**G-Means Clustering**

**Relevant C source code files:**

gmeans.c - Primary file for G-Means implementation containing main() function.

data.c, cluster.c, command.c - Relevant source code files from 'The C Clustering Library'.

seq\_kmeans.c - Relevant source code files from 'CUDA K-Means Clustering' library.

**Steps to execute programme: (Windows Systems)**

**Setup C environment:**

We need to have MinGW installed and integrated in Windows.

Please follow this link to setup MinGW : [Setup MinGW](http://www.multigesture.net/articles/how-to-install-mingw-msys-and-eclipse-on-windows/)

**Setup Project:**

Clone the repository using git (ensure [git](https://git-scm.com/book/en/v2/Getting-Started-Installing-Git) is installed on the system)

git clone https://github.ncsu.edu/sjgurav/G-Means.git

Go to the cloned repository.

cd G-Means

**Compile and Run Program:**

Use **git bash** command Prompt to execute the following commands.

Compile the necessary files:

gcc -std=c99 -c gmeans.c seq\_kmeans.c data.c command.c cluster.c

Link the object files to create an executable file called mygmeans

gcc -o mygmeans gmeans.o seq\_kmeans.o data.o command.o cluster.o -lm

Run the executable file (output.csv file contains the resultant cluster assignments):

./mygmeans.exe filepath/filename

**Example:**

./mygmeans.exe data/gmean3.csv

./mygmeans.exe data/gmean4.csv

./mygmeans.exe data/ecoli.csv

**Note:**

Please note that all files are in data folder, hence filepath is 'data'.

Make sure filepath is relative to the current folder in case if there is any. If file is directly in current folder, there is no need of filepath.

For above command filenames is:

gmean2.csv, gmean3.csv, gmean4.csv, gmean5.csv, ecoli.csv

**To plot the graphs to display generated clusters for synthetic and real-world data, run the following command:**

Rscript filename

(requires path to bin folder for R installation to be included in Windows PATH)

**Where the filename is either plot\_gmean5.R or plot\_ecoli\_gmeans.R, for generating plots for gmean5.csv and ecoli.csv datasets respectively.**

Above command will generate Rplots.pdf file with true cluster and G-Means cluster visualizations for the dataset.

Note: Running any R file for visualizing clusters requires mygmeans executable file for the corresponding dataset to be run right before it.