**COMP1011 Project Topic 2 Learning Linear Regression**

# **Problem statements:**

How to introduce concepts about Linear Regression for teaching purpose by programming?

# **Objective:**

1. Introduce concepts about the concept of linear regression to users.
2. Give out three example from beginner level to advanced level for illustration purpose.
3. Allows user to choose the sub-concepts to learn.
4. Input data (with guidelines by the program)
5. Visual representation of the calculation and the outcome.

# **Program design**

For Objective 1, I put all the content in the document “introduction.txt”, then we read file line by line and print on the terminal line by line in function introduction(). Also, I use the Sleep() function in windows.h to make sure user read every line in the interval time.

For Objective 2, we will have three different level of difficulties of dealing with datas. For the sake of uniformity, I put three set of data in three different txt file. And all data in txt file are in the form of matric. Then we need to read every file and processing data.

* Example 1: only three points for modelling Linear Regression.
* Example 2: more points for modelling Linear Regression.
* Example 3: same points as example 2 for Segmented Linear Regression.

***Example 1:***

*Before building an array for storing data from txt file, we should firstly calculate number of points, which is number of lines in the file. In the C++ program, it is calculate in the function getFileRows1().*

*文本, 信件

描述已自动生成As we know, for the Linear Regression Model, y = a x + b, a and b is calculated in this way.*

*文本

描述已自动生成We should first* *calculate the sum of x, sum of y, sum of (x \* y) and sum of (x\* x). Then we create an array data[rowNum][colNum] to store the coordinates of points from txt file. Then calculated by following equation.*

*文本

描述已自动生成But we did not have the final answer for the best-fit line, we still need to calculate the result by the function calc1Var(), which is the calculation for only one variables. Then we calculate the average of x and y by function average(). Then calculate a and b by the function A\_ba() and function B\_ba().屏幕上有字

描述已自动生成文本

描述已自动生成*

*These three function are all embedded in the function calc1Var(), then we can get the result for example 1.*

***Example 2:***

*Progressing data for example 2 is quite similar to example 1, so I will not repeat here again.*

***Example 3:***

*In example 3, the only difference is classifying all points into three groups with three central points, which are (100, 0), (50, 50), and (0, 100). So I will explain how I divide. Assume (0, 100) is point A, (50, 50) is point B, and (100, 0) is point C. Firstly, we can calculate the diatance between every data points and point A, B, C. So we will have three similar for loop. I create three array for distances to point A, B, C, which is distance\_a[rowNum], distance\_b[rowNum], and distance\_c[rowNum], put the newly calculated numbers into these three arrays.*

*Then we compare the same point to points A, B, C’s distances, and adding the points’ coordinates to the nearest point’s array(a\_array[rowNum][2], b\_array[rowNum][2], and c\_array[rowNum][2]) .*

*After classification of data points, we finally get into the calculation part, using the similar structure as example 1 for Cluster A, and use function calc1Var(), because there are a lot of data, so I stop 5 seconds allowing user to read clearly. Calculation process for Cluster B and Cluster C is similar.*

As for Objective 3, Allows user to choose the sub-concepts to learn. I also put all the content in the document “subconcept1.txt” and “subconcept2.txt”, then we read file line by line and print on the terminal line by line in function subcon1() and function subcon2(). I use the Sleep() function again like the way in the introduction part to make sure user read every line in the interval time. Then I use “if … else if … else” structure for user to choose sub-concepts they liked.

As for Objective 4, Input data (with guidelines by the program). In the function inputAndSum1(), I didn’t choose to store the data in the txt file and extract from file, it is more simple to just calculate from the first point. At first, ask user to enter number of point, and ask user for every coordinate and y-coordinate for every point using for loop. And then calculate the sum of x, sum of y, sum of (x \* y) and sum of (x\* x) again as mentioned in example 1. Then we use calc1Var() to calculate the final result.

For Objective 5, visual representation of calculation have already writen in the part of function calc1Var().文本

描述已自动生成

# **Structure of the program**

图示

描述已自动生成

**User manual**

1. Open the file main.cpp by Visual Studio Code, and directly click “run the code”, following the guideline provided by the program later.

2. If the program ask you to input any number, just think and knonk few of the key on the keyboard.

3. Read and study carefully following the program.

文本

描述已自动生成

# **Improvement**

1. Show the graph plot with the point and the best-fit line on the screen next time if I can draw the function using C++ language.
2. Layout of the printing style could be better.

# **Limitation**

According to my limited skill, I can only show the calculation process by simple “+” “-”,“\*” and “/” with a lot of complex calculation process instead of the accurate mathamatical notation on the screen for only one step.