PAM

Oscar Elí Bonilla Morales

2022-06-06

PARTITION AROUND MEDOIDS (PAM)

```
library(cluster)
```

Cargar la matriz de datos.

```
X<-as.data.frame(state.x77)
colnames(X)
```

```
## [1] "Population" "Income"     "Illiteracy" "Life Exp"   "Murder"
## [6] "HS Grad"    "Frost"      "Area"
```

Transformacion de datos

#Transformacion de las variables x1,x3 y x8 con la funcion de logaritmo.

```
X[,1]<-log(X[,1])
colnames(X)[1]<-"Log-Population"

X[,3]<-log(X[,3])
colnames(X)[3]<-"Log-Illiteracy"

X[,8]<-log(X[,8])
colnames(X)[8]<-"Log-Area"
```

Metodo PAM

#1.- Separacion de filas y columnas.

```
dim(X)
```

```
## [1] 50  8
```

```
n<-dim(X)[1]
p<-dim(X)[2]
```

2.- Estandarizacion univariante.

```
X.s<-scale(X)
```

3.- Aplicacion del algoritmo

```
pam.3<-pam(X.s,3)
```

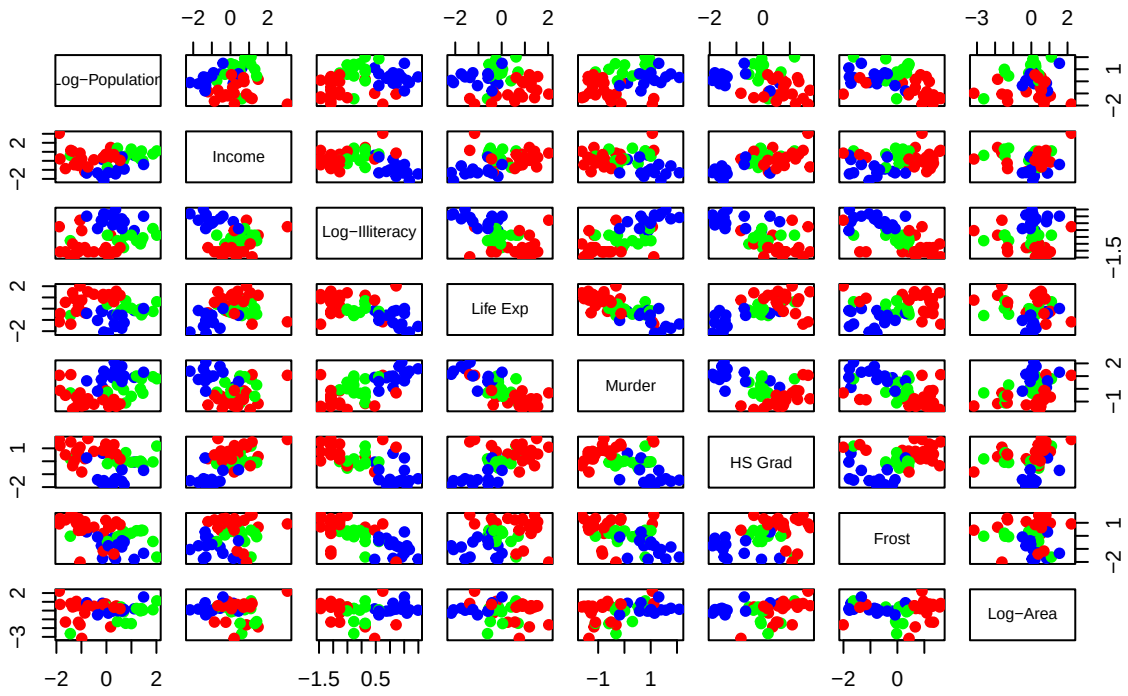
4.- Clusters

```
cl.pam<-pam.3$clustering
cl.pam
```

```
##        Alabama         Alaska        Arizona       Arkansas     California
##              1              2              1              1              3
##       Colorado    Connecticut       Delaware        Florida        Georgia
##              2              2              3              1              1
##         Hawaii          Idaho       Illinois        Indiana           Iowa
##              2              2              3              3              2
##         Kansas       Kentucky      Louisiana          Maine       Maryland
##              2              1              1              2              3
##  Massachusetts       Michigan      Minnesota    Mississippi       Missouri
##              3              3              2              1              3
##        Montana       Nebraska         Nevada  New Hampshire     New Jersey
##              2              2              2              2              3
##     New Mexico       New York North Carolina   North Dakota           Ohio
##              1              3              1              2              3
##       Oklahoma         Oregon   Pennsylvania   Rhode Island South Carolina
##              3              2              3              2              1
##   South Dakota      Tennessee          Texas           Utah        Vermont
##              2              1              1              2              2
##       Virginia     Washington  West Virginia      Wisconsin        Wyoming
##              1              2              1              2              2
```

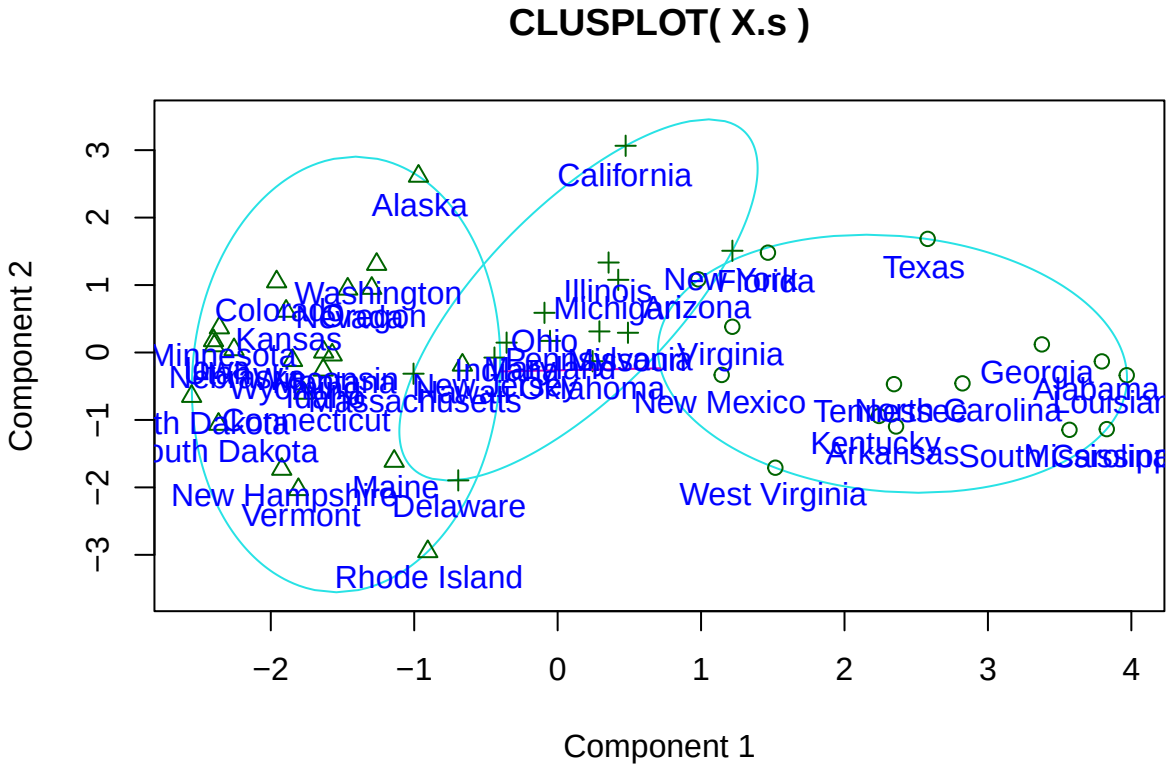
#5.- Scatter plot de la matriz con los grupos

```
col.cluster<-c("blue","red","green")[cl.pam]
pairs(X.s, col=col.cluster, main="PAM", pch=19)
```



Visualizacion con Componentes Principales

```
clusplot(X.s,cl.pam)
text(princomp(X.s)$scores[,1:2],
     labels=rownames(X.s),pos=1, col="blue")
```



Silhouette

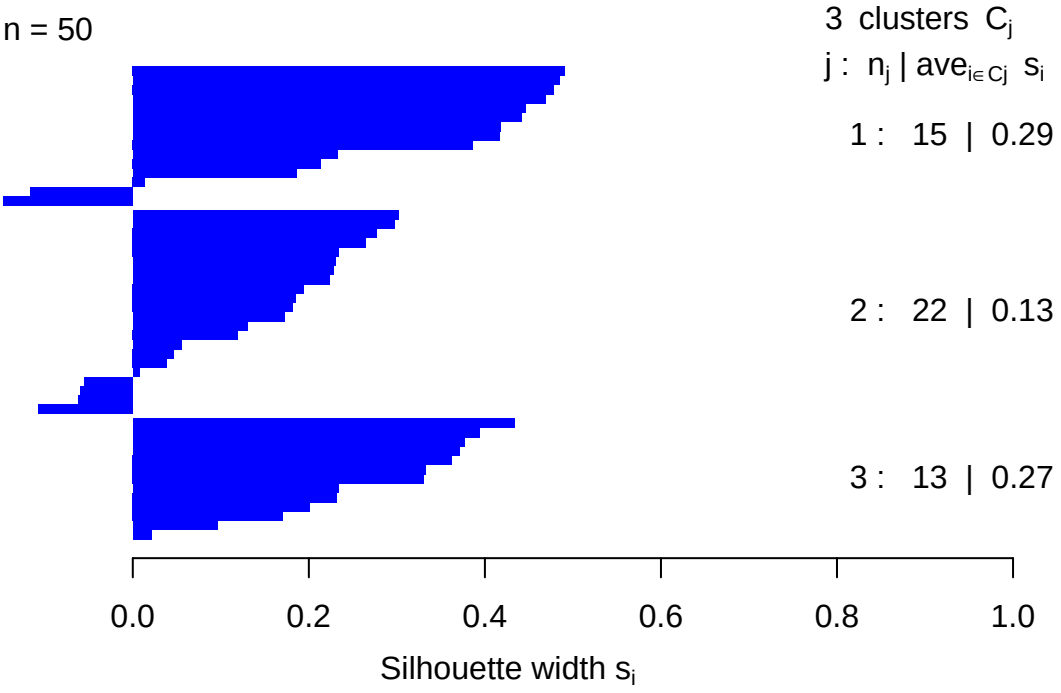
Representacion grafica de la eficacia de clasificacion de una observacion dentro de un grupo.

1.- Generacion de los calculos

```
dist.Euc<-dist(X.s, method = "euclidean")
Sil.pam<-silhouette(cl.pam, dist.Euc)
```

#2.- Generacion del grafico

```
plot(Sil.pam, main="Silhouette for PAM",
     col="blue")
```



Average silhouette width : 0.22

2

Nota:

Sugerir un nuevo numero de clusters

3.- Aplicacion del algoritmo

```
pam.3<-pam(X.s,7)
```

4.- Clusters

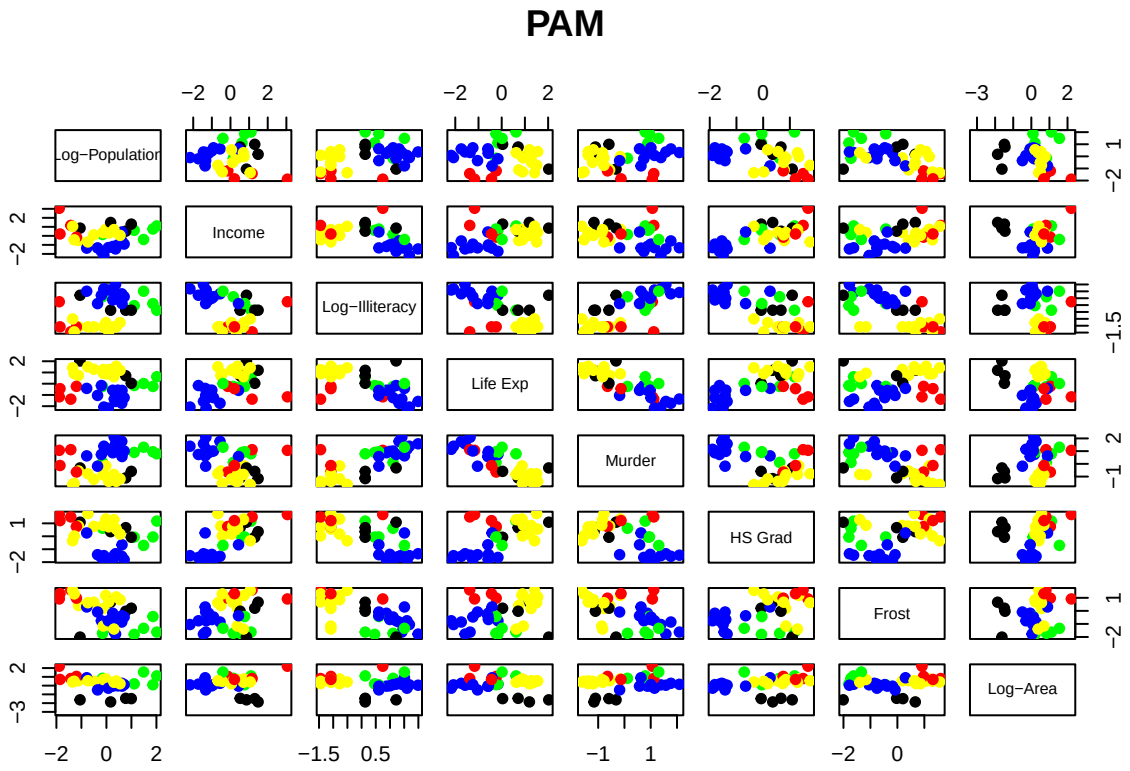
```
cl.pam<-pam.3$clustering
cl.pam
```

```
##        Alabama         Alaska        Arizona       Arkansas     California
##              1              2              3              1              3
##       Colorado    Connecticut       Delaware        Florida        Georgia
##              4              5              6              3              1
##         Hawaii          Idaho       Illinois        Indiana           Iowa
##              5              4              7              7              4
##         Kansas       Kentucky      Louisiana          Maine       Maryland
##              4              1              1              6              7
##  Massachusetts       Michigan      Minnesota    Mississippi       Missouri
##              5              7              4              1              7
```

```
##        Montana       Nebraska         Nevada  New Hampshire     New Jersey
##              2              4              2              6              5
##     New Mexico       New York North Carolina   North Dakota           Ohio
##              1              3              1              4              7
##       Oklahoma         Oregon   Pennsylvania   Rhode Island South Carolina
##              7              4              7              6              1
##   South Dakota      Tennessee          Texas           Utah        Vermont
##              4              1              3              4              6
##       Virginia     Washington  West Virginia      Wisconsin        Wyoming
##              1              4              1              4              2
```

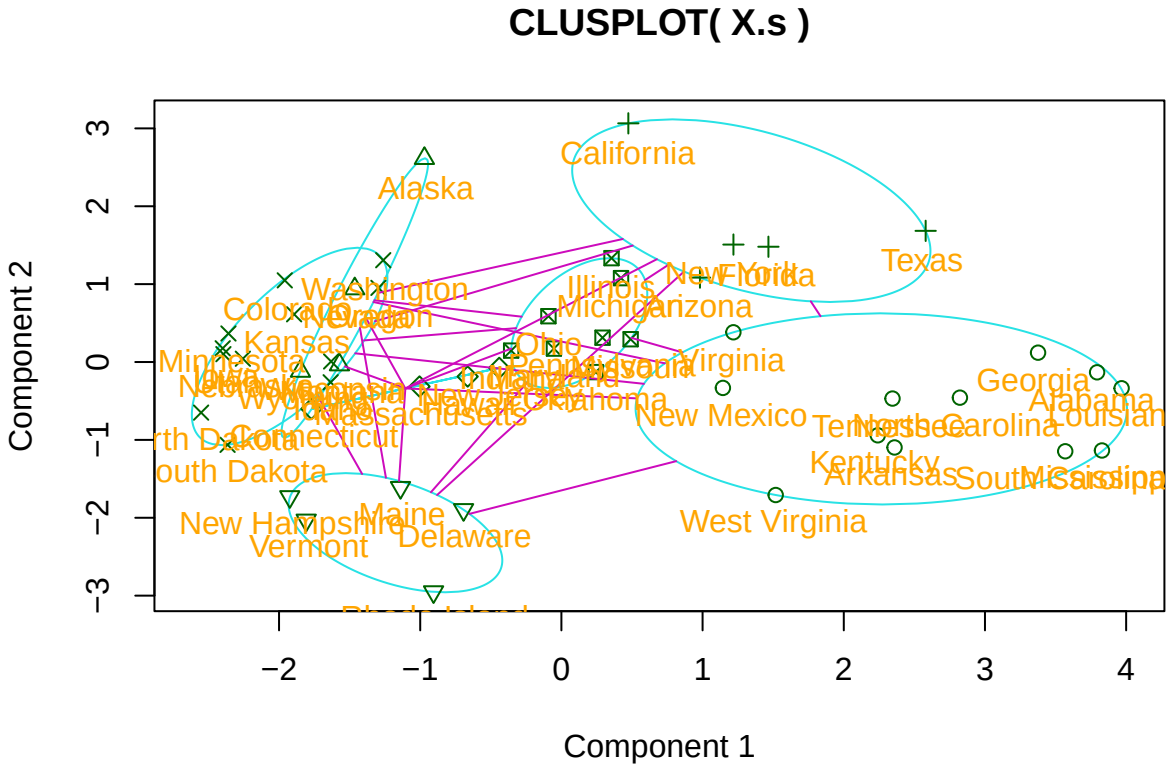
#5.- Scatter plot de la matriz con los grupos

```
col.cluster<-c("blue","red","green", "yellow", "black")[cl.pam]
pairs(X.s, col=col.cluster, main="PAM", pch=19)
```



Visualizacion con Componentes Principales

```
clusplot(X.s,cl.pam)
text(princomp(X.s)$scores[,1:2],
     labels=rownames(X.s),pos=1, col="orange")
```



These two components explain 62.5 % of the point variability.

Silhouette

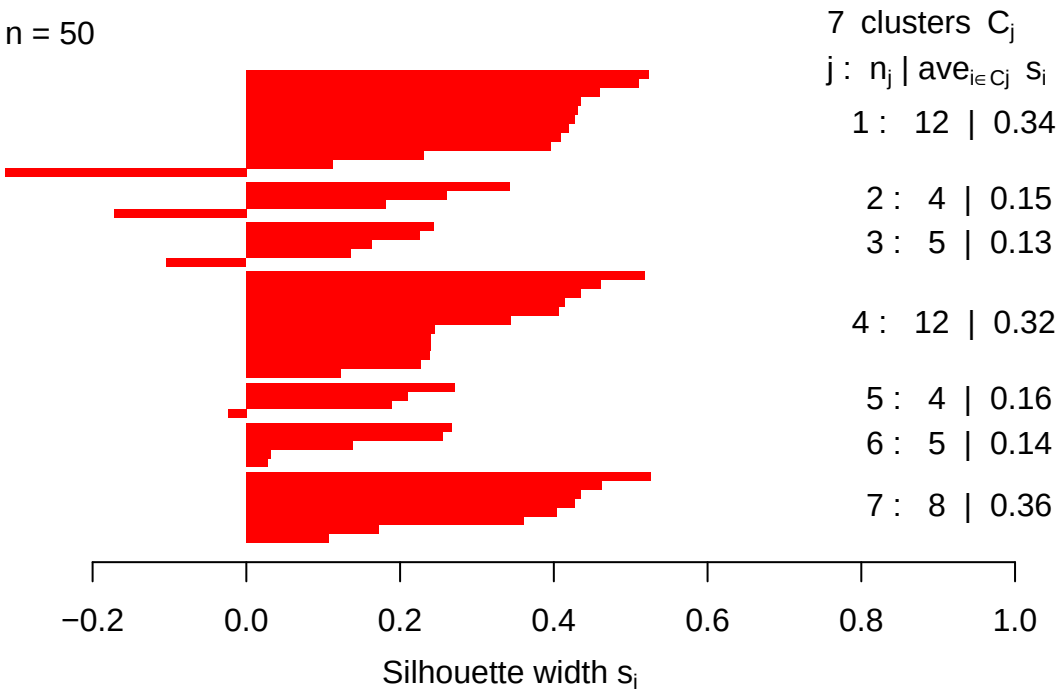
Representacion grafica de la eficacia de clasificacion de una observacion dentro de un grupo.

1.- Generacion de los calculos

```
dist.Euc<-dist(X.s, method = "euclidean")
Sil.pam<-silhouette(cl.pam, dist.Euc)
```

#2.- Generacion del grafico

```
plot(Sil.pam, main="Silhouette for PAM",
     col="red")
```



En este gráfico es posible apreciar que al realizar los cálculos con 7 clustes obtenemos valores un poco mas altos, aunque no son tan cercanos a 1