



Progressive Education Society's
Modern College of Engineering, Shivajinagar, Pune-05.
(An Autonomous Institute Affiliated to Savitribai Phule Pune University)
Department of MCA

PRACTICAL SUBMISSION RECORD- A.Y. 2025-26

Class: SYMCA Division : A Semester: III	Course Code: MCA01604 Course Name: Data Science Laboratory	Batch: S2
Name: Pranav Raju Malwatkar		Roll No: 52037
CO No: CO605.1		Assignment No: 7

Title : Load the tissue_gene_expression dataset.Run a k-means clustering on the data with K=7. Make a table comparing the identified clusters to the actual tissue types. Run the algorithm several times to see how the answer changes.

Code:

```
# Load necessary library
library(dslabs)

# Load dataset
data("tissue_gene_expression")

# View dataset structure
str(tissue_gene_expression)

# Extract data and tissue type labels
x <- tissue_gene_expression$x
y <- tissue_gene_expression$y

# Set seed for reproducibility
set.seed(123)

# Run K-means clustering with K = 7
k <- 7
kmeans_result <- kmeans(x, centers = k)

# Compare clusters with actual tissue types
comparison_table <- table(Cluster = kmeans_result$cluster, Tissue = y)
print(comparison_table)
```



Progressive Education Society's
Modern College of Engineering, Shivajinagar, Pune-05.
(An Autonomous Institute Affiliated to Savitribai Phule Pune University)

Department of MCA

Visualize the clustering using first two principal components

```
library(ggplot2)
pca <- prcomp(x)
df <- data.frame(PC1 = pca$x[,1], PC2 = pca$x[,2],
                  Cluster = as.factor(kmeans_result$cluster),
                  Tissue = y)
```

```
ggplot(df, aes(PC1, PC2, color = Cluster, shape = Tissue)) +
  geom_point(size = 2) +
  ggtitle("K-Means Clustering (K = 7) on Tissue Gene Expression Data")
```

Run the algorithm multiple times to see variation

```
set.seed(123)
for (i in 1:5) {
  km <- kmeans(x, centers = k)
  cat("\nRun", i, ":\n")
  print(table(Cluster = km$cluster, Tissue = y))
}
```

Output :

```
Console Terminal × Background Jobs ×
R 4.5.1 · ~/ 
> set.seed(123)
> for (i in 1:5) {
+   km <- kmeans(x, centers = k)
+   cat("\nRun", i, ":\n")
+   print(table(Cluster = km$cluster, Tissue = y))
+ }

Run 1 :
    Tissue
Cluster cerebellum colon endometrium hippocampus kidney liver placenta
  1          0     0           0         0      0     7     0
  2          0     0           0         0      0    17     0
  3         31     0           0         0      0     0     0
  4          2     0           0         0      2     2     0
  5          0    34           15        0      1     0     6
  6          5     0           0        31      0     0     0
  7          0     0           0         0     36     0     0

Run 2 :
    Tissue
Cluster cerebellum colon endometrium hippocampus kidney liver placenta
  1          31     0           0         0      0     0     0
```



Progressive Education Society's
Modern College of Engineering, Shivajinagar, Pune-05.
(An Autonomous Institute Affiliated to Savitribai Phule Pune University)

Department of MCA

Console Terminal × Background Jobs ×

R 4.5.1 · ~/

Run 2 :

Tissue

Cluster	cerebellum	colon	endometrium	hippocampus	kidney	liver	placenta
1	31	0	0	0	0	0	0
2	0	0	0	18	0	0	0
3	0	34	15	0	37	24	0
4	2	0	0	0	2	2	0
5	5	0	0	0	0	0	0
6	0	0	0	0	0	0	6
7	0	0	0	13	0	0	0

Run 3 :

Tissue

Cluster	cerebellum	colon	endometrium	hippocampus	kidney	liver	placenta
1	0	0	0	0	36	0	0
2	5	0	0	31	0	0	0
3	0	34	15	0	1	0	0
4	2	0	0	0	2	2	0
5	31	0	0	0	0	0	0
6	0	0	0	0	0	24	0
7	0	0	0	0	0	0	6

Console Terminal × Background Jobs ×

R 4.5.1 · ~/

Run 4 :

Tissue

Cluster	cerebellum	colon	endometrium	hippocampus	kidney	liver	placenta
1	0	0	0	0	0	24	0
2	0	0	0	0	36	0	0
3	0	0	15	0	0	0	0
4	0	34	0	0	0	0	6
5	5	0	0	31	0	0	0
6	31	0	0	0	0	0	0
7	2	0	0	0	3	2	0

Run 5 :

Tissue

Cluster	cerebellum	colon	endometrium	hippocampus	kidney	liver	placenta
1	0	0	0	0	21	0	0
2	6	0	0	0	0	0	0
3	0	0	0	31	0	0	0
4	32	0	0	0	0	0	0
5	0	0	0	0	18	0	0
6	0	0	0	0	0	26	0
7	0	34	15	0	0	0	6

> |



Progressive Education Society's
Modern College of Engineering, Shivajinagar, Pune-05.
(An Autonomous Institute Affiliated to Savitribai Phule Pune University)

Department of MCA

R Graphics: Device 2 (ACTIVE)

File History Resize

K-Means Clustering (K = 7) on Tissue Gene Expression Data

