Melissa LaHoud

Grand Canyon University

DSC 540: Machine Learning

Dr. Aiman Darwiche

4/21/2021

Decision Tree

People decide to drive a car or use public transportation to commute to work by weighing in such factors as weather and traffic. Table 1 describes the dataset *D*, containing data collected over a period of 14 days. The attributes are *Temperature* (x_1), *Wind* (x_2), *and Traffic-Jam* (x_3); and the target variable is *Car Driving Decision* (y).

Day $s^{(i)}$	Temperature x_1	Wind x_2	Traffic-Jam x_3	Car Driving y
$s^{(1)}$	hot	weak	long	no
$s^{(2)}$	hot	strong	long	no
$s^{(3)}$	hot	weak	long	yes
$s^{(4)}$	mild	weak	long	yes
$s^{(5)}$	cool	weak	short	yes
$s^{(6)}$	cool	strong	short	no
$s^{(7)}$	cool	strong	short	yes
$s^{(8)}$	mild	weak	long	no
$s^{(9)}$	cool	weak	short	yes
$s^{(10)}$	mild	weak	short	yes
$s^{(11)}$	mild	strong	short	yes
$s^{(12)}$	mild	strong	long	yes
$s^{(13)}$	hot	weak	short	yes
$s^{(14)}$	mild	strong	long	no

Table 1

Let's first look at a classification decision tree based on our data described in Table 1. We start with the root node which contains all the training data used to grow the tree. We want to determine which attribute to split on. In this example, we have Temperature, Wind or Traffic-Jams. We can look at each of them in Figure 1.

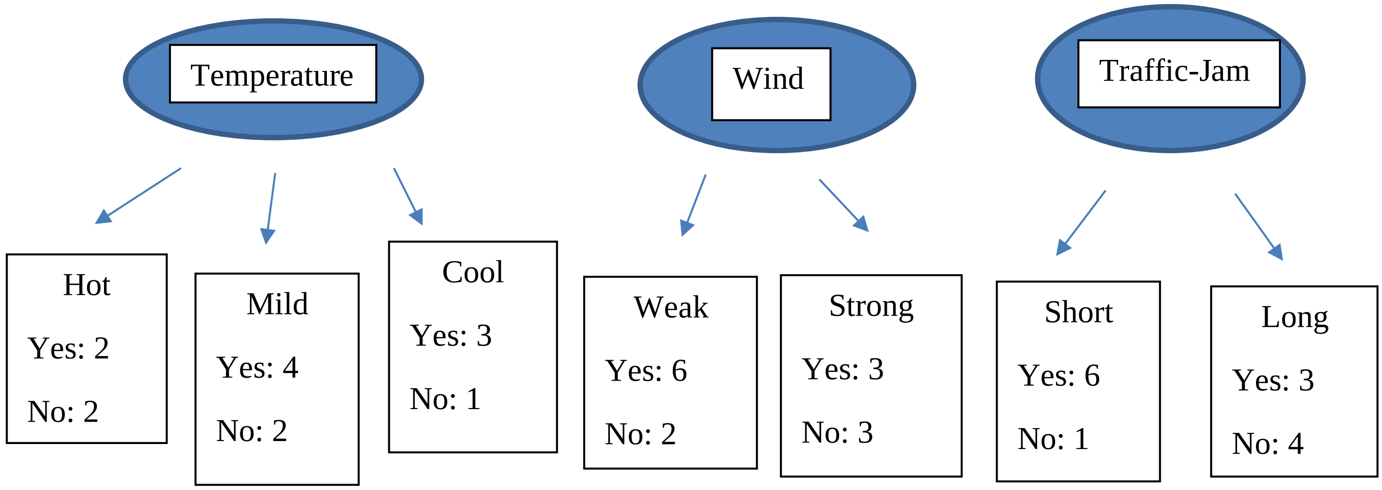


Figure 1

These will help us determine whose daughter nodes will achieve the leaf nodes to come up as soon as possible. In this case, I would choose Traffic-Jam as the splitting attribute at the root of the tree because it produces a daughter node that has a large number of instances of only on classification.

Next, we look at information gain and this is an idea for defining impurity. If a leaf is pure then the classes in the leaf can be described very easily. Entropy measures how disorganizes a system is. Below are the equations for both.

$$Info\ (D) = -\sum_{q=1}^{2} P_q \log_2(P_q)$$

$$Entropy\ (D) = -\sum_{q=1}^{2} P_q \log_2 P_q$$

For this specific example, we can see the information gain for each variable in Table 2 along with the entropy for D at 0.94 found in the tree in Figure 2.

	Information Gain
Temperature	0.261
Wind	0.048
Traffic-Jam	0.152

Table 2

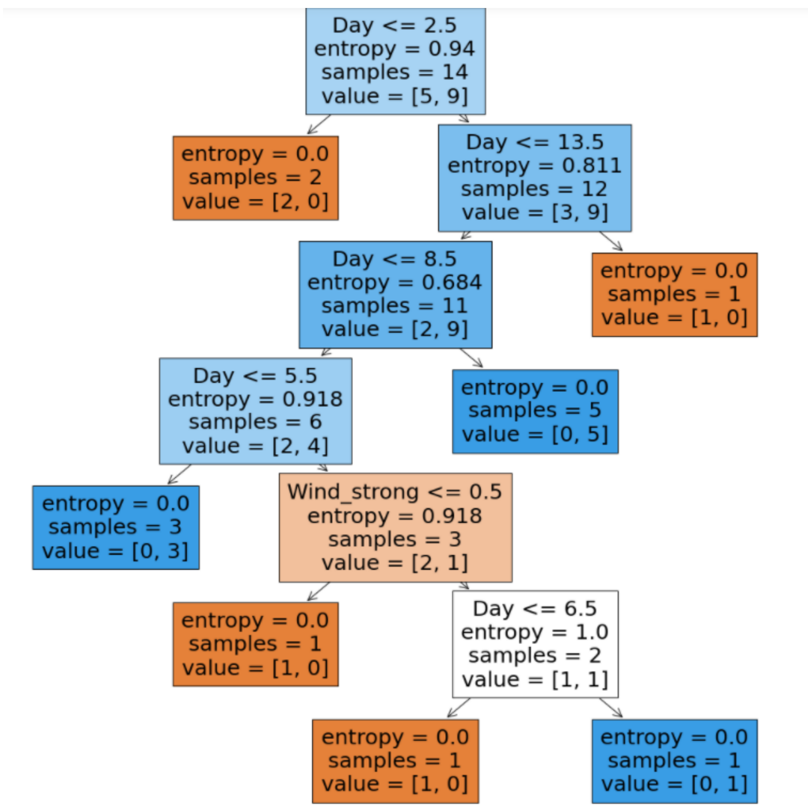


Figure 2

Figure 2 shows the final decision tree produced in python with each split until the leaf's are pure. It seems that this dataset did reach purity with the entropy being 0 at each of the ending nodes.

Fuzzy Decisions

Now we look at modifying the classifier tree so that it better mimics the way a human makes decisions. I accomplished this by transforming the decision tree into a *fuzzy decision tree*. To do this, I used the attributes Temperature (x1), Wind (x2), and Traffic-Jam (x3) from the dataset and selected a membership grade $\mu_D(s^{(i)})$ for all input examples. I also created membership values for each value of x1, x2, x3 that is shown in Table 3.

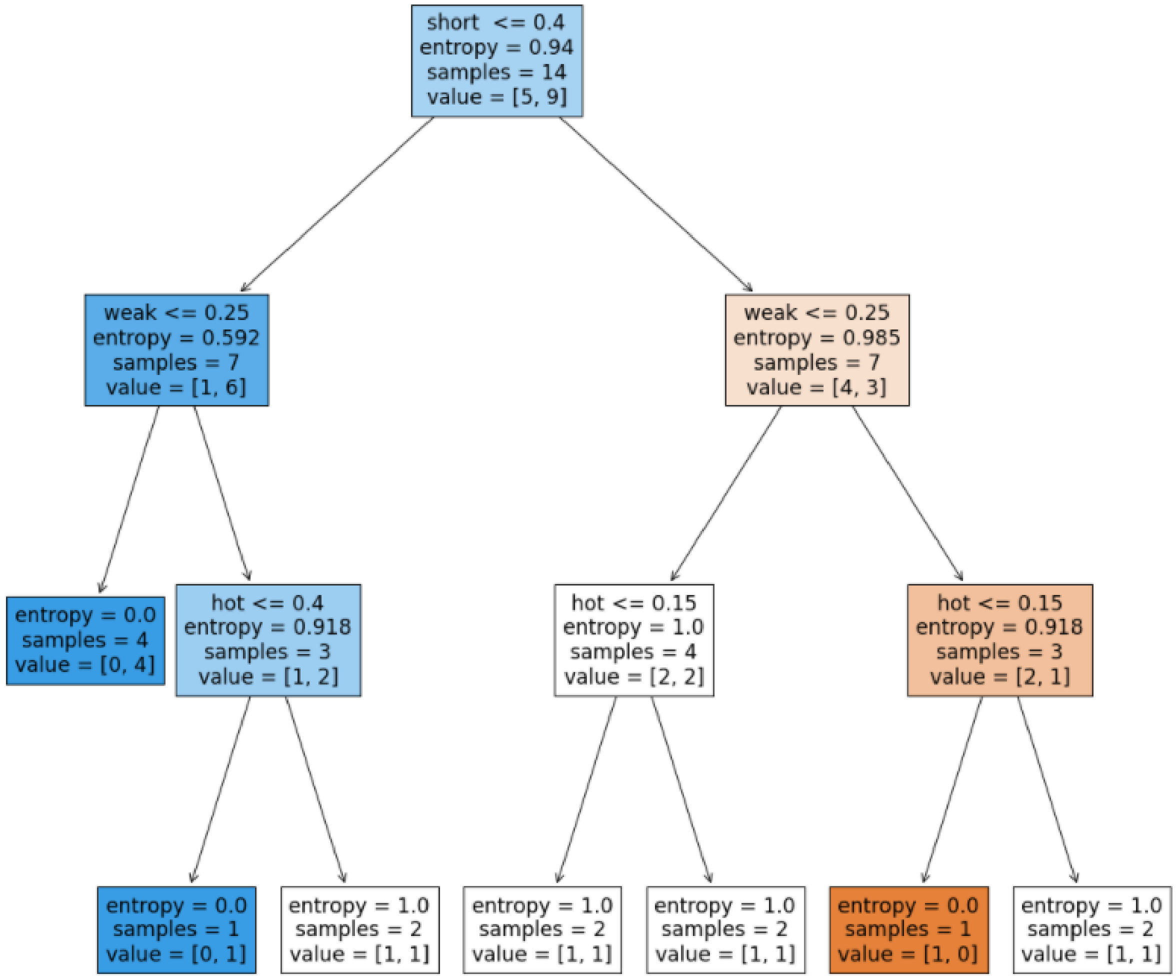
	hot	mild	cool	weak	strong	short	long	drive
0	1.0	0.0	0.0	1.0	0.0	0.2	0.8	0
1	1.0	0.0	0.0	0.5	0.5	0.2	0.8	0
2	1.0	0.0	0.0	1.0	0.0	0.2	0.8	1
3	0.7	0.3	0.0	1.0	0.0	0.2	0.8	1
4	0.0	0.5	0.5	1.0	0.0	1.0	0.0	1
5	0.0	0.5	0.5	0.5	0.5	1.0	0.0	0
6	0.0	0.5	0.5	0.5	0.5	1.0	0.0	1
7	0.7	0.3	0.0	1.0	0.0	0.2	0.8	0
8	0.0	0.5	0.5	1.0	0.0	1.0	0.0	1
9	0.7	0.3	0.0	1.0	0.0	1.0	0.0	1
10	0.7	0.3	0.0	0.5	0.5	1.0	0.0	1
11	0.7	0.3	0.0	0.5	0.5	0.2	0.8	1
12	1.0	0.0	0.0	1.0	0.0	1.0	0.0	1
13	0.7	0.3	0.0	0.5	0.5	0.2	0.8	0

Table 3

These are to help identify the memberships values for expansion of root nodes. Using these membership values, I was able to calculate to calculate the Entropy to be 0.703 using the equation below:

$$Entropy(\mathcal{D}) = -\beta_1 \log_2 \beta_1 - \beta_2 \log_2 \beta_2$$

The information gain is still used to determine what attribute to expand on. The highest value of the information gain is the one we use to expand first. Using the correct function in python, I was able to produce the fuzzy decision in Figure 3.



References

Gopal, M. (2019). Applied Machine Learing. New Delhi, India: McGraw-Hill Education.