

BoolFuncOnDB README

Introduction

The BoolFuncOnDB program takes in a dataset and a Boolean function to apply that Boolean function on each datapoint of the dataset. If a particular attribute of the dataset is not Boolean, then the program will automatically detect this and ask the user to define a range of two thresholds, or a minimum or maximum threshold, respectively. Values of “yes,” “true,” “t,” “1,” and others are accepted as Boolean. Otherwise, they will need a threshold. In the case of some value of something like “random_text,” the program will not work.

Definitions

Range: this is defined as an upper or lower bound. Therefore, a threshold of [2, 3] will define any value between 2 and 3 as true (1) and any number outside that threshold will be false.

Maximum threshold: for example, a maximum threshold of 2 will define any value less than or equal to 2 as true, and greater than 2 will be false.

Minimum threshold: for example, a minimum threshold of 2 will define any value greater than or equal to 2 as true, and lesser than 2 will be false.

Input/Output

The program takes in a CSV file as input. It simply has to be labeled “dataset.csv” in the executable’s directory. The first line of the file must be the names of the attributes. If there are more or less attributes than what is in the datapoints, or if the dataset is sparse. The program will likely not work.

The outputted results file will simply be labeled “results.csv.” The results file will have the Boolean function and thresholds (if there are any) located at the top of the file. Then, the datapoints will be listed along with the class. The results file is sorted so that datapoints with a class of 1 are given first. Moreover, the Boolean representation of each datapoint is given and located just to the right of each datapoint.

Examples

Here is an example of some user input:

```

Enter a Monotone Boolean function in the disjunctive normal form: x1x2x5 v x10
For the given dataset (dataset.csv in the current directory), the class will be appended to the end of each datapoint by
using the Boolean Function that was input.
Is the threshold for attribute bedrooms a range? (1/0): 0
Please enter a threshold for this attribute: 2
Is this threshold a max threshold or min threshold? (1/0): 0
Is the threshold for attribute bathrooms a range? (1/0): 0
Please enter a threshold for this attribute: 2
Is this threshold a max threshold or min threshold? (1/0): 0
Is the threshold for attribute stories a range? (1/0): 1
Please enter the min threshold: 1
Please enter the max threshold: 3
Is the threshold for attribute parking a range? (1/0): 0
Please enter a threshold for this attribute: 2
Is this threshold a max threshold or min threshold? (1/0): 0

```

Here is an example of some output:

x1x2 v x10										
x1 [2;infinity]										
x2 [2;infinity]										
x3 [1;3]										
x9 [2;infinity]										
bedrooms	bathroom	stories	mainroad	guestroom	basement	hotwater	airconditi	parking	prefarea	class
4	2	3	yes	no	no	no	yes	2	yes	1
4	4	4	yes	no	no	no	yes	3	no	1
3	2	2	yes	no	yes	no	no	2	yes	1
4	2	2	yes	no	yes	no	yes	3	yes	1
3	1	1	yes	yes	yes	no	no	2	yes	1
3	3	1	yes	no	yes	no	yes	2	yes	1
4	3	4	yes	no	no	no	yes	2	yes	1
Boolean representation										
1	1	1	1	0	0	0	1	1	1	1
1	1	0	1	0	0	0	1	1	0	1
1	1	1	1	0	1	0	0	1	1	1
1	1	1	1	0	1	0	1	1	1	1
1	0	1	1	1	1	0	0	1	1	1
1	1	1	1	0	1	0	1	1	1	1
1	1	0	1	0	0	0	1	1	1	1