HW 1: K-means Clustering

due: 9/16/03

Overview

The purpose of this assignment is to familiarize you with some basic Matlab functions in an exercise in coding the K-means clustering algorithm. K-means clustering partitions a dataset of any number of feature dimensions into K clusters. The ’memory’ of K-means system is the K centers that evolve over the iterations of the algorithm.

The stopping condition can be anything that works for the problem. One reasonable stopping condition would be the iteration at which the net change in center locations is less than some threshold. For this assignment, the centers should be stable after 20 iterations and this can be used as a simplified stopping condition.

Assignment

Use the dataset at www.cns.bu.edu/~timmck/cn550/hw1/kmdata.mat. kmdata contains 400 samples with two feature dimensions.

Create 5 clusters of the data. Repeat the algorithm for 8 trials, from different random starting centers, recording the ’memory’ created for each trial.

Create a plot illustrating the the clusters formed and the centers created.

Suggestions

Use Functions Implement the algorithm as a function that returns the centers and the cluster label for each sample point. Input parameters
could be the data and number of clusters desired. \([\text{centers, labels}] = \text{kmeans(data, K)}\) Another way is to use structs to move data in and out of a function. ex:

\[
\begin{align*}
%\text{classifier parameters} \\
\text{param.K} & = 5 ; \quad \text{param.iterations} = 20 ; \\
\text{res} & = \text{kmeans(data,param)} ; \\
%\text{returns: res.centers res.labels res.param}
\end{align*}
\]

Kmeans.m is then a tool that can be used whenever needed (like in the KNN or Radial Basis Function assignment in a few weeks.) Create a script that opens the datafile and calls the function.

**Useful Code Snippets**

- one way to compare a (sample) vector with an array (of centers) is to reproduce the vector into each of the K columns of the array \(s = (\text{ones}(K,1) * \text{sample})';\);
- to find the index(es) of the minimum of a vector \(\text{minidx} = \text{find}(d == \text{min}(d));\);

![Fig. 1: Evolution of centers using K-means clustering over 40 iterations](image-url)