Trinity College Dublin
Coláiste na Tríonóide, Baile Átha Cliath
The University of Dublin

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: Davy Nolan

Date: 01/05/2020

CSU33061 Artificial Intelligence 1 Exam 2020

Student name: Davy Nolan

Student number: 17330208

Question 2

(a)

Given an MDP **(S, A, p, r, y)**, for a set of states **S** and actions **A**, you can transition between the state s_n to $_{sn+1}$ by an action a_n. Therefore there are **SAS** possible transitions.

(b)

⇒ The transitions suggest the actions in the sequence are S-Deterministic as p(s, a, s') equals 1 for some s'. This implies that there is only one possible s' given a and s.

⇒ Sequence (3) is S-Deterministic since there is only one action for each state. Since this is the only sequence that is sampled from the MDP so our confidence decreases in this suggestion. A policy tells us what action to take. A policy for the MDP could increase our confidence if it states one available action from each state.

⇒ When trying to determine the probability and reward function it might be more useful to sample more sequences as the more sequences we have to sample from the more accurate the probability and reward functions will be. Overall the p and r functions will behave much better allowing the agent to learn more efficiently.

(c)

$S = \{s_1,s_2\}$ $A = \{a_1,a_2\}$ $y = \frac{1}{3}$

$q_0(s, a) = p(s, a, s_1)*r(s, a, s_1) + p(s, a, s_2)*r(s, a, s_2)$
$V_n(s) = max(q_n(s, a_1), q_n(s, a_2))$
$q_{n+1}(s, a) = q_0(s, a) + y[p(s, a, s_1)*V_n(s_1) + p(s, a, s_2)*V_n(s_2)]$

$q_1(s_1, a_2) = q_0(s_1, a_2) + \frac{1}{3}[p(s_1, a_2, s_1)*V_0(s_1) + p(s_1, a_2, s_2)* V_0 (s_2)]$
$= q_0 (s_1, a_2) + \frac{1}{3}[(0.7)* V_0 (s_1) + (0.3)* V_0 (s_2)]$

Calculating $V_0(s_1)$
$V_n = V_0 (s_1) = max(q_0 (s_1, a1), q_0 (s_1, a_2))$

$q_0(s_1, a_1) = p(s_1, a_1, s_1)*r(s_1, a_1, s_1) + p(s_1, a_1, s_2)*r(s_1, a_1, s_2) = (0.6)(7) + (0.4)(0) = 4.2$
$q_0 (s_1, a_2) = p(s_1, a_2, s_1)*r(s_1, a_2, s_1) + p(s_1, a_2, s_2)*r(s_1, a_2, s_2) = (0.7)*(0) + (0.3)(15) = 4.5$
$\therefore V_0 (s_1) = max(4.2, 4.5) = 4.5$

Calculating V_0(s2)
$V_n = V_0 (s_1) = max(q_0 (s_2, a_1), q_0 (s_2, a_2))$

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)(0)+(0.5)(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)(0)+(0.5)(2) = 1
∴ V_0 (s_2) = max(1.5, 1) = 1.5

q_0 (s_1, a_2) = 4.5
∴ q_1 (s_1, a_2) = 4.5 + $\frac{1}{3}$[(0.7)(4.5)+(0.3)(1.5)] = 5.7

(d)

Formula:
Q(s, a) = $\sum_{s' \in S} p(s, a, s') * (r(s, a, s') + y\ max_{a' \in A}(Q(s', a'))$

Absorbing formula:
Q(s, a) = r(s, a, s) + yV(s)
V(s) = $\frac{r_s}{1-y}$ where r_s =max r(s, a, s)

Using formula:
Q(s_1, a_1) = p(s_1, a_1, s_2)*(r(s_1, a_1, s_2)+$\frac{1}{2}$(max(Q(s_2, a_1), Q(s_2, a_2)))
Q(s_1, a_2) = p(s_1, a_2, s_3)*(r(s_1, a_2, s_3)+$\frac{1}{2}$(max(Q(s_3, a_1), Q(s_3, a_2)))
Q(s_2, a_1) = p(s_1, a_1, s_3)*(r(s_1, a_1, s_3)+$\frac{1}{2}$(max(Q(s_3, a_1), Q(s_3, a_2)))
Q(s_2, a_2) = p(s_1, a_2, s_3)*(r(s_1, a_2, s_3)+$\frac{1}{2}$(max(Q(s_3, a_1), Q(s_3, a_2)))

Using absorbing formula:
Q(s_3, a_1) = r(s_3, a_1, s_3) + y($\frac{r_{s3}}{1-y}$) = 4 + $\frac{1}{2}$ ($\frac{\max(4,4)}{1-\frac{1}{2}}$) = 4 + $\frac{1}{2}$ ($\frac{4}{1/2}$) = 8
Q(s_3, a_2) = r(s_3, a_2, s_3) + y($\frac{r_{s3}}{1-y}$) = 4 + $\frac{1}{2}$ ($\frac{\max(4,4)}{1-\frac{1}{2}}$) = 4 + $\frac{1}{2}$ ($\frac{4}{1/2}$) = 8

Backwards subbing into original formulas:
Q(s_2, a_2) = (1)*(2 + $\frac{1}{2}$(max(8, 8))) = (1)*(2+4) = 6
Q(s_2, a_1) = (1)*(2 + $\frac{1}{2}$(max(8, 8))) = (1)*(2+4) = 6
Q(s_1, a_2) = (1)*(1 + $\frac{1}{2}$(max(8, 8))) = (1)*(1+4) = 5
Q(s_1, a_1) = (1)*(1 + $\frac{1}{2}$(max(6, 6))) = (1)*(2+3) = 5

Q(s_1, a_1) = 5
Q(s_1, a_2) = 5
Q(s_2, a_1) = 6
Q(s_2, a_2) = 6
Q(s_3, a_1) = 8
Q(s_3, a_2) = 8

Question 3

(a)

TRUE

The learning rate determines to what extent new information overrides old information. A factor of 0 makes the agent learn nothing – exploiting prior knowledge, while a factor of 1 makes the agent consider only the most recent information – ignoring prior knowledge to explore possibilities.

Exploration is attempting to discover new features about the environment by performing an action which is different from past experience that indicates the correct decision given a certain0 environment. Exploitation is the concept of repeatedly performing same actions in the same environments because it results in the current maximum reward.

Since the learning rate controls whether the agent is learning more (exploring) or learning nothing (exploiting), the exploration-exploitation trade-off in Q-learning does depend on the learning rate.

(b)

TRUE

Arity is another name for the number of arguments that a predicate/functor has. Therefore for all predicates to have arity 0 means that all predicates have no arguments. If a mechanical procedure for logical consequence is goal-directed, this means that the agent will search through until a true statement has been reached.

Since all predicates have arity 0, all predicates have no arguments. Goal directed means the procedure will go through until a true statement is found. However, since the predicates have no arguments, the mechanical procedure in unable to complete. Therefore, this statement is true.

(c)

FALSE

Abduction is a form of reasoning where assumptions are made to explain observations. Deduction is a method of proving a hypothesis using scientific data without making any assumptions.

Abduction and deduction are both methods of reasoning for a hypothesis however abductions includes assumptions. This does not mean that abduction is the opposite/inverse of deduction i.e. abduction(deduction(x)) $\neq$ x.

(d)

TRUE

In a Bayesian network, each edge is a conditional dependency and each node is a random variable. A variable node s1 is independent from node s2 if s2 does not descend from s1. Since the cause node is the parent of the effect node, the cause node must come before the effect node. Therefore the random variables must be ordered so that the causes come before their effects.

(e)

TRUE

Moralization is the procedure for converting a Bayesian network into a Markov network. The moral graph M(G) of a Bayesian Network G = (V,E) is an undirected graph over V that contains an undirected edge between X_i and X_j if there is a directed edge between them and if X_i and X_j are both parents of the same node.

Let A, B and C be nodes where A and B are independent of each other but they both point to C (i.e. A and B are parent nodes of C). Moralizing this graph will "marry the parents" of the C node, adding an edge between them. Therefore A and B will lose their independence.