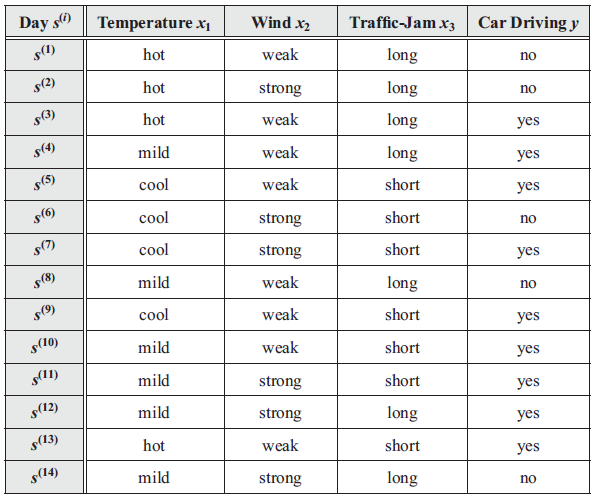
**Topic 5 Assignment: Fuzzify a Decision Tree**

**Description**: This assignment mixes theory and application, in the form of two problems. Perform the tasks described in each. Note that this (and other) assignments includes a few challenging research-related tasks. They are aimed at gradually building your capacity to tackle complex topics, familiarize yourself with academic discourse, and provide context and practice for the skills you will eventually need when working on your capstone thesis or project.

APA style is expected, as well as formal and rigorous scientific writing, using appropriate mathematical notation and references.

**Part 1 – Decisions**

People decide to drive a car or use public transportation to commute to work by weighing in such factors as weather and traffic. The following table describes the dataset *D*, containing data collected over a period of 14 days. The attributes are *Temperature (x1),* *Wind (x2)*, *and Traffic-Jam (x3)*; and the target variable is *Car Driving Decision (y)*.



Note that on 9 days, the decision is made to drive a car, whereas on the remaining 5 days, the decision is negative about driving a car.

You are tasked to create a decision tree that uses the above data to compute the decision on whether to drive a car, given the conditions depicted by the three variables in the table.

Consult the decision tree package, <https://scikit-learn.org/stable/modules/tree.html> and using a Jupyter notebook, implement the following in Python:

1. Calculate the information gain for x1, x2 and x3.

Text

Description automatically generated

1. Choose the root node for the decision tree.



1. Plot a partial decision tree from root node along with training examples sorted to each of its descendent nodes.

Diagram

Description automatically generated

1. You are required to continue with the decision-tree growing process until all nodes are pure (they contain examples that all have the same classification).

I have concluded that this is not possible.

1. Will it be possible to reach this situation for the given dataset? If not, why not? (Hint: Look for inconsistencies in the given dataset.)

Looking at the data one will notice that there are conflicting data entries. Situations where observations have identical features but result in opposing outcomes. This error creates uncertainty within the data, which pushes the model into the realm of probability, while this model is using set boundaries that are not created for probability.

1. Summarize the entire work in a technical report.

My technical report can be found here,

GitHub: <https://github.com/DouglasBui/GCU/tree/main/DSC-540/Assignment5>

**Part 2 – Fuzzy Decisions**

You are tasked to modify the classifier tree in Part 1 so that it better mimics the way a human makes decisions. You will accomplish this by transforming the decision tree into a *fuzzy decision tree*.

In addition to the readings assigned in this topic, you may find it useful to read the techniques described in the "Combined Application of Decision Tree and Fuzzy Logic Techniques for Intelligent Grading of Dried Figs."

Answer the following and provide necessary Python code to help visualize your model:

1. Explain how you will fuzzify the attributes x1, x2 and x3.

The application of fuzzifying such simple string attributes seems particularly strange. The idea behind fuzzification is to take a crisp input and subscribing memberships functions to it and then utilize rules that allow for a broadening of the interpretation of the data and then defuzzifying it to result in a crisp output. I took these 3 attributes and embellished the data to fit a more human interpretation of them. Once I then fuzzified these attributes I put them through a decisiontree() function.

1. Choose the root node for the fuzzy decision tree and defend your choice.

I would choose Temperature as the root node for a fuzzy decision tree, seeing how it is the only attribute that has a range that can be triangulated for a membership function.

1. Plot a partial fuzzy decision tree from root node along with training examples sorted to each of its descendent nodes. Assume the dataset to be crisp, i.e., all membership functions satisfy:  
   ***µD(s(i)) = 1***

Once we have calculated our fuzzy set from the attributes we can incorporate fuzzy rules into the decision tree, and these rules will be the deciding factor behind splits. Each feature will be split until the feature’s leaf nodes reach unity. With each split the rules must be reapplied to each child node.

1. As with Part 1, create a second report, detailing your approach to fuzzification of the decision tree.

GitHub: <https://github.com/DouglasBui/GCU/tree/main/DSC-540/Assignment5>