

地理建模实验 7 实验报告

42109232 吕文博 地信 2101 班

2024-06-16

层次聚类

baseR 函数 `hclust` 的 `method` 参数中:

- `ward.D` Ward 的最小方差方法, 基于欧氏距离的平方
- `ward.D2` Ward 的最小方差方法, 基于欧氏距离的平方和
- `single` 最小距离法
- `complete` 最大距离法
- `average` 类平均距离法
- `mcquitty` 类加权平均距离法
- `median` 中间距离法
- `centroid` 重心距离法

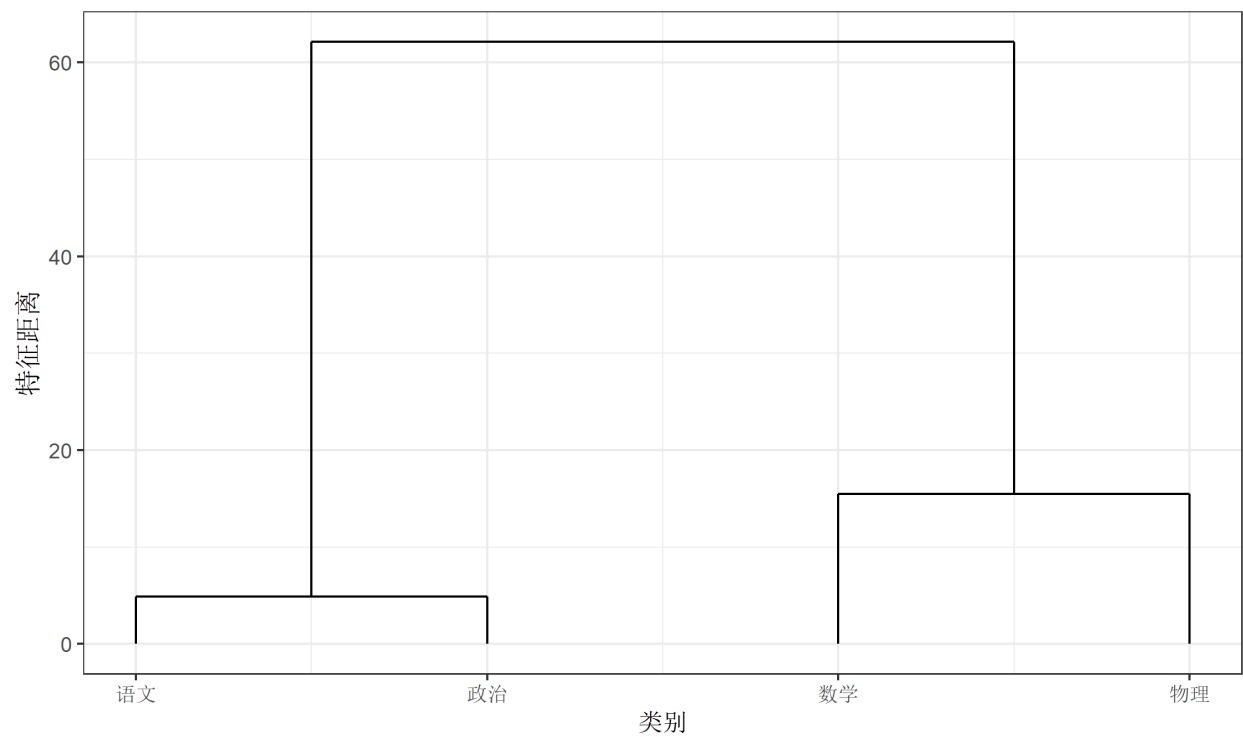
读取数据

```
scores = readxl::read_xls('../data/exp7/7.xls')
scores = t(as.matrix(scores[,-1]))
scores
```

##	[,1]	[,2]	[,3]	[,4]	[,5]	[,6]	[,7]	[,8]	[,9]	[,10]
## 数学	99	88	79	89	75	60	79	75	60	100
## 物理	98	89	80	78	78	65	87	76	56	100
## 语文	78	89	95	81	95	85	50	88	89	85
## 政治	80	90	97	82	96	88	51	89	90	84

最短距离聚类

```
d = dist(scores, method = 'euclidean', diag = TRUE)
clus1 = hclust(d, method = "single")
library(ggplot2)
library(ggdendro)
ggdendrogram(clus1, rotate = FALSE, size = 2) +
  xlab('类别') +
  ylab('特征距离') +
  theme_bw()
```



最远距离聚类

```
clus2 = hclust(d, method = "complete")
ggdendrogram(clus2, rotate = FALSE, size = 2) +
  xlab('类别') +
  ylab('特征距离') +
  theme_bw()
```

