

RL Course @ MSAI, fall 2021

Exam Program

Theoretical minimum:

- RL problem statement (environment, agent, etc. abstractions). Markov decision process and its properties. Reward, discounted reward.
- What is a Q-function and a Value-function? Relationship between them.
- How can RL be applied to NLP or CV tasks?
- What is an exploration-exploitation tradeoff?
- What is the difference between model-based and model-free RL?
- Value-based vs. Policy based methods (general idea)

Program:

1. Cross-entropy method (tabular and approximate case).
2. Value-based RL: state value and state-action value functions. Relationship between them. Bellman expectation and optimality equations.
3. Value iteration algorithm.
4. Policy iteration algorithm.
5. Model-free RL: Monte-Carlo vs. TD updates. Q-Learning algorithm.
6. SARSA, Expected-Value SARSA algorithms. On-policy vs. Off-policy methods.
7. Approximate Q-Learning. DQN, training details.
8. Policy-based methods. REINFORCE algorithm (without baselines, with derivation).
9. Baselines in Policy Gradient. Actor-critic algorithm.
10. Advanced Policy Gradient: A2C, A3C, GAE.
11. RL for NLP and CV. Self-critical sequence training algorithm.
12. Exploration in RL. Exploration strategies: eps-greedy, UCB, Thompson sampling. Metrics for exploration.
13. DDPG, TD3 algorithms.
14. Planning; Monte-Carlo tree search.