题目一.

(1) 若 $p(w_{max}|x) < \frac{1}{c}$，则 $p(w_i \neq w_{max}|x) < \frac{1}{c}$

$\sum_{i=1}^{c} p(w_i|x) < c \cdot \frac{1}{c} = 1$ 矛盾

$\therefore p(w_{max}|x) \geq \frac{1}{c}$

(2) $R(\alpha_i|x) = \sum_j R(\alpha_i|w_j)\, p(w_j|x) = 1 - p(w_i|x)$

$p(w_{max}|x) \geq p(w_i|x) \quad R(\alpha_{max}|x) \leq R(\alpha_i|x)$

$p(correct) = \int \max p(x|w_i)\,p(w_i)\,dx = \int \max p(w_i|x)\,p(x)\,dx$

$= \int p(w_{max}|x)\,p(x)\,dx$

$p(error) = 1 - p(correct) = 1 - \int p(w_{max}|x)\,p(x)\,dx$

(3) $p(error) = 1 - \int p(w_{max}|x)\,p(x)\,dx \leq 1 - \frac{1}{c}\int p(x)\,dx = \frac{c-1}{c}$

(4) 当 $p(w_1|x) = p(w_2|x) \cdots = p(w_c|x) = \frac{1}{c}$ 时 $p(w_{max}|x) = \frac{1}{c}$

$p(error) = \frac{c-1}{c}$

题目二、

(1) 贝叶斯风险最小决策

$R(\alpha_i|x) = \sum_{j=1}^{c} \lambda_{ij}\, p(w_j|x) \propto \sum_{j=1}^{c} \lambda_{ij}\, p(x|w_j)\,p(w_j)$

$\arg\min_i R(\alpha_i|x)$

$= \arg\min_i \sum_{j=1}^{c} \lambda_{ij}\, p(x|w_j)\,p(w_j)$

最小错误率决策

$R(\alpha_i|x) = 1 - p(w_i|x) \propto -p(x|w_i)\,p(w_i)$

$\arg\min R(\alpha_i|x)$

$= \arg\max p(x|w_i)p(w_i)$

(2) $R(\alpha_i|x) = \begin{cases} \lambda_s [1 - p(w_i|x)] & i=1,2\cdots c \\ \lambda_r & \text{reject} \end{cases}$

$\arg\min_i R(\alpha_i|x)$

$= \begin{cases} \arg\max_i p(w_i|x) & \text{if } \max p(w_i|x) > 1 - \frac{\lambda_r}{\lambda_s} \\ c+1 & \text{otherwise} \end{cases}$

$= \begin{cases} \arg\max_i p(x|w_i)p(w_i) & \text{if } \max p(x|w_i)p(w_i) > 1 - \frac{\lambda_r}{\lambda_s} \\ c+1 & \text{otherwise} \end{cases}$

题目三,

(1) $|\lambda I - \Sigma| = 0$

$\begin{vmatrix} \lambda-1 & 0 & 0 \\ 0 & \lambda-5 & -2 \\ 0 & -2 & \lambda-5 \end{vmatrix} = 0$

$(\lambda-1)[(\lambda-5)^2-4] = 0$

$\lambda_1 = 1 \quad \lambda_2 = 3 \quad \lambda_3 = 7$

$\lambda_1 = 1 \quad (I-\Sigma)x = 0 \quad \begin{bmatrix} 0 & 0 & 0 \\ 0 & -4 & -2 \\ 0 & -2 & -4 \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = 0$

$\begin{bmatrix} 0 & 1 & 2 \\ 0 & 0 & 6 \\ 0 & 0 & 0 \end{bmatrix} \quad \xi_1 = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}$

$\lambda_2 = 3 \quad (3I-\Sigma)x = 0 \quad \begin{bmatrix} 2 & 0 & 0 \\ 0 & -2 & -2 \\ 0 & -2 & -2 \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = 0$

$\begin{bmatrix} 2 & 0 & 0 \\ 0 & 1 & 1 \\ 0 & 0 & 0 \end{bmatrix} \quad \xi_2 = \begin{bmatrix} 0 \\ 1 \\ -1 \end{bmatrix}$

$\lambda_3 = 7 \quad (7I-\Sigma)x = 0 \quad \begin{bmatrix} 6 & 0 & 0 \\ 0 & 2 & -2 \\ 0 & -2 & 2 \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = 0$

$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & -1 \\ 0 & 0 & 0 \end{bmatrix} \quad \xi_3 = \begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix}$

$$\phi = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ 0 & -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{bmatrix} \quad \Lambda = diag(1, 3, 7)$$

$$A_w = \phi \Lambda^{-\frac{1}{2}} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ 0 & -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & \frac{1}{\sqrt{3}} & 0 \\ 0 & 0 & \frac{1}{\sqrt{7}} \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \frac{1}{\sqrt{6}} & \frac{1}{\sqrt{14}} \\ 0 & -\frac{1}{\sqrt{6}} & \frac{1}{\sqrt{14}} \end{bmatrix}$$

将此分布转化为 $N(0, I)$

$$X_w = A_w^T (X - \mu)$$

(2) $$X_w = A_w^T (X - \mu) = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \frac{1}{\sqrt{6}} & -\frac{1}{\sqrt{6}} \\ 0 & \frac{1}{\sqrt{14}} & \frac{1}{\sqrt{14}} \end{bmatrix} \begin{bmatrix} -0.5 \\ -2 \\ -1 \end{bmatrix} = \begin{bmatrix} -0.5 \\ -\frac{1}{\sqrt{6}} \\ -\frac{3}{\sqrt{14}} \end{bmatrix}$$

(3) $$d(x_0, \mu) = (x_0 - \mu)^T \Sigma^{-1} (x_0 - \mu)$$

$$= \begin{bmatrix} -0.5 & -2 & -1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & \frac{5}{21} & -\frac{2}{21} \\ 0 & -\frac{2}{21} & \frac{5}{21} \end{bmatrix} \begin{bmatrix} -0.5 \\ -2 \\ -1 \end{bmatrix}$$

$$= \frac{89}{84}$$

$$d(x_w, 0) = x_w^T x_w = \frac{1}{4} + \frac{1}{6} + \frac{9}{14} = \frac{89}{84}$$

$$\therefore d(x_0, \mu) = d(x_w, 0)$$

(4) $$x' = T^t x$$

$$\mu' = \sum_k x_k' = T^t \sum_k x = T^t \mu$$

$$Z = \begin{bmatrix} x_1^t - \mu^t \\ \vdots \\ x_n^t - \mu^t \end{bmatrix}$$

$$\Sigma' = \frac{1}{n-1} Z^T Z = \frac{1}{n-1} \begin{bmatrix} x_1 - \mu & \cdots & x_n - \mu \end{bmatrix} \begin{bmatrix} (x_1 - \mu)^t \\ \vdots \\ (x_n - \mu)^t \end{bmatrix}$$

$$= \frac{1}{n-1} \sum_{k=1}^{n} (x_k' - \mu')(x_k' - \mu')^t$$

$$= \frac{1}{n-1}\sum_{k=1}^{n}[T^t(x_k-\mu)][T^t(x_k-\mu)]^t$$

$$= T^t\left[\frac{1}{n-1}\sum_{k=1}^{n}(x_k-\mu)(x_k-\mu)^t\right]T$$

$$= T^t\Sigma T$$

$$p(x_0) = \frac{1}{(2\pi)^{d/2}|\Sigma|^{1/2}}\exp\left[-\frac{1}{2}(x_0-\mu)^t\Sigma^{-1}(x_0-\mu)\right]$$

$$p(T^t x_0) = \frac{1}{(2\pi)^{d/2}|\Sigma'|^{1/2}}\exp\left[-\frac{1}{2}(T^t x_0 - T^t\mu)^t(T^t\Sigma T)^{-1}(T^t x_0 - T^t\mu)\right]$$

$$= \frac{1}{(2\pi)^{d/2}|\Sigma'|^{1/2}}\exp\left[-\frac{1}{2}(x_0-\mu)^t T T^{-1}\Sigma^{-1}(T^t)^{-1}T^t(x_0-\mu)\right]$$

$$= \frac{1}{(2\pi)^{d/2}|\Sigma'|^{1/2}}\exp\left[-\frac{1}{2}(x_0-\mu)^t\Sigma^{-1}(x_0-\mu)\right]$$

若概率密度不变，则需满足 $|\Sigma| = |\Sigma'| = |T^t||\Sigma||T|$

$$|T||T^t| = 1$$

$$|T|^2 = 1$$

题目四

(1) $p(x|\omega_i) = \frac{1}{(2\pi)^{d/2}|\Sigma_i|^{1/2}}\exp\left[-\frac{1}{2}(x-\mu_i)^t\Sigma_i^{-1}(x-\mu_i)\right]$

(2) (a) $\Sigma_i \neq \Sigma_j$

$$g_i(x) = \ln p(x|\omega_i) + \ln p(\omega_i)$$

$$= -\frac{1}{2}(X-\mu_i)^t\Sigma_i^{-1}(X-\mu_i) - \frac{d}{2}\ln 2\pi - \frac{1}{2}\ln|\Sigma_i| + \ln p(w_i)$$

$$= -\frac{1}{2}(X-\mu_i)^t\Sigma_i^{-1}(X-\mu_i) - \frac{1}{2}\ln|\Sigma_i|$$

(b) $\Sigma_i = \Sigma$

$$g_i(x) = -\frac{1}{2}(X-\mu_i)^t\Sigma^{-1}(X-\mu_i)$$

(3) ① 利 PCA 对样本进行降维，使得在新的样本空间内协方差矩阵非奇异

② 对协方差矩阵进行正则化，$\Sigma' = \Sigma + \lambda I$

题目五.

(1) 利用 numpy 库中的函数生成多元高斯分布

(2) $$g_i(x) = -\frac{1}{2}(X-\mu_i)^t\Sigma_i^{-1}(X-\mu_i) - \frac{d}{2}\ln 2\pi - \frac{1}{2}\ln|\Sigma_i| + \ln p(w_i)$$

$g_1(x) = g_2(x)$ $\quad \Sigma_1 = \Sigma_2 = I \quad p(w_1) = p(w_2)$

$\Downarrow$

$$-\frac{1}{2}(X-\mu_1)^t\Sigma_1^{-1}(X-\mu_1) = -\frac{1}{2}(X-\mu_2)^t\Sigma_2^{-1}(X-\mu_2)$$

$$-\mu_1^tX - X^t\mu_1 + \mu_1^t\mu_1 = -\mu_2^tX - X^t\mu_2 + \mu_2^t\mu_2$$

$$-2\mu_1^tX + \mu_1^t\mu_1 = -2\mu_2^tX + \mu_2^t\mu_2$$

$$2(\mu_1^t - \mu_2^t)X = \mu_1^t\mu_1 - \mu_2^t\mu_2$$

$\mu_1^t = (1,0) \quad \mu_2^t = (-1,0)$

$\Downarrow$

$$W^tX + b = 0 \qquad W = \begin{bmatrix} 1 \\ 0 \end{bmatrix} \qquad b = 0$$

$$[1, 0][x_1\ x_2]^T = 0$$

$$x_1 = 0$$

$\Downarrow$

$x_1 > 0$ 判为 w_1

$x_1 < 0$ 判为 w_2

(3) $$\text{error} = \frac{n(x \in R_1, w_2) + n(x \in R_2, w_1)}{N} = 0.15$$

(4) n从100以步长100增加到1000时，误差波动较大，增加到10000时能看出误差收敛趋势