Metric multidimensional scaling Aplicado a la Cámara de Senadores

Rodrigo Murillo
Alain Cabrera
Stephane Keil

8 de marzo de 2017

Introducción

La técnica conocida como *Metric multidimensional scaling* (MDS) trasnforma una matriz de distancias en un conjunto coordenadas tales que las distancias (euclidianas) derivadas de estas coordenadas se aproximan lo más posible las distancias originales. La idea básica de MDS es transformar la matriz de distancias en un producto de matrices y encontrar la descomposición en valores singulares de ésta para obtener sus componentes principales (PCA).

El objetivo de este trabajo es aplicar MDS a los datos de la LXII Legislatura, utilizando como medida de similitud la diferencia entre los votos de cada senador.

Procedimiento

- Calcular la matriz de distancias D.
- Definir el vector de masa m cuyas entradas representan la masa de cada columna de la matriz D. Estas entradas deben ser positivas y sumar 1:
$$m^T\mathbf{1} = 1$$
- Definir la matriz centradora
$$H = I - \mathbf{1}m^T$$
- Calcular la matriz
$$S = -\frac{1}{2}HDH^T$$
- Obtener la factorización en valores singulares de la matriz S, es decir, escribir
$$S = U\Lambda U^T,$$
donde $UU^T = I$ y Λ es la matriz diagonal con los *eigenvalores* de S.
- Los nuevos valores para cada observación se calculan como
$$F = M^{-\frac{1}{2}}U\Lambda^{\frac{1}{2}},$$
con M la matriz diagonal de m.

Aplicación

Como existen senadores que no se presentan en algunas votaciones, únicamente se utilizarán en el análisis los datos de senadores con votos en común, esto para evitar tener *NA's* dentro de la matriz. Bajo ese supuesto, la base de datos con la que se trabajará cuenta con 59 observaciones.

Matriz de distancias

```
load("senado_votaciones.rda")

dist = as.data.frame(cor(senado_votaciones[c(4:137)]))
for(i in 1:134){
  for(j in 1:i){
    dist[i,j] = mean(senado_votaciones[,i+3] - senado_votaciones[,j+3], na.rm = TRUE)
    dist[j,i] = dist[i,j]
  }
}

# Quitar los NA
dist = na.omit(dist)
dist = dist[c(row.names(dist))]
D = as.matrix(dist)
```

Vector de masa

Tomaremos un vector que de la misma importancia a cada renglón de D, es decir, $m = \frac{1}{59}(1 \cdots 1)$.

```
m = as.vector(rep(1/59,59))
```

Matriz centradora

Ahora, se necesita construir la matriz centradora H

```
H = diag(59) - matrix(1,59,59)*m
```

Matriz S

A continuación, calculamos la matriz S como el producto de matrices

```
S = -.5*H%*%D%*%t(H)
```

Descomposición en valores singulares

Obtenemos la SVD deLa matriz S

```
svd = svd(S)
```

Aproximación

Calculamos los nuevos valores para cada individuo:

```
F = diag(m)^(-.5) %*% svd$u %*% diag(svd$d)^(.5)
```

Calidad de la aproximación

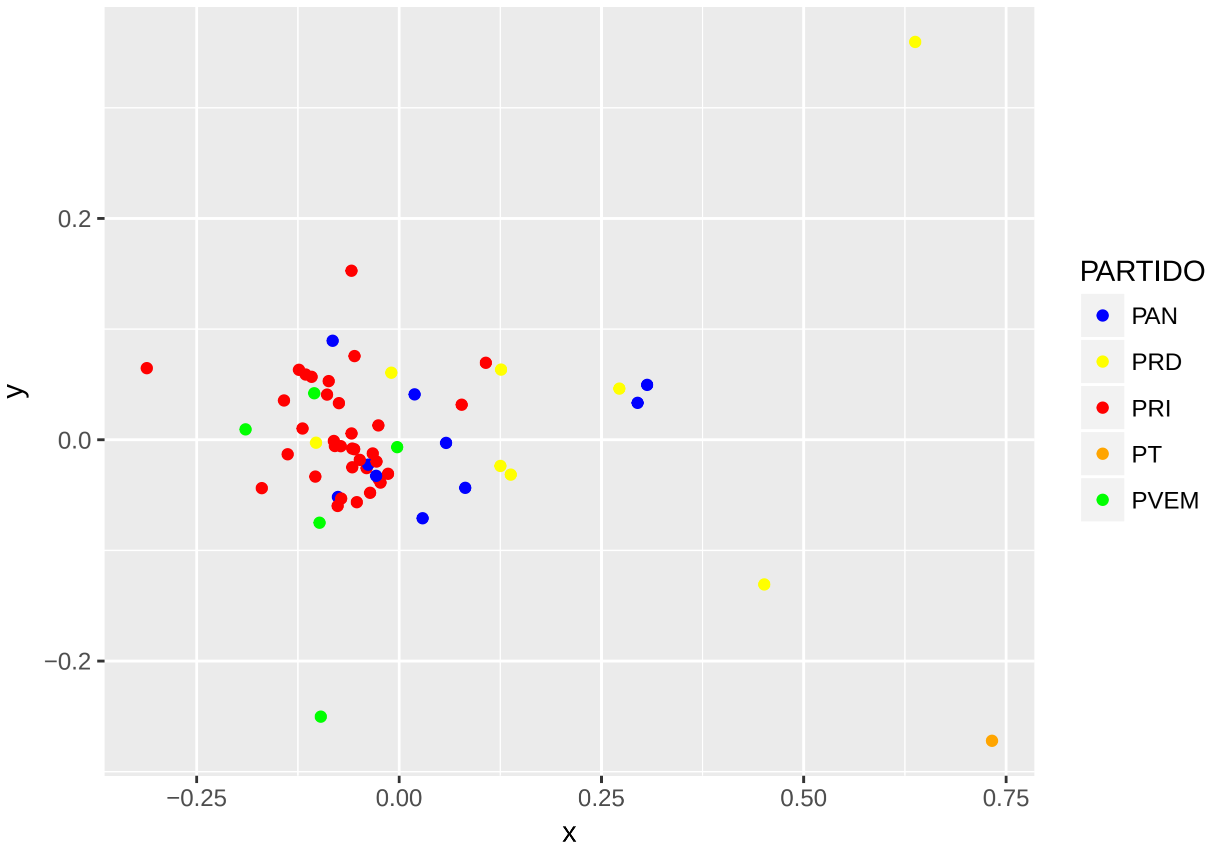
Para este ejemplo, únicamente elegiremos los primeros dos componentes los cuales explican el 32% de la varianza total.

```
sum(svd$d[1:2])/sum(svd$d)
```

```
## [1] 0.3202968
```

Visualización

A continuación se muestra un gráfico de dispersión de los senadores en las dos primeras componentes principales.



Conclusiones

La calidad de la aproximación es baja pues, al utilizar dos componentes principales, tan sólo se logra explicar el 32% de la varianza total. Analizando la primer componente principal, se puede observar que los senadores de los partidos de izquierda se localizan al lado derecho de los de de idiología de derecha.

Referencias

- Abdi, H. *Metric Multidimensional Scaling (MDS): Analyzing Distances Matrices* en Salkind, N. *Encyclopedia of Measurement and Statistics*, Thousand Oaks (CA): Sage (2007).