5.5 根据经验,今天与昨天的湿度差 x_1 及今天的压温差(气压与温度之差) x_2 是预报明天是否下雨的两个重要因素。今测得 $x_1=0.6$, $x_2=3.0$,假定两组的协方差矩阵相等。

(1) 试给出判别规则,并预报明天是否会下雨及用回代法来估计误判概率;

$$ar{oldsymbol{x}}_i = rac{1}{n_i} \sum_{j=1}^{n_i} oldsymbol{x}_{ij}$$

因为两组的协方差矩阵相等

$$oldsymbol{S}_p = rac{1}{n-k} \sum_{i=1}^k \left(n_i - 1
ight) oldsymbol{S}_i$$

$$n = n_1 + n_2 + \dots + n_k$$

$$oldsymbol{S}_i = rac{1}{n_i-1} \sum_{i=1}^{n_1} \left(oldsymbol{x}_{ij} - ar{oldsymbol{x}}_i
ight) \left(oldsymbol{x}_{ij} - ar{oldsymbol{x}}_i
ight)'$$

因为

$$egin{aligned} d^2(oldsymbol{x},\pi_i) &= (oldsymbol{x} - ar{oldsymbol{x}}_i)'oldsymbol{S}_p^{-1}(oldsymbol{x} - ar{oldsymbol{x}}_i'oldsymbol{S}_p^{-1}oldsymbol{x} + ar{oldsymbol{x}}_i'oldsymbol{S}_p^{-1}ar{oldsymbol{x}}_i \ &= oldsymbol{x}'oldsymbol{S}_p^{-1}oldsymbol{x} - 2(ar{oldsymbol{I}}_i'oldsymbol{x} + \hat{oldsymbol{c}}_i) \end{aligned}$$

其中
$$oldsymbol{I}_i = oldsymbol{S}_p^{-1}ar{oldsymbol{x}}_i$$
, $\hat{c}_i = -rac{1}{2}ar{oldsymbol{x}}_i'oldsymbol{S}_p^{-1}ar{oldsymbol{x}}_i$

$$\left\{egin{aligned} oldsymbol{x} \in \pi_1, & oldsymbol{I}_1'oldsymbol{x} + c_1 \geq oldsymbol{I}_2'oldsymbol{x} + c_2 \ oldsymbol{x} \in \pi_2, & oldsymbol{I}_1'oldsymbol{x} + c_1 < oldsymbol{I}_2'oldsymbol{x} + c_2 \end{aligned}
ight.$$

Linear Discriminant Function for type					
Variable	1	2			
Constant	-0.06870	-1.02533			
x1	0.03251	-0.04553			
x2	0.05118	0.22976			

由SAS输出的结果可知,

 $若0.03251x_1 + 0.05118x_2 - 0.06870 < -0.04553 + 0.22976x_2 - 1.02533$,则判 $x \in \pi_1$;

将 $x_1=0.6, x_2=3.0$ 代入该判别规则,得0.0104346>-0.363368,故判 $x\in\pi_1$,即预报明天会下雨。

Error Count Estimates for type						
	1	2	Total			
Rate	0.2000	0.1000	0.1500			
Priors	0.5000	0.5000				

 $\hat{P}(2|1) = 0.2, \hat{P}(1|2) = 0.1$.

(2) 假定两组的 $\mathbf{x} = (\mathbf{x}_1, \mathbf{x}_2)'$ 均服从二元正态分布,且根据其他信息及经验给出先验概率 $p_1 = 0.3$, $p_2 = 0.7$,试预报明天是否会下雨;

Posterior Probability of Membership in type					
Obs	Classifie	ed into type	1	2	
1	2		0.4062	0.5938	

 $P(\pi_1|\mathbf{x}) = 0.4062 < 0.5938 = P(\pi_2|\mathbf{x})$, 故判 $\mathbf{x} \in \pi_2$, 即预报明天不会下雨。

(3) 假如你现在考虑是否为明天安排一项活动,该活动不太适合在雨天进行,并在(2) 中假定的基础上还认为c(2|1)=3c(1|2),那么你今天是否应该为明天安排这项活动呢?

$$egin{align} \hat{m{a}} &= m{S}_p^{-1} \left(ar{m{x}}_1 - ar{m{x}}_2
ight) = egin{pmatrix} 0.078 \ -0.179 \end{pmatrix} \ &ar{m{x}} &= rac{1}{2} (ar{m{x}}_1 + ar{m{x}}_2) = egin{pmatrix} 0.27 \ 5.475 \end{pmatrix} \ &\hat{m{a}}'(m{x} - ar{m{x}}) = 0.468 > -0.251 = \ln \left[rac{c(1|2)p_2}{c(2|1)p_1}
ight] \end{aligned}$$

故判 $x \in \pi_1$,即预报明天会下雨,不应该安排这项活动。

SAS Code

```
1 data work.file_data;
2 input type x1 x2;
3 cards;
4 1 -1.9 3.2

5 1 -6.9 10.4

6 1 5.2 2.0

7 1 5.0 2.5
```

```
8 1 7.3 0.0
 9 1 6.8 12.7
10 | 1 0.9 -15.4
11 | 1 -12.5 -2.5
12 | 1 1.5 1.3
13 | 1 3.8 6.8
14 2 0.2 6.2
15 2 -0.1 7.5
16 2 0.4 14.6
17 2 2.7 8.3
18 2 2.1 0.8
19 2 -4.6 4.3
20 2 -1.7 10.9
21 2 -2.6 13.1
22 | 2 2.6 12.8
23 2 -2.8 10.0
24 ;
25
   run;
26 proc print data=work.file_data;
27 run;
28 | data work.data_test;
29 | input x1 x2;
30 cards;
31 0.6 3.0
32
   ,
33
   run;
   proc discrim data=work.file_data testdata=work.data_test list
34
    testlist;
35 class type;
36 var x1 x2;
37
   run;
    proc discrim data=work.file_data testdata=work.data_test list
38
    testlist method=normal pool=yes;
39 priors '1'=0.3 '2'=0.7;
40 class type;
41 var x1 x2;
42 run;
```