RBE595 - Week 7 Assignment

Keith Chester

Due date: February 22, 2023

Problem 1

Between DP (Dynamic Programming), MC (Monte-Carlo) and TD (Temporal Difference), which one of these algorithms use bootstrapping? Explain.

Problem 2

We mentioned that the target value for TD is $[R_{t+1} + \gamma V(S_{t+1})]$. What is the target value for Monte-carlo, Q-learning, SARSA and Expected-SARSA.

Problem 3

What are the similarities of TD and MC?

Problem 4

Assume that we have two states x and y with the current value of $V(x) = 10, V(y) = 1$. We run an episode of $\{x, 3, y, 0, y, 5, T\}$. What's the new estimate of $V(x), V(y)$ using TD (assume step size $\alpha = 0.1$ and discount rate $\gamma = 0.9$)

Problem 5

Can we consider TD an online (real-time) method and MC an offline method? Why?

Problem 6

Does Q-learning learn the outcome of exploratory actions? (Refer to the Cliff walking example).

Problem 7

What is the advantage of Double Q-learning over Q-learning?