Machine Learning: Programming Assignment 1: Concept Learning

Problem Statement:

Implement the Find-S concept learning algorithm (refer Chapter 2 of Tom Mitchell book) that finds the most specific hypothesis that is consistent with the given training data. Hypothesis can only be conjunction (AND) of literals. Literals are either attributes or their negations.

Data Set Description: Input Filename: data1.csv

Input File Format: Boolean input attributes (x1, x2, … , x8) in first 8 columns. The last (9^{th}) column represents the Boolean class label (y). Each row is a training instance. There are 20 training instances. We give a file data1.csv for training. Your program may be tested for other input files with same number of rows and columns.

Please STRICTLY follow the program input/output format specified below.

Input Format: *<filename>* (e.g., data1.csv)

Output Format: Hypothesis returned by the Find-S algorithm in the format: *n, l1, l2, li, …, ln* (comma separated values)

Here, *n* is the number of literals in the hypothesis, and *l1, l2, li*, …, *ln* are the literals present in the hypothesis (no need to output the don't cares). If a literal is an attribute, *li* is the corresponding attribute number. If the literal is a negation of an attribute, *li* is the negative of the corresponding attribute number. For example, if the output hypothesis is x2∧x3∧~x5 the output should be 3, 2, 3, -5.

Submission Guidelines:

You may use one of the following languages: c/C++/Java/Python. Your program should be standalone and should not use any *special purpose* library. The submitted single program file *should* have the following program header:

Roll

Name

Assignment number

Specific compilation/execution flags (if required)

*Please submit the program in moodle by **August 14, 2018 midnight** (hard deadline). Copying from friends/web will lead to strict penalties.*