Answer of PR.2014.final

李田所
lium15@njnu.edu.cn
December 28, 2021

Answer

1. 名词解释. 10 分

略。

2. K-L transform. 10 分

解法一（用 Fisher 判别分析，计算较简单）

已知两类情况下，K-L 变换求出的最优坐标轴方向就是就是 Fisher 线性判别中得到的最佳投影方向（课本 219 页）。

用 Fisher 线性判别解过程如下。

首先根据题目信息求均值向量

$$\boldsymbol{m}_1 = [0.75, 0.25, 0.75]^{\mathrm{T}} \tag{1}$$

$$\boldsymbol{m}_2 = [0.25, 0.75, 0.25]^{\mathrm{T}} \tag{2}$$

再分别计算两个类的类内离散度矩阵

$$\boldsymbol{S}_1 = \sum_{x_j \in \omega_1} (\boldsymbol{x}_j - \boldsymbol{m}_1)(\boldsymbol{x}_j - \boldsymbol{m}_1)^{\mathrm{T}} = \begin{bmatrix} 0.75 & 0.25 & -0.25 \\ 0.25 & 0.75 & 0.25 \\ -0.25 & 0.25 & 0.75 \end{bmatrix} \tag{3}$$

$$\boldsymbol{S}_2 = \sum_{x_j \in \omega_2} (\boldsymbol{x}_j - \boldsymbol{m}_2)(\boldsymbol{x}_j - \boldsymbol{m}_2)^{\mathrm{T}} = \begin{bmatrix} 0.75 & 0.25 & -0.25 \\ 0.25 & 0.75 & 0.25 \\ -0.25 & 0.25 & 0.75 \end{bmatrix} \tag{4}$$

总类内离散度矩阵为

$$\boldsymbol{S}_w = \boldsymbol{S}_1 + \boldsymbol{S}_2 = \begin{bmatrix} 1.5 & 0.5 & -0.5 \\ 0.5 & 1.5 & 0.5 \\ -0.5 & 0.5 & 1.5 \end{bmatrix} \tag{5}$$

则最优投影方向为

$$\begin{aligned} \boldsymbol{w}_1^* = \boldsymbol{S}_w^{-1}(\boldsymbol{m}_1 - \boldsymbol{m}_2) &= \begin{bmatrix} 1 & -0.5 & 0.5 \\ -0.5 & 1 & -0.5 \\ 0.5 & -0.5 & 1 \end{bmatrix} \begin{bmatrix} 0.5 \\ -0.5 \\ 0.5 \end{bmatrix} \\ &= \begin{bmatrix} 1 \\ -1 \\ 1 \end{bmatrix} \end{aligned} \tag{6}$$

即求得最优投影方向为 $[1, -1, 1]^{\mathrm{T}}$。

解法二（K-L transform，过程较繁琐）

K-L transform 中的总类内离散度矩阵 $\boldsymbol{S}_w$ 和类间离散度矩阵 $\boldsymbol{S}_b$ 与 Fisher 线性判别分析中有些微不同。我们先根据题目信息求解如下。

首先根据题目信息求均值向量

$$\boldsymbol{\mu}_1 = [0.75, 0.25, 0.75]^{\mathrm{T}} \tag{7}$$

$$\boldsymbol{\mu}_2 = [0.25, 0.75, 0.25]^{\mathrm{T}} \tag{8}$$

$$\boldsymbol{\mu} = \frac{1}{2}(\boldsymbol{\mu}_1 + \boldsymbol{\mu}_2) \tag{9}$$

$$= [0.5, 0.5, 0.5]^{\mathrm{T}} \tag{10}$$

之后求总类内离散度矩阵

$$\boldsymbol{S}_w = \sum_{x_j \in \omega_1} (\boldsymbol{x}_j - \boldsymbol{\mu}_1)(\boldsymbol{x}_j - \boldsymbol{\mu}_1)^{\mathrm{T}} + \sum_{x_j \in \omega_2} (\boldsymbol{x}_j - \boldsymbol{\mu}_2)(\boldsymbol{x}_j - \boldsymbol{\mu}_2)^{\mathrm{T}} = \begin{bmatrix} 0.75 & 0.25 & -0.25 \\ 0.25 & 0.75 & 0.25 \\ -0.25 & 0.25 & 0.75 \end{bmatrix} \tag{11}$$

类间离散度矩阵为

$$\boldsymbol{S}_b = \frac{1}{2}(\boldsymbol{\mu}_1 - \boldsymbol{\mu})(\boldsymbol{\mu}_1 - \boldsymbol{\mu})^{\mathrm{T}} + \frac{1}{2}(\boldsymbol{\mu}_2 - \boldsymbol{\mu})(\boldsymbol{\mu}_2 - \boldsymbol{\mu})^{\mathrm{T}} = \begin{bmatrix} 0.0625 & -0.0625 & 0.0625 \\ -0.0625 & 0.0625 & -0.0625 \\ 0.0625 & -0.0625 & 0.0625 \end{bmatrix} \tag{12}$$

对总类内离散度矩阵 $\boldsymbol{S}_w$ 做特征分解并对特征向量阵做 Gram-Schmidt 正交化，即

$$\boldsymbol{U}^{\mathrm{T}} \boldsymbol{S}_w \boldsymbol{U} = \boldsymbol{\Lambda} \tag{13}$$

其中，求解得

$$\boldsymbol{U} = \begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{3}} & -\frac{1}{\sqrt{6}} \\ \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{3}} & \frac{1}{\sqrt{6}} \\ 0 & \frac{1}{\sqrt{3}} & \frac{2}{\sqrt{6}} \end{bmatrix} \tag{14}$$

$$\boldsymbol{\Lambda} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0.25 & 0 \\ 0 & 0 & 1 \end{bmatrix} \tag{15}$$

令

$$\boldsymbol{B} = \boldsymbol{U}\boldsymbol{\Lambda}^{-\frac{1}{2}} = \begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{3}} & -\frac{1}{\sqrt{6}} \\ \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{3}} & \frac{1}{\sqrt{6}} \\ 0 & \frac{1}{\sqrt{3}} & \frac{2}{\sqrt{6}} \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 1 \end{bmatrix} = \begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{2}{\sqrt{3}} & -\frac{1}{\sqrt{6}} \\ \frac{1}{\sqrt{2}} & -\frac{2}{\sqrt{3}} & \frac{1}{\sqrt{6}} \\ 0 & \frac{2}{\sqrt{3}} & \frac{2}{\sqrt{6}} \end{bmatrix} \tag{16}$$

对类间离散度矩阵进行变换成为

$$\boldsymbol{S}_b' = \boldsymbol{B}^{\mathrm{T}} \boldsymbol{S}_b \boldsymbol{B} = \frac{1}{16} \begin{bmatrix} 0 & 0 & 0 \\ \sqrt{2} & -\sqrt{2} & \sqrt{2} \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} \sqrt{3} & 2\sqrt{2} & -1 \\ \sqrt{3} & -2\sqrt{2} & 1 \\ 0 & 2\sqrt{2} & 2 \end{bmatrix} = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0.75 & 0 \\ 0 & 0 & 0 \end{bmatrix} \tag{17}$$

显然 $\boldsymbol{S}_b'$ 相似于对角阵即自身，求解其对应于非零特征值 0.75 的特征向量

$$\boldsymbol{v} = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} \tag{18}$$

最佳方向为

$$\boldsymbol{w}_2^* = \boldsymbol{B}\boldsymbol{v} = \frac{1}{\sqrt{6}}\begin{bmatrix} 2\sqrt{2} \\ -2\sqrt{2} \\ 2\sqrt{2} \end{bmatrix} \tag{19}$$

和解法一中 Fisher 判别分析得到的最优投影方向 $[1,-1,1]^{\mathrm{T}}$ 一致。

3. Perceptron/MSE. 20 分

（1）用感知器算法求判别函数。

样本的增广化规范向量为

$$\boldsymbol{y}_1 = [1,1,1]^{\mathrm{T}}, \quad \boldsymbol{y}_2 = [1,1,2]^{\mathrm{T}}, \quad \boldsymbol{y}_3 = [1,2,1]^{\mathrm{T}}$$
$$\boldsymbol{y}_4 = [-1,0,1]^{\mathrm{T}}, \quad \boldsymbol{y}_5 = [-1,1,-1]^{\mathrm{T}}, \quad \boldsymbol{y}_6 = [-1,2,1]^{\mathrm{T}}$$

iteration	feature vector	$g(\boldsymbol{y}_i)$	weight vector
1	$[1,1,1]^{\mathrm{T}}$	$=0$	$[1,1,1]^{\mathrm{T}}$
	$[1,1,2]^{\mathrm{T}}$	>0	
	$[1,2,1]^{\mathrm{T}}$	>0	
	$[-1,0,1]^{\mathrm{T}}$	$=0$	$[0,1,2]^{\mathrm{T}}$
	$[-1,1,-1]^{\mathrm{T}}$	<0	$[-1,2,1]^{\mathrm{T}}$
	$[-1,2,1]^{\mathrm{T}}$	>0	
2	$[1,1,1]^{\mathrm{T}}$	>0	$[-1,2,1]^{\mathrm{T}}$
	$[1,1,2]^{\mathrm{T}}$	>0	
	$[1,2,1]^{\mathrm{T}}$	>0	
	$[-1,0,1]^{\mathrm{T}}$	>0	
	$[-1,1,-1]^{\mathrm{T}}$	>0	
	$[-1,2,1]^{\mathrm{T}}$	>0	$[-1,2,1]^{\mathrm{T}}$

表 1. 感知器算法求解过程。

感知器算法求出的判别函数为

$$g_1(\boldsymbol{x}) = 2x_1 + x_2 - 1 \tag{20}$$

分类线为

$$2x_1 + x_2 - 1 = 0 \tag{21}$$

（2）求 MSE 分类器判别函数。

列出方程组

$$\boldsymbol{Y}\boldsymbol{\alpha} = \boldsymbol{b} \tag{22}$$

其中

$$\boldsymbol{Y}=\begin{bmatrix}\boldsymbol{y}_1^{\mathrm{T}}\\ \vdots\\ \boldsymbol{y}_6^{\mathrm{T}}\end{bmatrix}=\begin{bmatrix}1&1&1\\1&1&2\\1&2&1\\-1&0&1\\-1&1&-1\\-1&2&1\end{bmatrix}\in\mathbb{R}^{6\times 3} \tag{23}$$

$$\boldsymbol{b}=[1,1,1,1,1,1]^{\mathrm{T}} \tag{24}$$

求得

$$\boldsymbol{\alpha}^*=\left(\boldsymbol{Y}^{\mathrm{T}}\boldsymbol{Y}\right)^{-1}\boldsymbol{Y}^{\mathrm{T}}\boldsymbol{b}=\begin{bmatrix}-0.2353\\0.4902\\0.3072\end{bmatrix} \tag{25}$$

则 MSE 分类器的判别函数为

$$g_1(\boldsymbol{x})=0.4902x_1+0.3072x_2-0.2353 \tag{26}$$

4. BP net.　10 分

为便于后续计算，令

$$\boldsymbol{W}^{(1)}=\left(\boldsymbol{W}^1\right)^{\mathrm{T}}=\begin{bmatrix}-1.0&1.0\\-0.5&1.5\\1.5&-0.5\end{bmatrix} \tag{27}$$

$$\boldsymbol{W}^{(2)}=\left(\boldsymbol{W}^2\right)^{\mathrm{T}}=\begin{bmatrix}1.0&-2.0&0.5\\0.5&-1.0&1.0\end{bmatrix} \tag{28}$$

则对这个三层模型，传播过程记为

$$\boldsymbol{a}^{(0)}=\boldsymbol{x} \tag{29}$$

$$\boldsymbol{z}^{(1)}=\boldsymbol{W}^{(1)}\boldsymbol{a}^{(0)},\quad \boldsymbol{a}^{(1)}=\boldsymbol{z}^{(1)} \tag{30}$$

$$\boldsymbol{z}^{(2)}=\boldsymbol{W}^{(2)}\boldsymbol{a}^{(1)},\quad \boldsymbol{a}^{(2)}=\boldsymbol{z}^{(2)} \tag{31}$$

其中

$$\boldsymbol{z}^{(1)}=\begin{bmatrix}0\\1\\1\end{bmatrix} \tag{32}$$

$$\boldsymbol{z}^{(2)}=\begin{bmatrix}-1.5\\0\end{bmatrix} \tag{33}$$

$$\tag{34}$$

关于反向传播算法直接有

$$\frac{\partial \mathcal{L}(\boldsymbol{y}, \hat{\boldsymbol{y}})}{\partial \boldsymbol{W}^{(l)}} = \delta^{(l)} \left(\boldsymbol{a}^{(l-1)}\right)^{\mathrm{T}} \in \mathbb{R}^{M_l \times M_{l-1}} \tag{35}$$

其中 M_l 为第 l 层神经元的个数，$\delta^{(l)}$ 为第 l 层的误差项，定义为

$$\begin{aligned} \delta^{(l)} &\triangleq \frac{\partial \mathcal{L}(\boldsymbol{y}, \hat{\boldsymbol{y}})}{\partial \boldsymbol{z}^{(l)}} \\ &= \frac{\partial \boldsymbol{a}^{(l)}}{\partial \boldsymbol{z}^{(l)}} \cdot \frac{\partial \boldsymbol{z}^{(l+1)}}{\partial \boldsymbol{a}^{(l)}} \cdot \frac{\partial \mathcal{L}(\boldsymbol{y}, \hat{\boldsymbol{y}})}{\partial \boldsymbol{z}^{(l+1)}} \\ &= 1 \cdot \left(\boldsymbol{W}^{(l+1)}\right)^{\mathrm{T}} \cdot \delta^{(l+1)} \\ &= \left(\boldsymbol{W}^{(l+1)}\right)^{\mathrm{T}} \delta^{(l+1)} \in \mathbb{R}^{M_l} \end{aligned} \tag{36}$$

则权重的更新方式为

$$\begin{aligned} \boldsymbol{W}^{(l)} &\leftarrow \boldsymbol{W}^{(l)} - \eta \cdot \frac{\partial \mathcal{L}(\boldsymbol{y}, \hat{\boldsymbol{y}})}{\partial \boldsymbol{W}^{(l)}} \\ &= \boldsymbol{W}^{(l)} - \eta \cdot \delta^{(l)} \left(\boldsymbol{a}^{(l-1)}\right)^{\mathrm{T}} \end{aligned} \tag{37}$$

注意到题中网络的输出为 $\boldsymbol{a}^{(2)} = \boldsymbol{z}^{(2)}$，这样对于输出层，损失函数为

$$\mathcal{L} = \frac{1}{2} \left\|\boldsymbol{a}^{(2)} - \boldsymbol{y}\right\|_2^2 = \frac{1}{2} \left\|\boldsymbol{W}^{(2)} \boldsymbol{a}^{(1)} - \boldsymbol{y}\right\|_2^2 \tag{38}$$

则损失函数关于 $\boldsymbol{W}^{(2)}$ 的梯度为

$$\frac{\partial \mathcal{L}}{\partial \boldsymbol{W}^{(2)}} = \left(\boldsymbol{a}^{(2)} - \boldsymbol{y}\right) \cdot \left(\boldsymbol{a}^{(1)}\right)^{\mathrm{T}} \tag{39}$$

其中第二层误差项为

$$\delta^{(2)} = \boldsymbol{a}^{(2)} - \boldsymbol{y} = \begin{bmatrix} -2.5 \\ 1 \end{bmatrix} \tag{40}$$

求得

$$\begin{aligned} \frac{\partial \mathcal{L}}{\partial \boldsymbol{W}^{(2)}} &= \delta^{(2)} \cdot \left(\boldsymbol{a}^{(1)}\right)^{\mathrm{T}} \\ &= \begin{bmatrix} -2.5 \\ 1 \end{bmatrix} \begin{bmatrix} 0 & 1 & 1 \end{bmatrix} \\ &= \begin{bmatrix} 0 & -2.5 & -2.5 \\ 0 & 1 & 1 \end{bmatrix} \end{aligned} \tag{41}$$

第二层的权重更新为

$$\begin{aligned} \boldsymbol{W}^{(2)} &\leftarrow \boldsymbol{W}^{(2)} - 0.1 \cdot \frac{\partial \mathcal{L}}{\partial \boldsymbol{W}^{(2)}} \\ &= \begin{bmatrix} 1.0 & -2.0 & 0.5 \\ 0.5 & -1.0 & 1.0 \end{bmatrix} - \begin{bmatrix} 0 & -0.25 & -0.25 \\ 0 & 0.1 & 0.1 \end{bmatrix} \\ &= \begin{bmatrix} 1.0 & -1.75 & 0.75 \\ 0.5 & -1.1 & 0.9 \end{bmatrix} \end{aligned} \tag{42}$$

同理可得第一层的误差项

$$
\begin{aligned}
\delta^{(1)} &= \left(\boldsymbol{W}^{(2)}\right)^{\mathrm{T}} \delta^{(2)} \\
&= \begin{bmatrix} 1.0 & 0.5 \\ -2.0 & -1.0 \\ 0.5 & 1.0 \end{bmatrix} \begin{bmatrix} -2.5 \\ 1 \end{bmatrix} \\
&= \begin{bmatrix} -2 \\ 4 \\ -0.25 \end{bmatrix}
\end{aligned} \tag{43}
$$

则 $\boldsymbol{W}^{(1)}$ 的梯度为

$$
\begin{aligned}
\frac{\partial \mathcal{L}}{\partial \boldsymbol{W}^{(1)}} &= \delta^{(1)} \cdot \left(\boldsymbol{a}^{(0)}\right)^{\mathrm{T}} \\
&= \begin{bmatrix} -2 \\ 4 \\ -0.25 \end{bmatrix} \begin{bmatrix} 1 & 1 \end{bmatrix} \\
&= \begin{bmatrix} -2 & -2 \\ 4 & 4 \\ -0.25 & -0.25 \end{bmatrix}
\end{aligned} \tag{44}
$$

第一层的权重更新为

$$
\begin{aligned}
\boldsymbol{W}^{(1)} &\leftarrow \boldsymbol{W}^{(1)} - 0.1 \cdot \frac{\partial \mathcal{L}}{\partial \boldsymbol{W}^{(1)}} \\
&= \begin{bmatrix} -1.0 & 1.0 \\ -0.5 & 1.5 \\ 1.5 & -0.5 \end{bmatrix} - \begin{bmatrix} -0.2 & -0.2 \\ 0.4 & 0.4 \\ -0.025 & -0.025 \end{bmatrix} \\
&= \begin{bmatrix} -0.8 & 1.2 \\ -0.9 & 1.1 \\ 1.525 & -0.475 \end{bmatrix}
\end{aligned} \tag{45}
$$

综上，该样本做一次修正后的权值为

$$
\boldsymbol{W}^{1} = \begin{bmatrix} -0.8 & -0.9 & 1.525 \\ 1.2 & 1.1 & -0.475 \end{bmatrix} \tag{46}
$$

$$
\boldsymbol{W}^{2} = \begin{bmatrix} 1.0 & 0.5 \\ -1.75 & -1.1 \\ 0.75 & 0.9 \end{bmatrix} \tag{47}
$$

5. Cost-sensitive L2-SVM. 10 分

Cost-sensitive L2-SVM 的原始问题为

$$\begin{aligned}
\min\quad & \frac{1}{2}\boldsymbol{w}^{\mathrm{T}}\boldsymbol{w}+\frac{1}{2}\sum_{i=1}^{l}C_i\xi_i^2 \\
\text{s.t.}\quad & y_i\left(\boldsymbol{w}^{\mathrm{T}}\boldsymbol{x}_i+b\right)\geq 1-\xi_i,\quad i=1,2,\cdots,l
\end{aligned} \tag{48}$$

写出原始问题的 Lagrange 函数为

$$L\left(\boldsymbol{w},b,\boldsymbol{\xi},\boldsymbol{\alpha}\right)=\frac{1}{2}\boldsymbol{w}^{\mathrm{T}}\boldsymbol{w}+\frac{1}{2}\sum_{i=1}^{l}C_i\xi_i^2-\sum_{i=1}^{l}\alpha_i\left[y_i\left(\boldsymbol{w}^{\mathrm{T}}\boldsymbol{x}+b\right)-1+\xi_i\right] \tag{49}$$

原问题等价于 Lagrange 函数的极小极大问题，而对偶问题是 Lagrange 函数的极大极小问题，为了得到对偶问题的解，首先求 $L\left(\boldsymbol{w},b,\boldsymbol{\xi},\boldsymbol{\alpha}\right)$ 对 $\boldsymbol{w},b,\boldsymbol{\xi}$ 的极小，由

$$\nabla_{\boldsymbol{w}}L\left(\boldsymbol{w},b,\boldsymbol{\xi},\boldsymbol{\alpha}\right)=\boldsymbol{w}-\sum_{i=1}^{l}\alpha_iy_i\boldsymbol{x}_i=0 \tag{50}$$

$$\nabla_{b}L\left(\boldsymbol{w},b,\boldsymbol{\xi},\boldsymbol{\alpha}\right)=-\sum_{i=1}^{l}\alpha_iy_i=0 \tag{51}$$

$$\nabla_{\boldsymbol{\xi_i}}L\left(\boldsymbol{w},b,\boldsymbol{\xi},\boldsymbol{\alpha}\right)=C_i\xi_i-\alpha_i=0 \tag{52}$$

得

$$\boldsymbol{w}=\sum_{i=1}^{l}\alpha_iy_i\boldsymbol{x}_i \tag{53}$$

$$\sum_{i=1}^{l}\alpha_iy_i=0 \tag{54}$$

$$C_i\xi_i=\alpha_i \tag{55}$$

将式53 ~ 式55代入式49，得

$$\min_{\boldsymbol{w},b,\boldsymbol{\xi}}\quad L\left(\boldsymbol{w},b,\boldsymbol{\xi},\boldsymbol{\alpha}\right)=\sum_{i=1}^{l}\alpha_i-\frac{1}{2}\sum_{i=1}^{l}\sum_{j=1}^{l}\alpha_i\alpha_jy_iy_j\boldsymbol{x}_i^{\mathrm{T}}\boldsymbol{x}_j-\frac{1}{2}\sum_{i=1}^{l}\frac{\alpha_i^2}{C_i} \tag{56}$$

再对 $\min_{\boldsymbol{w},b,\boldsymbol{\xi}}L\left(\boldsymbol{w},b,\boldsymbol{\xi},\boldsymbol{\alpha}\right)$ 求 $\boldsymbol{\alpha}$ 的极大，即得对偶问题：

$$\max_{\boldsymbol{\alpha}}\quad \sum_{i=1}^{l}\alpha_i-\frac{1}{2}\sum_{i=1}^{l}\frac{\alpha_i^2}{C_i}-\frac{1}{2}\sum_{i=1}^{l}\sum_{j=1}^{l}\alpha_i\alpha_jy_iy_j\boldsymbol{x}_i^{\mathrm{T}}\boldsymbol{x}_j \tag{57}$$

$$\text{s.t.}\quad \sum_{i=1}^{l}\alpha_iy_i=0 \tag{58}$$

$$\alpha_i\geq 0,\quad i=1,2,\cdots,l \tag{59}$$

对上面的对偶最优化问题进行变换，进行必要的化简并将对目标函数求极大转换为求极小，可得 Cost-sensitive

L2-SVM 的对偶问题为

$$\min_{\boldsymbol{\alpha}} \quad \frac{1}{2}\sum_{i=1}^{l}\sum_{j=1}^{l}\alpha_i\alpha_j y_i y_j\left(\boldsymbol{x}_i^{\mathrm{T}}\boldsymbol{x}_j+\frac{\delta_{ij}}{C_i}\right)-\sum_{i=1}^{l}\alpha_i,\ \text{where}\ \ \delta_{ij}=\begin{cases}1,\ i=j\\ 0,\ i\neq j\end{cases} \tag{60}$$

$$\text{s.t.} \quad \sum_{i=1}^{l}\alpha_i y_i = 0 \tag{61}$$

$$\alpha_i \geq 0, \quad i=1,2,\cdots,l \tag{62}$$

6. K-means. 10 分

已知

$$\boldsymbol{x}_1=[4,5]^{\mathrm{T}}, \quad \boldsymbol{x}_2=[1,4]^{\mathrm{T}}$$
$$\boldsymbol{x}_3=[0,1]^{\mathrm{T}}, \quad \boldsymbol{x}_4=[5,0]^{\mathrm{T}}$$

（1）$C_1=\{\boldsymbol{x}_1,\boldsymbol{x}_2\},C_2=\{\boldsymbol{x}_3,\boldsymbol{x}_4\}$

两类的均值为

$$\boldsymbol{m}_1=[2.5,4.5]^{\mathrm{T}}, \quad \boldsymbol{m}_2=[2.5,0.5]^{\mathrm{T}} \tag{63}$$

误差平方和为

$$J_e^{(1)}=18 \tag{64}$$

（2）$C_1=\{\boldsymbol{x}_1,\boldsymbol{x}_4\},C_2=\{\boldsymbol{x}_2,\boldsymbol{x}_3\}$

两类的均值为

$$\boldsymbol{m}_1=[4.5,2.5]^{\mathrm{T}}, \quad \boldsymbol{m}_2=[0.5,2.5]^{\mathrm{T}} \tag{65}$$

误差平方和为

$$J_e^{(2)}=18 \tag{66}$$

（3）$C_1=\{\boldsymbol{x}_1,\boldsymbol{x}_2,\boldsymbol{x}_3\},C_2=\{\boldsymbol{x}_4\}$

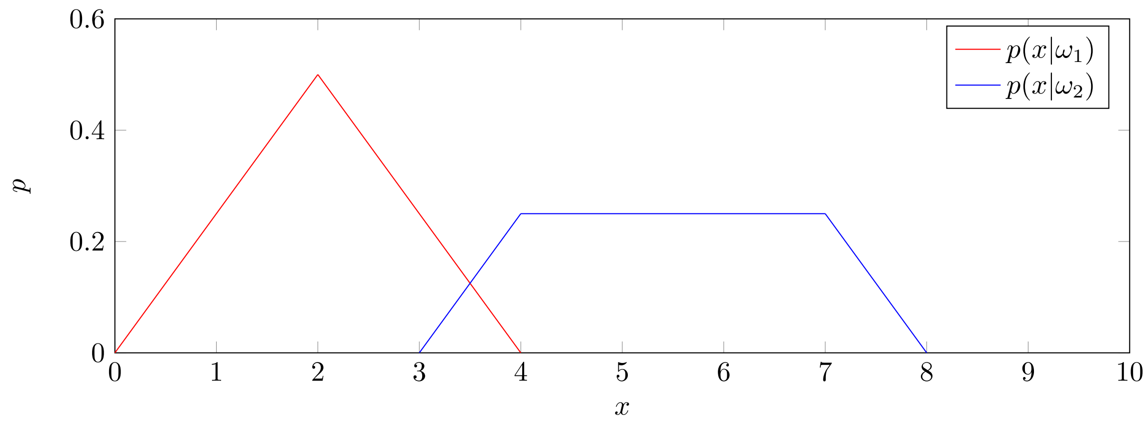
两类的均值为

$$\boldsymbol{m}_1=\left[\frac{5}{3},\frac{10}{3}\right]^{\mathrm{T}}, \quad \boldsymbol{m}_2=[5,0]^{\mathrm{T}} \tag{67}$$

误差平方和为

$$\begin{aligned}J_e^{(3)} &= \left\|\left[4-\frac{5}{3},5-\frac{10}{3}\right]^{\mathrm{T}}\right\|^2+\left\|\left[1-\frac{5}{3},4-\frac{10}{3}\right]^{\mathrm{T}}\right\|^2+\left\|\left[0-\frac{5}{3},1-\frac{10}{3}\right]^{\mathrm{T}}\right\|^2\\ &\approx 17.33\end{aligned}$$

7. Bayes decision. 20 分



（1）写出两个类的概率密度函数表达式。

两类的类条件概率密度为

$$p\left(x|\omega_1\right)=\begin{cases}0.25x, & 0\le x<2\\ -0.25x+1, & 2\le x<4\end{cases} \tag{68}$$

$$p\left(x|\omega_2\right)=\begin{cases}0.25x-0.75, & 3\le x<4\\ 0.25, & 4\le x<7\\ -0.25x+2, & 7\le x<8\end{cases} \tag{69}$$

（2）求最小错误率 Bayes 决策的决策边界。

令

$$\frac{p\left(x|\omega_1\right)}{p\left(x|\omega_2\right)}=\frac{P\left(\omega_2\right)}{P\left(\omega_1\right)} \tag{70}$$

解得决策边界为

$$x_{threshold}^{(1)}=\frac{17}{5}=3.4 \tag{71}$$

（3）求最小风险 Bayes 决策的决策边界。

令

$$\frac{p\left(x|\omega_1\right)}{p\left(x|\omega_2\right)}=\frac{P\left(\omega_2\right)}{P\left(\omega_1\right)}\cdot\frac{\lambda_{12}-\lambda_{22}}{\lambda_{21}-\lambda_{11}} \tag{72}$$

解得决策边界为

$$x_{threshold}^{(2)}=\frac{25}{7}\approx 3.57 \tag{73}$$

8. 简答题. 10 分

略。