_i x_i x_j w_i = x_j^T Y + x_j^T \varepsilon$

$\varepsilon \in N(X) \Rightarrow \varepsilon \perp x_j , x_i \perp x_j \Rightarrow x_j^T x_j \sum_i \delta_{ij} w_i = x_j^T Y \Rightarrow w_j = \dfrac{x_j^T Y}{x_j^T x_j}$

5

(1) $P(X \geq a) \leq \frac{E(X)}{a}$

$$E(x) = \int_0^\infty x f_x(x)dx = \int_0^a \overset{\geq 0}{x f_x(x) dx} + \int_a^\infty x f_x(x) dx$$

$$\Rightarrow E(x) \geq \int_a^\infty x f_x(x) dx \geq \int_a^\infty a f_x(x) dx = a\int_a^\infty f_x(x) dx = aP(X \geq a) \Rightarrow P(X \geq a) \leq \frac{E(x)}{a}$$

(2) $P(|z-\mu| \geq \varepsilon) = P((z-\mu)^2 \geq \varepsilon^2) \leq \frac{E[(z-\mu)^2]}{\varepsilon^2} = \frac{6^2}{\varepsilon^2}$

(3) $P(|\frac{X_{in}}{X_{all}} - \frac{\pi}{4}| \geq 10^{-2}) \leq \frac{Var(\frac{X_{in}}{X_{all}})}{10^{-4}} \leq 5 \times 10^{-2}$

$X_{in} = X_1 + X_2 + X_3 + \dots X_n$ where x_i's are i.i.d, $Ber(X_i | \theta), \theta = \frac{\pi}{4}$, $X_{all} = n$

$Var(X_{in}) = \sum_i Var(X_i) = n \frac{\pi}{4}(1-\frac{\pi}{4}) \Rightarrow Var(\frac{X_{in}}{X_{all}}) = \frac{1}{n} \frac{\pi}{4}(1-\frac{\pi}{4}) \simeq \frac{0.169}{n}$

$\Rightarrow \frac{0.169}{n} \leq 5 \times 10^{-6} \Rightarrow n \geq 3.37 \times 10^4$

⑥ • $A = U\Sigma V^T$, $A^{-1} = V\Sigma^{-1}U^T$, $AA^{-1} = I$

$\Sigma_{11} = \sigma_1 = \sigma_{max}(A)$, $\Sigma_{11}^{-1} = \sigma_1' = \sigma_{min}(A^{-1})$, $\sigma_1\sigma_1' = 1$ $\Rightarrow \sigma_{max}(A)\,\sigma_{min}(A^{-1}) = 1 \Rightarrow \sigma_{max}(A)\,\sigma_{max}(A^{-1}) \geq 1$

• $\|A\|_2^2 = \sigma_{max}^2$, $\|A\|_F^2 = trace(A^TA) = \sum_i \sigma_i^2 \Rightarrow \|A\|_2^2 \leq \|A\|_F^2$

$\sigma_{i \neq j} \leq \sigma_{max} \Rightarrow \sum_i \sigma_i^2 \leq rank(A)\,\sigma_{max} = rank(A)\,\|A\|_2^2$

$\Rightarrow \|A\|_2 \leq \|A\|_F \leq \sqrt{rank(A)}\,\|A\|_2$

⑦

$y(x,w) = w_0 + \sum_{j=1}^{n}\left[w_j\,\sigma\left(\frac{x-\mu_j}{s}\right)\right] = w_0 + \sum_{j=1}^{n}\left[w_j \frac{1}{1+e^{-\left(\frac{x-\mu_j}{s}\right)}}\right] = w_0 + \sum_{j=1}^{n}\left[w_j \frac{e^{\left(\frac{x-\mu_j}{s}\right)}}{e^{\left(\frac{x-\mu_j}{s}\right)}+1}\right]$

$y(x,w) = w_0 + \sum_{j=1}^{n}\left[w_j \tanh\left(\frac{x-\mu_j}{s}\right)\right] = w_0 + \sum_{j=1}^{n}\left[w_j \frac{e^{\left(\frac{x-\mu_j}{s}\right)} - e^{-\left(\frac{x-\mu_j}{s}\right)}}{e^{\left(\frac{x-\mu_j}{s}\right)} + e^{-\left(\frac{x-\mu_j}{s}\right)}}\right]$

$\Rightarrow w_{j_1} \frac{1}{1+e^{-\left(\frac{x-\mu_j}{s}\right)}} = w_{j_2} \frac{e^{2\left(\frac{x-\mu_j}{s}\right)} - 1}{e^{2\left(\frac{x-\mu_j}{s}\right)} + 1} \Rightarrow w_{j_1} = w_{j_2} \frac{e^{2\left(\frac{x-\mu_j}{s}\right)} - 1 + e^{\left(\frac{x-\mu_j}{s}\right)} + e^{-\left(\frac{x-\mu_j}{s}\right)}}{e^{2\left(\frac{x-\mu_j}{s}\right)} + 1}$