Reinforcement Learning
Homework 2
Markov Chain,
Markov Reward Process,
Markov Decision Process

February 15, 2025

Instructions

Total Points: 100
Deadline: Febryary 25, 2025
To calculate your homework grade, we will average the points earned from all homework assignments throughout the semester and take 15% of that Average.
Submission: Submit your solutions as a PDF document named "FirstName LastName Homework 2.pdf" by the deadline.

Problem 1:Markov Chains (10 Points).

Consider the following problem. We need to model and evaluate the state transitions of Football Club Ararat. The States of any football club could be $[Win, Loss, Draw]$ Let's say, FC Ararat has the following state transition probability matrix.

	W	L	D
W	0.6	0.1	0.3
L	0.3	0.4	0.3
D	0.4	0.2	0.4

Draw the Markov Chain graphically.

Problem 2: Markov Reward Process (MRP) (25 points).

Let's extend the Markov Chain by introducing the Rewards for each state. Let's say, that the club President gives the +10 Reward for winning the next game, +3 Reward for Draw on the next game, and 0 Reward for Losing the next game. The discount factor $\gamma = 0.9$.

- Draw the MRP graphically (5 points).

- Calculate the State values for each state (20 points).

Problem 3: Markov Decision Process (MDP) (65 points).

Let's extend the MRP into MDP by introducing the possible actions taken by the Coach and action-dependent State Transition Probabilities and Rewards.
The Coach can take one of the following actions regarding the next training session.

- Train as Usual (Action 0)

- Train harder (Action 1)

- Cancel Training (Action 2)

The State Transition Probabilities are:

Action 0	W	L	D	Action 1	W	L	D	Action 2	W	L	D
W	0.5	0.15	0.35	W	0.7	0.05	0.25	W	0.4	0.35	0.25
L	0.25	0.4	0.35	L	0.35	0.3	0.35	L	0.2	0.6	0.2
D	0.35	0.25	0.4	D	0.45	0.1	0.45	D	0.3	0.4	0.3

If the coach takes Action 0, the Rewards remain the same, if the coach chooses the action train harder, the club President appreciates it and gives the +15 Reward for winning the next game, +5 Reward for Draw on the next game, and 0 Reward for Losing the next match. If the coach cancels training then the club President will reduce the Reward: +8 Reward for winning the next game, +2 Reward for Draw on the next game, and 0 Reward for Losing the next match. The discount factor is the same. The coach will choose Action 0 with the probability 0.35, Action 1 with the probability 0.45 and Action 2 with the probability of 0.2.
Optional: you can choose other policy and compare the results.

- Draw the MDP graphically (5 points).

- Calculate the average state transition matrix and Reward (10 points)

- Calculate the state values for each state (25 points).

- Calculate action value functions for each state (25)