CSE 102 Spring 2024 – Computer Programming Assignment 2

Due on March 13, 2024 at 23:59

In this assignment, you'll be tasked with implementing a basic artificial intelligence (AI) algorithm using the C programming language. The goal is to gain a practical understanding of basic programming concepts such as file I/O, loops, and selection statements. Please use the most suitable selection mechanism when needed; otherwise, your grade for the assignment may be reduced. Don't use any array or user-defined function in this homework.

Part 1. [40pts] Data Exploration

Your first task is to write a C program that reads a text file named "data.txt," which contains data points in the format: X_coordinate Y_coordinate Class_label. Each row represents a data point with its X and Y coordinates, followed by its binary class label (1 or 0). You need to determine the maximum and minimum limits of X and Y coordinates from the data points read from the file and print the results.

Example:

An short example is given in the following "data.txt" file.

```
X Y C
3 3 1
5 5 0
1 2 1
4 6 0
```

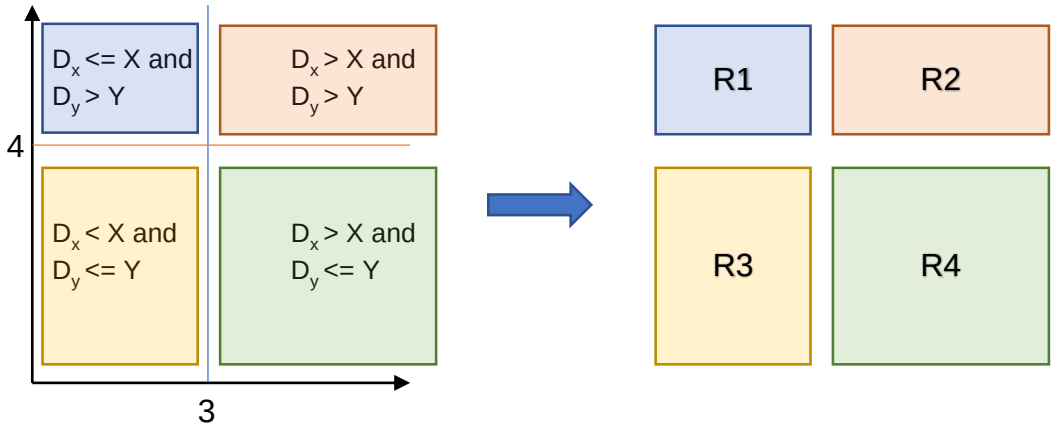
Your program should print:

```
Minimum and Maximum X and Y values:
Class 0: X in [1 3], Y in [2 3]
Class 1: X in [4 5], Y in [5 6]
```

Part 2. [30pts] AI Model Construction

You are asked to implement an AI algorithm using only loops and selection statements in C. The model decides whether a given data point belongs to class 1 or class 2. The model uses two separator coordinates, one for each dimension X and Y, to decide the class label of a given data point. The model compares the X and Y values of a given data point with the separator X and Y values and decides the label of this data point. Let's see the following examples for a better understanding:

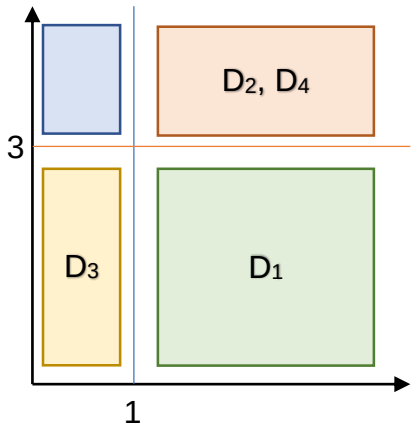
Separator X = 3, Y = 4 and data point coordinates are represented with D_x and D_y:



The separators divide the 2D space into R1, R2, R3, and R4 parts, as seen in the example. Each region is associated with one of the class labels (1 or 0). Thus, the model decides which class label is predicted for a given data point by comparing its X and Y coordinates with the separators. Your goal for training the model is to determine the best integer coordinates that lead to the minimum classification error. The classification error is calculated as the sum of misclassified points in the "data.txt" file by the separator coordinates. A point is misclassified if its predicted class label differs from its actual class label. Let's consider the data points given in question 1:

	X	Y	C
D1	3	3	1
D2	5	5	0
D3	1	2	1
D4	4	6	0

Example 1

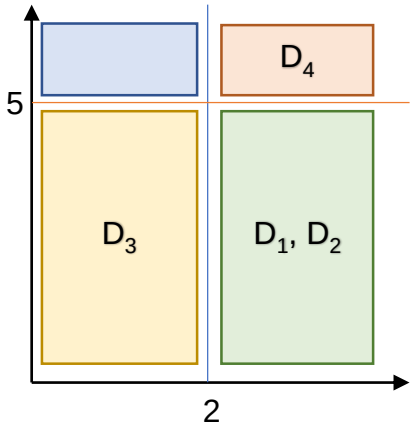


R2 = Class Label 0

R1, R3 and R4 = Class Label 1

Error = 0

Example 2



R2 = Class Label 0

R1, R3 and R4 = Class Label 1

Error = 1 (D_2 miss classified)

As in the examples, you are asked to find the separator coordinates X = 1 and Y = 3, which lead to the minimum error. You should consider each possible integer X and Y pairs between the minimum and maximum limits determined in question 1 to find the best coordinates for these

separators. Print the calculated separator coordinates with the corresponding error. Also, print the labels you assigned for each of the regions above (R1 – R4).

Example output:

```
Separator coordinates: X = 1, Y = 3
Training error of the model = 0
Labels of the regions: R1 = 1, R2 = 0, R3 = 1, R4 = 1
```

Part 3. [20pts] Testing and Evaluating

For the third question, you'll read a set of test data points from a file named "test.txt," following the same format as in the "data.txt" file. Using the AI model constructed in question 2, you'll predict the class labels for the test data points by using the determined X and Y separators. Print your predictions for the test data points with their X and Y coordinates. Calculate and print the success of your model using the following formula:

$$Success = \frac{Number\ of\ truly\ predicted\ class\ label\ for\ test\ data\ points}{Number\ of\ test\ data\ points}$$

In this formula, "$Number\ of\ truly\ predicted\ class\ label\ for\ test\ data\ points$" is calculated by counting the data points that have the same predicted and actual class labels.

Part 4. [10pts] Menu

Your task is to develop a program offering a menu that provides four options for exploring data, training a model, testing the model, and exiting the program.

1. Explore Data: Implement functionality to read and display the data points from the "data.txt" file, allowing the user to explore the dataset.
2. Train Model: Develop code to construct AI model using the data points read from the "data.txt" file. Use the algorithm implemented in Question 2.
3. Test Model: Create functionality to read test data points from the "test.txt" file, predict their class labels using the trained AI model, and display the results.
4. Exit: Provide an option for the user to exit the program.

Your menu should continuously prompt the user to choose an option until they select the "Exit" option. Ensure error handling for invalid inputs and appropriate messages for each menu option.

IMPORTANT NOTES:

- Submit your homework as a zip file named as your student id (StudentID.zip) and this file should include:
  - YourStudentID.c file
  - A pdf file named "YourStudentID.pdf" including a YouTube link and screenshots of your program outputs. In the video, you are expected to provide a demo of your assignment. For each requested functionality, you must explicitly explain your solution approach and

also execute and display the outputs. The video should not exceed 4 minutes. Please ensure that your camera is turned on during the recording.

- Do not use any library other than stdio.h.
- The output format must be as given, do not change it.
- Compile your work with given command “gcc --ansi your_program.c -o your_program”.
- Your work will be evaluated using gcc version 11.4.0.
- For any questions and problems, you can always contact me **via email** (b.koca@gtu.edu.tr), or you can find me in Room 119 during scheduled office hours on March 5 and 12, 2024, between 13:30 and 14:30.