CS 6375

ASSIGNMENT _____5_____

Names of students in your group:

SAI VIVEK KANAPARTHY (sxk163030)
HIMANSHU PARASHAR (hxp151330)

Number of free late days used: _____0_____

Note: You are allowed a **total** of 4 free late days for the **entire semester**. You can use at most 2 for each assignment. After that, there will be a penalty of 10% for each late day.

Part 1

We have implemented kMeans algorithm which iterates till the distance between centroids is not equal to zero or if number of iterations reaches 25 as per the termination condition given for this project

Output file is generated in the same folder.

A total of 5 runs with different with different k values. Below is the summary of 5 runs and their SSE values

Please note :- we are writing the summary of last iteration and attaching the output file along with it

Detail of Run 1 with k = 3

Iteration: 14
Iteration: 14
Cluster: K_0
Points: 5,7,9,15,23,30,31,35,43,44,45,46,49,62,66,67,76,79,81,86,87,90,
Cluster: K_1
Points:3,4,11,14,16,18,21,22,25,26,28,32,33,36,37,40,42,48,50,52,56,59,60,64,68,69,70,71,72,74,80,82,83,84,91,92,95,96,100,
Cluster: K_2
Points:1,2,6,8,10,12,13,17,19,20,24,27,29,34,38,39,41,47,51,53,54,55,57,58,61,63,65,73,75,77,78,85,88,89,93,94,97,98,99,
SSE : 1.864443021303258


output.txt

Detail of Run 2 with k = 5

Iteration: 25
Cluster: K_0
Points: 4,32,37,38,52,56,59,60,64,
Cluster: K_1
Points: 11,22,28,36,48,68,70,71,72,82,83,92,95,96,
Cluster: K_2
Points: 5,7,9,15,23,30,31,35,43,44,45,46,47,49,62,66,67,76,79,81,86,87,90,
Cluster: K_3
Points:1,2,6,8,10,12,13,17,19,20,24,27,29,34,39,41,51,53,54,55,57,58,61,63,65,73,75,77,78,85,88,89,93,94,97,98,99,

Cluster: K_4
Points: 3,14,16,18,21,25,26,33,40,42,50,69,74,80,84,91,100,
SSE : 1.37595255239899

output.txt

Detail of Run 3 with k = 7

Iteration: 25
Cluster: K_0
Points: 3,4,11,26,32,52,60,64,70,
Cluster: K_1
Points: 6,12,13,16,25,38,56,59,97,
Cluster: K_2
Points: 1,2,8,17,24,51,53,57,61,65,94,98,99,
Cluster: K_3
Points: 5,7,9,15,23,30,31,35,43,44,45,46,47,49,62,66,67,76,79,81,86,87,90,
Cluster: K_4
Points: 22,28,36,37,48,68,71,72,82,83,92,95,96,
Cluster: K_5
Points: 14,18,21,33,40,42,50,69,74,80,84,91,100,
Cluster: K_6
Points: 10,19,20,27,29,34,39,41,54,55,58,63,73,75,77,78,85,88,89,93,
SSE : 0.9806665819377991

output.txt

Detail of Run 4 with k = 2

Iteration: 13

Cluster: K_0
Points:
1,2,5,6,7,8,9,10,12,13,15,17,19,20,23,24,27,29,30,31,34,35,39,41,43,45,46,47,49,51,53,54,55,57,58,61,62,63,65,66,67,73,75,76,77,78,79,81,85,86,87,88,89,90,93,94,97,98,99,
Cluster: K_1
Points:
3,4,11,14,16,18,21,22,25,26,28,32,33,36,37,38,40,42,44,48,50,52,56,59,60,64,68,69,70,71,72,74,80,82,83,84,91,92,95,96,100,

SSE : 7.464315226724138

output.txt

Detail of Run 5 with k = 9

Iteration: 25
Cluster: K_0
Points: 6,12,13,20,24,38,41,65,75,77,93,97,
Cluster: K_1
Points: 11,22,28,37,48,68,70,71,72,83,92,95,96,
Cluster: K_2
Points: 36,82,
Cluster: K_3
Points: 5,7,35,43,62,67,87,
Cluster: K_4
Points: 2,10,19,27,29,34,39,51,53,54,55,57,58,63,73,78,85,88,89,
Cluster: K_5
Points: 3,14,16,18,21,25,26,33,40,42,50,69,74,80,84,91,100,
Cluster: K_6
Points: 4,32,52,56,59,60,64,
Cluster: K_7
Points: 1,8,17,31,61,79,94,98,99,
Cluster: K_8
Points: 9,15,23,30,44,45,46,47,49,66,76,81,86,90,
SSE : 0.6989832455808079

output.txt

Analysis –

After running the kmeans algorithm for different k values we can see the SSE is reducing when we increase the k value since centroids and clusters increases but it also increases the number of iterations and hence the running time. For large data we can run this algorithm with k= 3, By looking as our experiments