

判别分析

R 语言

徐平峰

长春工业大学

2021/3/30

距离判别

鸢尾花

- R 中iris（鸢尾花）数据，三种不同的鸢尾花的 150 个样品的花瓣、花萼长、宽的数据。

```
dim(iris)
```

```
## [1] 150   5
```

```
summary(iris)
```

```
##      Sepal.Length      Sepal.width      Petal.Length      Petal.width
##  Min.       :4.300    Min.       :2.000    Min.       :1.000    Min.       :0.100
##  1st Qu.:5.100    1st Qu.:2.800    1st Qu.:1.600    1st Qu.:0.300
##  Median :5.800    Median :3.000    Median :4.350    Median :1.300
##  Mean   :5.843    Mean   :3.057    Mean   :3.758    Mean   :1.199
##  3rd Qu.:6.400    3rd Qu.:3.300    3rd Qu.:5.100    3rd Qu.:1.800
##  Max.    :7.900    Max.    :4.400    Max.    :6.900    Max.    :2.500
##           Species
##  setosa      :50
##  versicolor:50
##  virginica   :50
##
##
##
```

- 距离判别

```
# =====计算每一类中的样本均值和样本方差
compute.mu.S.Shat<-function(data.train,class){
  ll<-levels(data.train[,class])
  nc=length(ll)#number of class
  index<-S<-vector("list", length=nc)
  mu<-matrix(0, nrow=ncol(data.train)-1, ncol=nc)
  for(k in 1:nc){
    index[[k]] = which(data.train[,class]==ll[k])
    mu[,k]<-colMeans(data.train[index[[k]],-which(colnames(data.t
    S[[k]]<-cov(data.train[index[[k]],-which(colnames(data.train)
  }
  Shat<-S[[1]]*(length(index[[1]])-1)
  for(k in 2:nc){
    Shat = Shat + S[[k]]*(length(index[[k]])-1)
  }
  Shat <- Shat/(nrow(data.train)-nc)
  muSShat<-list(mu=mu, S=S, Shat=Shat, levels=ll)
}
```

- 距离判别

```
# distance.Discriminant.Analysis
dda<-function(test.data, muSshat,linear=TRUE){#距离判别函数
  num.class = ncol(muSshat$mu)
  dist<-matrix(0, nrow=nrow(test.data), ncol=num.class)
  if(linear == TRUE){
    for(k in 1:num.class){
      dist[,k]=mahalanobis(x=test.data, center=muSshat$mu[,k],
                           cov=muSshat$shat, inverted = FALSE)
    }
  }
  else{
    for(k in 1:num.class){
      dist[,k]=mahalanobis(x=test.data, center=muSshat$mu[,k],
                           cov=muSshat$S[[k]], inverted = FALSE)
    }
  }
  class<-apply(X=dist,MARGIN = 1, which.min)
  list(class=muSshat$levels[class], dist=dist)
}
```

- 距离判别

```
muSShat = compute.mu.S.Shat(data.train=iris,class="Species")
dd1<-dda(test.data=iris[,-5], muSShat,linear=TRUE)#线性距离判别
table(iris$Species, dd1$class)
```

```
##
##          setosa versicolor virginica
## setosa          50           0         0
## versicolor       0          48         2
## virginica        0           1        49
```

```
dd2<-dda(test.data=iris[,-5], muSShat,linear=FALSE)#线性距离判别
table(iris$Species, dd2$class)
```

```
##
##          setosa versicolor virginica
## setosa          50           0         0
## versicolor       0          47         3
## virginica        0           0        50
```

天气数据

- 利用今天和昨天湿度差x1, 气温差x2, 预报今天x1=8.1, x2=2.0是否下雨
- 1 为下雨, 2为晴天

```
(w3=read.csv("E:/teaching_plan_notes/msa11091083/rmd/MVAPureData/
```

```
#      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
```

天气数据

```
w3$G<-as.factor(w3$G)
muSShat = compute.mu.S.Shat(data.train=w3,class="G")
dd3<-dda(test.data=w3[, -1], muSShat, linear=TRUE)#线性距离判别
table(w3$G, dd3$class)
```

```
#
#      1 2
#    1 9 1
#    2 1 9
```

```
dd4<-dda(test.data=w3[, -1], muSShat, linear=FALSE)#线性距离判别
table(w3$G, dd4$class)
```

```
#
#      1 2
#    1 10 0
#    2 2 8
```


蟹蟹

本幻灯片由 R 包 **xaringan** 生成；

查克拉来自于 **remark.js**、**knitr**、以及 **R Markdown**。