# Habib University
# CSE 351 - Artificial Intelligence
# Fall' 2021
# Assignment 3
50 Points

## Objective:

The objective of this assignment is to make student better understand of clustering and collaborative filtering techniques. Students will gain hands-on experience of clustering data using KMeans and of using gradient descent to perform matrix factorization.

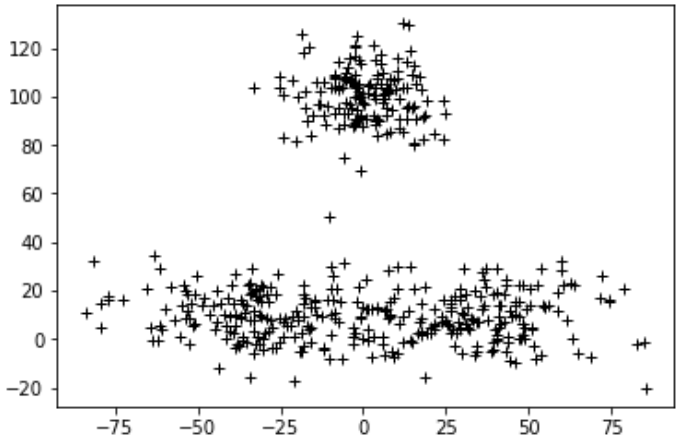
---

### Q 1 – Clustering [25 points]

You have to implement k-means algorithm to cluster a set of 1000 2D points. The data points (x,y) will be generated via Gaussian distribution using the sample code below:

```
points = []
def initializePoints(count):
    for i in range(int(count/3)):
        points.append([random.gauss(0,10),random.gauss(100,10)])
    for i in range(int(count/3)):
        points.append([random.gauss(-30,20),random.gauss(10,10)])
    for i in range(int(count/3)):
        points.append([random.gauss(30,20),random.gauss(10,10)])
```

This code generates following data points:




a) **[10 Point]** Implement K-means algorithm to cluster these points. K-means will take a set of points and number of clusters (K) to be formed and will cluster the points into given number of clusters. The algorithm will stop if there is no significant change coming in positions of centroids.

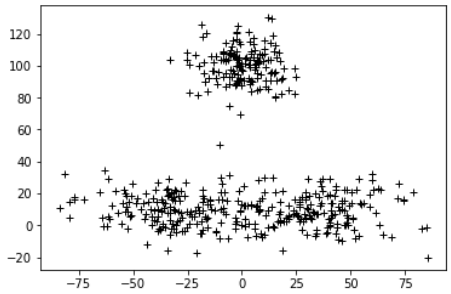
b) **[04 Points]** Provide graphical visualization of the process of formation of clusters (as shown on the next page).

c) **[06 Points]**Instead of running K-means once, run it 10 times and give the best clustering formation achieved. You need to identify the criteria to assess quality of clusters and then choose the best one.

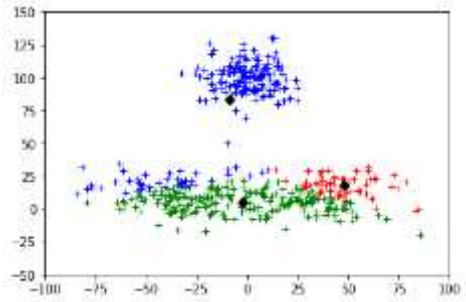
d) **[05 Points]** You will give different mean and variance to the Gaussian distribution to generate different input points and cluster them using your code. The clustering formation (interim and final) for each set of points will be shown in a pdf file.

The code will be written in the attached Q1_clustering.py file. The process of clustering will be shown as follows:

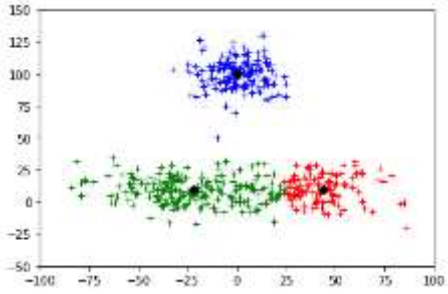
**Iteration 0**




**Iteration 1**




**Iteration 2**

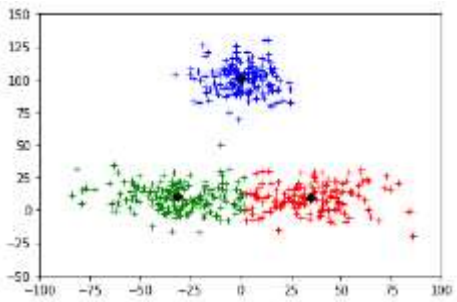



.

.

.

**Iteration 9**



## Q2 – Collaborative Filtering using Matrix Factorization

You are implementing collaborative filtering to make recommendations to users U for items I. You decided to use a model based approach that applies matrix factorization to factorize a rating matrix ($R$) into User features ($P$) and Item features ($Q$) such that,

$$P \times Q = \hat{R} \cong R$$

where P is a $m \times k$ matrix and Q is $k \times n$ matrix and $m$, $n$ and $k$ represent no. of users, no. of items and no. of latent factors respectively.

The predicted rating of an item $j$ by user $i$ is calculated as follows:

$$\hat{r}_{ij} = \sum_{k} p_{ik} q_{kj}$$

and error in this prediction is calculated as:

$$e_{ij}^2 = (r_{ij} - \sum_{k} p_{ik} q_{kj})^2$$

### Q2.1 – Computing gradients [5 points]

You are applying gradient descent to minimize mean square error. Compute gradient of error with respect to p and q values and derive formulas to update p and q values. The calculations of p would be:

$$\Delta p_{ij} = \frac{\partial e_{ij}^2}{\partial p_{ij}}$$

$$p'_{ij} = p_{ij} - \alpha \, \Delta p_{ij}$$

Where $p'$ is updated value of p and $\alpha$ is learning rate.

### Q2.2 - Adding biases [5 points]

There can be biases in user and item recommendations, which are handled by introducing user bias vector (bU) and item bias vector (bI). In the presence of these biases, rating $r_{ij}$ is predicted as follows:

$$\hat{r}_{ij} = bu_i + bi_j + \sum_k p_{ik} q_{kj}$$

Derive formulas to update p, q, and biases during gradient descent.

### Q 2.3 – Factorizing matrices using gradient descent [15 points]

In this question, you will implement gradient descent technique to perform matrix factorization. Given a rating matrix, your code will apply matrix factorization to determine User features (P), Item features (Q), User bias vector (bu) and Item bias vector (bu). The code will be written in the attached python file (Ass3.py) which provides basic skeleton of your program.

---

## Submission Instructions

The assignment will be done in pairs. Submissions will be made on Canvas by the due date (announced on Canvas). The submitted file should be in the form of a ZIP file named as **<studentid1>_<studentid2>_Ass3** containing a pdf document for Q1 (part d) and python code for both Q1 and Q2.