__/__/__

MDL Assignment 2 Part 1

Name: Abhijeeth Singam

Roll number: 2019101065

Reward: Arr[2019101065 % 15] = Arr[10] = 16.6

Question 1: Transition Table:

Starting State (s)	Action (a)	Final State (s')	Probability (T(s,a,s'))	Cost (R(s,a,s'))
A	Right	B	0.8	-1
A	Right	A	0.2	-1
A	Up	C	0.8	-1
A	Up	A	0.2	-1
B	Left	A	0.8	-1
B	Left	B	0.2	-1
B	Up	R	0.8	-4
B	Up	B	0.2	-1
C	Right	R	0.25	-3
C	Right	C	0.75	-1
C	Down	A	0.8	-1
C	Down	C	0.2	-1

Question 2:

I think the best path for a person at A is to first move right to B and then up to R. The main reason for this is the probility of success. The first step to be taken can be ignored as the probabilities and step costs are the same for both B and C, but to move to R from B and C the probilities are different. The reason I think B is a better choice than C is because of the much higher probability. Once you reach B, it is very likely that you reach R without wasting steps and remaining at B, while it is very likely that you waste at least a few steps at C trying to reach R. Thus I feel moving right to B and then up to R is the best path.

Question 3: Value iteration:

Initial state:

	A	B	C	R
V_0	0	0	0	16.6

→ Remains constant as there are no actions to be taken from R.

Iteration 1:

$$V_1(A) = \max\Big(0.8\big(-1+(0.2)\cdot 0\big) + 0.2\big(-1+(0.2)\cdot 0\big),$$
$$0.8\big(-1+(0.2)\cdot 0\big) + 0.2\big(-1+(0.2)\cdot 0\big)\Big)$$
$$= -1$$

~~$V_1(B) = \max\big(0.8(-1)$~~

$$V_1(B) = \max\Big(0.8\big(-4+(0.2)(16.6)\big) + 0.2\big(-1+(0.2)\cdot 0\big),$$
$$(0.8)\big(-1+(0.2)\times 0\big) + (0.2)\big(-1+(0.2)\cdot 0\big)\Big)$$
$$= 0.8\cdot(-0.68) - 0.2$$
$$= -0.744$$

$$V_1(C) = \max\Big((0.25)\cdot(-3+3.32) + 0.75(-1),$$
$$0.8\cdot(-1) + 0.2(-1)\Big)$$
$$= -0.67$$

	A	B	C
V_0	0	0	0
V_1	-1	-0.744	-0.67.

~~Not converge~~ Hasn't converged.

Iteration 2:

$\cancel{V_2(A) = \max\left(0.8(-1+0.2(-0.744)) + 0.2(-1),\ 0.8(-1+0.2(-0.67)) + 0.2(-1)\right)}$

$\cancel{= 0.8(-1-0.134) - 0.2}$

$\cancel{= -1.1072}$

$\cancel{V_2(B) = \max(0.8(-1+0.2(}$

$$V_2(A) = \max\Big(0.8(-1+0.2(-0.744)) + 0.2(-1+0.2(-1)),$$
$$0.8(-1+(0.2)(-0.67)) + 0.2(-1+0.2(-1))\Big)$$
$$= 0.8(-1-0.134) + 0.2(-1.2)$$
$$= 0.8(-1.134) - 0.24$$
$$= \underline{\underline{-1.1472}}$$

$$V_2(B) = \max\Big(0.8(-4+0.2(16.6)) + 0.2(-1+(0.2)\cdot(-0.744)),$$
$$0.8(-1+0.2(-1)) + 0.2(-1+0.2\cdot(-0.744))\Big)$$
$$= 0.8\cdot(-4+3.32) + 0.2(-1-0.1488)$$
$$= -0.544 - 0.22976$$
$$= \underline{\underline{-0.77376}}$$

$$V_2(C) = \max\Big(0.25(-3+0.2(16.6)) + 0.75(-1+0.2(\overset{-0.67}{\cancel{\ }})),$$
$$0.8(-1+0.2(-1)) + 0.2(-1+0.2(-0.67))\Big)$$
$$= 0.08 - 0.75(1.134)$$
$$= \underline{\underline{-0.7705}}$$

	A	B	C
V_1			
V_2			

	A	B	C
V_1	-1	-0.744	-0.67
V_2	-1.1472	-0.77376	-0.7705

~~Iteration 3:~~ ~~Since the values are becoming~~

$$\cancel{V_3(A) = \max\Big(0.8\big(-1+0.2(-0.77376)\big)+0.2\big(-1+0.2(-1.1472)\big),}$$

Approximately:

	A	B	C
V_1	-1	-0.74	-0.67
V_2	-1.15	-0.77	-0.77

Iteration 3:

$$V_3(A) = \max\Big(0.8\big(-1+0.2(-0.77)\big)+0.2\big(-1+0.2(-1.15)\big),$$
$$0.8\big(-1+0.2(-0.77)\big)+0.2\big(-1+0.2(-1.15)\big)\Big)$$

$$= 0.8(-1.154) + 0.2(-1.23)$$
$$= -0.9232 - 0.246$$
$$= -1.1692 = -1.17$$

$$V_3(B) = \max\Big(0.8\big(-4+0.2(16.6)\big)+0.2\big(-1+0.2(-0.77)\big),$$
$$0.8\big(-1+0.2(-1.15)\big)+0.2\big(-1+0.2(-0.77)\big)\Big)$$

$$= 0.8(-0.68) + 0.2(-1.154)$$
$$= -0.544 - 0.2308$$
$$= -0.7748 = -0.77$$

$$V_3(c) = \max\Big(0.25\big(-3+0.2(16.6)\big) + 0.75\big(-1+0.2(-0.77)\big),\ 0.8\big(-1+0.2(-1.15)\big) + 0.2\big(-1+0.2(-0.77)\big)\Big).$$

$$= 0.08 + 0.75(-1.154)$$

$$= 0.08 - 0.8655$$

$$= 0.7855 = 0.79.$$

	A	B	C
V_2	-1.15	-0.77	-0.77
V_3	-1.17	-0.77	-0.79

Converged (I won't be calculating future values of B since it has already converged).

Iteration 4:

$$V_4(A) = \max\Big(0.8\cdot\big(-1+0.2(-0.77)\big) + 0.2\big(-1+0.2(-1.17)\big),\ 0.8\cdot\big(-1+0.2(-0.79)\big) + 0.2\big(-1+0.2(-1.17)\big)\Big)$$

$$= 0.8\cdot(-1.154) + 0.2(-1.234)$$

$$= -0.9232 - 0.2468$$

$$= -1.17$$

$$V_4(c) = \max\Big(0.25\cdot\big(-3+0.2(16.6)\big) + 0.75\big(-1+0.2(-0.79)\big),\ 0.8\big(-1+0.2(-1.17)\big) + 0.2\big(-1+0.2(-0.79)\big)\Big)$$

$$= 0.08 + 0.75(-1.158)$$

$$= 0.08 - 0.8685$$

$$= -0.7885 = -0.79$$

	A	B	C
V_3	-1.17	-0.77	-0.79
V_4	-1.17	-0.77	-0.79

Converged

__/__/__

Question 4:

After value iteration, we have:

C	R
-0.79	16.6.
A	B
-1.17	-0.77

The optimal path for the person at A would be to go right from A to B and then up from B to R, assuming the reward is 16.6.

Yes, my guess was correct by the smallest of margins. My reward was luckily high enough to justify taking the higher costing step from B.

Question 5:

After observing the results for my value of reward, and thinking a little bit about what would happen if the reward was higher or lower, I came to the following conclusion:

As the reward becomes higher, the higher costing step from B becomes worth the cost due to its higher probability of success, and for lower ~~values, the would no longer warrant~~ values, the higher probability no longer warrants the higher cost. If we had a state D with a step costing 100 but with a 100% success or even an 81% chance, it will eventually become worth the cost, even if it is at reward values of 1000 or 1,000,000. ~~At lo~~

At lower reward values the high costs of the steps to R begin becoming less viable. Upto a certain point, ~~B~~ ~~the~~ the path ABR will be better than ACR but as the reward decreases, ACR will become more viable.

There will also be a point where these steps ~~&~~ are not worth their cost at all. i.e. the reward becomes so low, the step of cost 4 will become a loss and won't be worth it, and at an even lower value, the step of cost 3 will also not become worth it.

To guess these specific points, I'll just pick out the weird values from the given set.

8.8, 13.9, 16.5, 16.6.

Since I calculated for 16.6, I know that for 16.6 ABR is barely better than ACR.

So 16.5 is probably where ACR becomes better.

Among the other points of note. The ~~~~ value at which ABR becomes viable should be higher than that of ACR, so,

13.9 - viability of ABR

8.8 - viability of ACR.

(I do intend to verify the accuracy of these guesses because guessing is fun and this was actually pretty interesting)