

# Clustering de ideologías usando NMF

*Víctor Gallego y Roi Naveiro*

*01/04/2019*

En este ejercicio plantearemos un problema de factorización de una matriz no negativa que representará el número de veces que cierta persona ha votado a cierto partido político en los últimos años.

1. Construcción de la matriz de datos. Estos son los votos de cada persona:

- Irene: 1 Podemos, 2 PSOE.
- Pepe: 2 Podemos, 1 PSOE, 1 Ciudadanos.
- Francisco: 6 Vox.
- Luisa: 1 Ciudadanos, 3 PP, 2 Vox.
- Cayetano: 4 Ciudadanos, 1 PP.
- Eva: 2 Podemos.
- Eugenia: 1 Podemos.
- Mario: 1 Vox

```
## [1] "Matriz de datos"
```

```
##      irene pepe fran luisa cayetano eva eugenia mario
## Ps      1    2    0     0         0  2         1    0
## PSOE    2    1    0     0         0  0         0    0
## Cs      0    1    0     1         4  0         0    0
## PP      0    0    0     3         1  0         0    0
## Vox     0    0    6     2         0  0         0    1
```

2. Factorizar la matrix  $\mathbf{X}$  para distintas dimensiones latentes

```
library(NMF)
```

```
d = 3
```

```
pol_nmf <- nmf(X, d)
```

```
fit(pol_nmf)
```

```
FALSE <Object of class:NMFstd>
```

```
FALSE features: 5
```

```
FALSE basis/rank: 3
```

```
FALSE samples: 8
```

```
# Factores
```

```
print("Matrix de factores")
```

```
FALSE [1] "Matrix de factores"
```

```
W <- basis(pol_nmf)
```

```
print(W)
```

```
FALSE      [,1]      [,2]      [,3]
FALSE Ps    2.220446e-16 2.220446e-16 1.076589e+01
FALSE PSOE  2.220446e-16 2.220446e-16 5.382943e+00
FALSE Cs    7.451120e+00 2.220446e-16 2.220446e-16
FALSE PP    4.967413e+00 2.220446e-16 2.220446e-16
FALSE Vox   2.220446e-16 1.511064e+01 2.220446e-16
```

```
#dim(W)
```

```
print("Matrix de coeficientes")
```

```
FALSE [1] "Matrix de coeficientes"
```

```
H <- coef(pol_nmf)
```

```
print(H)
```

```
FALSE      irene      pepe      fran      luisa      cayetano
FALSE [1,] 2.220446e-16 8.052481e-02 2.220446e-16 3.220992e-01 4.026240e-01
FALSE [2,] 2.220446e-16 2.220446e-16 3.970713e-01 1.323571e-01 2.220446e-16
FALSE [3,] 1.857720e-01 1.857720e-01 2.220446e-16 2.220446e-16 2.220446e-16
FALSE      eva      eugenia      mario
FALSE [1,] 2.220446e-16 2.220446e-16 2.220446e-16
FALSE [2,] 2.220446e-16 2.220446e-16 6.617856e-02
FALSE [3,] 1.238480e-01 6.192400e-02 2.220446e-16
```

```
#dim(H)
```

3. Interpretar los clusters políticos.

4. Reconstruye la matriz y recomienda partidos a los que votar a las diferentes personas.

```
print("Matrix reconstruida")
```

```
FALSE [1] "Matrix reconstruida"
```

```
V.hat <- fitted(pol_nmf)
```

```
print(V.hat)
```

```
FALSE      irene      pepe      fran      luisa      cayetano
FALSE Ps    2.000000e+00 2.000000e+00 2.478674e-15 2.491416e-15 2.479907e-15
FALSE PSOE 1.000000e+00 1.000000e+00 1.283421e-15 1.296163e-15 1.284654e-15
FALSE Cs    1.695731e-15 6.000000e-01 1.742649e-15 2.400000e+00 3.000000e+00
FALSE PP    1.144237e-15 4.000000e-01 1.191155e-15 1.600000e+00 2.000000e+00
FALSE Vox   3.396485e-15 3.414365e-15 6.000000e+00 2.000000e+00 3.444635e-15
FALSE      eva      eugenia      mario
FALSE Ps    1.333333e+00 6.666667e-01 2.405201e-15
FALSE PSOE 6.666667e-01 3.333333e-01 1.209948e-15
FALSE Cs    1.681981e-15 1.668231e-15 1.669176e-15
FALSE PP    1.130487e-15 1.116737e-15 1.117682e-15
FALSE Vox   3.382735e-15 3.368985e-15 1.000000e+00
```