THE UNIVERSITY
OF ADELAIDE
AUSTRALIA

Semester 1 2015

Artificial Intelligence
COMPSCI 3007, 7059

Official Reading Time:	10 mins
Writing Time:	120 mins
Total Duration:	130 mins

Questions	Time	Marks
Answer all 6 questions	120 mins	120 marks
		120 Total

Instructions

- Examination material must not be removed from the examination room.
- No calculators allowed. Answers requiring numerical calculation may be left in the form of arithmetic expressions.
- English-other language paper dictionaries allowed.

Materials

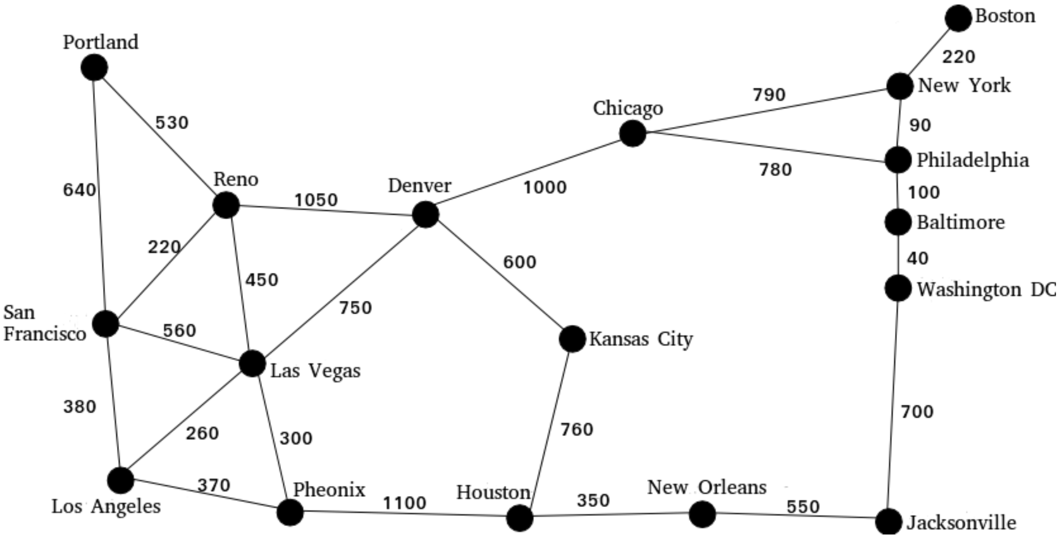
- 1 answer booklet

DO NOT COMMENCE WRITING UNTIL INSTRUCTED TO DO SO

Problem solving by searching

Question 1

(a) The following diagram shows a highway network that connects several cities in a certain country. The distance between any two cities connected by a highway is shown on the network.



You are currently in San Francisco and you wish to travel via road to Boston. You wish to use search methods to find a path that connects the two cities.

i. What is the branching factor of the search tree?

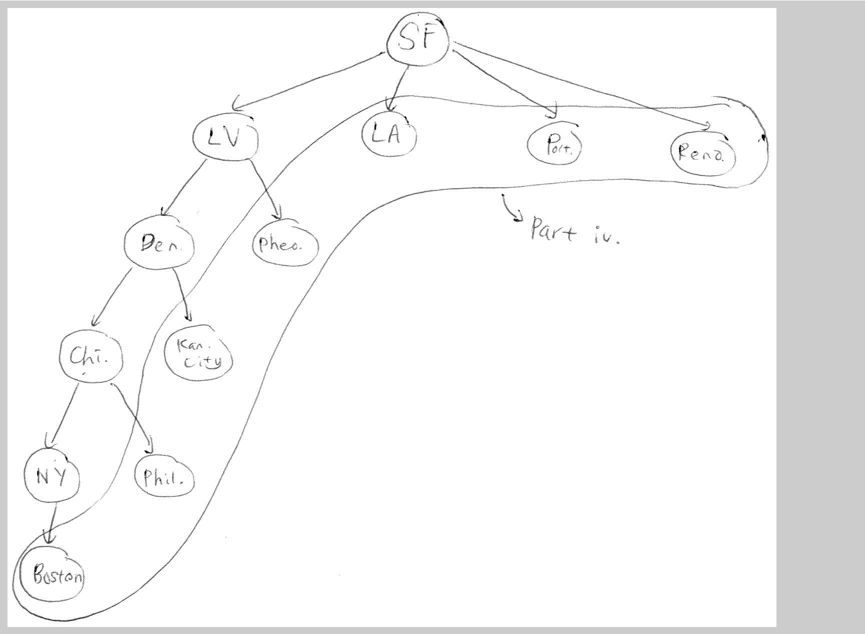
[1 mark]

Solution:
Branching factor is 5. [1 mark]

ii. Draw the search tree generated by depth first search (DFS) until a path between San Francisco and Boston is found. Remember to avoid repeated states. The order of expansion is based on the alphabetical order of the name of the cities.

[6 marks]

Solution:



iii. Give one advantage of depth first search (DFS) over breadth first search (BFS).

[2 marks]

Solution:
DFS has linear space/memory complexity while BFS has exponential space/memory complexity. [2 marks]
Can also accept the answers in big-O notation.

iv. In the course of generating the search tree by DFS in part ii above, what was the highest number of nodes you needed to store in the fringe (the list of yet to be expanded nodes)?

Outline in the tree you drew in part ii the set of nodes that exist in the fringe when its size was the largest.

[2 marks]

Solution:
See solution for part ii.

(b) The direct flight distance (DFD) (distance travelled by an air plane) between several cities and Boston is given in Table 1 in Page 3.

City	DFD to Boston
San Francisco	2700
Portland	2530
Reno	2520
Las Vegas	2370
Los Angeles	2600

Table 1: DFD between several cities and Boston.

For the other cities in the network, their DFD to Boston is unknown or not available.

A function $h(n)$ that takes as input a node n is defined as follows:

$$h(n) = \begin{cases} \text{DFD between city}(n) \text{ and Boston} & \text{if the DFD is available;} \\ 0 & \text{otherwise,} \end{cases}$$

where city(n) gives the city encapsulated in node n.

If the $h(n)$ defined above is used as a heuristic function in A* search to find a path between San Francisco and Boston, is the solution found guaranteed to be the shortest path? Justify your answer.

[3 marks]

Solution:

The solution is guaranteed to be the shortest path. [1 mark]

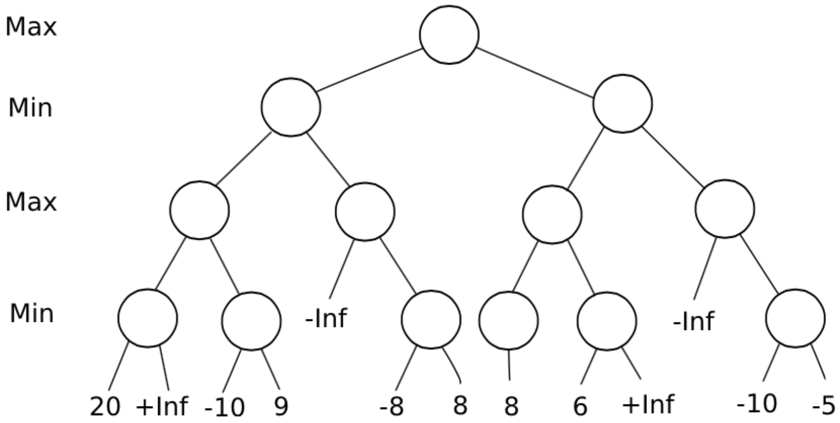
Although $h(n) = 0$ for some n, this still satisfies the admissibility condition. [2 marks]

[Total for Question 1: 14 marks]

Adversarial search

Question 2

(a) The following diagram shows a game tree to be searched based on the minimax algorithm. The first move belongs to Max. In the diagram, +Inf means $+\infty$ and -Inf means $-\infty$.

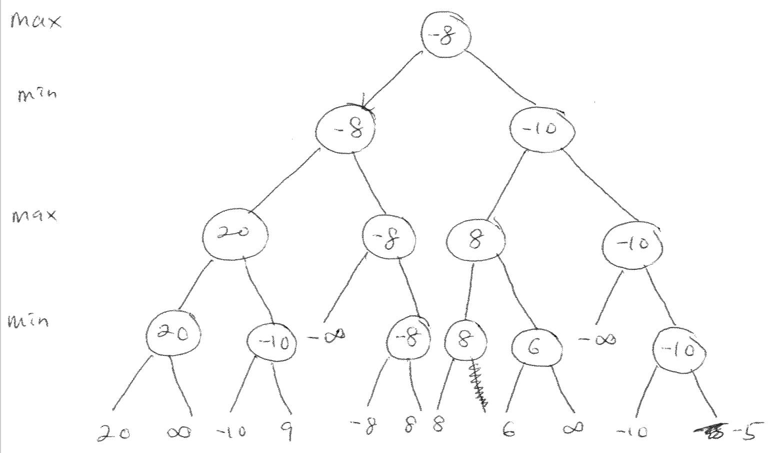


Copy this game tree into your answer book. Then,

i. Write the minimax value at each node.

[3 marks]

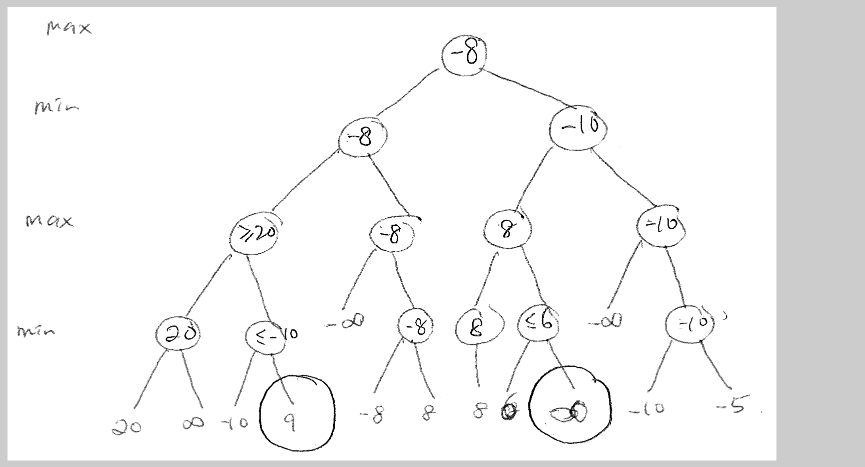
Solution:



ii. If *alpha-beta pruning* is to be used to search the game tree, clearly circle the branches that are pruned (i.e, branches that do *not* need to be searched). Note: at each node, search the child nodes from left to right.

[8 marks]

Solution:



(b) Here's a partially completed tic-tac-toe game. Player Max uses symbol X while player Min uses symbol O. The utility of Max wins is +1, draw is 0, and Min wins is -1.

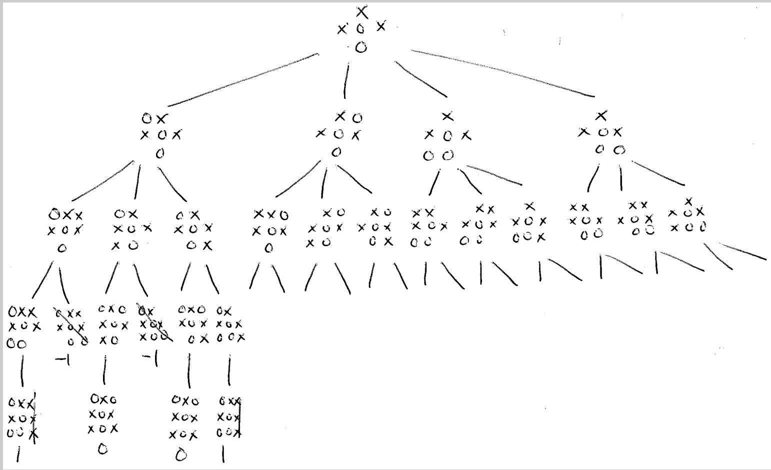
	X	
X	O	X
	O	

It's Min's turn to choose a move.

i. Draw the game tree starting from the state above until the end of the game. Write clearly the minimax value at each node of the tree.

[12 marks]

Solution:



ii. Starting from the state above, if both Min and Max play optimally, what is the best possible final outcome for Min?

[2 marks]

Solution:
Min will win.

[Total for Question 2: 25 marks]

Decision trees

Question 3

(a) Why is it important to conduct pruning in decision tree learning?

[1 mark]

Solution:
To prevent overfitting. [1 mark]
or
To improve the generalisation ability of the decision tree on new data. [1 mark]

(b) Post-pruning is a framework to prune decision trees. Briefly explain

i. What kinds of data are required to conduct post-pruning; and
ii. How to conduct post-pruning given the required data.

[4 marks]

Solution:

i. A set of training data [1 mark] and a set of validation/testing data [1 mark].

ii. First grow the tree fully using the training data [1 mark] and remove leaf nodes one-by-one if the prediction accuracy of the tree on the validation data improves [1 mark].

(c) Prof Dumbledore teaches the Artificial Intelligence (AI) course. In this semester, he wishes to predict whether the students will pass the course, before they take the final exam. He collected a set of data from 20 students who took the AI course last semester. Specifically, he recorded the marks they had obtained in the data structure course, the amount of time they had spent on studying for the AI course, and whether the students eventually passed AI. The following diagram plots the collected data.

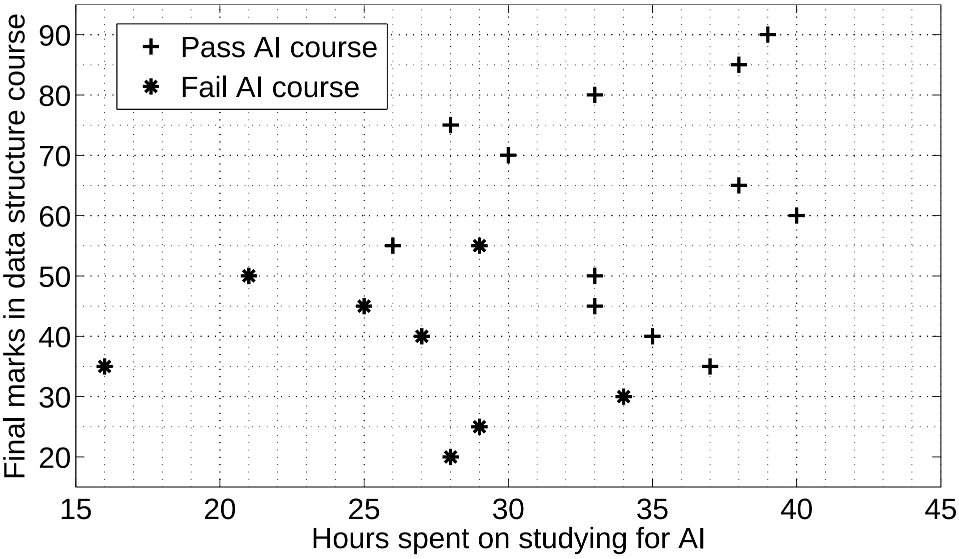
As his loyal student, help Prof Dumbledore to build a decision tree that can predict whether a student from this semester will pass the AI course, given the same type of data.

i. What is the information content at the root node of the decision tree? Leave your answer in the form of an arithmetic expression.

[4 marks]

Solution:

$$I(root) = -\left(\frac{8}{20}\right) log_2 \left(\frac{8}{20}\right) - \left(\frac{12}{20}\right) log_2 \left(\frac{12}{20}\right)$$



ii. Calculate the information gain for the following two candidate tests at the root node:

- Final marks in data structure course ≤ 57.
- Hours spent studying for AI ≤ 31.

Leave your answers in the form of arithmetic expressions.

[8 marks]

Solution:
For (Final marks in data structure course ≤ 57):

$$I(left) = -\left(\frac{8}{13}\right) log_2 \left(\frac{8}{13}\right) - \left(\frac{5}{13}\right) log_2 \left(\frac{5}{13}\right) \quad \text{[1 mark]}$$

$$I(right) = -\left(\frac{0}{7}\right) log_2 \left(\frac{0}{7}\right) - \left(\frac{7}{7}\right) log_2 \left(\frac{7}{7}\right) = 0 \quad \text{[1 mark]}$$

$$Gain = I(root) - \left(\frac{13}{20}\right) I(left) - \left(\frac{7}{20}\right) I(right) \quad \text{[2 marks]}$$

For (Hours spent studying for AI ≤ 31):

$$I(left) = -\left(\frac{7}{10}\right) log_2 \left(\frac{7}{10}\right) - \left(\frac{3}{10}\right) log_2 \left(\frac{3}{10}\right) \quad \text{[1 mark]}$$

$$I(right) = -\left(\frac{1}{10}\right) log_2 \left(\frac{1}{10}\right) - \left(\frac{9}{10}\right) log_2 \left(\frac{9}{10}\right) = 0 \quad \text{[1 mark]}$$

$$Gain = I(root) - \left(\frac{10}{20}\right) I(left) - \left(\frac{10}{20}\right) I(right) \quad \text{[2 marks]}$$

iii. If we split based on the attribute ‘Hours spent on studying for AI’ at the root node,

- What is the *minimum* number of unique split values that we need to examine in order to find the test that maximises the information gain?

- Write down a set of unique split values that we can examine to find the test that maximises the information gain.

[4 marks]

Solution:
14 unique split values to examine. [1 mark]
One number from each of the ranges below:

- (16,21)
- (21,25)
- (25,26)
- (26,27)
- (27,28)
- (28,29)
- (29,30)
- (30,33)
- (33,34)
- (34,35)
- (35,37)
- (37,38)
- (38,39)
- (39,40)

[3 marks if all correct, 1 or 2 marks if partially correct]

[Total for Question 3: 21 marks]

Clustering

Question 4

(a) i. What is unsupervised learning? Please describe in terms of the input (or features) and labels (class) of the training data.

[2 marks]

Solution: Training data have only input (or features). No labels (or class) are given.

ii. What is clustering? Is it unsupervised or supervised?

[2 marks]

Solution: Clustering find the groups of the data. (1 mark) No label/class information is given, thus it is unsupervised. (1 mark)

iii. Describe K-means algorithm. Suppose we know there are K clusters.

[4 marks]

Solution: (1) Randomly initialise K centroids (either picking them at random from the training data, or generate random points within the range of the training data) (1 mark); (2) Assign each training data point to its nearest centroid (1 mark); (3) Use the average of the points belonging to the cluster to update the centroid (1 mark); (4) Repeat step (2) and (4) until the centroids do not change much (1 mark).

iv. Name one problem with K-means algorithm (from many potential problems).

[2 marks]

Solution: Any one from below receives the full mark. (1) It requires the number of clusters K is known or given, which can be very difficult to obtain; (2) Results from K-means is non-deterministic. Different initialisations will give different results; (3) k-means assume the variance of the distribution of each attribute (variable) is spherical; (4) all variables have the same variance; (5) the prior probability for all k clusters are the same, i.e. each cluster has roughly equal number of observations.

(b) i. When Mean Shift is used in clustering, does it require to know how many clusters (like K-means does)? Please explain why.

[3 marks]

Solution: No (1 mark).
Mean Shift uses all data points as initial centroids. These centroids will move iteratively. Eventually, many centroids will overlap. The number of distinct (or at least not far apart) centroids is the number of clusters the data should have (2

marks).

ii. Assume there are N many data $\{x_i\}_{i=1}^N$, and the kernel function is $k(x_i, x_j)$. If an old mean vector is z, please write down the mathematical expression for the updated mean vector z_{new}.

[2 marks]

Solution:

$$z_{new} = \frac{\sum_{i=1}^N k(z, x_i) x_i}{\sum_{i=1}^N k(z, x_i)}$$

[Total for Question 4: 15 marks]

Probabilistic Graphical Models

Question 5

(a) The joint distribution for three boolean variables A, B, C is given in Table 2. Please compute the following probabilities.

a	b	c	0.01
a	b	$\neg c$	0.01
a	$\neg b$	c	0.06
a	$\neg b$	$\neg c$	0.02
$\neg a$	b	c	0.04
$\neg a$	b	$\neg c$	0.04
$\neg a$	$\neg b$	c	0.80
$\neg a$	$\neg b$	$\neg c$	0.02

Table 2: $P(A, B, C)$

i. What is $P(B = b, C = \neg c)$?

[3 marks]

Solution: 0.05

ii. What is $P(C = \neg c)$?

[3 marks]

Solution: 0.09

iii. What is $P(B = b|C = \neg c)$?

[4 marks]

Solution: 0.05/ 0.09 = 5/9
or equivalently 0.5556

(b) A Bayesian network with 3 boolean variables A, B, C has a graph in Figure 1 with the local (conditional) distributions provided in the Table 3.

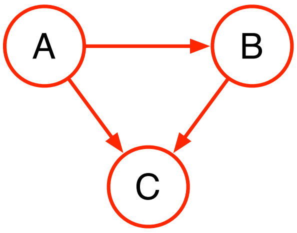


Figure 1: Bayesian Network

i. What is $P(A = a, B = \neg b, C = \neg c)$? Please write down the derivation and intermediate result instead of the final number alone.

[4 marks]

Table 3: Local (conditional) distributions $P(A), P(B|A), P(C|A,B)$

a	$\neg a$
0.3	0.7

	b	$\neg b$
a	0.6	0.4
$\neg a$	0.3	0.7

	c	$\neg c$
$a,\ b$	0.9	0.1
$a, \neg b$	0.4	0.6
$\neg a,\ b$	0.3	0.7
$\neg a, \neg b$	0.2	0.8

Solution:

$P(A = a, B = \neg b, C = \neg c)$
$= P(A = a)P(B = \neg b|A = a)P(C = \neg c|A = a, B = \neg b)$ (2 marks)
$= 0.3 \times 0.4 \times 0.6$
$= 0.072.$ (2 marks)

ii. What is $P(A = a|C = \neg c)$? Let $x = P(A = \neg a, C = \neg c)$. You can figure out the exact number of x from the table. But to save you some time, please express $P(A = a|C = \neg c)$ as a function of x (do not give us the number). Please write down the derivation and intermediate result instead of the final function alone.

[8 marks]

Solution:

$$
\begin{aligned}
&P(A=a,B=b,C=\neg c)\\
&=P(A=a)P(B=b|A=a)P(C=\neg c|A=a,B=b)\\
&=0.3\times 0.6\times 0.1\\
&=0.018.\texttt{(2 marks)}
\end{aligned}
$$

$$
\begin{aligned}
&P(A=a,C=\neg c)\\
&=P(A=a,B=\neg b,C=\neg c)+P(A=a,B=b,C=\neg c)\\
&=0.072+0.018\\
&=0.09.\texttt{(2 marks)}
\end{aligned}
$$

$$
\begin{aligned}
&P(C=\neg c)\\
&=P(A=a,C=\neg c)+P(A=\neg a,C=\neg c)\\
&=0.09+x.\texttt{(2 marks)}
\end{aligned}
$$

$$
\begin{aligned}
&P(A=a|C=\neg c)\\
&=\frac{P(A=a,C=\neg c)}{P(C=\neg c)}\\
&=\frac{0.09}{0.09+x}.\texttt{(2 marks)}
\end{aligned}
$$

iii. If the edge from A to B is deleted, will the local distribution tables be changed? If not, why? If yes, which one will be changed and becomes what?

[2 marks]

Solution: YES (1 mark).
The middle table $P(B|A)$ will be changed to $P(B)$ (1 mark).

iv. If the edge from A to B is deleted, is A independent to B? Prove it.

[4 marks]

Solution: Yes (2 marks).

$$
\begin{aligned}
P(A,B)&=\sum_C P(A,B,C)\\
&=\sum_C P(A)P(B)P(C|A,B)\\
&=P(A)P(B)\sum_C P(C|A,B)\\
&=P(A)P(B)
\end{aligned}
$$

(2 marks).

v. Which one of the following two inference methods can also be used to compute $P(C)$: max-product or sum-product?

[2 marks]

Solution: Sum-product.

[Total for Question 5: 30 marks]

Markov Decision Processes and Particle Filters

Question 6

(a) i. What is the Markov property?

[2 marks]

Solution: One of the following three versions is fine.
(1) The Markov property is the independence of the future from the past, given the present.
(2)The next state depends on the current state only (not the previous states).
(3) The state transition depends only on the current state not on the full history.

ii. In Markov Decision Processes, what is the meaning of a policy?

[1 mark]

Solution: A policy is a function that maps a state to an action.

iii. Let $[1, 2, 3]$ be the rewards for a sequence of states. Let the discount factor be 0.5. What is the discount utility for this sequence? What is the (additive) utility for this sequence?

[2 marks]

Solution: Discount utility: $1+0.5*2+0.5^2*3 = 2.75$ (1 mark)
Utility: $1 + 2 + 3 = 6$ (1 mark)

(b) Table 4 shows a map of a robot's world in which each grid square represents a discrete state of the robot. There are 5 states $a, b, ..., e$ in total where each state represents the robot's location. The rewards are shown on each state/location. Action: the robot can only move WEST or EAST, or EXIT (only available in exit states a and e).

5	0	0	0	1
a	b	c	d	e

Table 4: Rewards on a robot's map

i. If the discount factor $\gamma = 1$, what are the preferred actions (i.e. the output of the optimal policy) in states c and d?

[4 marks]

Solution: c: WEST (2 marks); d: WEST (2 marks)

ii. If the discount factor $\gamma = 0.1$, what are the preferred actions (i.e. the output of the optimal policy) in states c and d?

[4 marks]

Solution: c: WEST (2 marks); d: EAST (2 marks)

iii. What is the value of γ that makes WEST and EAST equally good in state d?

[2 marks]

Solution: $\gamma = \sqrt{(1/5)}$.

[Total for Question 6: 15 marks]

End of exam