CS3061: Final Exam

Conor McCauley - 17323203

May 1, 2020

Declaration

I understand that this is an individual assessment and that collaboration is not permitted. I have not received any assistance with my work for this assessment. Where I have used the published work of others, I have indicated this with appropriate citation. I have not and will not share any part of my work on this assessment, directly or indirectly, with any other student.

I have read and I understand the plagiarism provisions in the General Regulations of the University Calendar for the current year, found at http://www.tcd.ie/calendar. I have also completed the Online Tutorial on avoiding plagiarism 'Ready Steady Write', located at http://tcd-ie.libguides.com/plagiarism/ready-steady-write.

I understand that by returning this declaration with my work, I am agreeing with the above statement.

Name: Conor McCauley **Date:** 01/05/2020

Question 2

(a) From a state s_{i-1} the MDP can transition to any state $s_i \in S$ through any action $a_i \in A$, provided the probability of that transition occurring, $p(s_{i-1}, a_i, s_i)$, is greater than zero. In other words, the following transitions are possible:

$$s_{i-1} \xrightarrow{a_i} s_i, r_i \quad s_i \in S,\, a_i \in A,\, p(s_{i-1}, a_i, s_i) > 0$$

(b) The transitions in (3) suggest that there are positive rewards for the transitions as the sequence has not yet terminated and that there are few probabilities with values of 0 or 1 because, as before, the sequence has not yet terminated and has continued to progress.

If we were to discover that the same state, s_x, is occurring repeatedly we would have to reconsider both assumptions as $p(s_x, a, s_x)$ could be equal to 1, lowering our confidence in the latter assumption, and the reward for those transitions could be minimal, lowering our confidence in the former assumption.

Considering multiple sequences would allow us to determine which states can transition to others through certain actions but also what the probability of those transitions occurring is.

(c) The γ-discounted value of $q_1(s_1, a_2)$ can be found by evaluating the following expression:

$$q_1(s_1, a_2) = \sum_{s' \in S} p(s_1, a_2, s')(r(s_1, a_2, s') + \gamma \max_{a' \in A} q_0(s', a'))$$

First, we can find both maximum values of q_0: $\max_{a' \in A} q_0(s_1, a')$ and $\max_{a' \in A} q_0(s_2, a')$:

$$q_0(s_1, a_1) = p(s_1, a_1, s_1)r(s_1, a_1, s_1) + p(s_1, a_1, s_2)r(s_1, a_1, s_2) = 0.6 \cdot 7 + 0.4 \cdot 0 = 4.2$$

$$q_0(s_1, a_2) = p(s_1, a_2, s_1)r(s_1, a_2, s_1) + p(s_1, a_2, s_2)r(s_1, a_2, s_2) = 0.7 \cdot 0 + 0.3 \cdot 15 = 4.5$$

$$\max_{a' \in A} q_0(s_1, a') = \max(4.2, 4.5) = 4.5$$

$$q_0(s_2, a_1) = p(s_2, a_1, s_1)r(s_2, a_1, s_1) + p(s_2, a_1, s_2)r(s_2, a_1, s_2) = 0.5 \cdot 0 + 0.5 \cdot 3 = 1.5$$

$$q_0(s_2, a_2) = p(s_2, a_2, s_1)r(s_2, a_2, s_1) + p(s_2, a_2, s_2)r(s_2, a_2, s_2) = 0.5 \cdot 0 + 0.5 \cdot 2 = 1.0$$

$$\max_{a' \in A} q_0(s_2, a') = \max(1.5, 1.0) = 1.5$$

We can now rewrite our initial expression like so:

$$q_1(s_1, a_2) = p(s_1, a_2, s_1)(r(s_1, a_2, s_1) + \frac{1}{3} \cdot 4.5) + p(s_1, a_2, s_2)(r(s_1, a_2, s_2) + \frac{1}{3} \cdot 1.5)$$

$$= 0.7 \cdot (0 + 1.5) + 0.3 \cdot (15 + 0.5)) = 5.7$$

(d) Since all transition probabilities are 1 we can say both actions are s-deterministic and ignore the values of $p(s, a, s')$ in our calculations. Given that s_3 is an absorbing state, $p(s_3, a, s_3) = 1$ for all $a \in A$, we can calculate the following values for s_3:

$$Q(s_3, a_1) = r(s_3, a_1, s_3) + \gamma \frac{\max_{a' \in A} r(s_3, a', s_3)}{1 - \gamma} = 4 + 0.5 \cdot \frac{\max(4, 4)}{0.5} = 8$$

$$Q(s_3, a_2) = r(s_3, a_2, s_3) + \gamma \frac{\max_{a' \in A} r(s_3, a', s_3)}{1 - \gamma} = 4 + 0.5 \cdot \frac{\max(4, 4)}{0.5} = 8$$

We can focus solely on the rewards for s_2 as the probability is always 1. Using our results for s_3 we can calculate the following values for s_2:

$$Q(s_2, a_1) = \lim_{n \to \infty} q_n(s_2, a_1) = r(s_2, a_1, s_3) + \gamma \max_{a' \in A} Q(s_3, a') = 2 + 0.5 \cdot \max(8, 8) = 6$$

$$Q(s_2, a_2) = \lim_{n \to \infty} q_n(s_2, a_2) = r(s_2, a_2, s_3) + \gamma \max_{a' \in A} Q(s_3, a') = 2 + 0.5 \cdot \max(8, 8) = 6$$

Again, focusing solely on the rewards and using the values from our previous calculations, we can calculate the following values for s_1:

$$Q(s_1, a_1) = \lim_{n \to \infty} q_n(s_1, a_1) = r(s_1, a_1, s_2) + \gamma \max_{a' \in A} Q(s_2, a') = 2 + 0.5 \cdot \max(6, 6) = 5$$

$$Q(s_1, a_2) = \lim_{n \to \infty} q_n(s_1, a_2) = r(s_1, a_2, s_3) + \gamma \max_{a' \in A} Q(s_3, a') = 1 + 0.5 \cdot \max(8, 8) = 5$$

Question 3

(a) True. The learning rate, α, roughly determines how much we value previously discovered information versus new information. If the learning rate is high the program will be more likely to explore or consider new information but if the learning rate is low it will be more likely to exploit or depend on old information instead.
(b) True. As Datalog is not Turing-complete it is not guaranteed to find an answer (i.e. the goal) for a given program for which an answer exists. This is because Datalog is guaranteed to terminate without regard to whether an answer exists or not. Prolog, however, is Turing-complete.
(c) True. Abduction is, roughly speaking, the inference of facts using outputs while deduction is the inverse: the inference of outputs using facts.
(d) True. Bayes networks are represented via directed acyclic graphs (DAGs), and, as such, the direction (order) of the nodes (variables) in the graph determine the causes and their respective effects.
(e) True. The conditional independences in a Bayes network depend on there being determinable causes and effects via the aforementioned DAG. When a Bayes network is moralised to a Markov network any such direction or ordering is lost which causes the conditional independences to also be lost.