

《多元统计分析》第四次上机作业

判别分析

3180103000 许乐乐

- 实验目的与要求:

通过本试验项目, 使学生理解并掌握各类判别方法。

习题一

在天气预报中, 常根据当天天气的湿温差(x_1)和气温差 (x_2), 来预测第二天是否下雨。试利用观测到的天气数据ex5.2 (见附件), 判断当今天测得 $(x_1, x_2)=(8.1, 2.0)$ 或 $(7.5, 3.5)$ 时, 明天的天气应判断为下雨还是不下雨?

- 线性判别法

```
d<-read_csv("ex5.2.csv")
library(MASS)
results<-lda(G~x1+x2, d)
dp<-data.frame(
  x1=c(8.1, 7.5),
  x2=c(2.0, 3.5)
)
predict(results, dp)
```

```
$class
[1] 1 1
Levels: 1 2

$posterior
      1      2
1 0.9327428 0.06725717
2 0.8776066 0.12239335

$x
      LD1
1 -1.591809
2 -1.192497
```

因此 $(x_1, x_2)=(8.1, 2.0)$ 时, 明天的天气应判为下雨; $(x_1, x_2)=(7.5, 3.5)$ 时, 明天的天气应判为不下雨。

习题二

在研究沙基液化问题中, 选取7个因子。现从已液化和未液化的地层中分别抽了12个和23个样本, 其中1类表示已液化, 2类表示未液化。试用距离判别法对原来的35个样本进行回代分类并分析误判情况 (也就是对观测到的35个样本逐个进行判断, 得到的判断结果与原先的分类是否一致, 错误判断了多少个? 错误判断所占的比例为多少? 数据见附件ex5.4.)

- 线性判别法

```
# 导入数据
d<-read_csv("ex5.4.csv")
d$group<-as.factor(d$group) # 将Grop中的1, 2转换成因子变量
attach(d)
# 建立判别函数
library(MASS)
results<-lda(group~x1+x2+x3+x4+x5+x6+x7, d)
# 预测并与原数据进行比较
dp<-d[,3:9]
pre<-predict(results,dp) # 根据线性函数模型预测所属类别
# 或直接写为predict(results)
newGroup<-pre$class # 预测的所属类的结果
cbind(d$group,newGroup) # 显示预测前后分组结果
```

```
      [,1] [,2] [,3] [,4] [,5] [,6] [,7] [,8] [,9]
newGroup  1    1    1    1    1    1    1    1    1
      [,10] [,11] [,12] [,13] [,14] [,15] [,16]
newGroup  1    1    1    2    2    2    2
      [,17] [,18] [,19] [,20] [,21] [,22] [,23]
newGroup  2    2    2    2    2    2    2
      [,24] [,25] [,26] [,27] [,28] [,29] [,30]
newGroup  2    2    2    2    2    1    2
      [,31] [,32] [,33] [,34] [,35]
newGroup  2    2    2    2    2
```

```
# 对模型进行评价
tab<-table(d$group,newGroup) # 绘制混淆矩阵
```

```
newGroup
  1  2
1 11  1
2  1 22
```

```
erro<-1-sum(diag(prop.table(tab))) # 计算误判率
```

```
[1] 0.05714286
```

```
plot(tab) # 可视化
```

tab

