

实验四：判别分析

林泽钦 3160104013 统计 1601

目录

1 实验目的和要求	1
1.1 实验目的	1
1.2 实验内容	1
1.3 实验环境	1
2 实验过程与结果	2
2.1 第一部分	2
2.2 第二部分	4

1 实验目的和要求

1.1 实验目的

通过本试验项目，理解并掌握各类判别方法。

1.2 实验内容

1.2.1 第一部分

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？

1.2.2 第二部分

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.）

1.3 实验环境

- R-3.5.1
- RStudio

2 实验过程与结果

2.1 第一部分

2.1.1 数据导入与预处理

从数据文件 ex5.2 中删除了 G1 列，然后保存为 csv 文件以供读取。数据文件完整，不需要额外处理。

```
Data <- read.csv("Pro4Data1.csv")  
Data
```

```
##      G      x1      x2  
## 1  1 -1.9   3.2  
## 2  1 -6.9   0.4  
## 3  1  5.2   2.0  
## 4  1  5.0   2.5  
## 5  1  7.3   0.0  
## 6  1  6.8  12.7  
## 7  1  0.9  -5.4  
## 8  1 -12.5 -2.5  
## 9  1  1.5   1.3  
## 10 1  3.8   6.8
```

```
## 11 2    0.2  6.2
## 12 2   -0.1  7.5
## 13 2    0.4 14.6
## 14 2    2.7  8.3
## 15 2    2.1  0.8
## 16 2   -4.6  4.3
## 17 2   -1.7 10.9
## 18 2   -2.6 13.1
## 19 2    2.6 12.8
## 20 2   -2.8 10.0
```

2.1.2 编写判定函数

下面的函数使用距离判别来进行类别判定，其中距离是指马氏距离。参数中 Train1, Train2 为两个来自不同总体的训练样本集，Test 为待测试的样本，var.equal 用于指示两个总体的协方差矩阵是否相同。默认情况下，Test = NULL，这时我们用两个训练样本集作为测试集。

```
discrimisnate.dist <-function(Train1, Train2, Test = NULL, var.equal = FALSE) {
  # if test data is not assigned, use the train data as test data
  if (is.null(Test) == TRUE)
    Test <- rbind(Train1, Train2)
  # number of samples
  n.Test <- nrow(Test)
  n.Train1 <- nrow(Train1)
  n.Train2 <- nrow(Train2)
  # result of discrimination
  blong <- rep(0, n.Test)
  # sample mean
  mu1 <- colMeans(Train1)
  mu2 <- colMeans(Train2)
  # sample covariance
  S1 <- cov(Train1)
  S2 <- cov(Train2)
```

```

# the two covariance matrices are equal
if (var.equal == TRUE) {
  S <- ((n.Train1-1)*S1 + (n.Train2-1)*S2) / (n.Train1+n.Train2-2)
  w <- mahalanobis(Test, mu2, S) - mahalanobis(Test, mu1, S)
}
# the two covariance matrices are not equal
else
  w <- mahalanobis(Test, mu2, S2) - mahalanobis(Test, mu1, S1)
# discriminate
for (i in 1:n.Test) {
  if (w[i] > 0)
    blong[i] <- 1
  else
    blong[i] <- 2
}
return(blong)
}

```

2.1.3 判别结果

判别过程如下所示，结果显示，两组都应判别为下雨。

```

Train1 <- subset(Data, G==1)[2:3]
Train2 <- subset(Data, G==2)[2:3]
Test <- rbind(c(8.1,2.0), c(7.5, 3.5))
discrimisnate.dist(Train1, Train2, Test)

```

```
## [1] 1 1
```

2.2 第二部分

2.2.1 数据导入与预处理

无需预处理，直接导入即可，导入之后得到的数据框的第 10-20 行如下所示，可以看到导入成功。


```
}  
fail
```

```
## [1] 1
```