Due October 27, 2016

Introduction to CI – Fall 2016
Project #2 - Fuzzy Decision Trees

For this project, you will use fuzzy decision trees to help with the task of buying a car. A reduced version of the Automobile Data Set from the UCI Machine Learning Repository containing information about cars from 1985 (not the most up-to-date, but still good data) will be used. The data is saved in a comma-separated value format with 164 rows and 7 columns. Each row is a record for a different car containing the following attributes in order:

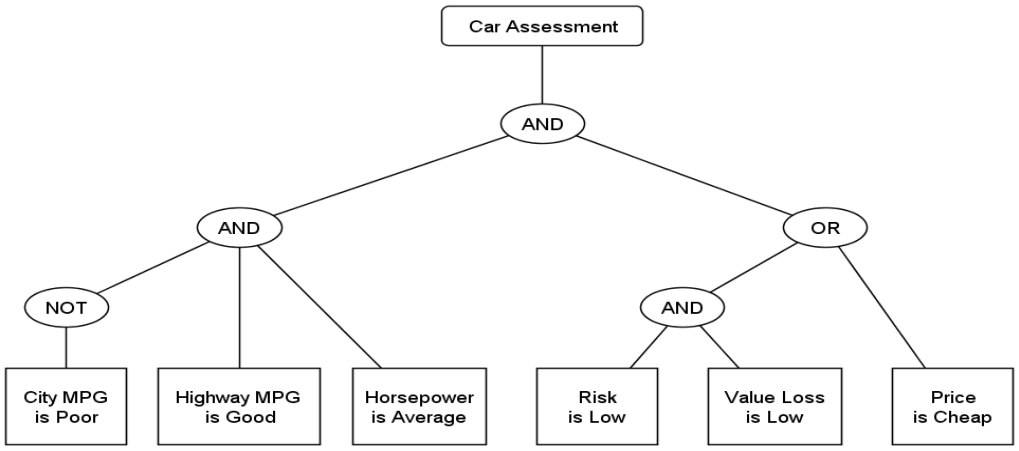
- **Car ID** – Unique identification number for each car
- **Risk** – Integer in the range [-3,3] representing the risk associated with purchasing the car. A value of -3 means very low risk, while a value of +3 means very high risk.
- **Value Loss** – Average loss of value in dollars per insured vehicle year.
- **Horsepower** – Horsepower of the engine.
- **City MPG** – Miles per gallon in the city.
- **Highway MPG** – Miles per gallon on the highway.
- **Price** – Cost of the vehicle in dollars.

Part A

Use MATLAB or a language of your choice to read in the data and build the following fuzzy decision tree. For each attribute (except the Car ID), define three membership functions as follows:

Risk Low = Trap(x; -3, -3, -2, 0) Average = Tri(x; -2, 0, 2) High = Trap(x; 0, 2, 3, 3)	City MPG Poor = Trap(x; 0, 0, 20, 30) Average = Tri(x; 20, 30, 40) Good = Trap(x; 30, 40, 60, 60)
Value Loss Low = Trap(x; 0, 0, 100, 120) Average = Tri(x; 100, 120, 200) High = Trap(x; 120, 200, 300, 300)	Highway MPG Poor = Trap(x; 0, 0, 20, 30) Average = Tri(x; 20, 30, 40) Good = Trap(x; 30, 40, 60, 60)
Horsepower Low = Trap(x; 0, 0, 60, 100) Average = Tri(x; 60, 100, 140) High = Trap(x; 100, 140, 250, 250)	Price Cheap = Trap(x; 0, 0, 7000, 10000) Average = Tri(x; 7000, 10000, 20000) Expensive = Trap(x; 10000, 20000, 40000, 40000)

Evaluate how well each car scores using the following decision tree with the standard fuzzy operators.



Questions

1) What is the highest scoring car, and how does it compare to the other cars? Create a histogram plot (MATLAB command hist) to see how the scores were distributed. Did many cars score highly or were there only a few?
2) Using the same decision tree, change the definition of the fuzzy operators to the bounded sum and difference. How does this affect the scoring of the cars and the resulting decision?
3) Again using the same decision tree, use the Yager family of fuzzy operators and the following values for *w*: $\{10^{-1}, 10^{0}, 10^{1}, 10^{2}, 10^{3}, 10^{4}\}$. How does *w* affect the distribution of scores?

Part B

Using the same data from Part A, design your own decision tree to find a car with features of your own choosing. Experiment with the number and types of membership functions, as well as your choice of fuzzy operators. Show plots of your membership functions (MATLAB command plot) as well as histograms of the score distributions. Do the membership function shapes have a significant effect on the resulting scores? Explain why (or why not) fuzzy set theory is a useful framework for this problem.

Be sure to design your program in a well-structured fashion and document the code appropriately.

Your report should contain the following sections:

1. Introduction / Overview / Problem Statement
2. Background information on implemented algorithms / methods
   Describe and outline algorithms and methods included in the project. Be precise. Equations and/or pseudo-code are recommended where they would be helpful.
3. Experiments and Results
   Carefully describe experiments conducted and the data sets used. Include a description of the purpose of each experiment.
4. Discussion / Analysis / Conclusion
   Describe any conclusions or things you learned from the project. The discussion/analysis/conclusions should follow from the experiments conducted.
5. Documented program listing.
   Your code should be clear and understandable to someone reading it. Comments are most likely necessary.

** Please bring a hard copy of your report to class on the due date. If turning your report in late, emailing a copy to the instructor is acceptable to establish the time it was turned in. However, a hard copy should still be turned in at the next class period.