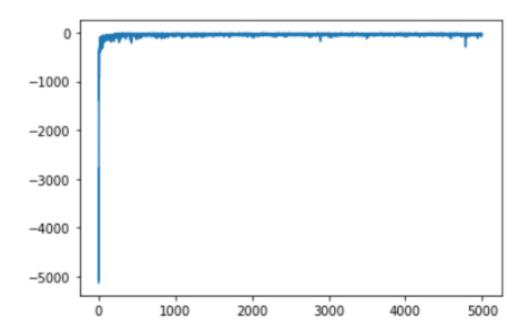
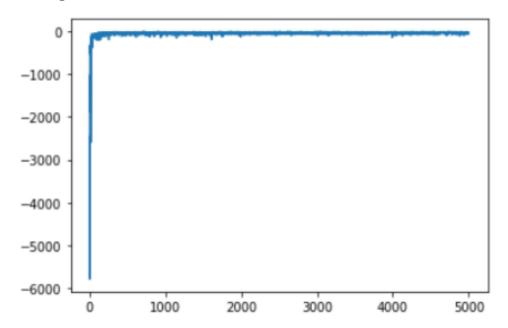
1. State action value Every-visit MC control

-254.86	-239.66	-249.32	-311.57
-284.11 -194.01			-373.05 -324.94
-132.65	-82.22	-88.11	-0.79
-213.10	-188.79	-165.17	0.00
	-71.44 -24.65	-34.76	0.00
-56.03	-33.40	-7.50	0.00
-134.61	-54.87	-17.53	-1.00
-101.84 -37.33		-15.17 -2.68	-5.03 -2.36
-108.03	-72.41	-6.20	-2.76



First-visit MC control

-207.95	-208.28	-225.50	-240.27
-201.28 -163.91	-216.21 -183.17	-257.45 -155.62	-356.74 -94.16
-60.72	-38.93	-57.35	-0.71
-171.70	-165.66	-151.96	0.00
-58.74 -38.21	-61.35	-28.63	0.00
-37.81	-28.93	-5.75	0.00
44.47	-38.18	-17.38	-1.00
-38.02	-39.64 -12.66	-13.74 -2.92	4.71 -3.20
46.03	-20.86	-7.70	-2.52



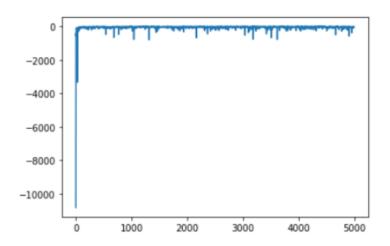
2.

是合理的,因為從 state action value 圖中可以觀察到,基本在 swamp 中,往下的動作和往右的動作價值最大,因為可以脫離沼澤或是更接近終點。而終點的動作價值都是 0 因為是在終點。而圖中可以看到起點的動作價值最大是往下動作可以離沼澤遠一點,因為 epsilon-greedy action selection,所以有機會往上走到沼澤,因此往下的 Reward 比往右的 Reward 平均會高一點。

3. State value

Every-visit MC control

	1	2	3	4
1	-176.16	-310.78	-76.38	-243.41
2	-33.57	-30.34	-14.85	0.0
3	43.16	-26.44	-8.31	-2.2



First-visit MC control

	1	2	3	4
1	-91.24	-131.57	-71.53	-502.71
2	-28.07	-31.78	-18.38	0.0
3	-20.07	-8.49	-6.17	-1.74

