

cluster jerarquico en R

Jaime Isaac

22/4/2021

cluster jerarquico en R

Introducción Aprenderá la visualización mejorada del dendrograma de agrupamiento con R studio. Este tema cubrirá los siguientes aspectos:

- Cálculo de la matriz de distancia
- Agrupación jerárquica
- Personalización de dendrograma

Datos de importacion

El archivo de datos utilizado aquí se obtiene del conjunto de datos de demostración de R **USArrests**. El uso de la función `head()` imprimirá las primeras seis filas del conjunto de datos **USArrests**. La función `str()` muestra la estructura interna del conjunto de datos.

```
head(USArrests)
```

```
##           Murder Assault UrbanPop Rape
## Alabama      13.2      236      58 21.2
## Alaska       10.0      263      48 44.5
## Arizona       8.1      294      80 31.0
## Arkansas      8.8      190      50 19.5
## California    9.0      276      91 40.6
## Colorado     7.9      204      78 38.7
```

```
str(USArrests)
```

```
## 'data.frame':   50 obs. of  4 variables:
## $ Murder   : num  13.2 10 8.1 8.8 9 7.9 3.3 5.9 15.4 17.4 ...
## $ Assault  : int  236 263 294 190 276 204 110 238 335 211 ...
## $ UrbanPop: int  58 48 80 50 91 78 77 72 80 60 ...
## $ Rape     : num  21.2 44.5 31 19.5 40.6 38.7 11.1 15.8 31.9 25.8 ...
```

Conjunto de datos escalado

Para escalar los datos, utilice la función `scale()`. `scale` es una función genérica cuyo método predeterminado centra y / o escala las columnas de una matriz numérica. Si el argumento `center` se establece en `VERDADERO`, el centrado se realiza tomando las desviaciones medias de cada columna. Si `scale = TRUE`, la escala se realiza dividiendo las columnas de datos (centradas) por sus desviaciones estándar.

```
data.scaled <- scale(x = USArrests,
                     center = TRUE,
                     scale = TRUE)
head(data.scaled)
```

```
##           Murder  Assault  UrbanPop      Rape
## Alabama    1.24256408 0.7828393 -0.5209066 -0.003416473
## Alaska     0.50786248 1.1068225 -1.2117642  2.484202941
## Arizona    0.07163341 1.4788032  0.9989801  1.042878388
## Arkansas   0.23234938 0.2308680 -1.0735927 -0.184916602
## California 0.27826823 1.2628144  1.7589234  2.067820292
## Colorado   0.02571456 0.3988593  0.8608085  1.864967207
```

Cálculo de la matriz de distancia.

La matriz de distancia se puede calcular usando la función `dist()`. Esta función calcula y devuelve la matriz de distancia calculada utilizando la medida de distancia especificada para calcular las distancias entre las filas de una matriz de datos. Para ejecutar esta función, asegúrese de que el paquete de estadísticas esté cargado usando la función `require()`.

El argumento `x` especifica una matriz numérica, un marco de datos o un objeto “dist”. El método del segundo argumento especifica la medida de distancia que se utilizará. El método debe ser uno de los siguientes:

“euclidiana”, “máximo”, “manhattan”, “canberra”, “binario” o “minkowski”

Calcule los resultados del objeto `res.dist` como una matriz utilizando la función `as.matrix()` y especifique el número de filas y columnas que se imprimirán entre corchetes.