**Exercise #1: Hierarchical clustering and Gaussian mixture model**

You are required to implement a designated data mining algorithm either in C++, Java, or Python. The implementation must be from *scratch*, i.e., it can use only functions from a standard library and the numpy package. Note that you are required to implement the models by yourself. Directly calling functions from any package is not allowed. For SVM, you can call functions from other packages to solve the primal/dual problem.

* C++: [MSDN library](http://msdn.microsoft.com/en-us/library/cscc687y(v=vs.80).aspx) or [GNU Library](http://www.gnu.org/software/libc/).
* Java: [Standard Edition 6](http://docs.oracle.com/javase/6/docs/api/).
* Python: [Check here](https://docs.python.org/3/library/)

**Goal**

Implement the Hierarchical clustering and the GMM methods. For GMM, you need to try different *k* and discuss how to choose the optimal one. Do some visualizations to compare the clustering results obtained by the two algorithms.

**Dataset (in folder ‘data2’)**

Download [here](https://www.cse.cuhk.edu.hk/~taoyf/course/cmsc5724/data/8gau.txt) (obtained from the data collection [here](http://cs.joensuu.fi/sipu/datasets/)). Each line has the following format:  
  
*x* *y*  
  
which represent the x- and y-coordinates of a point.

**Task**

Partition the dataset. You should try different number of clusters by yourself.

## Exercise#2: Collaborative Filtering

* **简介**：电影评价数据集，该数据集中包含600位用户对9000部电影的100,000条评分和3,600条标签。评价以5星制评分为基准，以半星递增（0.5星~5.0星）。
* **数据集介绍**：
  + [官网介绍](https://grouplens.org/datasets/movielens/latest/)
  + [README](https://files.grouplens.org/datasets/movielens/ml-latest-small-README.html)
* **下载地址**：
  + [官方下载地址](https://files.grouplens.org/datasets/movielens/ml-latest-small.zip)
* **GitHub**：可以参考一些开源项目作为实现协同过滤算法的例子。
  + 示例项目：[Surprise库示例](https://github.com/NicolasHug/Surprise) 或者 [其他推荐系统项目](https://github.com/search?q=movie+recommendation+system+movielens)

**目标**

* 实现经典的协同过滤算法。可以使用现有的软件包，如Surprise、scikit-learn等。
* 使用**k折交叉验证**方法验证算法有效性，**k=10**。
* 评价指标可包括但不限于**RMSE、MAE**等。

**Submission:**

* A zip package of your code with a readme file for execution instructions.
* A pdf report contains up to 1300 words.
* Please further zip your code package and the pdf report into a single file with the name “xxx-id-xxx-id”. Example: “王小明-22220-赵晓婷-22336”. Send your work to 张信宇（通过邮件：zhangxy869@mail2.sysu.edu.cn）。
* Deadline: Dec. 30th.