Homework #5

1) Show that an action-value version of (6.6) holds for the action-value form of the TD error $\delta_t = \mathrm{R}_{t+1} + \gamma\, \mathrm{Q}(S_{t+1}, A_{t+1}) - \mathrm{Q}(\mathrm{S}_t, A_t)$, assuming that the values don't change from step to step.

2) Implement SARSA, Q-learning and expected SARSA algorithms on the noisy gridworld of problem and for the Gymnasium Cart Pole problem.

3) What are the update equations for the Double Expected SARSA using ϵ-greedy target policy?

4) Implement the SARSA(n) algorithm and compare the performance of different "n" (1, 2, 4, 8, 16, 32, MC) against the RME of all the states, and present the result just like the graph below presented in class.