Assignment6 – Decision Trees

Given: June 6 Due: June 11

Problem 6.1 (Sunbathing)

Eight people go sunbathing. They are categorized by the attributes Hair and Lotion and the result of whether they got sunburned.

Name	Hair	Lotion	Result: Sunburned
Sarah	Light	No	Yes
Dana	Light	Yes	No
Alex	Dark	Yes	No
Annie	Light	No	Yes
Julie	Light	No	No
Pete	Dark	No	No
John	Dark	No	No
Ruth	Light	No	No

1. Which quantity does the information theoretic decision tree learning algorithm use to pick the attribute to split on?

Solution: Information gain.

---

2. Compute that quantity for the attributes Hair and Lotion. (Simplify as much as you can without computing logarithms.)

Solution:

$$E_0 := I(\langle \frac{2}{8}, \frac{6}{8} \rangle) = -\frac{2}{8}\log_2(\frac{2}{8}) - \frac{6}{8}\log_2(\frac{6}{8}) \approx 0.81$$

$$\text{Gain}(Hair) = E_0 - \underbrace{\frac{5}{8} I(\langle \frac{2}{5}, \frac{3}{5} \rangle)}_{Light} - \underbrace{\frac{3}{8} I(\langle 0, 1 \rangle)}_{Dark} \qquad \approx 0.20$$

$$\text{Gain}(Lotion) = E_0 - \underbrace{\frac{2}{8} I(\langle 0, 1 \rangle)}_{Yes} - \underbrace{\frac{6}{8} I(\langle \frac{2}{6}, \frac{4}{6} \rangle)}_{No} \qquad \approx 0.12$$

Entropy is undefined for 0. If we were to continue simplifying, we'd use $0 \cdot \log_2 0 = 0$.

---

3. Assuming the logarithms are computed, how does the algorithm pick the attribute?

Solution: It picks the one with the highest information gain (in this case Hair).

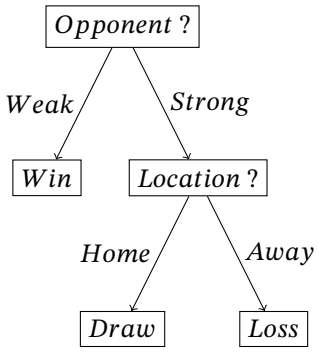
Problem 6.2 (Decision Trees)

You observe the values below for 6 different football games of your favorite team. You want to construct a decision tree that predicts the result.

#	Day	Weather	Location	Opponent	Result
1	Monday	Rainy	Home	Weak	Win
2	Monday	Sunny	Home	Weak	Win
3	Friday	Rainy	Away	Strong	Loss
4	Sunday	Sunny	Home	Weak	Win
5	Friday	Cloudy	Home	Strong	Draw
6	Sunday	Sunny	Home	Strong	Draw

1. Assume you choose attributes in the order
$Opponent, Location, Weather, Day$.
Give the resulting decision tree.

Solution: The tree is



---

2. Without using the above observations, give the formula for the information gain of the attribute $Opponent$.

Solution: $\text{Gain}(Opponent) = I(P(Result)) - P(Opponent = Strong) \cdot I(P(Result \mid Opponent = Strong)) - P(Opponent = Weak) \cdot I(P(Result \mid Opponent = Weak))$

---

3. Using the above observations, give the results of
   - $I(P(Result)) =$
   - $P(Result = Loss \mid Opponent = Strong) =$

   You do not have to compute irrational logarithms.

Solution: $I(P(Result)) = -1/2 \log_2 1/2 - 1/3 \log_2 1/3 - 1/6 \log_2 1/6$ and $P(Result = Loss \mid Opponent = Strong) = 1/3$.

Problem 6.3 (Decision Tree Learning in Python)

Implement the *Decision Tree Learning algorithm* (*DTL*) in Python using the files at `https://kwarc.info/teaching/AI/resources/AI2/dtl`.

Solution: See `https://kwarc.info/teaching/AI/resources/AI2/dtl`
