National University of Computer and Emerging Sciences

Peshawar Campus

Reinforcement Learning (CS3018)

Sessional-I Exam

Total Time (Hrs): 1
Total Marks: 30
Total Questions: 4

Date: 24th February, 2025

Course Instructor(s)

Ms. Sara Rehmat

Roll No _____ Section BS(AI) Student Signature _____

Do not write below this line

Attempt all the questions.

CLO 1: Explain the basic concepts and mathematical tools of reinforcement learning.

Q1: Differentiate between the following terms: **[5 marks]**

a. Reinforcement Learning and Supervised Learning
b. Exploration and Exploitation
c. Rewards and Returns
d. Policy and Goal
e. Stationary and Non-stationary problem

Q2: Consider the multi-armed bandit problem. **[2+3=5 marks]**

a. A greedy agent will always choose an action whose current estimate is highest. Give an example of a scenario where this will not lead to choosing the optimal action.
b. Describe the ways an agent can be made to try actions other than that having the highest estimate. (Hint: There is more than one way.)

Q3: Devise an example task of your own that fits into the Markov Decision Process framework. Identify its states, actions, and rewards. Also draw the transition graph. **[5+5=10 marks]**

CLO 2: Implement basic RL algorithms to solve standard benchmark problems.

Q4: Write pseudo code for estimating the action values in the multi-armed bandit problem using **Upper Confidence Bound** algorithm. **[10 marks]**

Consider the 10-armed test bed in which true value of each action is sampled from a normal distribution with mean 0 and variance 1 and each reward for each action is sampled from a normal distribution with mean equal to true value of that action and variance equal to 1.

Evaluate the performance of UCB by determining the average rewards and number of times the optimal action is taken by the agent in 1000 time steps over 3000 independent runs of the test bed.