

Questions

“Training”

1. How is reinforcement learning different from supervised learning?
2. What is the “Reward Hypothesis” in Reinforcement learning?
3. Define the following terms: action, state, observation, reward, return, environment, agent
4. What is the difference between model-based and model-free reinforcement learning agents?
5. What is the discount factor γ ? What happens if $\gamma = 0$? What about $\gamma = 1$?
6. Suppose you have a value-based agent. How can we obtain a policy from the learned value function?

“Generalization”

1. Suppose we are training a policy based RL agent. What happens if we optimize the *reward* (rather than the *return*) of the actions produced by the agent?
2. Why is it (generally) more preferable to learn a *stochastic* policy function rather than a deterministic one? (consider exploration-exploitation)
3. Suppose you have a value-based agent. How can we obtain a **stochastic** policy from the learned value function?
4. In what kind of RL tasks would you use a small discount factor γ when computing returns?
5. In what kind of RL tasks would you use a large discount factor γ when computing returns?