**CSE3315 Assignment 2  
Classification and Clustering for the Microarray Data Analysis**

**README**

12191656 이채연

**Programming language**

C++ programming language를 사용하여 K-nearest neighbor (KNN) classification과 K-means clustering(KMC)를 구현하였다.

**How to run the program**

**1) K-nearest neighbor (KNN) classification**

K-nearest neighbor (KNN) classification 프로그램을 구현한 source code의 이름은 knn.cpp이고, executable file의 이름은 knn이다.

knn executable file을 실행하기 위해서 먼저 knn executable file이 실행될 위치에 ribo-data.txt 와 nonribo-data.txt파일이 있어야 한다. 그리고 터미널을 열어 knn executable file이 실행될 위치로 이동 후, ‘./knn’ 을 입력하면 knn executable file이 실행된다.

K-nearest neighbor (KNN) classification을 수행하는 이 프로그램은 사용자로부터 parameter K와 parameter p를 입력 받는다.

Parameter K와 parameter p를 입력하면 터미널 창에 output이 출력되며, knn executable file을 실행한 위치에 있는 knn.out file에도 output이 출력된다.

텍스트이(가) 표시된 사진

자동 생성된 설명

텍스트이(가) 표시된 사진

자동 생성된 설명

**2) K-means clustering (KMC)**

K-means clustering (KMC) 프로그램을 구현한 source code의 이름은 kmc.cpp이고, executable file의 이름은 kmc이다.

kmc executable file을 실행하기 위해서 먼저 kmc executable file이 실행될 위치에 ribo-data.txt 와 nonribo-data.txt 파일이 있어야 한다. 그리고 터미널을 열어 kmc executable file이 실행될 위치로 이동 후, ‘./kmc’ 을 입력하면 kmc executable file이 실행된다.

K-means clustering (KMC)을 수행하는 이 프로그램은 사용자로부터 cluster의 개수를 의미하는parameter K와 starting center로 지정할 data point 번호를 입력받는다.

* ribo.txt와 nonribo.txt에서 first data point를 starting center로 지정하고 싶은 경우

첫 번째 data point이므로 1을 입력하면 된다.

텍스트이(가) 표시된 사진

자동 생성된 설명

* ribo.txt와 nonribo.txt에서 random data point를 starting center로 지정하고 싶은 경우

0이 random data point를 의미하므로 0을 입력하면 된다.

텍스트이(가) 표시된 사진

자동 생성된 설명

parameter K와 starting center로 지정할 data point 번호를 입력하면 다음 정보들이 터미널 창에 출

력된다.

* ribosomal gene cluster에 있는 ribosomal gene들의 비율
* non-ribosomal gene cluster에 있는 ribosomal gene들의 비율
* non-ribosomal gene으로 잘못 clustering된 ribosomal gene number