Stanford CME 241 (Winter 2022) - Assignment 11

Assignments:

1. We have written the function mc_prediction in rl/monte_carlo.py as an implementation of Monte-Carlo Prediction with function approximation. You have also learnt that since Tabular MC Prediction is a special case of MC Prediction with Function Approximation and so, writing a separate function for Tabular MC Prediction is not necessary. But for a learning experience, it's a great idea to write a function for Tabular MC Prediction **from scratch**. Think about what the input and output types must be. Be sure to reduce the learning rate appropriately as a function of number of updates (or as a function of number of episodes).

2. We have written the function td_prediction in rl/td.py as an implementation of Temporal-Difference Prediction with function approximation. You have also learnt that since Tabular TD Prediction is a special case of TD Prediction with Function Approximation and so, writing a separate function for Tabular TD Prediction is not necessary. But for a learning experience, it's a great idea to write a function for Tabular TD Prediction **from scratch**. Think about what the input and output types must be. Be sure to reduce the learning rate appropriately as a function of number of updates.

3. Test your above implementations of Tabular MC Prediction and Tabular TD Prediction on SimpleInventoryMRPFinite (from rl/chapter2/simple_inventory_mrp.py) by ensuring that your Value Function output matches that produced by the function approximation versions of MC Prediction and TD Prediction.

4. **Optional** Extend RandomWalkMRP (in rl/chapter10/random_walk_mrp.py) to RandomWalkMRP2D which is a random walk in 2-D with states $\{(i, j)|0 \leq i \leq B_1, 0 \leq j \leq B_2\}$ with terminal states as $(0, j)$ and (B_1, j) for all j, $(i, 0)$ and (i, B_2) for all i, and with reward of 0 for all $(0, j)$ and for all $(i, 0)$, reward of 1 for all (B_1, j) and for all (i, B_2), and with discrete probabilities of 4 movements - UP, DOWN, LEFT, RIGHT from any non-terminal state. Analyze the convergence of MC and TD on this RandomWalkMRP2D much like how we analyzed it for RandomWalkMRP, along with plots of similar graphs.