Design and Analysis of Algorithms
10.4 Clustering

Daniel Shannon

June 10th, 2022

10.4.2

Design a local search algorithm for finding clusters.

You could pick k random points in the data. Then find the scores off all the other points for each cluster k. If we imagine scores of clusters as a topology, as we get further away from the selected the score will fall, so maybe we can take the topologies of each random cluster and see where the "level" point is. There would be a rim of all the points that intersect and who's scores even out. Then we can start adjusting the random points until the rim of each cluster encompasses the right number of points.

10.4.4

- Do the results you get from running the k-means algorithm depend on your choice of starting point?

- Either come up with an explanation for why not, or come up with an example.

Intuitively, yes, the results of the k-means algorithm does depend on the starting point. Given a set of N points and k clusters, say we choose k starting points that are all very close to eachother. The clusters will move away from eachother as we iterate through the algorithm. What if we have an outlier set of points and we choose one of our starting points in that outlier, will it move away from the outlier? If we have the same set of points with outliers and we don't choose a starting point in the outlier group, I could see how the outlier might be included in a cluster, but wouldn't be the cluster itself.