

<u>SURNAME:</u>		<u>NAME</u>		<b><u>B1</u></b>
<u>STUDENT ID:</u>				
<u>PROFESSOR:</u>				

Question 1	Answers
<p>Given the following numbers in 2's Complement (2C) of 8 bits, displayed in the hexadecimal notation:</p> <ul style="list-style-type: none"><li>- N1: A3</li><li>- N2: 3F</li></ul> <p>calculate their sum and tell if an overflow occurs.</p>	<p>Result of the sum:</p>          <p>Overflow (yes/no):</p>
<p>Report ALL steps</p>	

<b>Question 2</b>
Calculate the truth table of the following Boolean expression: $(Z \text{ and } (X \text{ and not}(Y))) \text{ or } (\text{not}(X) \text{ and } Z)$
Report all significant steps

<b>Question 3</b>
What are similarities and differences between RAM and ROM memory?

**Question 4 (PROGRAMMING)**

A classification algorithm is a kind of machine learning algorithm which is used to assign a label to an object, based on the knowledge of similar objects. Write a C program able to implement the classification algorithm 1-NN.

The program receives as input two files `<train>` and `<test>`, whose names are passed as arguments from the command line:

The `<train>` file contains a set of N objects (one for each line), each one with the corresponding label. This file is used to train the classification algorithm, which learns how to recognize the known objects. Each line of the file contains the name of the object represented as a string at most 5 characters, a set of M real numbers  $T_1 T_2 T_3 \dots T_M$ , followed by an integer L. The space character is used to separate the elements. The M real numbers are the features that represents an object. The integer L is the object label.

The `<test>` file contains a set of objects to be classified. The format is similar to the `<train>` file, but the number of rows is **NOT KNOWN IN ADVANCE**.

Assume that the values M and N are known in advance and are defined as symbolic constants by means of the `#define` directive.

The distance between two objects O and T is defined as follows:

$$D_{O-T} = \sqrt{\sum_{i=1}^M (O_i - T_i)^2} \text{ where } O_i \text{ and } T_i \text{ are the real numbers used to represents the objects.}$$

The program shall predict the label of each object O contained in the file `<test>`, according to the following rules:

- Calculate the distance between the object O and every object T of the file `<train>`.
- The label to be assigned to O is the one of the object T that is the nearest to O.

For each classification, the program shall print a message on the screen indicating the name of the object, the label that has been predicted for the object, and its actual label (as reported in the file `<test>`).

Finally, the program shall print on screen the accuracy of the classification algorithm, which is calculated as the ratio between the number of labels predicted properly divided by the total number of classified objects.

**Example:**

M=2 N=9

Train.txt				Test.txt				Distances				Distances			
T1	1.3	3.8	1	O1	1.3	2.5	2	O1 - T1	1.30			O2 - T1	2.21		
T2	1.6	3.9	1	O2	3.5	4	3	O1 - T2	1.43			O2 - T2	1.90		
T3	1.5	3.7	1					O1 - T3	1.22			O2 - T3	2.02		
T4	4.0	1.0	3					O1 - T4	3.09			O2 - T4	3.04		
T5	4.1	1.1	3					O1 - T5	3.13			O2 - T5	2.96		
T6	4.2	1.4	3					O1 - T6	3.10			O2 - T6	2.69		
T7	2.5	2.5	2					O1 - T7	1.20			O2 - T7	1.80		
T8	2.3	2.4	2					O1 - T8	1.00	O1 label = 2		O2 - T8	2.00		
T9	2.6	2.6	2					O1 - T9	1.30	(correct)		O2 - T9	1.66	O2 label = 2	
														(wrong)	

C:\>classify.exe train.txt test.txt

Object O1: predicted label 2 – actual label 2

Object O2: predicted label 2 – actual label 3

The accuracy is equal to 0.50

<u>SURNAME:</u>		<u>NAME</u>		<b><u>B2</u></b>
<u>STUDENT ID:</u>				
<u>PROFESSOR:</u>				

Question 1	Answers
<p>Given the following numbers in pure binary on 8bits, displayed in the hexadecimal notation:</p> <ul style="list-style-type: none"><li>- N1: A3</li><li>- N2: 3F</li></ul> <p>calculate their sum, and tell if an overflow occurs.</p>	<p>Result of the sum:</p>          <p>Overflow (yes/no):</p>
<p>Report ALL steps</p>	

<b>Question 2</b>
Calculate the truth table of the following Boolean expression: $(\neg Z) \wedge (X \wedge Y) \vee (X \wedge Z)$
Report all significant steps

<b>Question 3</b>
Briefly describe the differences between main memory and mass memory

**Question 4 (PROGRAMMING)**

A classification algorithm is a kind of machine learning algorithm which is used to assign a label to an object, based on the knowledge of similar objects. Write a C program able to implement the classification algorithm 1-NN.

The program receives as input two files `<train>` and `<test>`, whose names are passed as arguments from the command line:

The `<train>` file contains a set of  $N$  objects (one for each line), each one with the corresponding label. This file is used to train the classification algorithm, which learns how to recognize the known objects. Each line of the file contains the name of the object represented as a string at most 5 characters, a set of  $M$  real numbers  $T_1 T_2 T_3 \dots T_M$ , followed by an integer  $L$ . The space character is used to separate the elements. The  $M$  real numbers are the features that represents an object. The integer  $L$  is the object label.

The `<test>` file contains a set of objects to be classified. The format is similar to the `<train>` file, but the number of rows is **NOT KNOWN IN ADVANCE**.

Assume that the values  $M$  and  $N$  are known in advance and defined as symbolic constants by means of a `#define` directive.

The distance between two objects  $O$  and  $T$  is defined as follows:

$$D_{O-T} = \sum_{i=1}^M |O_i - T_i| \text{ where } O_i \text{ and } T_i \text{ are the real numbers used to represent the objects.}$$

The program shall predict the label of each object  $O$  contained in the file `<test>`, according to the following rules:

- Calculate the distance between the object  $O$  and every object  $T$  of the file `<train>`.
- The label to be assigned to  $O$  is the one of the object  $T$  that is the nearest to  $O$ .

For each classification, the program shall print a message on the screen indicating the name of the object, the label that has been predicted for the object, and its actual label (as reported in the file `<test>`).

Finally, the program shall print on screen the accuracy of the classification algorithm, which is calculated as the ratio between the number of labels predicted properly divided by the total number of classified objects.

**Example:**

$M=2$   $N=9$

Train.txt				Test.txt				Distances				Distances			
T1	1.3	3.8	1	O1	1.3	2.5	2	O1 - T1	1.30			O2 - T1	2.40		
T2	1.6	3.9	1	O2	3.5	4	3	O1 - T2	1.70			O2 - T2	2.00	O2 label = 1	
T3	1.5	3.7	1					O1 - T3	1.40			O2 - T3	2.30	(wrong)	
T4	4.0	1.0	3					O1 - T4	4.20			O2 - T4	3.50		
T5	4.1	1.1	3					O1 - T5	4.20			O2 - T5	3.50		
T6	4.2	1.4	3					O1 - T6	4.00			O2 - T6	3.30		
T7	2.5	2.5	2					O1 - T7	1.20			O2 - T7	2.50		
T8	2.3	2.4	2					O1 - T8	1.10	O1 label = 2		O2 - T8	2.80		
T9	2.6	2.6	2					O1 - T9	1.40	(correct)		O2 - T9	2.30		

C:\>classify.exe train.txt test.txt

Object O1: predicted label 2 – actual label 2

Object O2: predicted label 1 – actual label 3

The accuracy is equal to 0.50