Aleksi Artturi Korkee

ID:12180067

Programming language: **C++**

Program can be run from **kmc.exe** in folder **hw2/x64/Release/**.

Source code for the **kmc.exe** can be found in folder **hw2/kmc/**.

Header files: **Data.h, FileManagement.h**

Source files: **kmc.cpp, Data.cpp, FileManagement.cpp**

Running the program:

* Run **kmc.exe**
* Program asks if you want to assign starting centers randomly or choose them yourself.
* Then input the k.
* If you chose random it will calculate the centers randomly.
* If you chose to pick the centers yourself, the program asks you to input k number of gene indexes to get the centers from. The ribosomal genes are indexed here from 2347 to 2467.
* Program will output how many ribosomal genes are in each of the clusters.
* Program then asks if you want to run it again.

Pseudocode:

**Function names are not the same in the source code. Here the function names just tell roughly what is happening in the real code.**

**kmc.cpp:**

**main()**{

Data data = readInputFiles();

isRandom = randomOrChooseCenters();

k = whatIsK();

if(isRandom){

centers = chooseRandomCenters();

}

else{

centers = userChoosesCenters();

}

clusters = runKMeansClustering();

outputResults();

}

**kmc()**{

for(50 times){

for(each gene){

minDistance = min(calculateDistanceToEachCenter());

assignGeneToClosestCluster();

newCenter += geneExpressionVector;

if(gene == lastGene){

newCenter = newCenter/numberOfGenes;

}

}

If(distanceBetweenClusters == oldDistanceBetweenClusters){

Break; //no change in clusters

}

}

Return clusters;

}

**randomCenters()**{

for(k times){

randomNumber = GenerateRandomNumberBetween(1,2467);

center = getGeneExpressionVector(randomNumber);

centers.add (center)

}

Return centers;

}

**chooseCenters()**{

for(k times){

index = inputNumber();

center = getGeneExpressionVector(randomNumber);

centers.add (center)

}

Return centers;

}

**Data.cpp:**

**Data::euclideanDistance()**{

sum = (a(0)-b(0))2+(a(1)-b(1))2+…+(a(78)-b(78))2;

Return sqrt(sum);

}