CSE 6331 Cloud Computing
Summer 2015, © DL, UTA, 2015

Programming Assignment 8
Machine Learning (Cloud)
Due: July 3, 2015, 04:00 (UTA time)

Description:

1. On a cloud provider you will use a K-Means clustering tool to find clusters of data.
2. You will obtain the dataset from Blackboard and find 1 x 1 clusters.
3. K-means has been implemented in Python modules, Weka, and R (statistics), Java, and many others.
4. Your program will allow a user to specify number of clusters (and optionally datasets, and attributes)
5. You will display centroids and number of points in clusters, and other metrics (distances between clusters, etc.)
6. Please submit your results (text, html, or doc file, a few pages), and all code and Configuration files, but NOT executables, binaries or raw data files.

Regardless of the number of people in your "team", you will need to understand the implementation and results of this assignment.

You may work in groups of up to 3 people. All referenced information, including sources of data must be cited and referenced in your write-ups and in each relevant section of your code.

Please, Email ONLY to the class account. All work must be your own.
You must e-mail this lab, working (or partially) by the due date. The e-mail subject should clearly state the lab number.
You may (optionally) demonstrate this lab, working (or partially) to the GTA before the due date.
Your program should be well commented and documented, make sure the first few lines of your program contain your name, this course number, and the lab number.
Your comments should reflect your design and issues in your implementation.
Your design and implementation should address error conditions.

Please, Email ONLY to the class account (CloudAtUTA@gmail.com).
All work must be your own, you may reference web sites, books, or my code but You MUST site the references.