COMPUTING SUBJECT: Machine Learning

TYPE: WORK ASSIGNMENT

IDENTIFICATION: Chapter 9 Unsupervised Learning

COPYRIGHT: *Michael Claudius*

DEGREE OF DIFFICULTY: Easy

TIME CONSUMPTION: 1 hours

EXTENT: < 100 lines

OBJECTIVE: Basic understanding of clustering.
                              K-Means, K-Means++

COMMANDS:

IDENTIFICATION: Chapter 9 Unsupervised Learning/MICL

The Mission
To understand the idea behind Clustering and some of the clustering algorithms.

Precondition
You must have read the second part of Training Models in chapter 4 and SVM in chapter 6.

Useful links

https://sklearn.org/modules/clustering.html#clustering
https://sklearn.org/modules/neighbors.html#unsupervised-nearest-neighbors
https://thomasdelatte.com/2020/04/kmeans/
https://towardsdatascience.com/k-means-clustering-with-scikit-learn-6b47a369a83c

Theoretical Assignment on clustering

1. What is clustering about ?

2. Give few examples from real life where clustering should/could be used.

3. Name a few clustering algorithms besides K-Means.

4. What is the idea behind the classic K-Means algorithms ?

5. Is it important to scale? When and Why ?

6. Initizlization of centroids is a problem in classic K-means. Why?
   What can one do instead of just one random selection of centroids?
   - number of random initializations
   - Position of centroids distant from each others K-Means++
   Explain!

7. What is inertia ?

8. What is the idea behind MiniBatchKMeans?
   State the advantage and disadvantage behind using mini-batches.

9. Optimal number of centroids can be be found by "elbow" techniques.
   How does it work?

10. Optimal number of centroids can be be found by "silhoutte" techniques.
    How does it work?

11. What is/are good alternative(s) to K-Means and when to use it ?

Congratulation: You are now an beginners-level 1 in Unsupervised Clustering ☺