Aleksi Artturi Korkee

ID:12180067

Programming language: **C++**

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

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

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

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

-Source files(.cpp and .h) can be opened in any text editor. I used visual studio to create and edit the code.

Running the program:

* Run **knn.exe**
* Program asks the user to input k.
* Then it asks to input p.
* Program will run 6 fold cross validation.
* Program will print results and save them in **knn.out** file in the same folder as knn.exe.
* The program will also save a file in folder **hw2/x64/Release/knn\_out** during each cross validation containing indexes of nonribosomal genes that are misclassified as ribosomal.

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.**

**knn.cpp:**

**crossValidation()**{

NonRibosomalToRandomOrder();

RibosomalToRandomOrder();

DivideRiboToGroups();

DivideNonRiboToGroups();

For(6 times){

test\_set = pickOneOfTheGroups();

train\_set = pickTheRestOfTheGroups();

result = knn(test\_set, train\_set);

results.add(result);

}

calculateAccuracy();

calculateSensitivity();

calsculateSpecificity();

outputResults();

}

**knn()**{

for(each test\_data\_point){

for(each train\_data\_point){

distance = distanceBetween(train\_data\_point, test\_data\_point);

if(distance < kSmallestDistances.largestDistance()){

kSmallestDistances.add(distance);

kSmallestDistance.remove(largestDistance);

}

}

ribosomals = numberOfRibosomal(kSmallestDistances)

predicted\_class = ribosomals/k;

result.addPredictedVSTruth(predicted\_class, true\_class); //TP, TN, FP, FN

return result;

}