Solutions for Pattern Classification CH.2

Building C303, Harbin Institute of Technology, Shenzhen University Town, Xili, Shenzhen,Guangdong Province, P.R. China
Yonghui Wu, Yaoyun Zhang; yonghui.wu@gmail.com

Problem 12

- a)
  Suppose that $P(\omega_{max}|x) < \frac{1}{c}$, so we get :

$$P(\omega_i|x) \leq P(\omega_{max}|x) < \frac{1}{c}, i = (1, \cdots, c)$$

  we obtain:

$$\sum_{i=1}^{c} P(\omega_i|x) < \sum_{i=1}^{c} \frac{1}{c} = 1$$

  contradict with: $\sum_{i=1}^{c} P(\omega_i|x) = 1$

- b)
  According to minimum-error-rate:

$$\begin{aligned} P(error) =& \sum P(error|x)p(x)dx \\ =& \sum_{i=1,\omega_i \neq \omega_{max}}^{c} P(\omega_i|x)p(x)dx \\ =& \sum (1 - P(\omega_{max}|x))p(x)dx \\ =& 1 - \sum P(\omega_{max}|x)p(x)dx \end{aligned}$$

- c)
  see above:

$$\begin{aligned} P(error) =& 1 - \sum P(\omega_{max}|x)p(x)dx \\ \leq& 1 - \sum \frac{1}{c}p(x)dx = 1 - \frac{1}{c} = \frac{c-1}{c} \end{aligned} \tag{1}$$

- d)
  When $P(\omega_1|x) = P(\omega_2|x) = \cdots = P(\omega_c|x)$,that's to say:$P(\omega_{max}|x) = \frac{1}{c}$, then we obtain:

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

Problem 14

- a

  Case 1: $g_i(x)$ $i \neq c+1$, therefore

$$\begin{cases} g_i(x) > g_j(x) & i \neq j \ and \ j \neq c+1 \\ g_i(x) > g_{c+1}(x) \end{cases} \Rightarrow \begin{cases} p(x|\omega_i)P(\omega_i) & > & p(x|\omega_j)P(\omega_j) \\ p(x|\omega_i)P(\omega_i) & > & \frac{\lambda_s \ - \ \lambda_r}{\lambda_s}\Sigma_{i=1}^{c}p(x|\omega_i P(\omega_i)) \end{cases}$$

$$\Rightarrow \begin{cases} P(\omega_i|x) & > & P(\omega_j|x) \\ P(\omega_i|x) & > & 1 - \frac{\lambda_r}{\lambda_s} \end{cases}$$

  This is correspondent with the decision rule of Problem 13, thus in case 1 discriminant function can get the minimum risk, that is it is optimal.

  Case 2: decide $g_{c+1}(x)$, therefore

  $g_{c+1}(x) \ > \ g_i(x) \ j \neq \ c+1$

  $\Rightarrow \frac{\lambda_s \ - \ \lambda_r}{\lambda_s} \sum_{i=1}^{c} p(x|\omega_i)P(\omega_i) \ > \ p(x|\omega_i)P(\omega_i) \ \geq \ p(x|\omega_i)P(\omega_i) \ \Rightarrow \ \frac{\lambda_s \ - \ \lambda_r}{\lambda_s} \ > \ P(\omega_i|x)$

  This result is correspondence with the decision rule of Problem 13, thus in Case 2 discriminant function can get the minimum risk, that is it is optimal

- b

$$\begin{aligned} g_1(x) &= \frac{1}{2}p(x|\omega_1) \\ g_2(x) &= \frac{1}{2}p(x|\omega_2) \\ g_3(x) &= \frac{3}{8}(p(x|\omega_1 \ + \ p(x|\omega_2) \end{aligned}$$

- c

  R_3 is changed from $(-\infty, +\infty)$ to $\emptyset$

- d

$$\begin{aligned} g_1(x) &= \frac{1}{3\sqrt{2\pi}}e^{\frac{-(x-1)^2}{2}} \\ g_2(x) &= \frac{4}{3\sqrt{2\pi}}e^{-2x^2} \\ g_3(x) &= \frac{1}{6\sqrt{2\pi}}e^{\frac{-(x-1)^2}{2}} \ + \ \frac{2}{3\sqrt{2\pi}}e^{-2x^2} \Rightarrow \end{aligned}$$

$$R_1: \; g_1(x) \;\geq\; g_2(x) \; and \; g_1(x) \;\geq\; g_3(x)$$
$$R_2: \; g_2(x) \;\geq\; g_1(x) \; and \; g_2(x) \;\geq\; g_3(x)$$
$$R_3: \; g_3(x) \;\geq\; g_1(x) \; and \; g_3(x) \;\geq\; g_2(x) \Rightarrow$$

$$R_1 \qquad : \; x < \frac{-1-2\sqrt{1+3\ln 2}}{3}, \; x > \frac{-1+2\sqrt{1+3\ln 2}}{3}$$
$$R_2 \qquad : \frac{-1-2\sqrt{1+3\ln 2}}{3} < x < \frac{-1+2\sqrt{1+3\ln 2}}{3}$$

Problem 22

- a

$$H(P(x)) = -\int (p(x)\ln(p(x))dx = \varepsilon[\ln(\frac{1}{p(x)})]$$
$$\varepsilon[\ln(\frac{1}{p(x)})] = \frac{d}{2}\ln(2\pi) + \ln(|\Sigma|^{\frac{1}{2}}) + \frac{1}{2}\varepsilon[(\overrightarrow{x} - \overrightarrow{\mu})^t\Sigma^{-1}(\overrightarrow{x} - \overrightarrow{\mu})]$$

As we know, when x_i are independent, we have this:

$$(\overrightarrow{x} - \overrightarrow{\mu})^t\Sigma^{-1}(\overrightarrow{x} - \overrightarrow{\mu}) = \sum_{i=1}^{d}(\frac{x_i - \mu_i}{\delta_i})^2 \sim N(0,1)$$

therefore,

$$\varepsilon[(\overrightarrow{x} - \overrightarrow{\mu})^t\Sigma^{-1}(\overrightarrow{x} - \overrightarrow{\mu})] = \varepsilon[\sum_{i=1}^{d}(\frac{x_i - \mu_i}{\delta_i})^2] = \sum_{i=1}^{d}\varepsilon[(\frac{x_i - \mu_i}{\delta_i})^2 - 0^2] = d$$

so, $H(p(x)) = \frac{d}{2}\ln(2\pi) + \frac{1}{2}\ln(|\Sigma|) + \frac{1}{2}d$

Problem 23

- a

$$x_0 = (0.5, 0, 1)^t$$
$$p(x_0|\omega) = \frac{1}{(2\pi)^{\frac{3}{2}}|\Sigma|^{\frac{1}{2}}} exp\left[-\frac{1}{2}(x_0 - \mu)^t\Sigma^{-1}(x_0 - \mu)\right]$$
$$|\Sigma| = \begin{vmatrix} 1 & 0 & 0 \\ 0 & 5 & 2 \\ 0 & 2 & 5 \end{vmatrix} = \begin{vmatrix} 5 & 2 \\ 2 & 5 \end{vmatrix} = 21$$
$$\Sigma^{-1} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 5 & 2 \\ 0 & 2 & 5 \end{pmatrix}^{-1} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \frac{5}{21} & \frac{-2}{21} \\ 0 & \frac{-2}{21} & \frac{5}{21} \end{pmatrix}$$

$$(x_0-\mu)^t\Sigma^{-1}(x_0-\mu) = \left[\begin{pmatrix}0.5\\0\\1\end{pmatrix} - \begin{pmatrix}1\\2\\2\end{pmatrix}\right]^t \begin{pmatrix}1 & 0 & 0\\0 & \frac{5}{21} & \frac{-2}{21}\\0 & \frac{-2}{21} & \frac{5}{21}\end{pmatrix} \left[\begin{pmatrix}0.5\\0\\1\end{pmatrix} - \begin{pmatrix}1\\2\\2\end{pmatrix}\right]$$
$$= \begin{pmatrix}-0.5\\-\frac{8}{21}\\-\frac{1}{21}\end{pmatrix}^t \begin{pmatrix}-0.5\\-\frac{8}{21}\\-\frac{1}{21}\end{pmatrix}$$
$$=1.06$$

$$p(x_0|\omega) = \frac{1}{(2\pi)^{\frac{3}{2}}(21)^{\frac{1}{2}}}e^{-\frac{1}{2}\times 1.06} = 8.16\times 10^{-3}$$

- b

$$|\Sigma - \lambda I| = 0$$

Then we calculate the eigenvalues:

$$\begin{vmatrix}1-\lambda & 0 & 0\\0 & 5-\lambda & 2\\0 & 2 & 5-\lambda\end{vmatrix} = (1-\lambda)\left[(5-\lambda)^2-4\right] = 0$$

$$\lambda = 1, \lambda = 3, \lambda = 7$$
$$\Lambda = \begin{pmatrix}1 & 0 & 0\\0 & 3 & 0\\0 & 0 & 7\end{pmatrix}$$

then, we calculate its eigenvetors:

$$\sum e_1 = \lambda_1 e_1$$

$$\begin{pmatrix}1 & 0 & 0\\0 & 5 & 2\\0 & 2 & 5\end{pmatrix}\begin{pmatrix}x_1\\x_2\\x_3\end{pmatrix} = \begin{pmatrix}x_1\\x_2\\x_3\end{pmatrix}, \begin{pmatrix}x_1\\5x_2+2x_3\\2x_2+5x_3\end{pmatrix} = \begin{pmatrix}x_1\\x_2\\x_3\end{pmatrix}$$

$$5x_2+2x_3 = x_2$$
$$2x_2+5x_3 = x_3$$

let $x_1 = 1$, we get:

$$x_2 = 0, x_3 = 0$$

then we get:

$$e_1 = \begin{pmatrix}1\\0\\0\end{pmatrix}$$

Similarly,

$$\begin{pmatrix} x_1 \\ 5x_2+2x_3 \\ 2x_2+5x_3 \end{pmatrix} = \begin{pmatrix} 3x_1 \\ 3x_2 \\ 3x_3 \end{pmatrix}, then\ we\ get : x_1 = 0, x_2 = -x_3, let x_2 = 1, then :$$

$$e_2 = \begin{pmatrix} 0 \\ 1 \\ -1 \end{pmatrix}, by\ nomalization, e_2 = \begin{pmatrix} 0 \\ \frac{1}{\sqrt{2}} \\ -\frac{1}{\sqrt{2}} \end{pmatrix}$$

$$\begin{pmatrix} x_1 \\ 5x_2+2x_3 \\ 2x_2+5x_3 \end{pmatrix} = \begin{pmatrix} 7x_1 \\ 7x_2 \\ 7x_3 \end{pmatrix}, then\ we\ get : x_1 = 0, x_2 = x_3, let x_2 = 1, then :$$

$$e_3 = \begin{pmatrix} 0 \\ 1 \\ 1 \end{pmatrix}, by\ nomalization, e_3 = \begin{pmatrix} 0 \\ \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{pmatrix}$$

Hence:

$$\Phi = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ 0 & -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{pmatrix}$$

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

- c

$$\begin{aligned} x_w =& A_w^t(x_0 - \mu) \\ =& \begin{pmatrix} 1 & 0 & 0 \\ 0 & \frac{1}{\sqrt{6}} & \frac{1}{\sqrt{14}} \\ 0 & \frac{-1}{\sqrt{6}} & \frac{1}{\sqrt{14}} \end{pmatrix}^t \left[\begin{pmatrix} 0.5 \\ 0 \\ 1 \end{pmatrix} - \begin{pmatrix} 1 \\ 2 \\ 2 \end{pmatrix} \right] \\ =& \begin{pmatrix} 1 & 0 & 0 \\ 0 & \frac{1}{\sqrt{6}} & \frac{1}{\sqrt{14}} \\ 0 & \frac{-1}{\sqrt{6}} & \frac{1}{\sqrt{14}} \end{pmatrix}^t \begin{pmatrix} -0.5 \\ -2 \\ -1 \end{pmatrix} \\ =& \begin{pmatrix} -0.5 \\ \frac{-1}{\sqrt{6}} \\ -\frac{3}{\sqrt{14}} \end{pmatrix} \end{aligned}$$

- d The square of the Mahalanobis distance from x_0 to μ is
$$r^2 = (x_0 - \mu)^t \Sigma^{-1}(x_0 - \mu) = 1.06$$
The square of the Mahalanobis distance from x_w to 0 is
$$r_w^2 = x_w^t x_w = \begin{pmatrix} -0.5 & \frac{1}{\sqrt{6}} & -\frac{3}{\sqrt{14}} \end{pmatrix} \begin{pmatrix} -0.5 \\ \frac{1}{\sqrt{6}} \\ -\frac{3}{\sqrt{14}} \end{pmatrix} = 1.06$$
Thus, $r = r_w$.
- e
$$p(x_0|N(\mu, \Sigma)) \sim p(x_0) = \frac{1}{(2\pi)^{\frac{d}{2}}|\Sigma|^{\frac{1}{2}}} exp[-\frac{1}{2}(x_0 - \mu)^t \Sigma^{-1}(x_0 - \mu)]$$
if $x^{'} = T^t x_0$, then
$$\mu^{'} = \frac{1}{n}\sum_{k=1}^{n} x_k^{'} = \frac{1}{n}\sum_{k=1}^{n} x_{0k}^{'} = \frac{1}{n}T^t \sum_{k=1}^{n} x_{0k}^{'} = T^t \mu$$
$$\begin{aligned}\Sigma^{'} &= \sum_{k=1}^{n}(x_k^{'} - \mu^{'})(x_k^{'} - \mu^{'})^t \\ &= \sum_{k=1}^{n} T^t(x_{0k} - \mu)(x_{0k} - \mu)^t T \\ &= T^t\left[\sum_{k=1}^{n}(x_{0k} - \mu)(x_{0k} - \mu)^t\right] T \\ &= T^t \Sigma T\end{aligned}$$
Thus, we have $p(T^t x_0|N(T^t\mu, T^t\Sigma T))$.

$$\begin{aligned}p(T^t x_0|N(T^t\mu, T^t\Sigma T)) &= \frac{1}{(2\pi)^{\frac{d}{2}}|\Sigma^{'}|^{\frac{1}{2}}} \exp -\frac{1}{2}(T^t x - T^t\mu)^t (T^t\Sigma T)^{-1}(T^t x - T^t \mu) \\ &= \frac{1}{(2\pi)^{\frac{d}{2}}|\Sigma^{'}|^{\frac{1}{2}}} \exp -\frac{1}{2}(x^t T - \mu^t T)(T^{-1}(T^t\Sigma)^{-1}(T^t x - T^t\mu) \\ &= \frac{1}{(2\pi)^{\frac{d}{2}}|\Sigma^{'}|^{\frac{1}{2}}} \exp -\frac{1}{2}(x^t - \mu^t)TT^{-1}\Sigma^{-1}(T^t)^{-1}T^t(x - \mu) \\ &= \frac{1}{(2\pi)^{\frac{d}{2}}|T^t\Sigma T|^{\frac{1}{2}}} \exp -\frac{1}{2}(x - \mu)^t\Sigma^{-1}(x - \mu) \\ &= \frac{1}{(2\pi)^{\frac{d}{2}|\Sigma|^{\frac{1}{2}}}|T|} \exp -\frac{1}{2}(x - \mu)^t\Sigma^{-1}(x - \mu)\end{aligned}$$
thus for some $|T| = 1$, we have:
$$p(x_0|\mu, \Sigma) = p(T^t x_0|N(T^t\mu, T^t\Sigma T))$$

- f

  Since $\Sigma\Phi = \Phi\Lambda$, so $\Sigma = \Phi\Lambda\Phi^{-1}$, meanwhile, Φ is a symmetric matrix, thus, $\Phi^{-1} = \Phi^t$.

$$\begin{aligned}A_w^t \Sigma A_w =&(\Phi\Lambda^{-\frac{1}{2}})^t \Sigma - \frac{1}{2})\\ =&\Lambda^{-\frac{1}{2}}\Phi^t\Phi\Lambda\Phi^t\Phi\Lambda - \frac{1}{2}\\ =&\Lambda^{-\frac{1}{2}}\Lambda\Lambda^{-\frac{1}{2}}\\ =&I\end{aligned}$$

Problem 25

- a

  From Eq. 59

$$\begin{aligned}g_i(x) =& -\frac{1}{2}(x-\mu_i)^t(\sum)^{-1}(x-\mu_i) + \ln p(\omega_i)\\ =&\frac{1}{2}X^t(\sum)^{-1}X + \frac{1}{2}X^t(\sum)^{-1}\mu_i + \frac{1}{2}\mu_i^t(\sum)^{-1}X - \frac{1}{2}\mu_i^t(\sum)^{-1}\mu_i + \ln p(\omega_i)\end{aligned}$$

  The quadratic term is independent of i, and $\sum^{-1}$ is a symmetrical matrix, it can be rewrite as:

$$\begin{aligned}g_i(x) =&\frac{1}{2}X^t(\sum)^{-1}\mu_i + \frac{1}{2}\mu_i^t(\sum)^{-1}X - \frac{1}{2}\mu_i^t(\sum)^{-1}\mu_i + \ln p(\omega_i)\\ =&\frac{1}{2}(\mu_i^t(\sum)^{-1}X)^t + \frac{1}{2}\mu_i^t(\sum)^{-1}X - \frac{1}{2}\mu_i^t(\sum)^{-1}\mu_i + \ln p(\omega_i)\\ =&\mu_i^t(\sum)^{-1}X - \frac{1}{2}\mu_i^t(\sum)^{-1}\mu_i + \ln p(\omega_i)\\ =&((\sum)^{-1}\mu_i)^t X - \frac{1}{2}\mu_i^t(\sum)^{-1}\mu_i + \ln p(\omega_i)\\ =&W_i^t X + W_{i0}\end{aligned}$$

- b

  The decision surface for a linear machine is defied by:

$$g_i(x) - g_j(x) = 0$$

that is:

$$
\begin{aligned}
&((\sum)^{-1}\mu_i)^t X - \frac{1}{2}\mu_i^t(\sum)^{-1}\mu_i + \ln p(\omega_i) - ((\sum)^{-1}\mu_j)^t X + \frac{1}{2}\mu_j^t(\sum)^{-1}\mu_j - \ln p(\omega_j) \\
=&[((\sum)^{-1}\mu_i)^t - ((\sum)^{-1}\mu_j)^t]X - \frac{1}{2}(\mu_i^t(\sum)^{-1}\mu_i - \mu_j^t(\sum)^{-1}\mu_j) + \ln\frac{p(\omega_i)}{p(\omega_j)} \\
=&[(\sum)^{-1}(\mu_i - \mu_j)]^t X - \frac{1}{2}(\mu_i - \mu_j)^t(\sum)^{-1}(\mu_i + \mu_j) + \frac{p(\omega_i)}{p(\omega_j)} \\
=&[(\sum)^{-1}(\mu_i - \mu_j)]^t (X - \frac{1}{2}(\mu_i + \mu_j) + \frac{\ln\frac{p(\omega_i)}{p(\omega_j)}(\mu_i - \mu_j)}{[(\sum)^{-1}(\mu_i - \mu_j)]^t(\mu_i - \mu_j)}) \\
=&[(\sum)^{-1}(\mu_i - \mu_j)]^t (X - \frac{1}{2}(\mu_i + \mu_j) + \frac{\ln\frac{p(\omega_i)}{p(\omega_j)}(\mu_i - \mu_j)}{(\mu_i - \mu_j)^t(\sum)^{-1}(\mu_i - \mu_j)}) \\
=&W^t(X - X_0)
\end{aligned}
$$

Problem 43

- a

  P_{ij} represents the probability of the ith component of x in the state of nature ω_j

- b)

  Proof

  According to section 2.4.1, the minimum-error-rate classification can be achieved by use of the discriminant functions:

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

  X is binary-valued, we obtain:

  $$
  \begin{aligned}
  g_i(x) &= \ln p(x|\omega_i) + \ln P(\omega_i) \\
  &= \ln \prod_{i=1}^{d} P_{ij}^{X_i}(1 - P_{ij})^{1-X_i} + \ln P(\omega_i) \\
  &= \sum_{i=1}^{d} X_i \ln\frac{P_{ij}}{1 - P_{ij}} + \sum_{i=1}^{d} X_i \ln(1 - P_{ij}) + \ln P(\omega_i)
  \end{aligned}
  $$