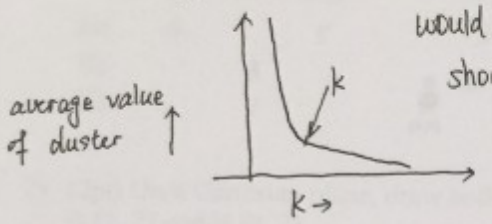
Name: _____ USC ID: _____

2. (2 pts) Discuss effective ways to determine an appropriate value of k to use in **k-means clustering**.

① Because the measure of the clustering quality like the average diameter... would be changed very rapidly at the begining as the graph shows. and then with the increasing of k, the trend ~~will~~ would become slow. So we can choose that point before ~~that would not change so much~~ the change is stable.



② Also we can use a sample of data points to first determine the number of k.

2

3. (5 pts) Are each of the following True or False?
   1) K-means clustering is more computationally efficient than hierarchical clustering. T
   2) The dendrogram is read from right to left. F
   3) Clustering should be done on samples of 300 or more. F
   4) In cluster analysis, objects with larger distances between them are more similar to each other than are those at smaller distances. F
   5) The centroid method is a variance method of hierarchical clustering in which the distance between two clusters is the distance between their centroids (means for all the variables). T
   6) Nonhierarchical clustering is faster than hierarchical methods. T
   7) One method of assessing reliability and validity of clustering is to use different methods of clustering and compare the results. T
   8) To reduce the number of variables, a large set of variables can often be replaced by the set of cluster components. T
   9) The most commonly used measure of similarity is the Euclidean distance or its square. T
   10) Non-hierarchical clustering is a clustering procedure characterized by the development of a tree-like structure. F

5