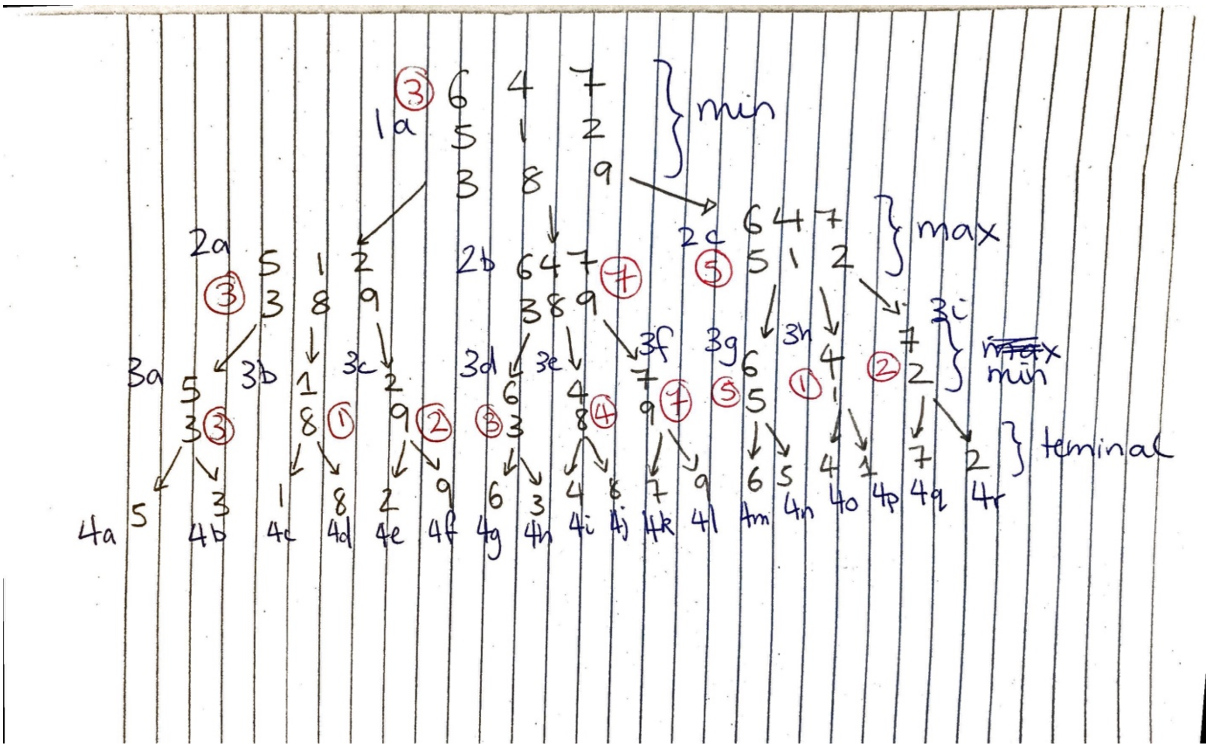
Homework 3

A)



b)

Step	Node	α(in)	β (in)	Value (out)	Children Skipped
1	1a	- ∞	+ ∞	3	
2	2a	- ∞	+ ∞	3	
3	3a	- ∞	+ ∞	3	
4	4a	- ∞	+ ∞	5	
5	4b	- ∞	5	3	
6	3b	3	+ ∞	1	4d
7	4c	3	+ ∞	1	
8	3c	3	+ ∞	2	4f
9	4e	3	+ ∞	2	
10	2b	- ∞	3	3	3e, 3f
11	3d	- ∞	3	3	
12	4g	- ∞	3	6	
13	4h	- ∞	3	3	
14	2c	- ∞	3	5	3h, 3i
15	3g	- ∞	3	5	
16	4m	- ∞	3	6	
17	4n	- ∞	3	5	

c)

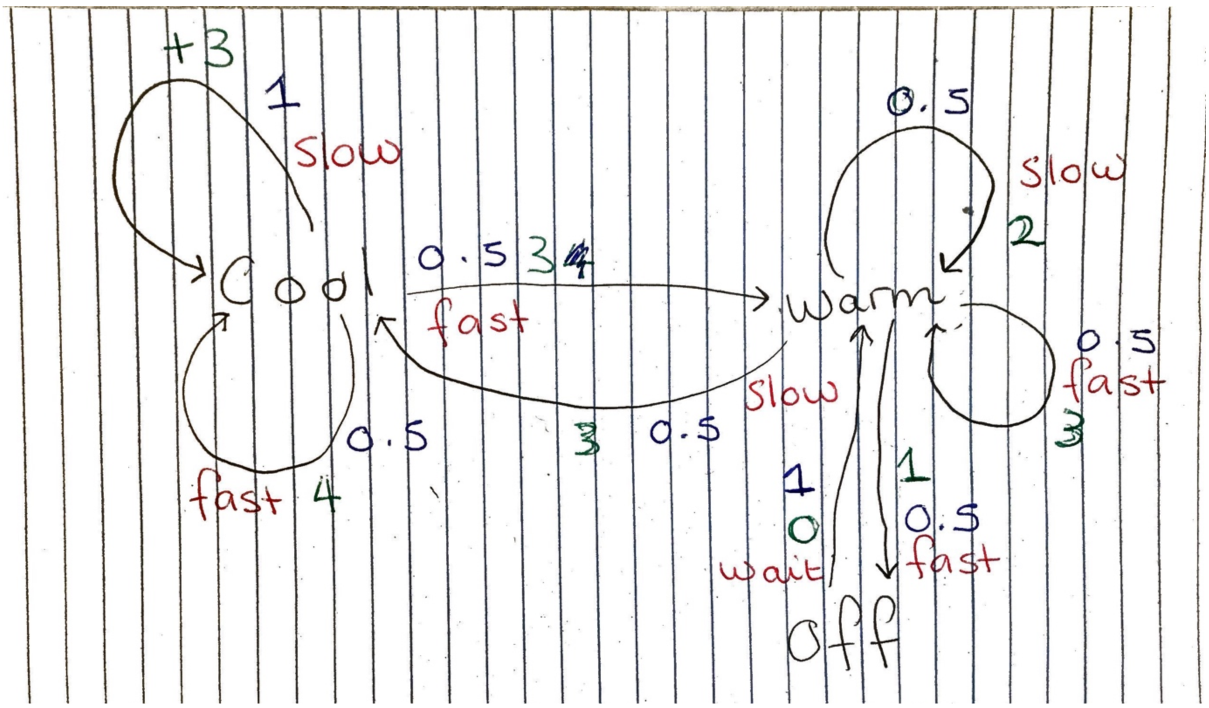
What has changed from the game tree is that there is no rational value to choose, and there is an equal chance that any of the values are picked. That means that P1 could greater minimise their score if the column [1, 8] or [2, 9] was chosen. The best move for P1 would be the same move as from the alpha-beta search to choose to remove the first row. However, the expected utility at the end of the game would no longer be 3, but 2.

d)

For the evaluation function 1) – the player would choose 2c as this has the lowest expected utility value of 25/6 = 4.17 (2dp)

For the evaluation function 2) – There are now two equally viable actions that can be made. The first row can be eliminated or the third row can be eliminated, and both actions return a result of 1.

<u>Question 2</u>



b)

$V^{\pi f}(\text{cool}) = 0.5(4 + 0.9(c)) + 0.5(3 + 0.9(w)) = 2 + 0.45c + 0.45w = 2 + 0.45(c + w)$

$V^{\pi f}(\text{warm}) = 0.5(3 + 0.9(w)) + 0.5(1 + 0.9(o)) = 1.5 + 0.45w + 0.5 + 0.45o = 2 + 0.45(w + o)$

$V^{\pi f}(\text{off}) = 1(0 + 0.9w) = 0.9w$

If $o = 0.9w$ ->

then $V^{\pi f}(\text{warm}) = 2 + 0.45(w + 0.9w) = 2 + 0.45(1.9w)$ => $0.145w = 2$

$V^{\pi f}(\text{warm}) = 13.79$ (2dp)

If w = 13.79 ->

Then $V^{\pi f}(\text{cool}) = 2 + 0.45(c + 13.79)$ => $0.55c = 9.706\ldots$ =>

$V^{\pi f}(\text{cool}) = 17.64$

Then $V^{\pi f}(\text{off}) = 0.9(13.79) = 12.41$

$V^{\pi f}(\text{cool}) = 17.64$

$V^{\pi f}(\text{warm}) = 13.79$ (2dp)

Then $V^{\pi f}(\text{off}) = 12.41$

c)

$\pi_{i+1}(\text{cool}) = \text{argmax}((\text{cool, fast}), (\text{cool, slow})) = \text{argmax}(17.64, 1(3 + 0.9(17.64)) =$ $\text{argmax}(17.64, 18.88 \text{ (2dp)}) = 18.88$

Therefore, the action that is returned is for the car to choose the slow action at the cool state.

$\pi_{i+1}(\text{warm}) = \text{argmax}((\text{warm, fast}), (\text{warm, slow}))$

$= \text{argmax}(13.79, 0.5(2 + 0.9(13.79)) + 0.5(3 + 0.9(17.64)) = \text{argmax}(13.79, 16.64) = 16.64$

Therefore, the action that is returned is for the car to choose the slow action at the warm state.

$\pi_{i+1}(\text{off}) = \text{argmax}(\text{off, wait}) = \text{argmax}(12.41) = 12.41$

There is only one policy for the car in the off state, which means that that is the only policy it will ever take.

d)

$V_1(\text{cool}) = \max(3 + 0.9(V_0(C)), 0.5(4 + 0.9(V_0(c)) + 0.5(3 + 0.9(V_0(w))) = \max(3, 3.5) = 3.5$

$V_1(\text{Warm}) =$

$\max(0.5(3.+ 0.9(V_0(w)) + 0.5(1 + 0.9(V_0(\text{off})), 0.5(2 + 0.9(V_0(w)) + 0.5(3 + 0.9(V_0(c))$

$= \max(2, 2.5) = 2.5$

$V_1(\text{off}) = \max(1(0 + 0.9(V_0(w)) = \max(0) = 0$

$V_2(\text{cool}) = \max(3 + 0.9(3.5), 0.5(4 + 0.9(3.5)) + 0.5(3 + 0.9(2.5))) = \max(6.15, 6.2)$

$V_2(\text{warm}) =$

$\max(0.5(3.+ 0.9(2.5) + 0.5(1 + 0.9(0)), 0.5(2 + 0.9(2.5) + 0.5(3 + 0.9(3.5)$

$= \max(2.875, 3.6625) = 3.6625$

$V_2(\text{Off}) = \max(1(0 + 0.9(2.5)) = 2.25$

State	$V_0(s)$	$V_1(s)$	$V_2(s)$
Cool	0	3.5	6.2
Warm	0	2.25	3.6625
Off	0	0	2.25

e)

T^(s, a, s'):

T(cool, slow, cool) = 1
T(cool, fast, cool) = 1/3
T(cool, fast, warm) = 2/3
T(warm, slow, cool) = 1
T(warm, fast, off) = 1
T(off, wait, off) = 0.5
T(off, wait, warm) = 0.5

R^(s, a, s'):

T(cool, slow, cool) = 2
T(cool, fast, cool) = 2
T(cool, fast, warm) = 4
T(warm, slow, cool) = 2
T(warm, fast, off) = 0
T(off, wait, off) = 0
T(off, wait, warm) = 0