Expert Data Mining:

Nested Functions

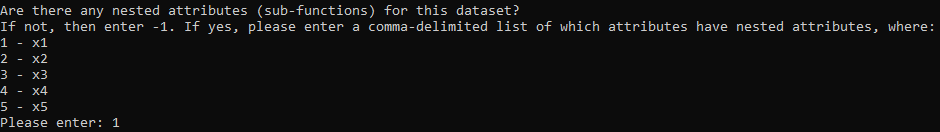
# Introduction

The Expert Data Mining software is now able to use nested functions. Essentially, any attributes can have sub-attributes, meaning that there can be multiple sequences of asking questions to the user. A common function that we like to use for testing is “x1x3x5 v x2x3x5 v x4x5.” Any one of these attributes can have a couple sub-attributes. Let’s say that attribute x3 has two sub-attributes, x3.1 and x3.2. These two attributes will have their own Boolean function which makes the attribute x3 true or false. This sub-function could be something like “x3.1 v x3.2.” This sub-function can be subbed into the parent Boolean function to get

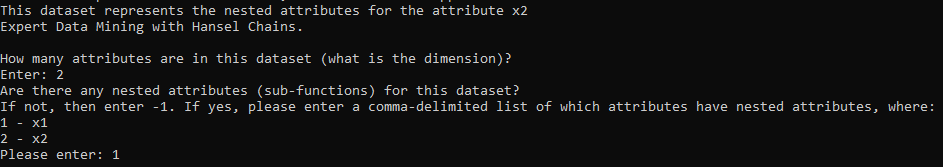
“x1(x3.1 v x3.2)x5 v x2(x3.1 v x3.2)x5 v x4x5.”

# Implementation

The implementation is quite simple: before we ask the user any questions, we ask if there are any sub-attributes.

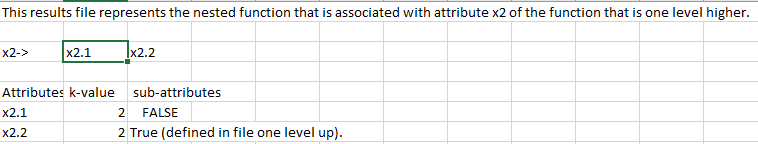


In this example, we entered one so that the attribute x1 would have sub-attributes. Multiple attributes can be designated to have sub-attributes. After this, the sequence of questions will be the exact same as it was in its previous implementations. Once all the questions have been answered, we move onto asking questions about the sub-attributes, starting with how many of them there are.



We entered two so that that there are 2 attributes in the dataset (dimension of 2), and then specified that there are sub-sub-attributes. The process is iterative, so any depth of sub-attributes could be used, although it would likely be better to group the attributes in a more efficient way if the depth gets too deep.

Furthermore, the output file of the results uses a slightly different nomenclature and format to account for the sub-functions and sub-attributes. First, the file is named “results\_of\_function\_*n*.csv” where n starts at the number 0 and increases. If n is equal to 0, then that results file represents the parent Boolean function. If *n* is equal to 1, it is for the sub-function. If it is equal to 2, it is for the sub-sub-function, and the pattern continues. If there are multiple attributes with sub-attributes in the same depth, then the results will be printed in the same file. An example of what is new in the output file is shown below



This file is associated with the function that is in the results file that is labeled with *n* – 1, so if this file is “results\_of\_function\_*1*.csv” (which it is in the attached files), then the parent will be “results\_of\_function\_0.csv.” The attributes are named slightly differently as well. Since the attributes in the picture represent its parent attribute “x2,” they are given the name “x2.1” and “x2.2.” for sub-sub-attributes, this can be quite confusing, so multiple levels aren’t considered in this aspect. For a sub-attribute of “x1.2,” within the results file, the parent attribute will just be called “x2,” for example.