K MEAN CLUSTERING

Suppose that the data mining task is to cluster points into three clusters, where the points are A1(2, 10), A2(2, 5), A3(8, 4), B1(5, 8), B2(7, 5), B3(6, 4), C1(1, 2), C2(4, 9). The distance function is Euclidean distance. Suppose initially we assign A1, B1, and C1 as the center of each cluster, respectively.

Data points			(2 ,10)	(5,8)	1,2	Clusters	New Cluster	
A1	2	10	0.0	3.61	8.06	1		
A2	2	5	5	4.24	3.16	3		
A3	8	4	8.49	5	7.28	2		
B1	5	8	3.61	0.0	7.21	2		
B2	7	5	7.07	3.61	6.71	2		
B3	6	4	7.21	4.12	5.39	2		
C1	1	2	8.06	7.21	0.0	3		
C2	4	9	2.24	1.41	7.62	2		

NEW CENTROIDS

A1(2,10)

B1(6,6)

C1(1.5, 3.5)

Data points			(2 ,10)	(6,6)	1.5, 3.5	Clusters	New Cluster	
A1	2	10	0.0	5.66	6.52	1	1	
A2	2	5	5	4.12	1.58	3	3	
A3	8	4	8.49	2.83	6.52	2	2	
B1	5	8	3.61	2.24	5.70	2	2	
B2	7	5	7.07	1.41	6.71	2	2	
B3	6	4	7.21	2.0	4.53	2	2	
C1	1	2	8.06	6.40	1.58	3	3	
C2	4	9	2.24	3.61	6.04	2	1	

NEW CENTROIDS

A1(3, 9.5)

B1(6.5,5.25)

C1(1.5, 3.5)

Repeat the process again and until previous clusters and new cluster are not equal

NEW CENTROIDS

A1(3.67, 9)

B1(7, 4.33)

C1(1.5, 3.5)

R CODE

```r
# Load necessary library
library(ggplot2)

# Define the dataset
points <- data.frame(
  x = c(2, 2, 8, 5, 7, 6, 1, 4),
  y = c(10, 5, 4, 8, 5, 4, 2, 9),
  labels = c("A1", "A2", "A3", "B1", "B2", "B3", "C1", "C2")
)

# Initial cluster centers (A1, B1, C1)
initial_centers <- data.frame(
  x = c(2, 5, 1),
  y = c(10, 8, 2)
)
```

R CODE

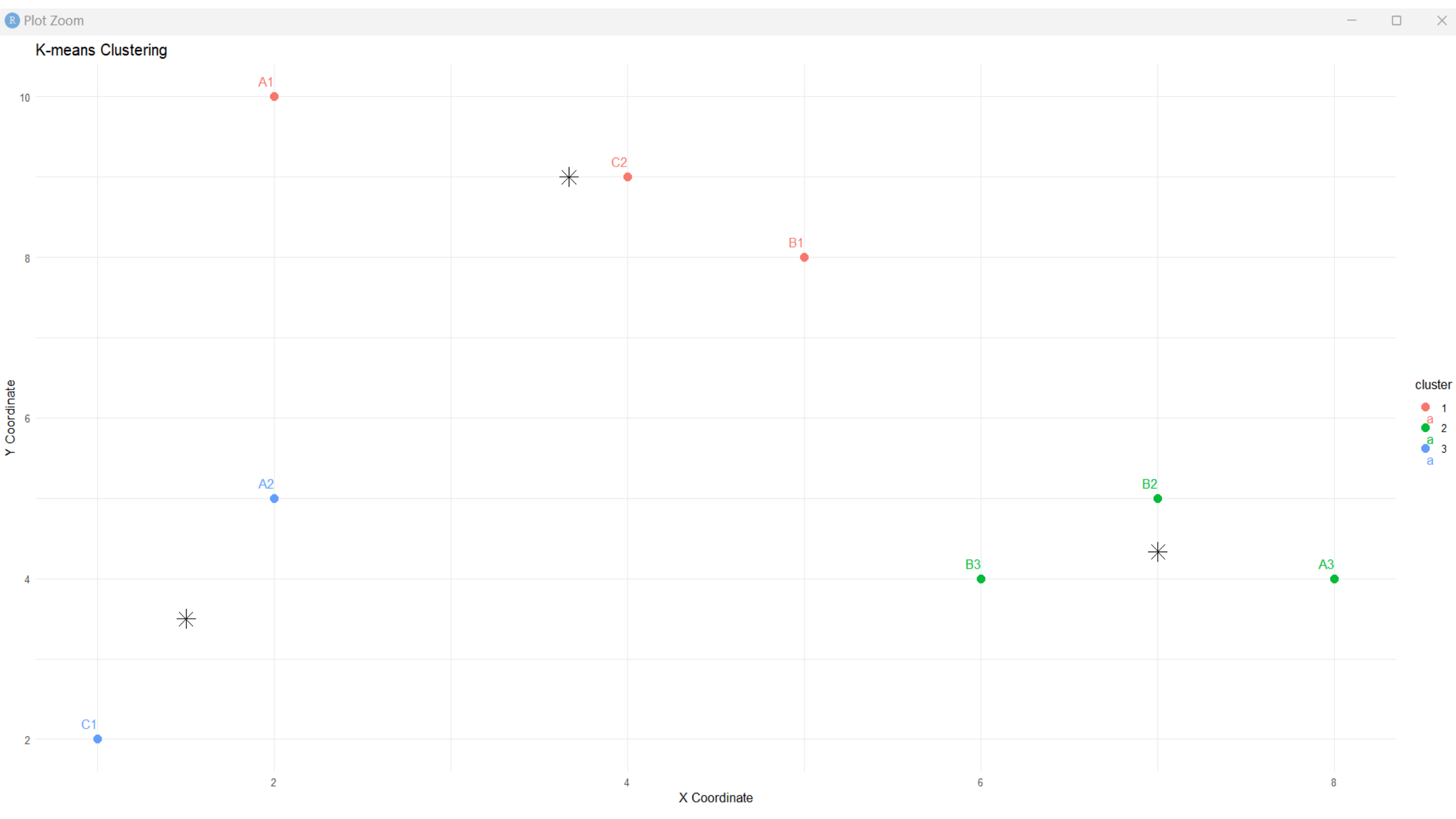
```r
set.seed(123)
kmeans_result <- kmeans(points[, c("x", "y")], centers = initial_centers, iter.max = 10)

# Add cluster assignment to data
points$cluster <- as.factor(kmeans_result$cluster)

# Ensure the centers have column names 'x' and 'y'
centers <- data.frame(x = kmeans_result$centers[, 1], y = kmeans_result$centers[, 2])
```

R CODE

```r
# Visualize the clusters
ggplot(points, aes(x = x, y = y, color = cluster)) +
  geom_point(size = 3) +
  geom_text(aes(label = labels), vjust = -1, hjust = 1) +
  geom_point(data = centers, aes(x = x, y = y), color = "black", size = 5, shape = 8) +
  labs(title = "K-means Clustering",
       x = "X Coordinate",
       y = "Y Coordinate") +
  theme_minimal()
```


Plot Zoom
K-means Clustering
A1
C2
B1
A2
B2
B3
A3
C1
10
8
6
4
2
2
4
6
8
Y Coordinate
X Coordinate
cluster
1
2
3

K MEAN CLUSTERING

Lecture 13

MATCH YOUR RESULT WITH THE GIVEN OUTPUT