Problem 1:

- Prediction : $\hat{i} = \arg\max_i \; p(Y=i|x) = \frac{p(x|y=i)\,p(y=i)}{p(x)} \equiv \arg\max_i \frac{p(x|y=i)\,p(y=i)}{\sum_{j=1}^{3} p(x|y=j)\,p(y=j)}$

$\rightarrow \hat{c} = \arg\max_i \; p(x|y=i)\,p(y=i) \equiv \arg\max_i \; p(x|y=i)$, $p(y=i) = 1/3$

- $p(\underline{x}|y=i) = \frac{1}{(2\pi)\sqrt{\Sigma_i}} \exp\left(-\frac{1}{2}(\underline{x}-\underline{\mu})^T \Sigma^{-1} (\underline{x}-\underline{\mu})\right)$

$|\Sigma_1| = 0.49 \rightarrow \Sigma_1^{-1} = \frac{1}{0.49}\begin{bmatrix} 0.7 & 0 \\ 0 & 0.7 \end{bmatrix}$

$|\Sigma_2| = 0.07 \rightarrow \Sigma_2^{-1} = \frac{1}{0.07}\begin{bmatrix} 0.2 & -0.3 \\ -0.3 & 0.8 \end{bmatrix}$

$|\Sigma_3| = 0.52 \rightarrow \Sigma_3^{-1} = \frac{1}{0.52}\begin{bmatrix} 0.8 & -0.2 \\ -0.2 & 0.7 \end{bmatrix}$

الف) $\underline{x} = \begin{bmatrix} 50 \\ 0.5 \end{bmatrix}$

- $p(x|y=1) = \frac{1}{2\pi(0.7)} \exp\left[-\frac{1}{2}\begin{bmatrix} 50 & 0.5 \end{bmatrix}\begin{bmatrix} 0.7 & 0 \\ 0 & 0.7 \end{bmatrix}\begin{bmatrix} 50 \\ 0.5 \end{bmatrix}\frac{1}{0.49}\right] = \frac{1}{2\pi(0.7)} \exp[-1785.892]$

- $p(x|y=2) = \frac{1}{2\pi\sqrt{0.07}} \exp\left[\frac{-1}{2\times 0.07}\begin{bmatrix} 49 & -0.5 \end{bmatrix}\begin{bmatrix} 0.2 & -0.3 \\ -0.3 & 0.8 \end{bmatrix}\begin{bmatrix} 49 \\ -0.5 \end{bmatrix}\right] = \frac{1}{2\pi\sqrt{0.07}} \exp[3536.9285]$

- $p(x|y=3) = \frac{1}{2\pi\sqrt{0.52}} \exp\left[\frac{-1}{2\times 0.52}\begin{bmatrix} 49 & -0.5 \end{bmatrix}\begin{bmatrix} 0.8 & -0.2 \\ -0.2 & 0.7 \end{bmatrix}\begin{bmatrix} 49 \\ -0.5 \end{bmatrix}\right] = \frac{1}{2\pi\sqrt{0.52}} \exp[-1856.51442]$

$\Rightarrow \hat{c} = \arg\max_i \; p(x|y=i) = \underline{1}$

ب) $\underline{x} [0.5 \;\; 0.5]^T$

- $p(x|y=1) = \frac{1}{2\pi(0.7)} \exp(-0.3571)$

- $p(x|y=2) = \frac{1}{2\pi(\sqrt{0.07})} \exp(-0.7142)$

- $p(x|y=3) = \frac{1}{2\pi(\sqrt{0.52})} \exp(-0.2644)$

$\Rightarrow \hat{c} = \arg\max_i \; p(x|y=i) = \underline{2}$

Problem 2:

- $\tilde{E}_D(\omega) = \frac{1}{2}\sum_{n=1}^{N}\left(y[x_n+\varepsilon_n, \omega] - y_n\right)^2 \quad ; \quad \varepsilon_i \sim \mathcal{N}(0, \sigma^2 I)$

$$= \frac{1}{2}\sum_{n=1}^{N}\left(\omega_0 + \sum_{i=1}^{D}\omega_i(x_{ni}+\varepsilon_i) - y_n\right)^2 = \frac{1}{2}\sum_{n=1}^{N}\left(\omega_0 + \sum_{i=1}^{D}\omega_i x_{ni} - y_n\right)^2 + \left(\sum_{i=1}^{D}\omega_i\varepsilon_i\right)^2 + 2\sum_{i=1}^{D}\omega_i\varepsilon_i\left(\omega_0 + \sum_{i=1}^{D}\omega_i x_{ni} - y_n\right)$$

- $E\left\{\tilde{E}_D(\omega)\right\} = \frac{1}{2}\sum_{n=1}^{N}\left(\omega_0 + \sum_{i=1}^{D}\omega_i x_{ni} - y_n\right)^2 + E\left\{\sum_{i=1}^{D}\omega_i\varepsilon_i\right\} - 2\sum_{i=1}^{D}\omega_i \cancel{E\{\varepsilon_i\}}\left(\omega_0 + \sum_{i=1}^{D}\omega_i x_{ni} - y_n\right)$

$$E\left\{\sum_{i=1}^{D}(\omega_i\varepsilon_i)^2 + \sum_{\substack{i,j=1\\ i\neq j}}^{D}\omega_i\varepsilon_i\omega_j\varepsilon_j\right\} = E\left\{\sum_{i=1}^{D}(\omega_i\varepsilon_i)^2\right\}$$

$\rightarrow E\left\{\tilde{E}_D(\omega)\right\} = \frac{1}{2}\sum_{n=1}^{N}\left(\left(\omega_0 + \sum_{i=1}^{D}\omega_i x_{ni} - y_n\right)^2 + \sum_{i=1}^{D}\omega_i^2\sigma^2\right)$

$$= \frac{1}{2}E_D(\omega) + \frac{N\sigma^2}{2}\sum_{i=1}^{D}\omega_i^2$$

Problem 3

a) k class $\rightarrow y \in \{1,0\}^k$ and $\sum_{i=1}^{k} y_i = 1$ $\quad y^{(i)} \in \{1, \dots, k\}$

notation: $T(y) = y \quad T(1) = \begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix}, \dots, T(k) = \begin{bmatrix} 0 \\ \vdots \\ k \\ 0 \end{bmatrix}$

$1\{True\} = 1, \quad 1\{False\} = 0 \rightarrow (T(y))_i = 1\{y = i\}$

$\varphi_1, \dots, \varphi_k \rightarrow$ specify the prob. of each outcome, $\sum_{i=1}^{k} \varphi_i = 1$

$$P(y;\varphi) = \varphi_1^{1\{y=1\}} \cdots \varphi_k^{1\{y=k\}} = \varphi_1^{(T(y))_1} \cdots \varphi_k^{(T(y))_k = 1 - \sum_{j=1}^{k-1} (T(y))_j}$$

$$= \exp\left[(T(y))_1 \log \varphi_1 + \dots + \left(1 - \sum_{j=1}^{k-1} (T(y))_j\right) \log \varphi_k \right]$$

$$= \exp\left[(T(y))_1 \log \frac{\varphi_1}{\varphi_k} + \dots + (T(y))_{k-1} \log \frac{\varphi_{k-1}}{\varphi_k} + \log \varphi_k \right]$$

write in Exponential Family Format $p(y;\eta) = b(y) \exp\{\eta^T T(y) - a(\eta)\}$

$$\rightarrow \underline{\eta} = \begin{bmatrix} \log\left(\frac{\varphi_1}{\varphi_k}\right) \\ \vdots \\ \log\left(\frac{\varphi_{k-1}}{\varphi_k}\right) \end{bmatrix}, \quad T(y)_i = \begin{bmatrix} 0 \\ \vdots \\ 1 \\ \vdots \end{bmatrix} \leftarrow i\text{th} \quad \rightarrow a(\eta) = -\log \varphi_k$$

$$\eta_i = \log \frac{\varphi_i}{\varphi_k}, \quad \eta_k = 0 \rightarrow \varphi_k e^{\eta_i} = \varphi_i \rightarrow \varphi_k = \frac{1}{\sum_{i=1}^{k} e^{\eta_i}} \overset{\eta_k = 0}{=} \frac{1}{\sum_{i=1}^{k-1} e^{\eta_i} + 1}$$

$$\Rightarrow \varphi_i = \frac{e^{\eta_i}}{\sum_{j=1}^{k-1} e^{\eta_j} + 1}, \quad w_i^T x = \eta_i \quad \rightarrow E\{y | x; \theta\} = h_w(x)$$

$$\Rightarrow P(y = i | x; \theta) = \frac{\exp(\underline{w}_i^T \underline{x})}{\sum_{j=1}^{k-1} e^{(\underline{w}_j^T x)} + 1}; \quad P(y = k | x; w) = \frac{1}{1 + \sum_{j=1}^{k-1} e^{(w_j^T x)}}$$

b)

$$L(w_1, \dots, w_{k-1}) = \sum_{i=1}^{n} \ln \prod_{j=1}^{k} \left(\frac{e^{\underline{w}_j^T \underline{x}^{(i)}}}{\sum_{l=1}^{k} e^{w_l^T \underline{x}^{(i)}}} \right)^{1\{y^{(i)} = j\}} = \sum_{i=1}^{n} \sum_{j=1}^{k} 1\{y^{(i)} = j\} \ln \frac{e^{w_j^T \underline{x}^{(i)}}}{\sum_{l=1}^{k} e^{w_l^T \underline{x}^{(i)}}}$$

3

c) $\frac{\partial L(\underline{w})}{\underline{w}_k} = \sum_{i=1}^{n} \mathbb{1}\{y^{(i)}=k\} \frac{\partial}{\partial w_k} \frac{e^{\underline{w}_j^T x^{(i)}}}{\sum_{l=1}^{K} e^{\underline{w}_l^T \underline{x}^{(i)}}} = \sum_{i=1}^{n} \mathbb{1}\{y^{(i)}=k\} \frac{\sum_{l=1, l\neq k}^{K} e^{\underline{w}_j^T x^{(i)}}}{\sum_{l=1}^{K} e^{\underline{w}_l^T x^{(i)}}} \underline{x}^{(i)}$

$= \sum_{i=1}^{n} \left(\mathbb{1}\{y^{(i)}=k\} - f_k(x^{(i)}; \underline{w}_k)\right)\underline{x}^{(i)}$; $f_k(\underline{x}^{(i)}; \underline{w}_k) \triangleq \frac{e^{\underline{w}_k^T x^{(i)}}}{\sum_{l=1}^{K} e^{\underline{w}_l^T x^{(i)}}}$

d) $\frac{\partial f(\underline{w})}{\partial w_k} = \frac{\partial L(\underline{w})}{\partial w_k} - \frac{\lambda}{2} \frac{\partial}{\partial w_k} \sum_{j=1}^{K-1} \underline{w}_j^T \underline{w}_j = \sum_{i=1}^{n} \left(\mathbb{1}\{y^{(i)}=k\} - f_k(\underline{x}^{(i)}; \underline{w}_k)\right) \underline{x}^{(i)} - \lambda \|\underline{w}_k\|^2$

Problem 4:

a) learn only on $\omega_j^{(i)} \rightarrow \hat{y} = \omega_j x_j^{(i)}$

$$J(\omega_j) = \frac{1}{2n} \sum_{i=1}^{n} (\hat{y}_i - y_i)^2 = \frac{1}{2n} \sum_{i=1}^{n} [\omega_j x_j^{(i)} - y_i]^2$$

$$\frac{\partial J(\omega_j)}{\partial \omega_j} = 0 \rightsquigarrow \frac{1}{n} \sum_{i=1}^{n} [\omega_j x_j^{(i)} - y_i] x_j^{(i)} = 0$$

$$\rightarrow \omega_j \sum_{i=1}^{n} (x_j^{(i)})^2 = \sum_{i=1}^{n} y_i x_j^{(i)} \rightarrow \omega_j \underline{x}_j^T \underline{x}_j = x_j^T \underline{y}$$

$$\rightarrow \omega_j = \frac{x_j^T y}{\underline{x}_j^T \underline{x}_j}$$

b) $J(\underline{\omega}) = \frac{1}{2} (X\underline{\omega} - \underline{y})^T (X\underline{\omega} - \underline{y})$

$\nabla_{\underline{\omega}} J(\underline{\omega}) = 0 \rightsquigarrow X^T X \underline{\omega} - X^T \underline{y} = 0 \rightarrow \underline{\omega} = (X^T X)^{-1} X^T \underline{y}$

$\underline{x}_i \perp x_j \; (i \neq j) \rightarrow X^T X$ Diagonal $\rightsquigarrow$

$$\begin{bmatrix} a_1 & & 0 \\ & \ddots & \\ 0 & & a_m \end{bmatrix} \begin{bmatrix} \omega_1 \\ \vdots \\ \omega_m \end{bmatrix} = \begin{bmatrix} b_1 \\ \vdots \\ b_m \end{bmatrix}$$

$(X^T X)_i \triangleq a_i$

$(X^T y)_i = C_i$

$$\Rightarrow a_i \omega_i = C_i \rightsquigarrow \omega_i = \frac{C_i}{a_i} = \frac{(X^T y)_i}{(X^T X)_i}$$

c) estimator $E\{x\} \simeq \frac{1}{n} \sum_{i=1}^{n} x^{(i)}$, estimator Var $\simeq \frac{1}{n} \sum_{i=1}^{n} (x^{(i)} - E\{x\})^2$

$y^{(i)} = \omega_0 + \omega_j x_j^{(i)}$

$$J(\underline{\omega}) = \frac{1}{2n} \sum_{i=1}^{n} (\hat{y}^{(i)} - y^{(i)})^2 = \frac{1}{2n} \sum_{i=1}^{n} [\omega_0 + \omega_j x_j^{(i)} - y^{(i)}]^2$$

- $\dfrac{\partial J(w)}{\partial w_0} = 0 \rightarrow \dfrac{1}{n} \sum_{i=1}^{n} [w_0 + w_j x_j^{(i)} - y_i] = 0 \rightarrow w_0 = \frac{1}{n} \sum_{i=1}^{n} y_i - \frac{1}{n} w_j \sum_{i=1}^{n} x_j^{(i)}$

$$\Rightarrow w_0 = E\{y\} - w_j E\{x_j\}$$

- $\dfrac{\partial J(w)}{\partial w_j} = 0 \rightarrow \dfrac{1}{n} \sum_{i=1}^{n} (w_0 + w_j x_j^{(i)} - y_i)(x_j^{(i)}) = 0$

$$\Rightarrow \left(w_0 \sum_{i=1}^{n} x_j^{(i)} + w_j \sum_{i=1}^{n} (x_j^{(i)})^2 - \sum_{i=1}^{n} y^{(i)} x_j^{(i)} \right) \frac{1}{n} = 0$$

$$\Rightarrow \frac{1}{n^2} \sum_{i=1}^{n} y^{(i)} \sum_{i=1}^{n} x_j^{(i)} - \frac{1}{n^2} w_j \sum_{i=1}^{n} x_j^{(i)} \sum_{i=1}^{n} x_j^{(i)} + \frac{1}{n} w_j \sum_{i=1}^{n} (x_j^{(i)})^2 - \frac{1}{n} \sum_{i=1}^{n} y^{(i)} x_j^{(i)} = 0$$

$$\Rightarrow E\{y\}E\{x_j\} - w_j E^2\{x_j\} + w_j E\{(x_j^{(i)})^2\} - E\{y x_j\} = 0$$

$$\Rightarrow w_j = \frac{E\{y\}E\{x_j\} - E\{y x_j\}}{Var(x_j)} = \frac{Cov(x_j, y)}{Var(x_j)} \quad \blacksquare$$

Problem 5:

a) let x be positive Random Variable

$$\rightarrow E\{x\} = \int_{-\infty}^{+\infty} x f(x)\, dx = \int_{0}^{+\infty} x f(x)\, dx \geqslant \int_{\alpha}^{+\infty} x f(x)\, dx \geqslant \int_{\alpha}^{+\infty} \alpha f(x)\, dx$$

(α is positive $\alpha \geqslant 0$) $\quad \{x \geqslant \alpha\}$

$$= \alpha\, p(X \geqslant \alpha) \Rightarrow \frac{E\{x\}}{\alpha} \geqslant p(X \geqslant \alpha)$$

b) let z is RV $\rightarrow$ $Y \triangleq (z - E\{z\})^2$ $\rightsquigarrow$ so Y is positive RV

markove $\rightsquigarrow$ $p(Y \geqslant \varepsilon^2) \leqslant \dfrac{E\{Y\}}{\varepsilon^2}$

$$E\{y\} = E\{(z - E\{z\})\} = Var(z) = \sigma^2$$

$$p(Y \geqslant \varepsilon^2) = p((z - E\{z\})^2 \geqslant \varepsilon^2) = p(|z - \mu| \geqslant \varepsilon)$$

$$\Rightarrow p(|z - \mu| \geqslant \varepsilon) \leqslant \frac{\sigma^2}{\varepsilon^2}$$

7

c)

z_i is Random Variable that determines the poin falls in Circle or not

$p(z_i=1) = \frac{\pi}{4}$, $P(z_i=0) = 1-\pi/4$

$E\{z_i\} = 1 \times \frac{\pi}{4} + 0(1-\pi/4) = \frac{\pi}{4}$

$Var(z_i) = E\{z_i^2\} - E^2\{z_i\} = \frac{\pi}{4} - \frac{\pi^2}{16}$

the estimator is $\hat{\pi} = \frac{4}{n} \sum z_i$ $\qquad E\{\hat{\pi}\} = E\{\frac{4}{n} \sum_{i=1}^{n} z_i\} = \frac{4}{n} \frac{n\pi}{4} = \pi$

$$Var\{\hat{\pi}\} = \frac{16}{n^2} Var(\sum_{i=1}^{n} z_i) = \frac{16}{n^2} \cdot n \times (\frac{\pi}{4} - \frac{\pi^2}{16})$$

$$= \frac{\pi}{n}[4-\pi]$$

chebyshev's inequality $\rightarrow$ $p(|\hat{\pi} - \pi| < 0.01) \geq 0.95 \rightarrow p(|\hat{\pi}-\pi| \geq 0.01) \leq 0.05$

$\frac{\sigma^2}{\varepsilon^2} \leq 0.05 \rightarrow \frac{\pi}{n}[4-\pi] \times 10^4 \leq 0.05 \rightarrow \frac{\pi}{n}[4-\pi] \leq 5 \times 10^{-6}$

$\varepsilon = 0.01$

$n \geq 539353.25 \rightarrow n_{min} = 539354$

Problem 6:

a) $AA^{-1} = I \sim \sigma(A^{-1}) = \frac{1}{\sigma(A)} \sim \sigma_{max}(A^{-1}) = \frac{1}{\sigma_{min}(A)}$

$$\sigma_{max}(A)\sigma_{max}(A^{-1}) = \frac{\sigma_{max}(A)}{\sigma_{min}(A)} \geq 1 \blacksquare$$

b) $\|A\|_2 = \sigma_{max}(A)$, $\|A\|_F = \sqrt{tr(A^H A)} = \sqrt{\sum_{i=1}^{rank(A)} \sigma_i^2} = \sqrt{\sigma_{max}^2 + \cdots \sigma_{min}^2}$

$\rightarrow \|A\|_2 = \sigma_{max} \leq \sqrt{\sigma_{max}^2 + \cdots \sigma_{min}^2} = \|A\|_F \leq \sqrt{(rank(A)\,\sigma_{max}^2} = \sqrt{rank(A)}\,\sigma_{max}$

$= \sqrt{rank(A)}\,\|A\|_2$

8

Problem 7:

$$\sigma(z) = \frac{1}{1+e^{-z}} \qquad \tanh(z) = \frac{e^{z}-e^{-z}}{e^{z}+e^{-z}} = \frac{e^{z}-e^{-z}-e^{z}+e^{z}}{e^{z}+e^{-z}}$$

$$= -1 + \frac{2e^{z}}{e^{z}+e^{-z}} = \frac{2}{e^{-2z}+1} - 1 = 2\sigma(2z)-1$$

$$\therefore\ y(x, \omega) = \omega_0 + \sum_{j=1}^{n}\left[\omega_j \tanh\left(\frac{x-\mu_j}{s}\right)\right] = \omega_0 + \sum_{j=1}^{n} \omega_j \left[2\sigma\left(\frac{x-\mu_j}{s/2}\right)-1\right]$$

$$= \underbrace{\omega_0 - \sum_{j=1}^{n} \omega_j}_{\omega_0'} + \sum_{j=1}^{n} \underbrace{2\omega_j}_{\omega_j'} \sigma\left(\frac{x-\mu_j}{\underbrace{s/2}_{s'}}\right)$$

$$= \omega_0' + \sum_{j=1}^{n} \omega_j' \sigma\left(\frac{x-\mu_j}{s'}\right)$$