



Instituto Tecnológico
de Buenos Aires

Diplomatura en Ciencia de Datos

Módulo 2: Machine Learning - TP N°3

Profesor: Marcela Leticia Riccillo

Nombre: Ivan Kim

DNI: 37.687.040

Trabajo Práctico 3

Aprendizaje No Supervisado

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.

Taste of Cheddar cheese

Description

In a study of cheddar cheese from the LaTrobe Valley of Victoria, Australia, samples of cheese were analyzed for their chemical composition and were subjected to taste tests. Overall taste scores were obtained by combining the scores from several tasters.

Format

A data frame with 30 observations on the following 4 variables.

taste

a subjective taste score

Acetic

concentration of acetic acid (log scale)

H2S

concentration of hydrogen sulfide (log scale)

Lactic

concentration of lactic acid

Source

David S. Moore and George P. McCabe (1993) Introduction to the Practice of Statistics, W. H. Freeman and company, second edition.

Básicamente analizan los diferentes tipos de queso según su composición química y prueba de su sabor.

4 Variables a definir.

- Taste - El gusto del queso
- Acetic - Medir el ácido acético
- H2S - Concentración de sulfuro de hidrógeno
- Láctica - 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)
```

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

```
> km$size
[1] 14  7  9
```

El kmeans lo separo en 3 grupos con mi dni que termina en 0, que son

GRUPO 1 - 14

GRUPO 2 - 7

GRUPO 3 - 9

Dando un total de 30 observaciones.

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

```
> km$cluster
[1] 2 1 3 3 2 1 3 1 1 1 3 3 2 1 3 3 1 2 1 3 1 1 1 3 1 2 1 2 1 2
```

Pertenece al GRUPO 2

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

```
      taste  Acetic      H2S   Lactic
1 20.392857 5.377643 5.741714 1.385000
2  6.114286 5.234571 3.793714 1.194286
3 45.300000 5.890222 7.923667 1.723333
```

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

Analizando la información de los centroides basados en gusto (*taste*), acidez(*acetic*), niveles de azufre(*h2s*) y lactosa (*lactic*).

Una de las variables con mayor diferencia entre los grupos es el gusto (*taste*), donde las medias varían ampliamente, reflejando distintos niveles de sabor percibido.

En segundo lugar, el sulfuro de hidrógeno (*h2s*) presenta una variación más marcada. En el contexto del análisis de quesos y otros alimentos fermentados, este compuesto volátil de azufre contribuye al aroma intenso o característico del producto.

Por su parte, el ácido acético (*acetic*) mide la acidez volátil generada durante la fermentación láctica o bacteriana. Valores más altos indican un sabor más ácido o picante, mientras que valores menores reflejan un sabor más suave y cremoso.

Como último análisis, la acidez junto con la lactosa (*lactic*) muestra variaciones leves, manteniéndose relativamente estable entre los grupos.

En síntesis, los centroides permiten identificar tres tipos de queso:

- **Grupo 1:** quesos de sabor moderado y acidez media (balanceado).
- **Grupo 2:** quesos suaves, con baja acidez y menor intensidad aromática.
- **Grupo 3:** quesos intensos, aromáticos y de sabor pronunciado.

ANEXO CÓDIGO R

```
library(faraway)
data(cheddar)
dim(cheddar)
head(cheddar)
summary(cheddar)
?cheddar
set.seed(37687040);km=kmeans(cheddar,3)
km$size
km$cluster
km$centers
```