Assignment6 – MDP, Decision Trees

Given: June 2   Due: June 11

Problem 6.1 (Sunbathing) 0 pt

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?
2. Compute that quantity for the attributes Hair and Lotion. (Simplify as much as you can without computing logarithms.)
3. Assuming the logarithms are computed, how does the algorithm pick the attribute?

Solution:

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

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

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

   Entropy is undefined for 0. If we were to continue simplifying, we'd use $0 \cdot \log_2 0 = 0$.
3. It picks the one with the highest information gain (in this case Hair).

Problem 6.2 (Decision Trees) 40 pt

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.

2. How does the information-theoretic algorithm choose an attribute?

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

4. Using the above observations, give the results of

   - $I(P(Result)) =$
   - $P(Result = Loss|Opponent = Strong) =$

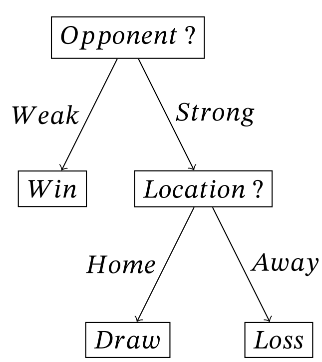
   You do not have to compute irrational logarithms.

5. Give a minimal set A of attributes such that $A \succ Result$ holds for the above observations.

6. Explain *why or why not* the determination $Day, Weather \succ Result$ holds for the above observations.

Solution:

1. The tree is



2. The algorithm chooses the attribute with the highest information gain.
3. $Gain(Opponent) = I(P(Result)) - P(Opponent = Strong) \cdot I(P(Result|Opponent = Strong)) - P(Opponent = Weak) \cdot I(P(Result|Opponent = Weak))$
4. $I(P(Result)) = -1/2 \log_2 1/2 - 1/3 \log_2 1/3 - 1/6 \log_2 1/6$ and $P(Result = Loss|Opponent = Strong) = 1/3$.
5. $A = \{Location, Opponent\}$ or $A = \{Weather, Opponent\}$
6. It does not hold. Games 4 and 6 agree on Day and $Weather$ but not on $Result$.

Problem 6.3 (Overfitting) 20 pt

Explain what *overfitting* means and why we want to avoid it.

Solution: Overfitting is a modeling error that occurs when the chosen hypothesis is too closely fit to a sample set of data points. It picks an overly complex hypothesis that also explains idiosyncrasies and errors in the data. A simpler hypothesis that fits the data less exactly is often a better match for the underlying mechanisms.

Problem 6.4 (Loss) 40 pt

Our goal is to find a linear approximation $h(x) = ax$ for the series of square numbers $0, 1, 4, 9, 16$.

1. Model this situation as an inductive learning problem.
2. Assuming all 5 possible examples are equality probable, compute the generalized loss using the squared error loss function. (This is a function of h.)
3. Find h^*.
4. What is the error rate of h^*?

Solution:

1. The inductive learning problem is $(\mathcal{H}, f)$ where
   - the hypothesis space $\mathcal{H}$ is the set containing all functions $h(x) = ax$ with $\mathrm{dom}(h) = \{0, \dots, 4\}$ for $a \in \mathbb{R}$
   - the target function is $f(x) = x^2$ with $\mathrm{dom}(f) = \{0, 1, \dots, 4\}$
2. Each example (x, x^2) has probability $1/5$. For each x, the loss is $L_2(x^2, ax) = (x^2 - ax)^2$. Thus for each $h(x) = ax$, we have

   $$GenLoss(h) = \sum_{x=0,\dots,4} (x^2 - ax)^2 \cdot 1/5 = ((1-a)^2 + (4-2a)^2 + (9-3a)^2 + (16-4a)^2)/5 = (354 - 200a + 30a^2)/5$$

3. We need to find the a that minimizes the loss. The derivative of $GenLoss$ for a is $(60a - 200)/5$. So the minimum is at $a = 10/3$.
4. The error rate is $4/5 = 1$ because $h^*(x) = 10x/3$ predicts 4 out of 5 examples incorrectly. (E.g., $h(x) = x$ would have better error rate $3/5$ despite having higher generalized loss.)
