CS 584 – Data Mining
CA1 - Programming ID3

The Problem

In this assignment, you are to write a C++ program (using Visual C++ 2017 or above) that reads training data in WEKA arff format and generates ID3 decision tree in a format similar to that of the tree generated by Weka (run J48 which is C4.5, an extension of ID3) or my sample program ID3.exe. Please note the following:

1. Your algorithm will use the entire data set to generate the tree. You may assume that the attributes (a) are of nominal type (i.e., no numeric data), and (b) have no missing values.

2. In general, the basic ID3 algorithm uses entropy measure to select the best attribute to divide the data set. It continues to select attribute for further branching (based on the subset of data belong to that branch) until either (a) all attributes have been used, or (b) all instances under a node belong to the same class. This ensures a 0% error rate on the training set although it may not work the best with future data due to over-fitting.

Sample Input

weather.nominal.arff
contact-lenses.arff
restaurant.arff

Sample Output

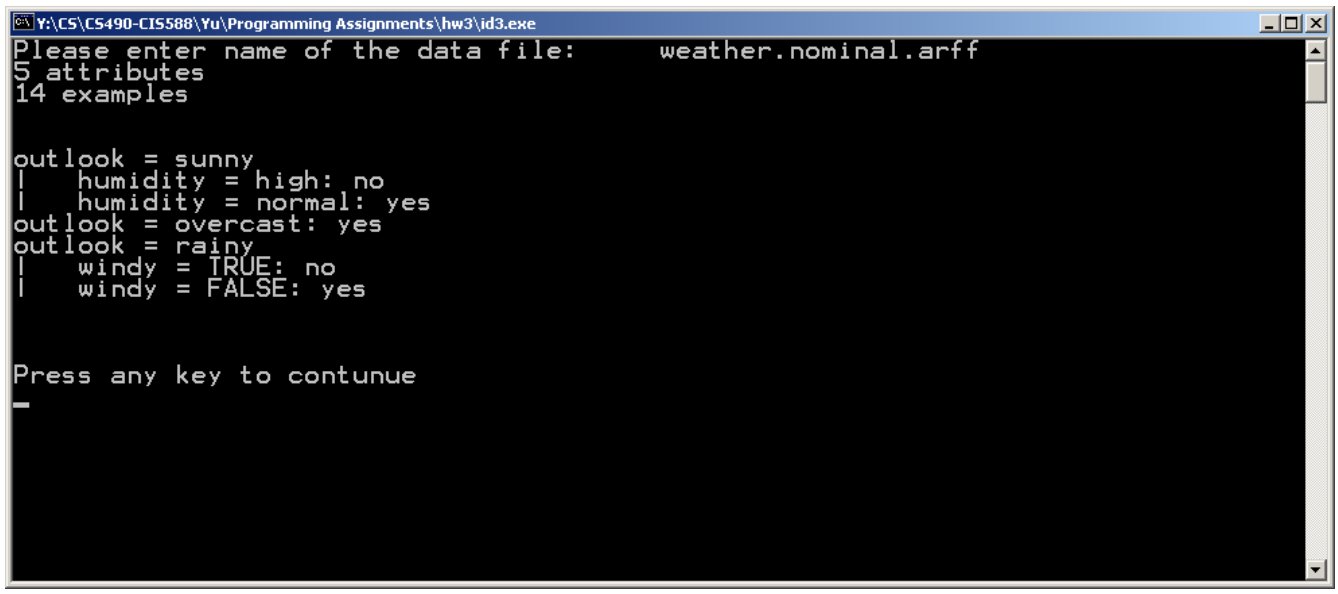
Run ID3.exe on the sample input file. Sample output screenshots are included on the next page.

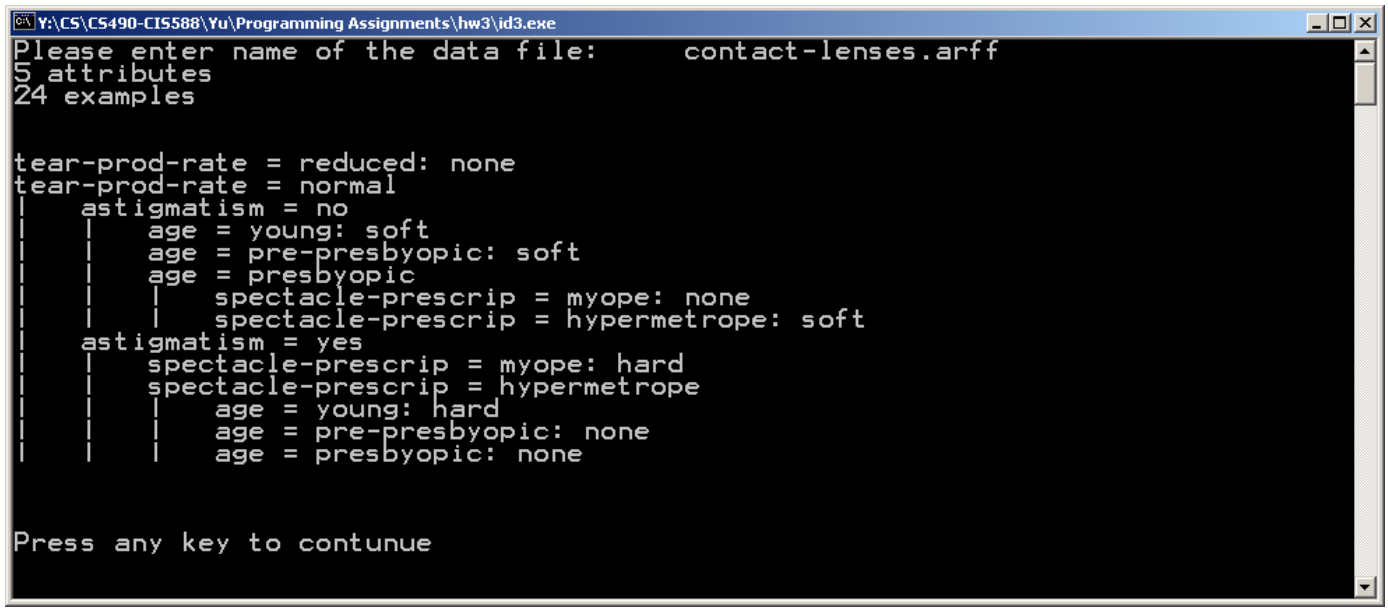
Required Comments

For this assignment, you must **also include a performance summary** detailing how well you program performs on each of the data sets. The performance summary is worth 10 points.

Due Time

As mentioned previously, you may work in team of two for this and all future assignments. The assignment is *due Tuesday (3/4) at 10 PM*. Submit a zip file which contains your *entire project folder* (or simply your .cpp file(s)) on Moodle by the due time. Late submission will be accepted until Saturday (3/8) at 10 pm. Standard late penalty applies.


Y:\CS\CS490-CIS588\Yu\Programming Assignments\hw3\id3.exe
Please enter name of the data file: weather.nominal.arff
5 attributes
14 examples
outlook = sunny
| humidity = high: no
| humidity = normal: yes
outlook = overcast: yes
outlook = rainy
| windy = TRUE: no
| windy = FALSE: yes
Press any key to contunue


Y:\CS\CS490-CIS588\Yu\Programming Assignments\hw3\id3.exe
Please enter name of the data file: contact-lenses.arff
5 attributes
24 examples
tear-prod-rate = reduced: none
tear-prod-rate = normal
| astigmatism = no
| | age = young: soft
| | age = pre-presbyopic: soft
| | age = presbyopic
| | | spectacle-prescrip = myope: none
| | | spectacle-prescrip = hypermetrope: soft
| astigmatism = yes
| | spectacle-prescrip = myope: hard
| | spectacle-prescrip = hypermetrope
| | | age = young: hard
| | | age = pre-presbyopic: none
| | | age = presbyopic: none
Press any key to contunue