# Clustering con K-medias

## Caso k=2

Los resultados obtenidos en este caso son:

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
|  | hall | offers | research | science | students | history | Class |
| 1 | 0.287682 | 0.0 | 0.916291 | 1.049.822 | 0.287682 | 0.0 | A |
| 3 | 0.287682 | 0.0 | 0.916291 | 1.049.822 | 0.287682 | 0.0 | A |
| 4 | 0.287682 | 0.0 | 0.0 | 1.049.822 | 0.287682 | 0.0 | A |
| 6 | 0.287682 | 0.0 | 0.0 | 1.049.822 | 0.287682 | 0.0 | A |
| 8 | 0.0 | 0.0 | 0.916291 | 0.0 | 0.0 | 0.0 | A |
| 10 | 0.0 | 0.0 | 0.0 | 1.049.822 | 0.287682 | 0.0 | A |
| 12 | 0.287682 | 0.798508 | 0.916291 | 1.049.822 | 0.287682 | 0.0 | A |
| 16 | 0.287682 | 0.0 | 0.916291 | 0.0 | 0.287682 | 0.0 | A |
| 17 | 0.287682 | 0.0 | 0.916291 | 1.049.822 | 0.287682 | 0.0 | A |
| 18 | 0.287682 | 0.798508 | 0.916291 | 0.0 | 0.287682 | 0.0 | A |
| 19 | 0.287682 | 0.798508 | 0.916291 | 0.0 | 0.287682 | 0.0 | A |
| 2 | 0.287682 | 0.798508 | 0.0 | 0.0 | 0.287682 | 0.0 | B |
| 5 | 0.0 | 0.798508 | 0.0 | 0.0 | 0.287682 | 0.0 | B |
| 7 | 0.0 | 0.0 | 0.0 | 0.0 | 0.287682 | 0.0 | B |
| 9 | 0.287682 | 0.798508 | 0.0 | 0.0 | 0.0 | 0.0 | B |
| 11 | 0.0 | 0.798508 | 0.0 | 0.0 | 0.0 | 189.712 | B |
| 13 | 0.287682 | 0.798508 | 0.0 | 0.0 | 0.0 | 0.0 | B |
| 14 | 0.287682 | 0.0 | 0.0 | 0.0 | 0.287682 | 189.712 | B |
| 15 | 0.287682 | 0.798508 | 0.0 | 0.0 | 0.0 | 189.712 | B |
| 20 | 0.287682 | 0.0 | 0.0 | 0.0 | 0.287682 | 0.0 | B |

Al hacer el clustering con k=2 tenemos una división con una asignación de instancias del 100%.

Se han creado dos instancias 0 y 1 que se corresponden con A y B ajustándose a los valores esperados.

|  |  |  |
| --- | --- | --- |
| 0 | 9 (45%) | B |
| 1 | 11 (55%) | A |

## Caso k=3

En este caso se han creado tres clusters y hay un error de 4 instancias, 20%.

Los centroides se han dividido según esta tabla:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  |  | Cluster | | |
| Attribute | Full Data | 0 | 1 | 2 |
| hall | 0.2158 | 0.1918 | 0.2466 | 0.2158 |
| offers | 0.3593 | 0.5323 | 0.1141 | 0.3993 |
| research | 0.3665 | 0 | 0.5236 | 0.9163 |
| science | 0.3674 | 0 | 10.498 | 0 |
| students | 0.2158 | 0.1598 | 0.2877 | 0.2158 |
| history | 0.2846 | 0.6324 | 0 | 0 |

Las instancias creadas son:

|  |  |  |
| --- | --- | --- |
| 0 | 9 (45%) | B |
| 1 | 7 (35%) | A |
| 2 | 4 (40%) | A (no class) |

La tercera clase parece que se ha intentado asignar a la clase A o también dejarla sin clase. En esta nueva clase se ha metido: hall, offers, research, students. SI no fuera por la inclusión de la palabra research se podría decir que esta nueva división hace referencia a estudiantes en general. También lo ha asignado a la clase A ya que contiene research que hace referencia a ciencias.

## Caso k=4

Ahora se han creado cuatro clases. El error obtenido es de 7 instancias (35%). En este caso los centroides de los clusters han quedado como:

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  |  | Cluster |  |  |  |
| Attribute | Full Data | 0 | 1 | 2 | 3 |
| hall | 0,22 | 0,25 | 0,29 | 0,22 | 0,00 |
| offers | 0,36 | 0,57 | 0,13 | 0,40 | 0,27 |
| research | 0,37 | 0,00 | 0,61 | 0,92 | 0,00 |
| science | 0,37 | 0,00 | 10498,00 | 0,00 | 0,35 |
| students | 0,22 | 0,12 | 0,29 | 0,22 | 0,29 |
| history | 0,28 | 0,81 | 0,00 | 0,00 | 0,00 |

La división final de clases es:

|  |  |  |
| --- | --- | --- |
| 0 | 7 (35%) | B |
| 1 | 6 (30%) | A |
| 2 | 4 (20%) | A (no class) |
| 3 | 3 (15%) | A y 2B (no class) |

Se puede ver que hay dos clases que pueden estar dentro de A o B o ser clases nuevas. Es curioso que el tercer cluster se puede dividir entre las clases A y B.

La clase 2 contiene: hall, offers, research, students y la 3: offers, science, students. Se puede ver que las dos clases en cierta medida corresponden con A que se refiere a ciencias por eso las dos se han asignado a esta. La clase 3 se ha asignado también a la 0 lo cual no tiene mucho sentido ya que tiene science y esta corresponde con history.