CS504 - Assignment 4

(k means)

REWARD: 200 points

DUE: Thu Mar 30 at 5PM PT

TEST DATA: testDataA4.tar

the task

Write code to perform k-means analysis. The first input is the number of expected groups called k. The second input number is the number of times the **center points** of each cluster or group is computed. This is followed by a matrix of **input points**, one per row in standard input format of: first the number of rows, then the number of columns then the data. Note that the input points may be in high dimensional space, not just two dimensions.

Your program should then run the k-means algorithm to determine the k center points that are the centers of each group of input points. It should begin by surveying the input and initializing k random points such that each of the values in a given dimension lie between the min and max of the input points in that dimension. Part of the algorithm is to collect all the input points that are considered closest to point into a group and average the group to find the a new location for that center point. (See the algorithm.) It is possible that this process will find that for some center point the process finds no points in its group. In that case, reinitialize that center point randomly as it did at the beginning.

Run the regroup/average part of the algorithm the number of times specified to try to converge toward the center points for the clusters.. Print the center points at the beginning and after each averaging to create new center points.

submission

Tar up the `kmeans` code with a makefile to build the program `kmeans` that reads from stdin as described above. Remember it has to run on the test machine which is like wormulon so be sure to invoke python using the unix `env` tool: `#!/usr/bin/env python3` for python3 and `#!/usr/bin/env python` for python2.x. Submit a tar for the assignment. The tar can be the same or different depending on what you decide to do. Homework will be submitted as an **uncompressed** tar file to the homework submission page linked from the class web page. You can submit as many times as you like. **The LAST file you submit BEFORE the deadline will be the one graded.** For all submissions you will receive email at your uidaho.edu mail address giving you some automated feedback on the unpacking and compiling and running of code and possibly some other things that can be autotested.

Have fun.