



**EDUCACIÓN**  
SECRETARÍA DE EDUCACIÓN PÚBLICA



TECNOLÓGICO  
NACIONAL DE MÉXICO®

TECNOLÓGICO NACIONAL DE MÉXICO

INSTITUTO TECNOLÓGICO DE TIJUANA

SUBDIRECCIÓN ACADÉMICA

DEPARTAMENTO DE SISTEMAS Y COMPUTACIÓN

NOMBRE DE LOS ALUMNOS:

RAYMUNDO HIRALES LAZARENO (N. CONTROL: 17212339)

GALAVIZ LONA OSCAR EDUARDO (N.CONTROL: 17212993)

Carrera: Ingeniería Informática

Semestre: 9no

MATERIA: Minería de datos

PROFESOR: JOSE CHRISTIAN ROMERO HERNANDEZ

TRABAJOS: Practica evaluatoria U4

FECHA: 13/12/21

## Introduccion

Para esta practica se utilizo el modelo kmeans, el modelo es un metodo de agrupamiento que tiene como objetivo la particion de un conjunto de n observaciones en k grupos en el que la observacion pertenece al grupo con un valor medio mas cercano, ademas de ser de forma agil y controlada, esta utiliza operaciones estadisticas que para minimizar los errores y mostrar diferentes puntos aleatorios de k llamados centroides y cada punto de k se le asigna una observacion en estos casos busca el centroide con menos errores y por ultimo este modelo nos mostrara los datos de una forma diferentes a otros en este caso en forma de puntos y a diferencia de los demas este busca que tengan menos errores posibles en cada punto

## Desarrollo

```
dataset = read.csv('iris.csv')
dataset = dataset[1:4]
install.packages("cluster")
library(cluster)
set.seed(101)
irisCluster <- kmeans(dataset[,1:4], center=3, nstart=20)
irisCluster
```

	sepal_length	sepal_width	petal_length	petal_width
1	5.1	3.5	1.4	0.2
2	4.9	3.0	1.4	0.2
3	4.7	3.2	1.3	0.2
4	4.6	3.1	1.5	0.2
5	5.0	3.6	1.4	0.2
6	5.4	3.9	1.7	0.4
7	4.6	3.4	1.4	0.3
8	5.0	3.4	1.5	0.2
9	4.4	2.9	1.4	0.2
10	4.9	3.1	1.5	0.1
11	5.4	3.7	1.5	0.2
12	4.8	3.4	1.6	0.2
13	4.8	3.0	1.4	0.1
14	4.3	3.0	1.1	0.1
15	5.8	4.0	1.2	0.2
16	5.7	4.4	1.5	0.4
17	5.4	3.9	1.3	0.4
18	5.1	3.5	1.4	0.3
19	5.7	3.8	1.7	0.3
20	5.1	3.8	1.5	0.3
21	5.4	3.4	1.7	0.2

K-means clustering with 3 clusters of sizes 62, 38, 50

Cluster means:

```

sepal_length sepal_width petal_length petal_width
1      5.901613      2.748387      4.393548      1.433871
2      6.850000      3.073684      5.742105      2.071053
3      5.006000      3.418000      1.464000      0.244000

```

Clustering vector:

```

[1] 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3
[34] 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 1 1 2 1 1 1 1 1 1 1 1 1 1 1 1 1
[67] 1 1 1 1 1 1 1 1 1 1 1 1 2 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
[100] 1 2 1 2 2 2 2 1 2 2 2 2 2 2 1 1 2 2 2 2 1 2 1 2 1 2 2 1 1 2 2 2
[133] 2 1 2 2 2 2 1 2 2 2 1 2 2 2 1 2 2 1

```

within cluster sum of squares by cluster:

```

[1] 39.82097 23.87947 15.24040
(between_SS / total_SS = 88.4 %)

```

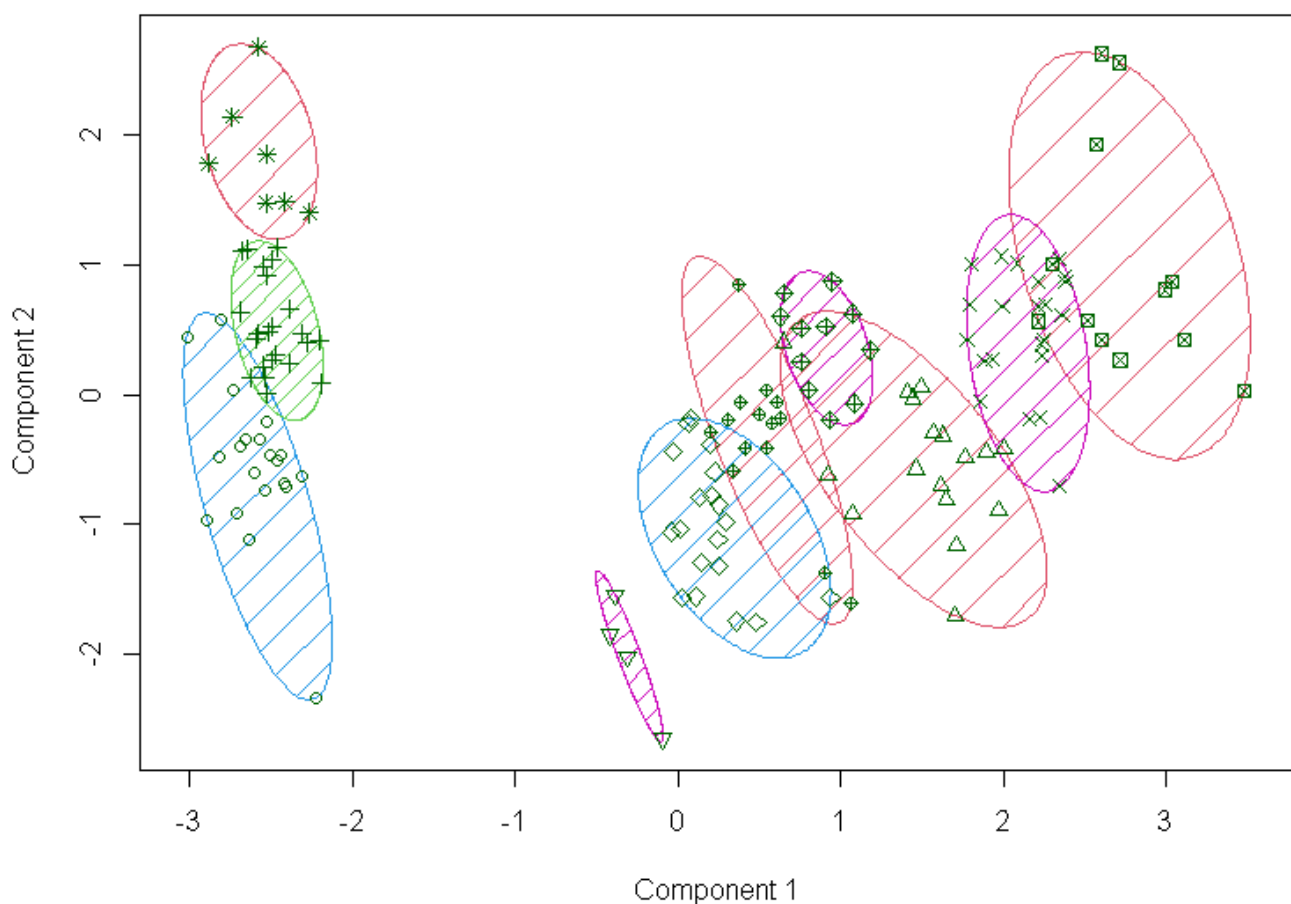
Available components:

```
[1] "cluster"      "centers"      "totss"        "withinss"
[5] "tot.withinss" "betweeness"   "size"         "iter"
[9] "ifault"
```

Name	Type	Value
irisCluster	list [9] (S3: kmeans)	List of length 9
cluster	integer [150]	10 6 6 6 10 4 ...
centers	double [10 x 4]	7.475 6.568 5.969 5.529 6.655 4.690 3.125 3.086 2.862 4.043 3.045 3.085 6.300 5. ...
totss	double [1]	680.8244
withinss	double [10]	4.655 4.315 3.114 0.834 1.182 2.567 ...
tot.withinss	double [1]	25.81339
betweeness	double [1]	655.011
size	integer [10]	12 22 16 7 11 20 ...
iter	integer [1]	4
ifault	integer [1]	0

```
library(cluster)
clusplot(iris, irisCluster$cluster, color=T, shade=T, labels=0, lines=0)
```

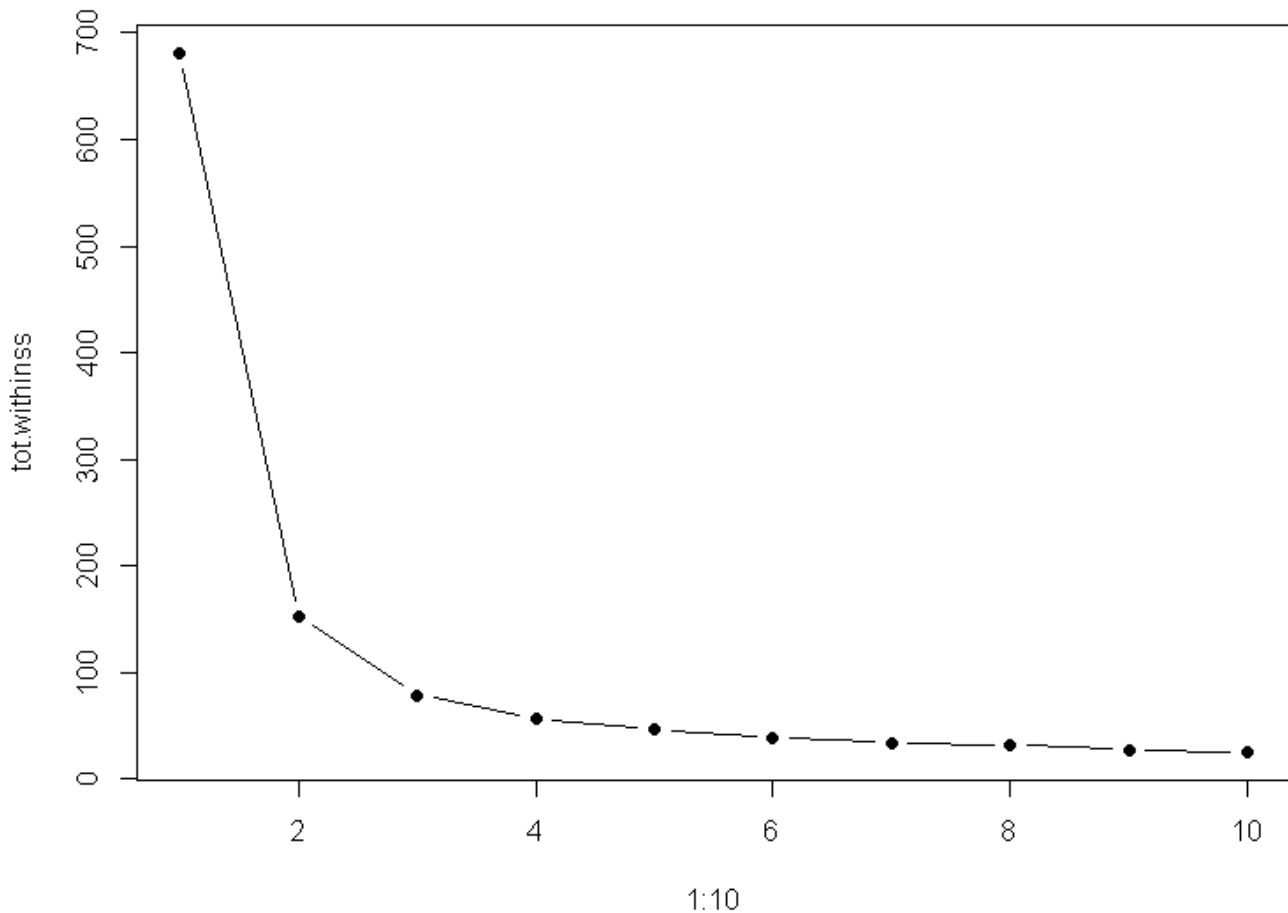
### CLUSPLOT(iris)



These two components explain 95.02 % of the point variability.

```
tot.withinss <- vector(mode="character", length=10)
for (i in 1:10){
  irisCluster <- kmeans(dataset[,1:4], center=i, nstart=20)
  tot.withinss[i] <- irisCluster$tot.withinss
}

plot(1:10, tot.withinss, type="b", pch=19)
```



## Análisis de resultados

Aquí podemos ver Implementado el metodo K-means al ver agrupados los datos del repositorio .csv notamos que los componentes del cluster estan dispersos ya que los dos componentes explican la variabilidad de 95.02% en el repositorio lo que al relacionarlo con la grafica nos da una reduccion en los valores de los grupos mientras mas avanzamos.

### Conclusion

Esta practica quizas no fue tan larga pero demuestra que tiene cosas que enseñar tratandose de un modelo que nos permite agrupar datos y buscar los puntos con menos errores significa que es un modelo que se preocupa por su exactitud, este modelo funciona para el agrupamiento de datos, asi como identificar los

diferentes puntos intermedios del grupo de datos, en conclusion es un modelo que tiene una exactitud bastante buena y que para la clasificacion es de uno de los modelos a utilizar