

TAAHA HUSSAIN KHAN

L1F21BSCS0917

F7

ASSIGNMENT 04

BIG DATA ANALYTICS

R CODE:

```
1 data <- data.frame(  
2   A = c(1.0, 1.5, 3.0, 5.0, 3.5, 4.5, 3.5),  
3   B = c(1.0, 2.0, 4.0, 7.0, 5.0, 5.0, 4.5)  
4 )  
5  
6  
7 initial_centers <- data.frame(A = c(1.0, 5.0), B = c(1.0, 7.0))  
8 rownames(initial_centers) <- c("Cluster1", "Cluster2")  
9  
10 kmeans_iteration <- function(data, centers) {  
11  
12   dists <- as.matrix(dist(rbind(centers, data)))  
13   dists <- dists[-c(1:2), 1:2]  
14   clusters <- max.col(-dists, ties.method = "first")  
15  
16   new_centers <- data.frame(A = numeric(2), B = numeric(2))  
17   for (k in 1:2) {  
18     cluster_points <- data[clusters == k, ]  
19     new_centers[k, ] <- colMeans(cluster_points)  
20   }  
21  
22   list(clusters = clusters, centers = new_centers)  
23 }  
24  
25  
26 iteration1 <- kmeans_iteration(data, initial_centers)  
27 iteration1_centers <- iteration1$centers  
28 iteration1_clusters <- iteration1$clusters  
29  
30  
31 iteration2 <- kmeans_iteration(data, iteration1_centers)  
32 iteration2_centers <- iteration2$centers  
33 iteration2_clusters <- iteration2$clusters  
34  
35  
36 cat("Iteration 1 Centers:\n")  
37 print(iteration1_centers)  
38 cat("Iteration 1 Clusters:\n")  
39 print(iteration1_clusters)  
40 cat("Iteration 2 Centers:\n")  
41 print(iteration2_centers)  
42 cat("Iteration 2 Clusters:\n")  
43 print(iteration2_clusters)  
44
```

CONSOLE OUTPUT:

```
Console Terminal Background Jobs  
R 4.4.0 - ~/ASSIGNMENT 4/  
Type 'demo()' for some demos, 'help()' for on-line help, or  
'help.start()' for an HTML browser interface to help.  
Type 'q()' to quit R.  
  
> source("~/active-rstudio-document")  
Iteration 1 Centers:  
      A      B  
1 1.833333 2.333333  
2 4.125000 5.375000  
Iteration 1 Clusters:  
[1] 1 1 1 2 2 2  
Iteration 2 Centers:  
      A      B  
1 1.25 1.5  
2 3.90 5.1  
Iteration 2 Clusters:  
[1] 1 1 2 2 2 2  
>
```