



Instituto Tecnológico
de Buenos Aires

Inteligencia Artificial y Machine Learning Certificación Avanzada en Ciencia de Datos

Trabajo Práctico 3

Aprendizaje no supervisado

Profesora: Dra Marcela Riccillo

Alumno: Federico José Zarategui

Dni: 27.691.042

1) Abra la base cheddar de la librería faraway
library(faraway)
data(cheddar)

Escriba ?cheddar e indique de qué tratan cada una de las variables

El dataset está compuesto por 4 variables:

a) taste: una puntuación subjetiva de gusto.

b) Acetic: concentración de ácido acético (escala logarítmica).

c) H2S: concentración de sulfuro de hidrógeno (escala logarítmica).

d) Lactic: concentración de ácido láctico.

2) Muestre un summary de la base.

taste	Acetic	H2S	Lactic
Min. : 0.70	Min. :4.477	Min. : 2.996	Min. :0.860
1st Qu.:13.55	1st Qu.:5.237	1st Qu.: 3.978	1st Qu.:1.250
Median :20.95	Median :5.425	Median : 5.329	Median :1.450
Mean :24.53	Mean :5.498	Mean : 5.942	Mean :1.442
3rd Qu.:36.70	3rd Qu.:5.883	3rd Qu.: 7.575	3rd Qu.:1.667
Max. :57.20	Max. :6.458	Max. :10.199	Max. :2.010

3) Considere su DNI (completo) para el seteo de la semilla y realice un agrupamiento kmeans.

Además, si su DNI termina en 0, 1, 2 ó 3

Cantidad de Grupos=3

Si su DNI termina en 4, 5, 6 ó 7

Cantidad de Grupos=4

Si su DNI termina en 8 ó 9

Cantidad de Grupos=5

Deje el resto de los parámetros por defecto. Indique el código R utilizado.

```
set.seed(DNI);km=kmeans(cheddar,cantGrupos)
```

```
set.seed(27691042);km=kmeans(cheddar,3)
```

4) ¿Cuántos elementos quedaron en cada grupo? `km$size`

Quedaron 9 en el grupo 1, 17 en el grupo 2 y 4 en el grupo 3.

5) Muestre `km$cluster` ¿A qué grupo pertenece el primer elemento de la base?

```
[1] 2 2 1 3 2 1 1 2 2 2 1 3 2 1 3 1 2 2 2 1 2 2 1 3 2 2 1 2 2 2
```

El primer elemento pertenece al grupo 2.

6) Muestre una captura de pantalla de los centroides `km$centers`

```
      taste    Acetic      H2S    Lactic
1 33.47778  5.738556  7.522778  1.535556
2 12.82353  5.240176  4.599588  1.305882
3 54.17500  6.052750  8.088750  1.810000
```

7) Determine alguna característica de alguno de los grupos con respecto a los otros grupos (tip: analice la información de los centroides).

El grupo 2 obtuvo la peor puntuación por el gusto.

Anexo con el código en R utilizado:

```
library(faraway)
data(cheddar)
?cheddar
base=cheddar
summary(base)
set.seed(27691042); km=kmeans(cheddar, 3)
km$size
km$cluster
km$centers
```