Quick Start

Clément Viricel

September 16, 2017

1 Input File

Let's take a look at the file which the program eat. It has the following format:

```
Nb_of_point Nb_of_dimension Nb_of_Clusters Nb_of_Iteration HasName
x1 x2 ... (Name)
x1 x2 ... (Name)
x1 x2 ... (Name)
.
.
.
```

For example, if you are trying to run the program for 1000 iteration over one hundred $\mathbb{R}^3$ coordinates in 4 clusters the file will look like:

```
100 3 4 1000 0
x y z
x y z
.
.
.
```

It's an easy and common way to represent the data but the parser can be changed if needed (with more work).
Adding a name to the point can allow user to see if there is a pattern with its data. For exemple, one can cluster 1000 measures of any dimension of N different types of any kind and clusterize it into N clusters to see if the measure are strongly correlated to the type.

2 Output File

The output file produce by the program is as simple as the input file. It will write in "*Results.txt*" each cluster with its points and its central value (that correspond to te barycenter of the cluster, aka central value, aka representative value :

```
Cluster i
Point p : x1 x2 ... (Name)
.
.
Central point
```

3 The program

To compile the program please enter the following command:

```
g++ src/*.cpp -o kmeans
```

To run the program, it's really easy, just enter:

```
./kmeans [input_file]
```

Enjoy !