Assignment6 – Learning

Given: June 5 Due: June 10

Problem 6.1 (Value Iteration for Navigation)

Implement value iteration for an agent navigating worlds like the 4x3 world from the *lecture notes*. The agent has four possible actions: *right*, *up*, *left*, *down*. The probability of actually moving in the intended direction is p and the probability of moving in one of the orthogonal directions is $\frac{1-p}{2}$ respectively. For example, if $p = 0.8$ and the chosen action is *up*, the agent will actually move up with a probability of $p = 0.8$ and will move left and right with a probability of 0.1 each. If the agent ends up moving in a direction that has no free adjacent square, it will remain on its current square instead. For example, if the agent is on square $(0, 0)$ with the action *up*, it will end up on square $(0, 1)$ with a probability of p, on square $(1, 0)$ with a probability of $\frac{1-p}{2}$ and on square $(0, 0)$ with a probability of $\frac{1-p}{2}$.

(0, 2) -0.040 → 0.647	(1, 2) -0.040 → 0.753	(2, 2) -0.040 → 0.855	(3, 2) 1.000 T 1.000
(0, 1) -0.040 ↑ 0.557	(1, 1) W	(2, 1) -0.040 ↑ 0.569	(3, 1) -1.000 T -1.000
(0, 0) -0.040 ↑ 0.465	(1, 0) -0.040 ← 0.386	(2, 0) -0.040 ↑ 0.451	(3, 0) -0.040 ← 0.230

Results for 4x3 world with $p = 0.8$, $\gamma = 0.95$, $\varepsilon = 0.001$.

A skeleton *implementation* with technical instructions can be found at `https://kwarc.info/teaching/AI/resources/AI2/mdp/`. It also allows the visualization of the computed utilities and policy (see figure above): Each square is annotated with the coordinates, the reward, the computed policy and the computed utility. Walls and terminal nodes don't have a policy and are marked with *W* and *T* respectively.

Hint: You will also have to compute a policy based on the utilities obtained from value iteration. For that, you should pick the actions that *maximize* the expected utility. A common mistake is the assumption that the best policy is always to go in the direction of the square with the maximal utility.

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

Problem 6.2 (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:

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

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

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

   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).

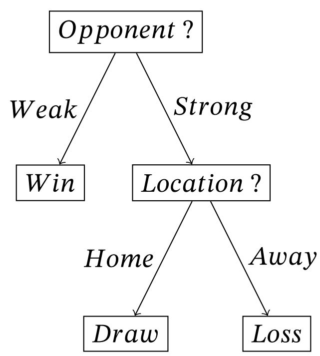
Problem 6.3 (Decision Trees)

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 6.4 (Decision Trees)

Consider a *word* W chosen uniformly from $\{bad, bed, bend, pend, pad, ped\}$. You are allowed to ask the following questions about W:

- A length of the word
- B first letter of the word
- C second letter of the word
- D last letter of the word

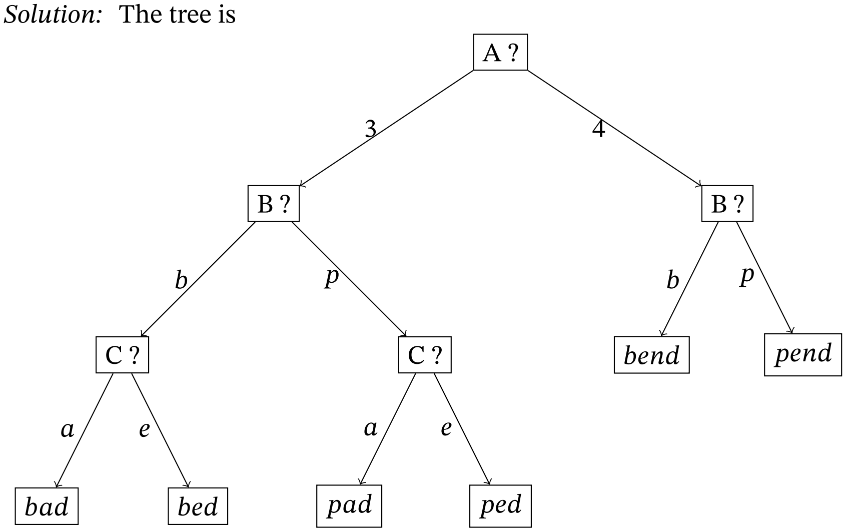
1. Show that there is no *decision tree* for W of depth 2.

Solution: All questions have at most 2 possible answers, so a *decision tree* of *depth* 2 has at most 4 *leaves*. But we need at least 5 *leaves* to cover all options for W.

---

2. Draw the *decision tree* for W that arises from asking the questions in the order A,B,C,D. (Do not ask additional questions if the *word* can already be identified.)

Solution: The tree is



---

3. Calculate the *information gain* for all 4 questions.

Solution: A, C: $-1/3\log_2 1/3 - 2/3\log_2 2/3$
B: $-1/2\log_2 1/2 - 1/2\log_2 1/2 = 1$
D: $-1\log_2 1 = 0$

---

4. Which question would the *information gain algorithm* ask first?

Solution: B

We write $Q \succ A$ if the set Q of attributes determines the attribute A, i.e., examples that agree on all attributes in Q also agree on A.

5. Give all minimal sets $Q \subseteq \{A, B, C, D\}$ of questions for which the *determination* $Q \succ W$ holds?

Solution: $\{A, B, C\}$

---

6. Give two *words* such that removing them from the choices for W makes the *determination* $\{A, B\} \succ W$ hold.

Solution: Any pair out of $\{bad, bed\} \times \{pad, ped\}$
