(CS 5008) Reinforcement Learning : Assignment 4

Temporal Difference Learning

You are expected to submit a Python notebook and an accompanying report. You are free to play around with what you want to demonstrate, and it is expected that the report will provide a description of the same.

Q1) Implement a function **create-MDP** which accepts grid size n and probability of success of an action p_{succ} as inputs, and outputs an MDP. The number of actions k to be implemented is left to your imagination. The output should be a $n \times n \times n \times n \times k$ array specifying the state transition probabilities and a $n \times n$ array specifying the reward mapping. [2 Marks]

Q2) Implement a function **gen-rand-SDP** that accepts the MDP as input gives a random stationary deterministic policy (SDP) as output. The output policy should change from run to run. [2 Marks]

Q3) Implement a function **Gauss-Elimination** to solve a given system of linear equations. [2 Marks]

Q4) Implement a function **Value-Via-Inv** to compute V_u of a policy u. The function accepts probability transition, reward map and the discount factor $\gamma \in (0, 1)$ as input and outputs V_u using the function **Gauss-Elimination**. [2 Marks]

Q5) Implement a function **One-Step** which accepts a V as input and $T_u V$ as output (the dimensions of the quantities V and $T_u V$ are understood from the problem instance). [2 Marks]

Q6) Implement a function **Value-Eval** which uses **One-Step** to compute V_u. [2 Marks]

Q7) Implement a function **Random-Walk** which accept trajectory length $L > 0$, policy, transition probability and reward map as input, and outputs the data set $(s_t, r_t, s_{t+1})_{t=0}^{L}$. [2 Marks]

Q8) Compute d_u from the data set and show that it satisfies the property $d_u^\top = d_u^\top \mathbb{P}_u$. [2 Marks]

Q9) Compute the maximum allowable step size $\alpha > 0$ so that TD(0) algorithm will be stable. [2 Marks]

Q10) Run TD(0) on the data set. [5 Marks]

Q11) Compute theoretically the rate of decay of the bias terms and the amount of noise. Does it match with the simulation runs? [2 Marks]