Docente: Dra. Marcela Riccillo

Alumno Marcelo Berra
Parameters Grupo 3

Trabajo Práctico Final – Data Mining y Machine Learning

---

1. Abra la base Vehicle de la librería mlbench y renómbrela como “base”.

```
> library(mlbench)
> data(Vehicle)
> base=Vehicle
> summary(base)
> str(base)
```

2. Indique de qué trata el problema, comente las variables, cantidad de registros.

```
A partir de una silueta de un vehiculo clasificarla en 4 modelos posibles:
  Autobus de 2 pisos     bus
  Camioneta Chevrolet    van
  Saab 9000              saab
  Opel Manta 400         opel
```

Info del Data Frame

	Cantidad
Registros	846
Variables	19

	CantVariables
Categoricas	18
Decision	1

bus	opel	saab	van
218	212	217	199

Definicion de las Variables

Variable	Definicion		Tipo
Categoricas			
Comp	Compacidad		Cuantitativa/Discreta
Circ	Circularidad		Cuantitativa/Discreta
D.Circ	Circularidad a Distancia		Cuantitativa/Discreta
Rad.Ra	Ratio Radio		Cuantitativa/Discreta
Pr.Axis.Ra	Ratio Pr.Axix aspecto		Cuantitativa/Discreta
Max.L.Ra	Ratio Maxima Longintud Aspecto		Cuantitativa/Discreta
Scat.Ra	Ratio de Dispersion		Cuantitativa/Discreta
Elong	Alargamiento		Cuantitativa/Discreta
Pr.Axis.Rect	Rectangularidad Pr.Axis		Cuantitativa/Discreta
Max.L.Rect	Rectangularidad Maxima Longitud		Cuantitativa/Discreta
Sc.Var.Maxis	Longitud del Eje Mayor, varianza		Cuantitativa/Discreta
Sc.Var.maxis	Longitud del Eje Menor, varianza		Cuantitativa/Discreta
Ra.Gyr	Radio de Giro		Cuantitativa/Discreta
Skew.Maxis	Sesgo sobre el eje mayor		Cuantitativa/Discreta
Skew.maxis	Sesgo sobre el eje menor		Cuantitativa/Discreta
Kurt.maxis	kurtosis sobre el eje menor		Cuantitativa/Discreta
Kurt.Maxis	kurtosis sobre el eje mayor		Cuantitativa/Discreta
Holl.Ra	Ratio de huecos		Cuantitativa/Discreta
		Valores Posibles	
Decision			
Class	Tipo de Auto	Bus / Opel / Saab / Van	Cualitativa/Nominal

3. Muestre un head de la base.

```
> head(base)
```

```
  Comp Circ D.Circ Rad.Ra Pr.Axis.Ra Max.L.Ra Scat.Ra Elong Pr.Axis.Rect Max.L.Rect Sc.Var.Maxis Sc.Var.maxis
1   95   48     83    178         72        10     162    42           20        159          176          379
2   91   41     84    141         57         9     149    45           19        143          170          330
3  104   50    106    209         66        10     207    32           23        158          223          635
4   93   41     82    159         63         9     144    46           19        143          160          309
5   85   44     70    205        103        52     149    45           19        144          241          325
6  107   57    106    172         50         6     255    26           28        169          280          957

  Ra.Gyr Skew.Maxis Skew.maxis Kurt.maxis Kurt.Maxis Holl.Ra Class
1    184         70          6         16        187     197   van
2    158         72          9         14        189     199   van
3    220         73         14          9        188     196  saab
4    127         63          6         10        199     207   van
5    188        127          9         11        180     183   bus
6    264         85          5          9        181     183   bus
```

4. Borre la variable Class y muestre nuevamente un head de la base.

```
> base$Class=NULL
> head(base)
```

```
  Comp Circ D.Circ Rad.Ra Pr.Axis.Ra Max.L.Ra Scat.Ra Elong Pr.Axis.Rect Max.L.Rect Sc.Var.Maxis Sc.Var.maxis
1   95   48     83    178         72        10     162    42           20         159          176          379
2   91   41     84    141         57         9     149    45           19         143          170          330
3  104   50    106    209         66        10     207    32           23         158          223          635
4   93   41     82    159         63         9     144    46           19         143          160          309
5   85   44     70    205        103        52     149    45           19         144          241          325
6  107   57    106    172         50         6     255    26           28         169          280          957

  Ra.Gyr Skew.Maxis Skew.maxis Kurt.maxis Kurt.Maxis Holl.Ra
1    184         70          6         16        187     197
2    158         72          9         14        189     199
3    220         73         14          9        188     196
4    127         63          6         10        199     207
5    188        127          9         11        180     183
6    264         85          5          9        181     183
```

5. Setee la semilla=8 y realice un Agrupamiento K-means con cantidad de **grupos = 3**
Indique el código R utilizado.

```
> set.seed(8)
> km=kmeans(base,3)  # separo en 3 grupos
```

6. Muestre una imagen de los centroides.

```
> km$centers
```

```
       Comp     Circ    D.Circ   Rad.Ra Pr.Axis.Ra Max.L.Ra  Scat.Ra    Elong Pr.Axis.Rect Max.L.Rect Sc.Var.Maxis
1 104.13333 53.11429 102.78095 201.6238   62.00000 9.747619 217.6905 30.64286     24.44762   166.5524     230.6952
2  96.29333 45.06667  88.25333 195.5533   65.40667 8.933333 179.4267 36.65333     21.22000   146.9933     202.5467
3  88.35391 41.23251  71.24486 146.6049   60.41564 7.944444 144.4630 46.70165     18.71605   140.2922     166.1502

  Sc.Var.maxis   Ra.Gyr Skew.Maxis Skew.maxis Kurt.maxis Kurt.Maxis  Holl.Ra
1     703.9905 214.5571   72.69048   7.314286   15.74286   187.9000 196.1810
2     487.6000 174.9933   69.10667   6.053333   13.64000   193.7800 200.3133
3     311.0844 157.3930   73.39918   6.072016   10.91975   187.8827 193.9506
```

7. ¿Cuántos elementos quedaron en cada grupo?

```
> km$size
[1] 210 150 486
```

Grupo 1: 210
Grupo 2: 150
Grupo 3: 486

8. ¿A qué grupo pertenece el cuarto elemento de la base?

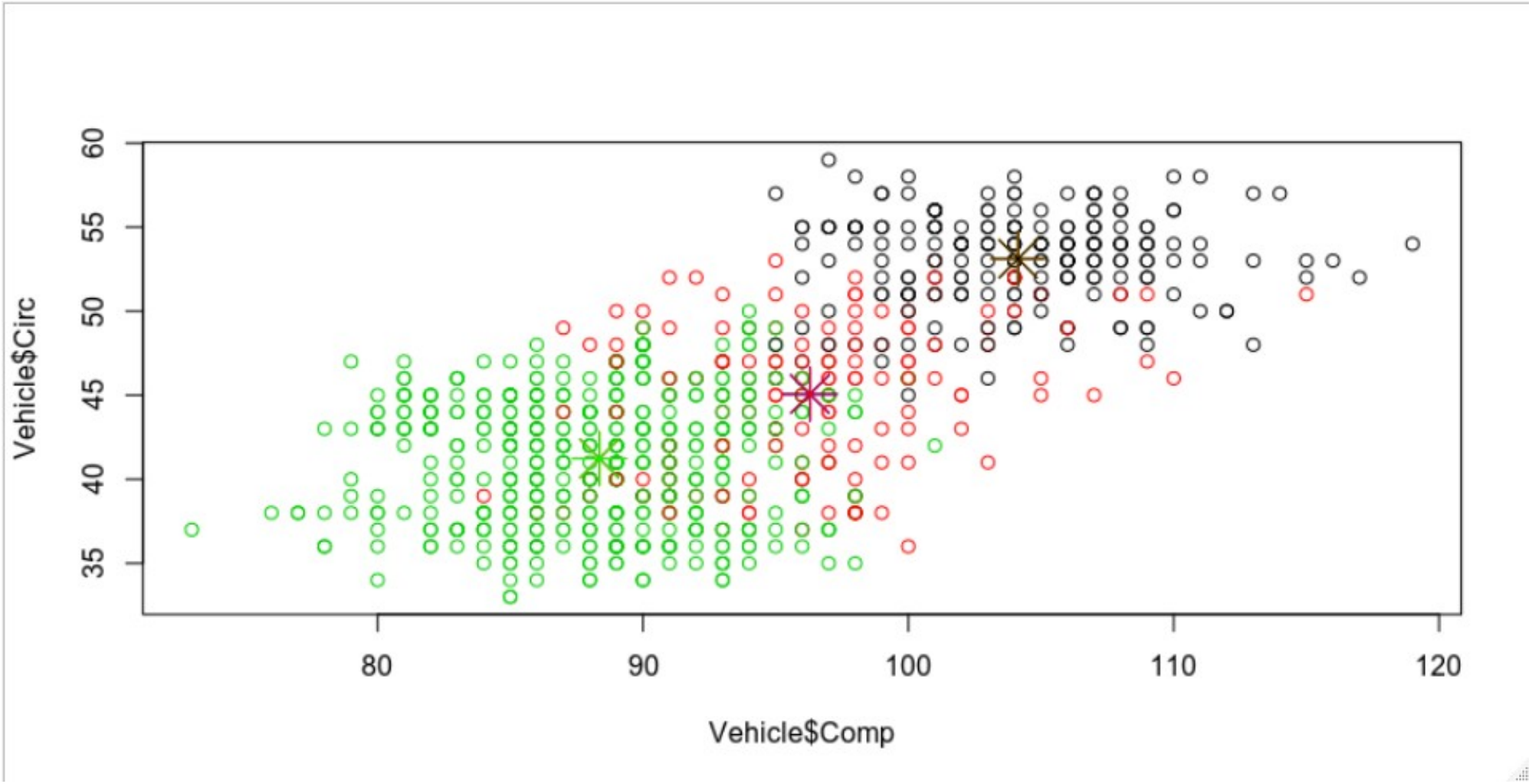
```
> km$cluster
  1   2   3   4   5   6   7   8   9  10 ...
  3   3   1   3   3   1   3   3   3   2 ...
```

El cuarto elemento pertenece al grupo 3. Otra forma de verlo es:

```
> km$cluster[4]
 4
 3
```

9. Realice un gráfico con dos variables coloreado por los grupos formados.

```
plot(Vehicle$Comp, Vehicle$Circ, col=km$cluster)
points(km$centers[,c("Comp", "Circ")], col=c("black", "red", "green"), pch=8, cex=3)
```



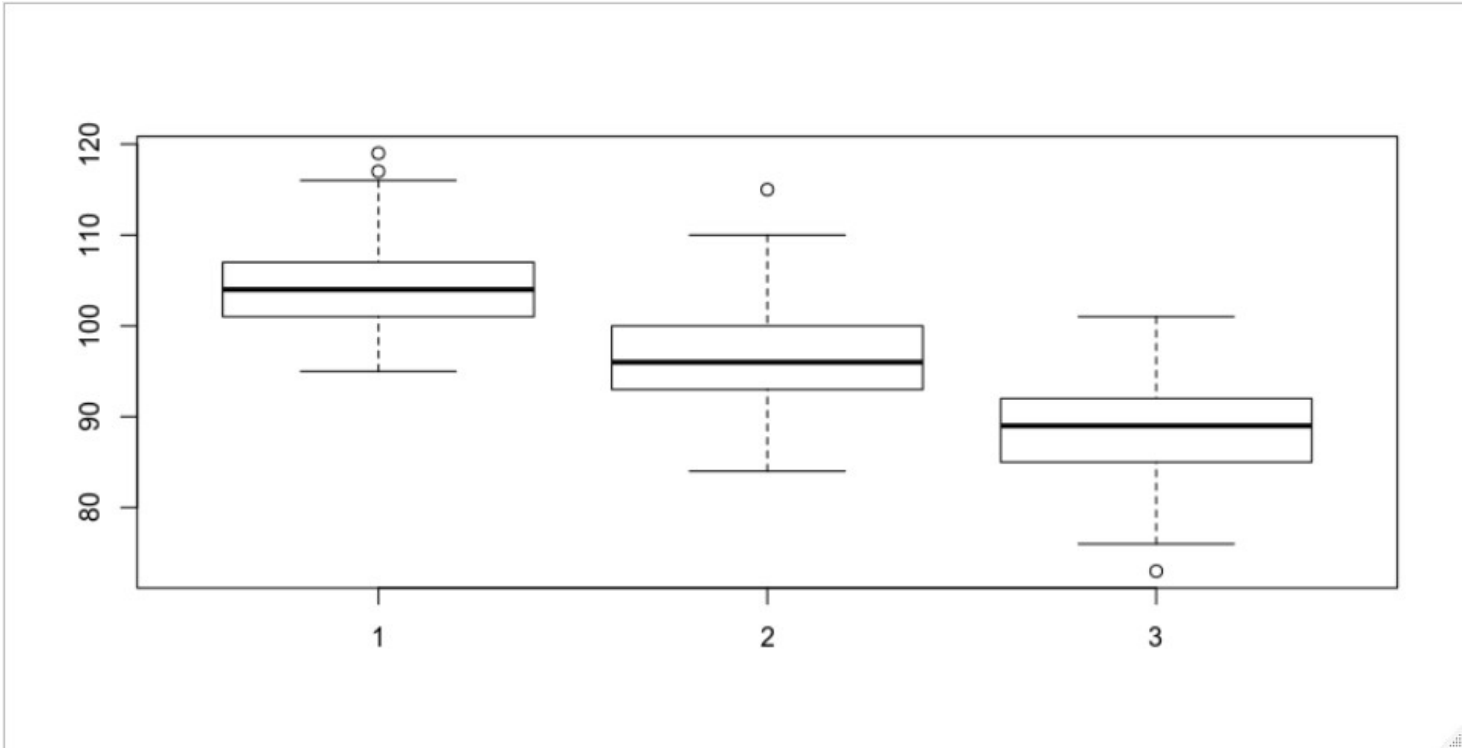
10. ¿Puede determinar alguna característica de alguno de los grupos?

```
# en general, los vehiculos del grupo 1(negro) tienen mayor Compacidad y Circularidad
  que los del grupo 3 (verde)

# las siguientes variables categoricas son mayores en los vehiculos del grupo 1, disminuyen en el
  grupo 2 y vuelven a disminuir en el grupo 3
```

Comp	Rad.Ra	Pr.Axis.Rect	Sc.Var.maxis
Circ	Max.L.Ra	Max.L.Rect	Ra.Gyr
D.Circ	Scat.Ra	Sc.Var.Maxis	Kurt.maxis

```
# analizando la variable categorica Compacidad,  vemos que las medianas de los 3 grupos
  estan bien diferenciadas entre si.
```



Anexo Codigo R

```
library(mlbench)
data(Vehicle)
base=Vehicle
summary(base$Class)
str(base)
#-------------------------------------------------------------------------------
head(base)
base$Class=NULL
head(base)
#-------------------------------------------------------------------------------
set.seed(8)
km=kmeans(base,3)
km$centers
km$size
km$cluster[4]
#-------------------------------------------------------------------------------
plot(Vehicle$Comp, Vehicle$Circ, col=km$cluster)
points(km$centers[,c("Comp", "Circ")], col=c("black", "red", "green"), pch=8, cex=3)
#-------------------------------------------------------------------------------
boxplot(Vehicle$Comp~km$cluster)
#-------------------------------------------------------------------------------
```