

多元分析：第十四周作业

蒋翌坤 20307100013

《实用多元统计分析》P510: 11.24

a

(x_1, x_2) , (x_1, x_3) , (x_1, x_4) 的平面散布图如图 1 所示。从图中可以看出, (x_1, x_2) , (x_1, x_3) , (x_1, x_4) 都显示了二元正态性。

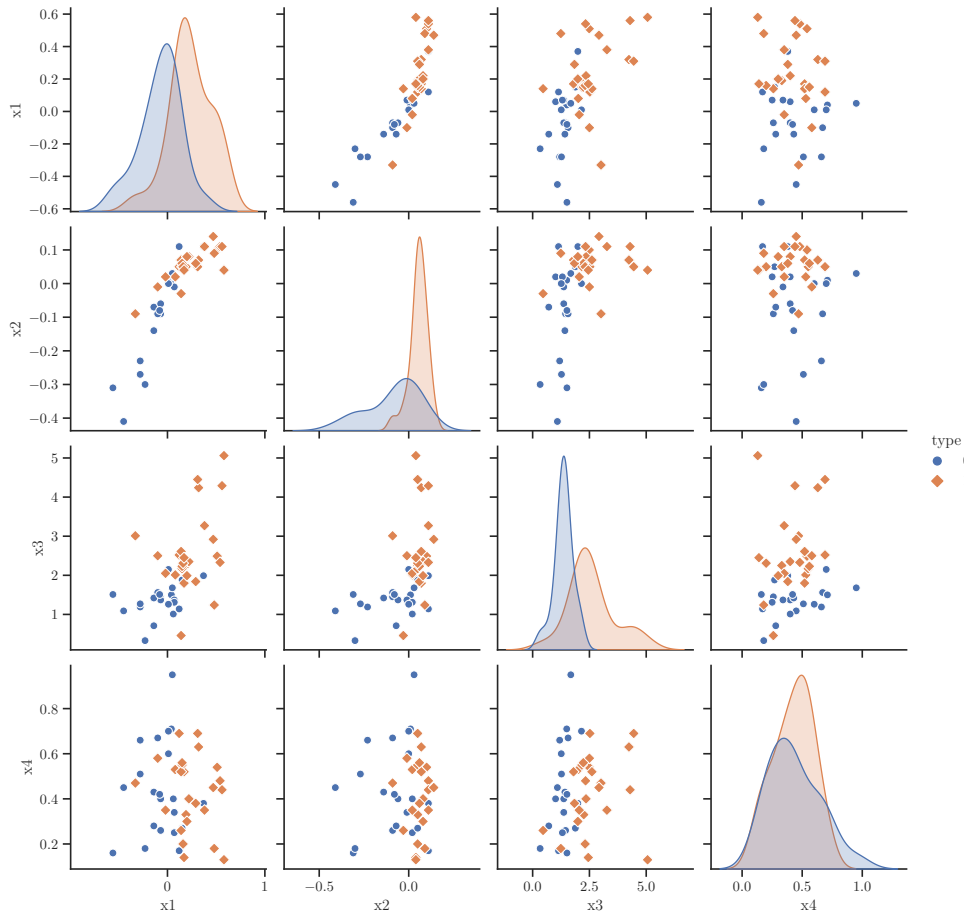


图 1: 变量间两两平面散布图

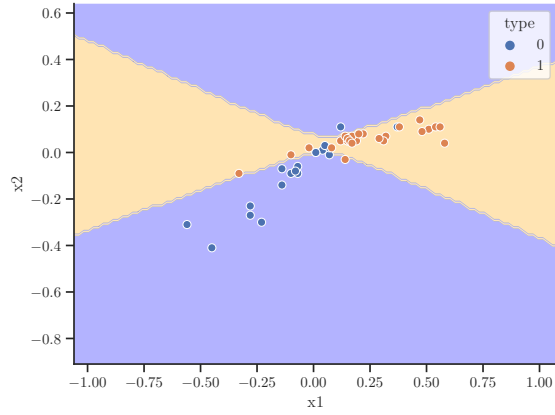
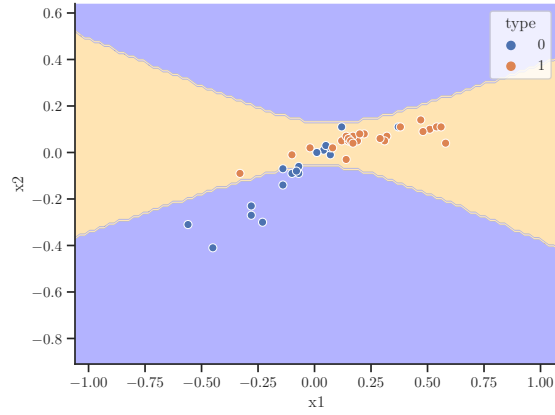
b

计算得

$$\bar{x}_1 = \begin{bmatrix} -0.069 \\ -0.081 \end{bmatrix} \quad \bar{x}_2 = \begin{bmatrix} 0.235 \\ 0.056 \end{bmatrix} \quad S_1 = \begin{bmatrix} 0.044 & 0.028 \\ 0.028 & 0.021 \end{bmatrix} \quad S_2 = \begin{bmatrix} 0.047 & 0.009 \\ 0.009 & 0.002 \end{bmatrix}$$

c

在 $p_1 = p_2$, $c(1|2) = c(2|1)$ 时, 分类法则如图 2a 所示。

(a) $p_1 = p_2$ (b) $p_1 = 0.05, p_2 = 0.95$ 图 2: (x_1, x_2) 的分类法则

d

计算得 $\text{APER} = 0.196, \hat{\text{E}}(\text{AER}) = 0.217$

e

在 $p_1 = 0.05, p_2 = 0.95$, $c(1|2) = c(2|1)$ 时, 分类法则如图 2b 所示。

计算得 $\text{APER} = 0.261, \hat{\text{E}}(\text{AER}) = 0.216$ 。这个先验概率假定不合理, 因为 S_1 和 S_2 差距不是很大, 两者被错误分类的概率应该差不多。

g

对于 (x_1, x_3) , 计算得

$$\bar{x}_1 = \begin{bmatrix} -0.069 \\ -0.081 \end{bmatrix} \quad \bar{x}_2 = \begin{bmatrix} 1.367 \\ 2.594 \end{bmatrix} \quad S_1 = \begin{bmatrix} 0.044 & 0.034 \\ 0.034 & 0.164 \end{bmatrix} \quad S_2 = \begin{bmatrix} 0.047 & 0.075 \\ 0.075 & 1.047 \end{bmatrix}$$

$$\text{APER} = \begin{cases} 0.109 & p_1 = p_2 \\ 0.370 & p_1 = 0.05, p_2 = 0.95 \end{cases} \quad \hat{\text{E}}(\text{AER}) = \begin{cases} 0.130 & p_1 = p_2 \\ 0.391 & p_1 = 0.05, p_2 = 0.95 \end{cases}$$

对于 (x_1, x_4) , 计算得

$$\bar{x}_1 = \begin{bmatrix} -0.069 \\ -0.081 \end{bmatrix} \quad \bar{x}_2 = \begin{bmatrix} 0.438 \\ 0.427 \end{bmatrix} \quad S_1 = \begin{bmatrix} 0.044 & 0.004 \\ 0.004 & 0.045 \end{bmatrix} \quad S_2 = \begin{bmatrix} 0.047 & -0.007 \\ -0.007 & 0.026 \end{bmatrix}$$

$$\text{APER} = \begin{cases} 0.174 & p_1 = p_2 \\ 0.391 & p_1 = 0.05, p_2 = 0.95 \end{cases} \quad \hat{\text{E}}(\text{AER}) = \begin{cases} 0.217 & p_1 = p_2 \\ 0.457 & p_1 = 0.05, p_2 = 0.95 \end{cases}$$

$p_1 = p_2$ 、 $p_1 = 0.05, p_2 = 0.95$ 时, 分类法则如图 3 所示。在这里, (x_1, x_3) 、 (x_1, x_4) 先验概率假定也不合理, 因为 \mathbf{S}_1 和 \mathbf{S}_2 差距不是很大, 两者被错误分类的概率应该差不多。用 (x_1, x_3) 、 (x_1, x_4) 作为分类量看起来较好, 这是因为在评估分类法则的失误率时, 他们的失误率较小。

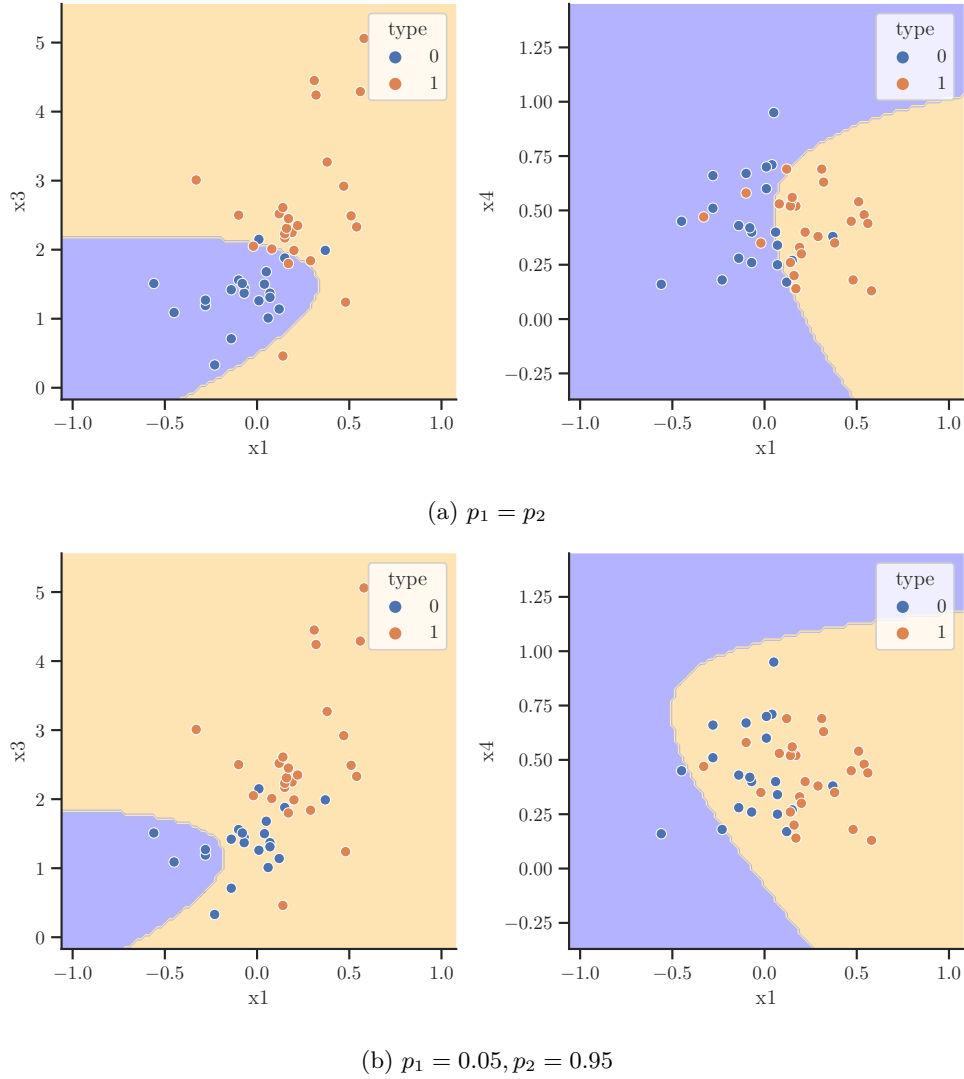


图 3: (x_1, x_3) 、 (x_1, x_4) 的分类法则

h

对于 (x_2, x_3) , 计算得

$$\bar{\mathbf{x}}_1 = \begin{bmatrix} -0.081 \\ 0.056 \end{bmatrix} \quad \bar{\mathbf{x}}_2 = \begin{bmatrix} 1.367 \\ 2.594 \end{bmatrix} \quad \mathbf{S}_1 = \begin{bmatrix} 0.021 & 0.026 \\ 0.026 & 0.164 \end{bmatrix} \quad \mathbf{S}_2 = \begin{bmatrix} 0.002 & 0.009 \\ 0.009 & 1.047 \end{bmatrix}$$

$$\text{APER} = \begin{cases} 0.109 & p_1 = p_2 \\ 0.543 & p_1 = 0.05, p_2 = 0.95 \end{cases} \quad \hat{\text{E}}(\text{AER}) = \begin{cases} 0.174 & p_1 = p_2 \\ 0.543 & p_1 = 0.05, p_2 = 0.95 \end{cases}$$

对于 (x_2, x_4) , 计算得

$$\bar{x}_1 = \begin{bmatrix} -0.081 \\ 0.056 \end{bmatrix} \quad \bar{x}_2 = \begin{bmatrix} 0.438 \\ 0.427 \end{bmatrix} \quad S_1 = \begin{bmatrix} 0.021 & 0.003 \\ 0.003 & 0.045 \end{bmatrix} \quad S_2 = \begin{bmatrix} 0.002 & 0.000 \\ 0.000 & 0.026 \end{bmatrix}$$

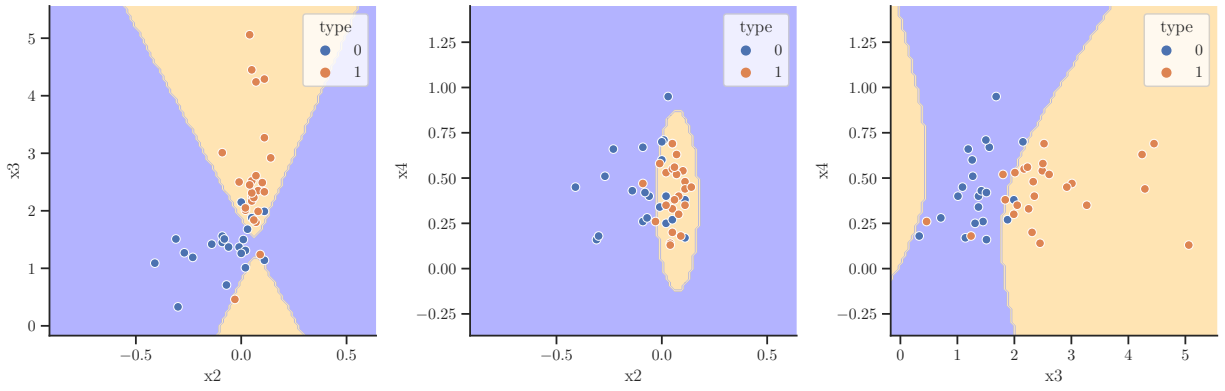
$$\text{APER} = \begin{cases} 0.239 & p_1 = p_2 \\ 0.413 & p_1 = 0.05, p_2 = 0.95 \end{cases} \quad \hat{E}(\text{AER}) = \begin{cases} 0.261 & p_1 = p_2 \\ 0.413 & p_1 = 0.05, p_2 = 0.95 \end{cases}$$

对于 (x_3, x_4) , 计算得

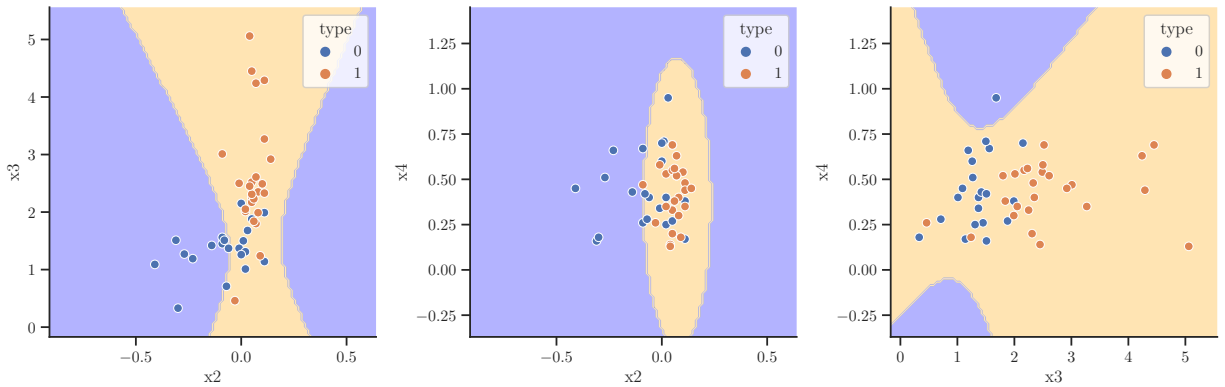
$$\bar{x}_1 = \begin{bmatrix} 1.367 \\ 2.594 \end{bmatrix} \quad \bar{x}_2 = \begin{bmatrix} 0.438 \\ 0.427 \end{bmatrix} \quad S_1 = \begin{bmatrix} 0.164 & 0.033 \\ 0.033 & 0.045 \end{bmatrix} \quad S_2 = \begin{bmatrix} 1.047 & 0.033 \\ 0.033 & 0.026 \end{bmatrix}$$

$$\text{APER} = \begin{cases} 0.130 & p_1 = p_2 \\ 0.717 & p_1 = 0.05, p_2 = 0.95 \end{cases} \quad \hat{E}(\text{AER}) = \begin{cases} 0.174 & p_1 = p_2 \\ 0.717 & p_1 = 0.05, p_2 = 0.95 \end{cases}$$

$p_1 = p_2$ 、 $p_1 = 0.05, p_2 = 0.95$ 时, 分类法则如图 4 所示。在这里, (x_2, x_3) 、 (x_2, x_4) 、 (x_3, x_4) 先验概率假定也不合理, 因为 S_1 和 S_2 差距不是很大, 两者被错误分类的概率应该差不多。



(a) $p_1 = p_2$



(b) $p_1 = 0.05, p_2 = 0.95$

图 4: (x_2, x_3) 、 (x_2, x_4) 、 (x_3, x_4) 的分类法则