**CPSC 483 PROJECT 2 100 points**

**DUE DATE: Wednesday, Dec 11, 11:59 PM**

**GROUPS ARE ALLOWED (MAXIMUM OF 2 (TWO) STUDENTS PER GROUP). EACH STUDENT IN THE GROUP WILL GET THE SAME POINTS.**

**YOU CAN ALSO CHOOSE TO DO THE PROJECT INDIVIDUALLY.**

**WORK SHOULD BE YOUR OWN AND NOT COPIED FROM THE INTERNET OR ANYWHERE ELSE.**

PACKAGES LIKE sklearn.cluster.KMeans (python), kmeans (R) or any other such package **ARE NOT TO BE USED.**

Helper packages like numpy, math module or any other such package or module for data storage and manipulation and facilitating min/max operations, etc., **ARE TOTALLY ALLOWED.**

**The idea is to CODE ALL PARTS of KMeans algorithm yourself.**

**The number of clusters, K = 3. The dataset is uploaded in project directory. The initial centroids are ideally random, but can taken to be [3,3], [6,2], [8,5] for this project. No need to calculate and subsequently minimize the cost. Simply run the KMeans algorithm 10 times, and consider it converged at that point.**

1. **The initial point distribution should be depicted in the 2-D space, alongside 3 initial centroid points. The latter should be in 3 different colors and specific markers indicating they are centroids and not points. The dataset points should be in any 4th color. So essentially 2 markers and 4 colors in total. 1 plot submission.**
2. **Subsequent to the above step, once the points are assigned to its closest centroid, color each point to its assigned centroid color. The centroid should still show in same color, but different marker. So now essentially 2 markers and 3 colors in total. 1 plot submission.**
3. **Then only changing centroids should be printed after each iteration. No plots.**
4. **Show the final point distribution, with each point colored according to its final centroid. So essentially 2 markers and 3 colors in total. 1 plot submission.**

**Deliverables:**

1. **Kmeans code**
2. **3 plot figures (Output from running the code)**
3. **Changing centroids at each iteration. (Output printed from running the code)**

Submit item 1), 2) and 3) above, in Titanium by **Wednesday, Dec 11, 11:59 PM**.

**KMeans algorithm is available in slides**