Plan du travail :

Jeu de données : (il faut s’assurer qu'il y a “simulé” et “réel”)

* Pen Digits : <http://archive.ics.uci.edu/ml/machine-learning-databases/pendigits/>
* Tic Tac Toe : <http://archive.ics.uci.edu/ml/datasets/Tic-Tac-Toe+Endgame>
* ~~Bike sharing :~~ [~~http://archive.ics.uci.edu/ml/datasets/Bike+Sharing+Dataset~~](http://archive.ics.uci.edu/ml/datasets/Bike+Sharing+Dataset) ~~(Regression)~~
* ~~Metro :~~ [~~http://archive.ics.uci.edu/ml/datasets/Metro+Interstate+Traffic+Volume~~](http://archive.ics.uci.edu/ml/datasets/Metro+Interstate+Traffic+Volume) ~~(Regression)~~

Référence :

* <https://smorbieu.gitlab.io/fr/mod%C3%A8le-m%C3%A9lange-gaussien-k-moyennes.html#:~:text=L'algorithme%20CEM%20appliqu%C3%A9%20pour,'algorithme%20des%20k%2Dmoyennes.&text=Nous%20obtenons%20les%20m%C3%AAmes%20r%C3%A9sultats%20qu'avec%20les%20k%2Dmoyennes>.

Theoretical Introduction :

Plan de travail : (Ayman) (Aymen MB)

* Lecture des données (data preprocessing)
* Implémentation (algorithme EM)
* Implémentation (algorithme CEM)
* Comparaison avec k-means
  + Pour les données simulées, on comparera tout d’abord les deux algorithmes lorsque les classes sont très séparées (err ~ 0:05 c.à.d. chevauchement des classes de 5%) ; pour cela, on étudiera des situations correspondant à des proportions égales, différentes ou très différentes et des situations correspondant à des volumes égaux, différents ou très différents. On fera les mêmes comparaisons lorsque les classes sont moyennement séparées (err ~ 0:15) et très séparées (err ~ 0:30).
* Comment peut-on utiliser les modèles de mélange pour faire de l’apprentissage supervisé (classement) ? Illustrer votre réponse par une application de votre choix.
* Proposer ainsi une autre méthode d’apprentissage supervisé, la Présenter (avec synthèse) et la comparer avec les modèles de mélange.