**In-Class Exercise – “EM Algorithm”**

Due Day: YYYY/MM/DD, dddd, HH:MM

objective: To practice how EM Algorithm works

Explanations:

1. Given 2 datasets (test and real data), please analyze them using the EM Algorithm.
   * Simplified datasets using test\_case\_1a.csv
   * Real datasets using raw\_real\_data.csv
2. Use any Python IDE (we could use Google Colab) to **run the code.** Then **add screenshots** to report in Word files.
3. The report should contain:
   * Screenshots of the modified code and the result (simplified and real dataset) from point 1 and 2 above
   * The answer to these questions:
     + Explain the reason why we use the “EM algorithm” to estimate machine yield rate?
     + After the manufacturer has the results of machine yield rate estimation, what can they do using that result?
     + What do you learn from this lab?
   * Any questions and/or difficulties in following the EM Algorithm topic?
4. Compress (\*.zip) the following file with the name of the group (e.g. EM\_Labs\_Group\_1.zip)
   * Code (\*.py, \*ipynb, or another python file extension)
   * Result (test\_report.csv, and realworld\_report.csv)
   * Report(\*.doc)

Note:

* Each group only require one report
* Upload the file on e-Learning site to complete. Make sure to upload before the deadline (see in lms)
* To update the contents of the report, can be directly re-upload file with the name of version (e.g. EM\_Labs\_Group\_1\_v2.zip)