

Problem Statement: Software Development Engineer I

Write a simple cpp code implementing a k-means clustering algorithm for given sets of points. The algorithm should be a simple kmeans implementation without going into fancy kd-tree speedup or nearest neighbor approximations.

Kindly refer to Andrew NG course on coursera in case you don't know the algorithm.

The code should be able to take in a text file specifying the data points in D dimensional space along with number of clusters demanded and should write out the centroid of the clusters in an output text file.

Example format of text file will be (for 3 dimension) - points.txt

0.1 0.4 0.2

0.6 0.1 0.5

0.4 1.6 0.9

1.2 3.4 6.5

9.2 9.6 8.6

Your code should infer the dimension as well as the number of data points given. Also it should take in number of cluster point that the user wants.

Output of the cluster centroids should be (for k=2 clusters) - clusters.txt

0.91 0.4 0.2

0.76 0.34 1.2

As part of the submission you need to give your original cpp implementation along with compilation and running instruction which should compile at our end. We will run it at our end with the following command

`./prog points.txt 2`

where prog is your binary file, points.txt is a file with aforementioned format and 2 is the number of clusters.