Dr. Rakesh Kanji

JAYPEE UNIVERSITY OF INFORMATION TECHNOLOGY, WAKNAGHAT

TEST -2 EXAMINATIONS-2022

B.Tech-VIII Semester (CS/IT)

COURSE CODE: 19B1WCI837 MAX. MARKS: 25

COURSE NAME: REINFORCEMENT LEARNING

COURSE CREDITS: 3 MAX. TIME: 1 Hour 30 Min

***Note:** All questions are compulsory. Marks are indicated against each question in square brackets.*

Q1.What is more general method Monte carlo or Markov decision process and why? [3]

Q2. How you model tic toe game with Monte carlo or Markov decision process? Any advantage of choosing your option? [3+2]

Q3. Explain dynamic programming to solve bellman equation. [4]

Q4. Why asynchronous dynamic programming can find the optimal value of states of bellman equation? [2]

Q5. Compare value and policy iteration algorithm. [3]

Q6. Probability of choosing state from any state (transition) with reward +1, -1, and 0 are 0.8, 0.1 and 0.1 respectively. Consider discounting rate is 0.9 and action or policy as left (dotted arrow) and right (bold arrow) movement. Please maximize $V_1(3,3)$ with bellman update rule and mention the corresponding action. V_0 is given below as the initial reward. In $V_i\ (p, q)$, i *and* (p, q) represents the index of iteration and states respectively. [5]

V_0	$q = 1$	$q = 2$	$q = 3$	$q = 4$
$p = 1$	0	0	0	0
$p = 2$	0	0	1 ↑	-1
$p = 3$	0	0 ←	0 →	1 →

Q7. Consider a1, a2 as action or policy and 1, -1 as reward. Compute $V^{\pi}(1)$ with interaction of agent and environment as given below. Consider discount rate is γ [3]