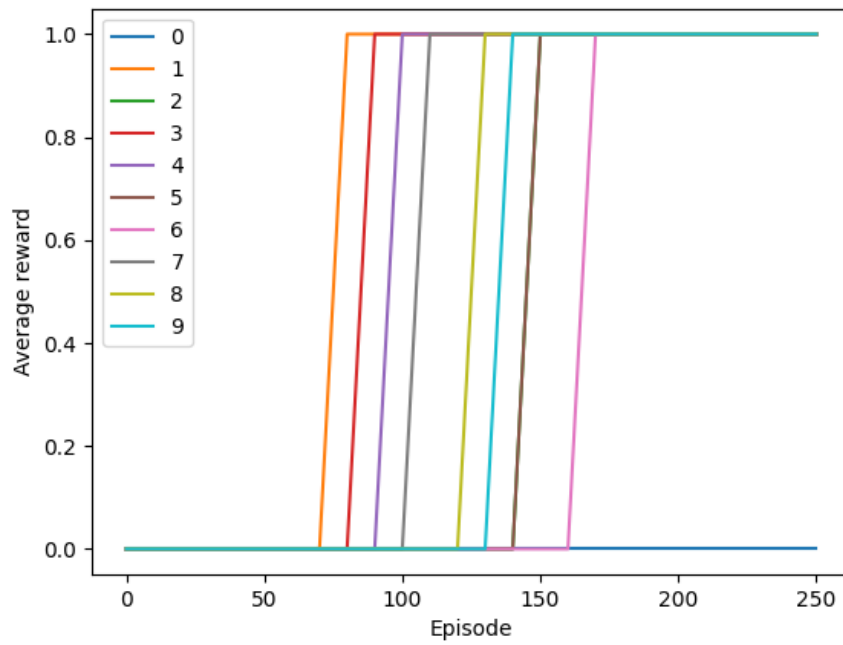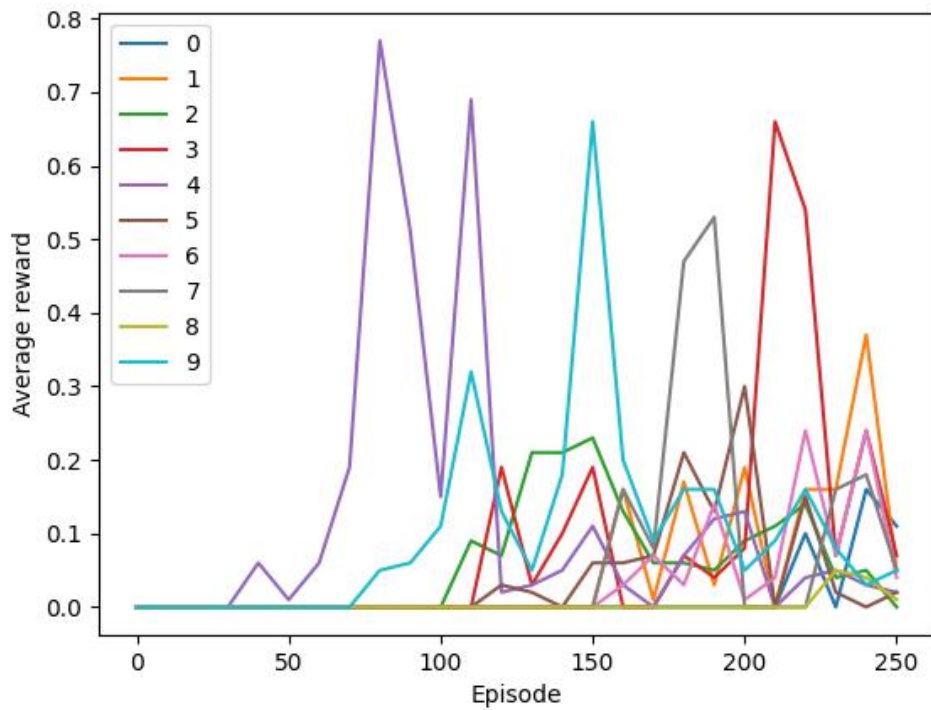a) Deterministic case. 250 episodes, 150 max steps, 0.9 discount factor

b) Non-deterministic. 250 episodes, 150 max steps, 0.9 discount factor

c) Non-deterministic with the non-deterministic update rule. 250 episodes, 150 max steps, 0.9 discount factor, 0.5 learning rate