

11 Abra la base cheddar de la libreria faraway.

21 Muestre un summary de la base.

31 Seleccione 2 variables y realice un gráfico coloreado por los grupos formados.

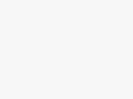
41 ¿Cuántos elementos quedaron en cada grupo?

51 Muestre km\$cluster

61 Muestre una captura de pantalla de los Centroides.

71 Determine alguna característica de alguno de los grupos con respecto a los otros grupos.

Anexo:



# Diplomatura en BIG DATA

## Machine Learning & Data Mining

Escalada Christian, DNI: 33549575

10-06-2022



### Trabajo Práctico N° 3

#### Aprendizaje No Supervisado

##### 1.1. Abra la base cheddar de la libreria faraway.

```
# Importamos Libreria
library(faraway)

# Cargamos Dataset
data(cheddar)

# Abrimos la base
head(cheddar)
```

```
##   taste Acetic   H2S Lactic
## 1  12.3  4.543 3.135   0.86
## 2  20.9  5.159 5.043   1.53
## 3  39.0  5.366 5.438   1.57
## 4  47.9  5.759 7.496   1.81
## 5   5.6  4.663 3.807   0.99
## 6  25.9  5.697 7.601   1.09
```

##### 1.2. Indique de qué tratan cada una de las variables.

- Las variables se tratan acerca de:
  - Taste**: Es una puntuación de gusto subjetivo.
  - Acetic**: Es la concentración de ácido acético medido en escala logarítmica.
  - H2S**: Es la concentración de sulfuro de hidrógeno medido en escala logarítmica.
  - Lactic**: Es la concentración de ácido láctico.

##### 2.1. Muestre un summary de la base.

```
# Instanciamos Dataset
base = cheddar

# Hacemos Summary
summary(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
```

##### 2.2. Considere los 3 últimos dígitos de su DNI para el seteo de la semilla.

- Además**:
  - Si su DNI termina en 0, 1 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**
- Mis 3 últimos números de mi DNI son: **575**
- Mi último número de DNI es **5**
- La **Cantidad de Grupos** asignado será: **4**

##### 2.3. Realice un agrupamiento kmeans. Indique el código R utilizado.

```
# Código R Utilizado:
set.seed(575); km=kmeans(cheddar,4)
```

##### 3.1 Seleccione 2 variables y realice un gráfico coloreado por los grupos formados.



- Las Variables que elegi para realizar el gráfico fueron:
  - Taste**.
  - Acetic**.

##### 4.1 ¿Cuántos elementos quedaron en cada grupo?

```
# Verificamos los elementos.
km
```

```
## K-means clustering with 4 clusters of sizes 13, 6, 7, 4
##
## Cluster means:
##   taste Acetic   H2S   Lactic
## 1 19.500000 5.371692 5.472462 1.380769
## 2 37.166667 5.709333 8.033333 1.618333
## 3  6.114286 5.234571 3.793714 1.194286
## 4 54.175000 6.052750 8.088750 1.810000
##
## Clustering vector:
## [1] 3 1 2 4 3 1 2 1 1 1 2 4 3 1 4 2 1 3 1 2 1 1 2 4 1 3 1 3 1 3
##
## Within cluster sum of squares by cluster:
## [1] 281.04499  68.90523 134.22596  62.47347
## (between_SS / total_SS =  93.0 %)
##
## Available components:
## [1] "cluster"      "centers"      "totss"      "withinss"    "tot.withinss"
## [6] "betweenss"   "size"        "iter"      "ifault"
```

- Podemos observar que los grupos quedaron conformados por las siguientes cantidades:
  - Grupo N°1: **13** unidades.
  - Grupo N°2: **6** unidades.
  - Grupo N°3: **7** unidades.
  - Grupo N°4: **4** unidades.

##### 5.1 Muestre km\$cluster.

```
km$cluster
```

```
## [1] 3 1 2 4 3 1 2 1 1 1 2 4 3 1 4 2 1 3 1 2 1 1 2 4 1 3 1 3 1 3
```

##### 5.2 ¿A qué grupo pertenece el tercer elemento de la base?

- el **Tercer** elemento de la Base pertenece al **Grupo N°2**.

##### 6.1 Muestre una captura de pantalla de los Centroides.

```
# Cargamos Centroides
km$centers
```

```
##      taste Acetic   H2S   Lactic
## 1 19.500000 5.371692 5.472462 1.380769
## 2 37.166667 5.709333 8.033333 1.618333
## 3  6.114286 5.234571 3.793714 1.194286
## 4 54.175000 6.052750 8.088750 1.810000
```

```
# Mostramos Centroides
attach(base)
plot(taste,Acetic,pch=19,col=km$cluster,main = "Gráfico de Grupos con Centroides")
points(km$centers[,c("taste","Acetic")],pch="0",col=1:4,cex=4)
```



##### 7.1 Determine alguna característica de alguno de los grupos con respecto a los otros grupos.

- tip**: Analice la información de los Centroides.

```
# Cargamos Centroides
km$centers
```

```
##      taste Acetic   H2S   Lactic
## 1 19.500000 5.371692 5.472462 1.380769
## 2 37.166667 5.709333 8.033333 1.618333
## 3  6.114286 5.234571 3.793714 1.194286
## 4 54.175000 6.052750 8.088750 1.810000
```

- Podemos observar que en general los valores que pertenecen a **taste** son los más grandes en cada grupo, por contraparte los valores que pertenecen a **Lactic** son los más bajos en cada grupo, mientras que los valores de **H2S** y **Acetic** son valores muy similares. En general podemos afirmar que el **Grupo 4** presenta valores más grandes, mientras que el **Grupo 3** presentan los valores más chicos.

##### Anexo:

###### Librerías:

```
library(faraway)
library(caret)
```

###### Código:

```
# Importamos Libreria
library(faraway)

# Cargamos Dataset
data(cheddar)

# Abrimos la base
head(cheddar)

# Instanciamos Dataset
base = cheddar

# Hacemos Summary
summary(base)

# Código R Utilizado:
set.seed(575); km=kmeans(cheddar,4)

# Importamos Libreria
library(caret)

# Graficamos
xyplot(base$Acetic~base$taste,base,groups=km$cluster,auto.key=TRUE,
       par.settings=simpleTheme(pch=c(3,4,5,8)),pch=c(3,4,5,8),
       main="Gráfico de Grupos",
       xlab="Taste", ylab="Acetic")

# Verificamos los elementos.
km

# Mostramos cluster
km$cluster

# Mostramos Centroides
attach(base)
plot(taste,Acetic,pch=19,col=km$cluster,main = "Gráfico de Grupos con Centroides")
points(km$centers[,c("taste","Acetic")],pch="0",col=1:4,cex=4)

# Cargamos los Centroides
km$centers
```

###### Bibliografía:

- plot Function
- Clustering
- R Plot()
- Pch symbols
- xyplot()
- Cluster I
- Cluster II
- K-means Wikipedia
- Cheese emoji

iMuchas Gracias!

