Week 1 to 6

Q1.What is multi arm bandit problem? Explain explore, exploit, regret, reward terms related to MAB.

Q2. Write Q function in the stationary and non-stationary updates in the multi-armed bandit problem?

Q3. Define Explore- vs Exploit-only vs Epsilon-Greedy vs Upper Confidence Bounds

Q4. Define Probably Approximately (PAC) Learnable with its use cases.

Q5. What is reinforcement learning? Explain with diagram

Q6. Describe median elimination PAC with a case study.

Q7. Differential stationary and non stationary bandits.

Q8. Write mathematical model of policy based reinforcement learning and explain?

Q9. How contextual bandits make a decision?

Q10. Explain the concepts of the Markov Decision Process, Bellman Equation and Policies?

Q11. How Does the Markov Decision Process Work?

Q12.Give example of Examples of the Markov Decision Process.

Q13. How do iterative methods like value iteration, q-learning, and advanced methods converge when training?

Q14. How do iterative methods like value iteration, q-learning, and advanced methods converge when training?

Q15. what is policy iteration and value iteration?

Q16. How policy improvement process takes place.

Q17. In Generalized policy iteration (GPI) how two processes — policy evaluation and policy improvement, to interact.

Week 7 to 9

Q18 How to replace traces updates in SARSA(lemda)?

Q19. What are eligibility traces and how are they controlled.

Q20. Write ϵ-greedy exploration strategy .

Q21. Explain off policy TD(lemda).

Q22 In off-policy TD(λ), how to update eligibility traces.

Q23.What is role of function approximation in RL.

Q24.How incremental updates for a linear approximation of the value function takes place.

Q25. Compare and contrast LSTD and LSTDQ algorithms?

Q26. State coding method of state aggregation for grid world problems.

Q28. Design high level DQN workflow.

Q29. Write Policy Gradient Theorem for average reward formulation

Q30. Monte Carlo policy gradient methods typically converge faster than the actor-critic methods, given that we use similar parameterisations and that the approximation to the Qπ used in the actor-critic method satisfies the compatibility criteria.Comment on above?

Q31. What are actor critic methods. Write a case study on it.

Q32. Write common architectures for Hierarchical reinforcement learning and explain them?

Q33.Write update equation for SMDP Q-learning.

Q34.Differentiate makov and semi markov policies.

Q35.









