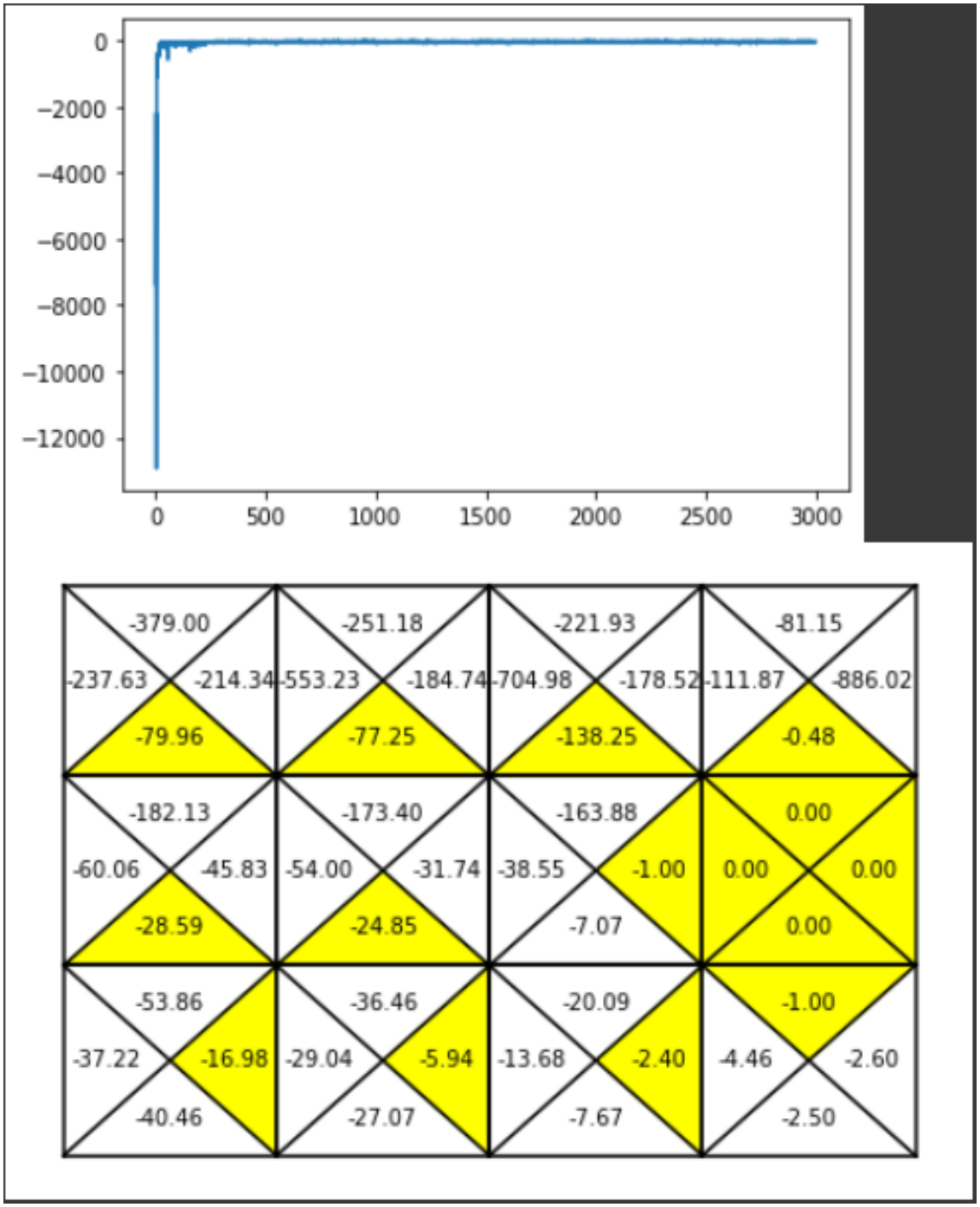
Experiments and Analysis (40%)

1. Plot the q_values in your result. (20%)

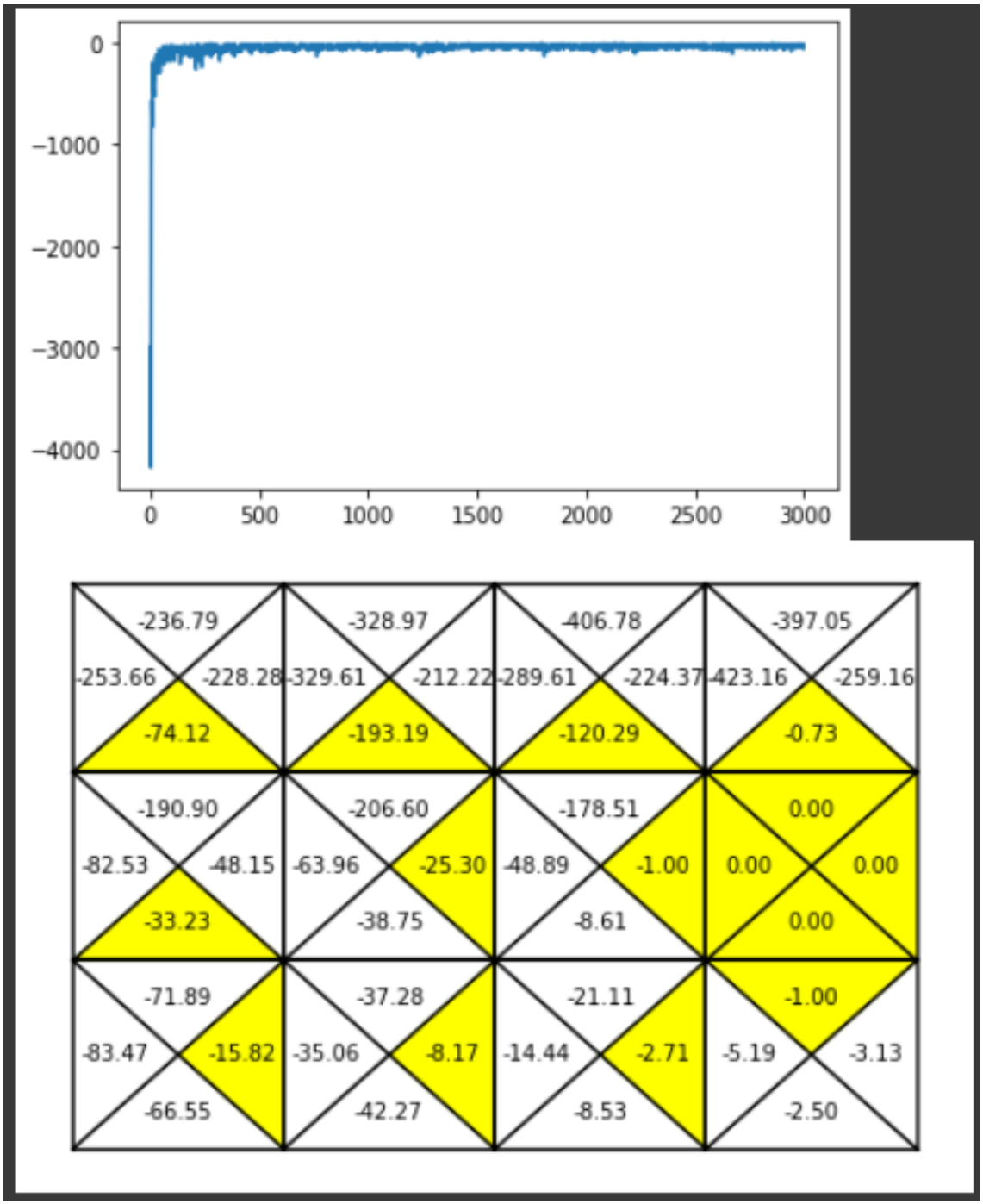
First-visit MC control:

State-action values and learning curve



Every-visit MC control:

State-action values and learning curve



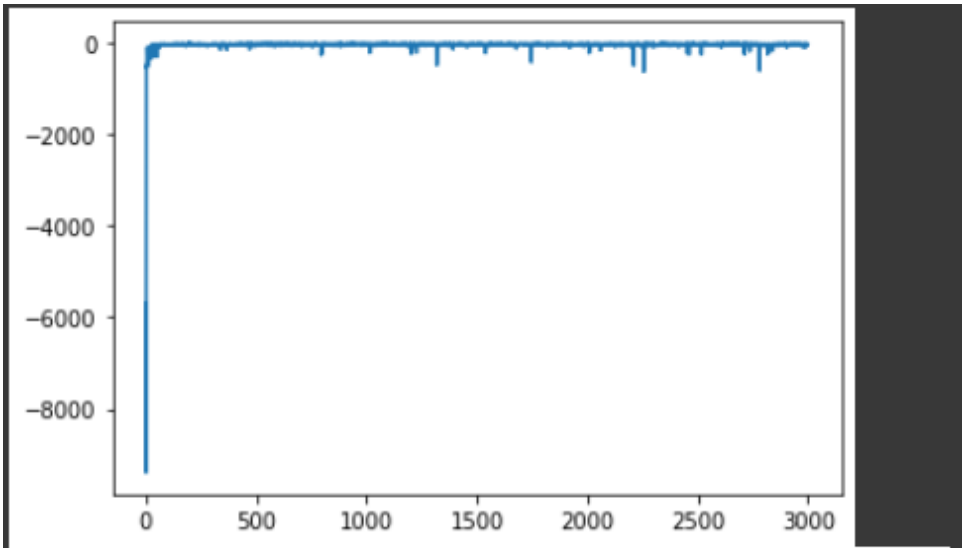
2. Whether q_values are reasonable? Explain your result. (10%)

合理，可以從上圖中得知，他會選擇最遠的距離進行移動，因為他是蒙地卡羅法是使用 on-policy 的方法，而在沼澤地因為 reward 為-100，所以他會全部往下走，離開沼澤地，所以 q_value 向下的較大。

3. Transfer state-action values to state values and plot it. (10%)

First-visit MC predict:

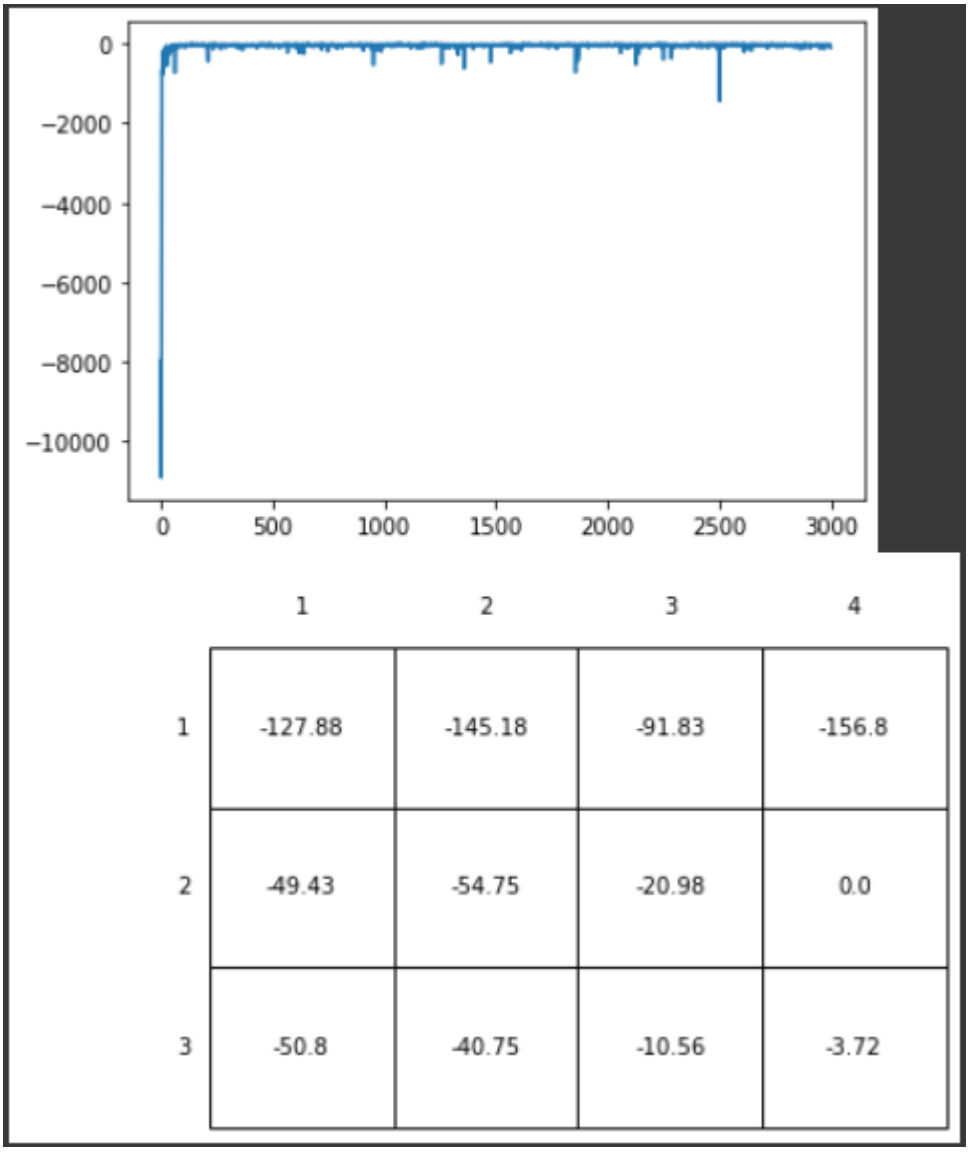
State values and learning curve



	1	2	3	4
1	-107.24	-170.08	-115.82	-175.2
2	-35.92	-34.36	-21.39	0.0
3	-25.57	-14.84	-9.48	-1.39

Every-visit MC predict:

State values and learning curve



	1	2	3	4
1	-127.88	-145.18	-91.83	-156.8
2	-49.43	-54.75	-20.98	0.0
3	-50.8	-40.75	-10.56	-3.72