Exercise 5

Task 1 - Random Walk

- The first episode terminates in the left state.
- The states are not yet visited or the ones visited are updated with TD error = 0 (→ the values are not changing at all.)
- $V_{t+1}(A) = \alpha * (1 - V_t(A))$ in our case: $0.1 * (1 - 0.5) = 0.05$

Task 2 - Sarsa and Q-learning on the FrozenLake

a)

see Figure 1 & Figure 2 & Figure 3.

→	↑	←	↑
←	←	←	←
↓	↓	←	←
←	↓	↑	←

Table 1: Sarsa policy

b)

see Figure 4 & Figure 5 & Figure 6.

↓	↓	↓	↓
↓	←	↓	←
→	←	↓	←
←	→	↓	←

Table 2: Q-Learning policy

c)

No exploration is done, so the best neighbor is always chosen.

d)

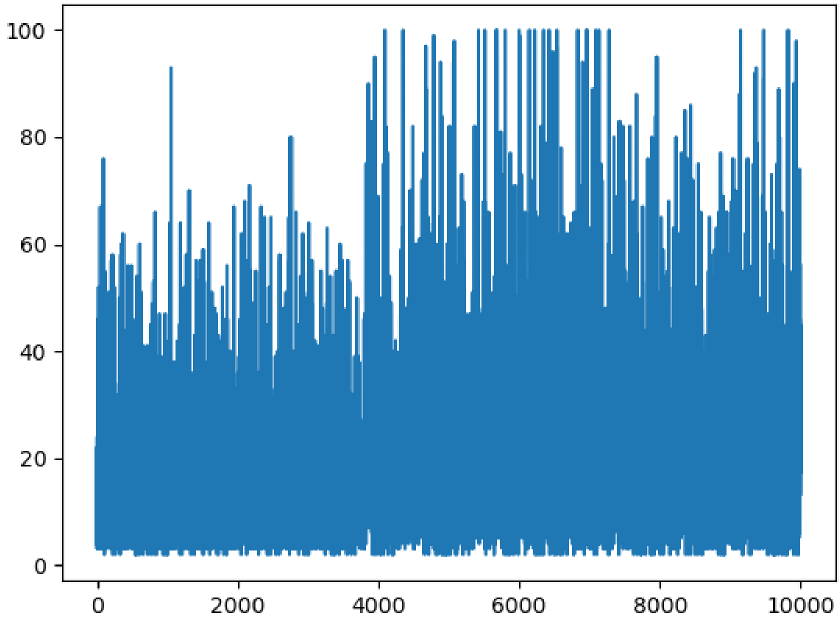
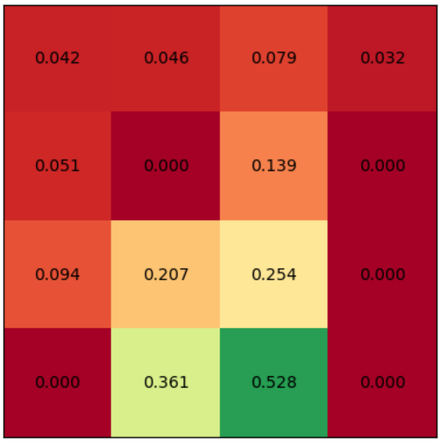


Figure 1: Sarsa training length



0.042	0.046	0.079	0.032
0.051	0.000	0.139	0.000
0.094	0.207	0.254	0.000
0.000	0.361	0.528	0.000

Figure 2: Sarsa V

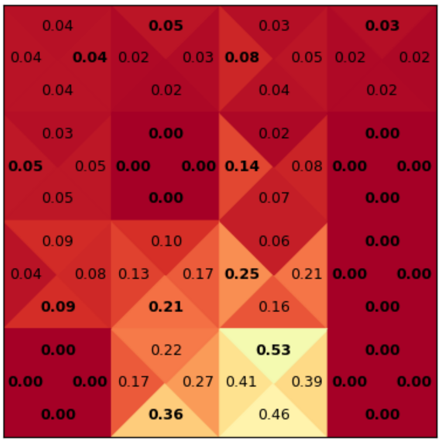


Figure 3: Sarsa Q

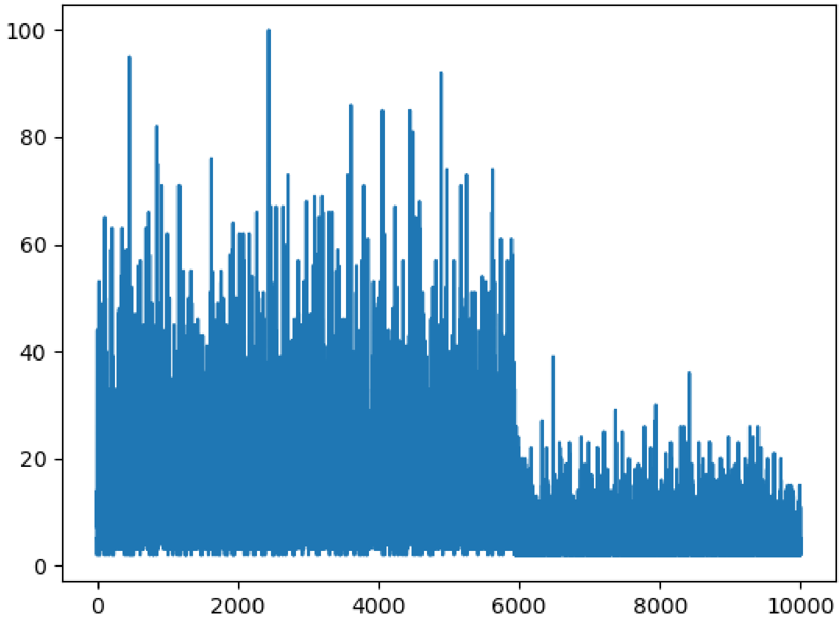


Figure 4: Q-Learning training length

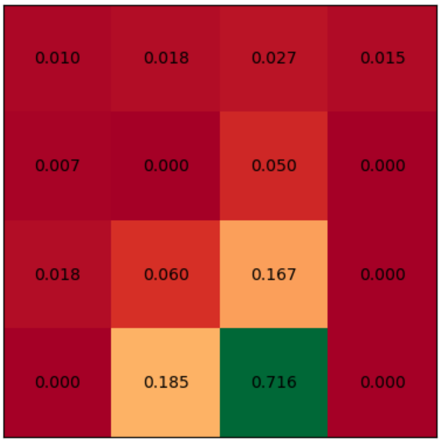


Figure 5: Q-Learning V

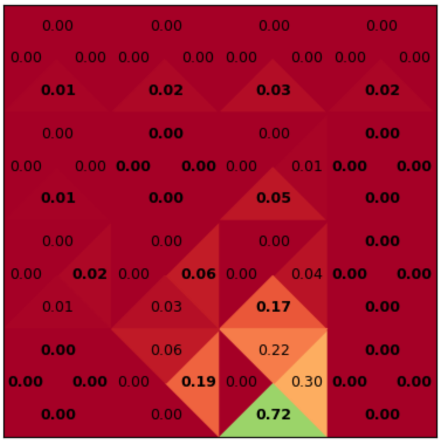


Figure 6: Q-Learning Q