Policy Iteration

测验, 7 个问题

1 point

1。
What are the two main steps in value-based approach to Reinforcement Learning?

- [ ] 2 - extract a value function from the reward function.
- [ ] 1 - build a policy function.
- [x] 1 - build a value function.
- [ ] 2 - extract a reward function from the value function.
- [x] 2 - extract a policy function from the value function.
- [ ] 2 - extract a value function from policy.
- [ ] 1 - estimate a reward function.

---

1 point

2。
What is true about policy improvement? Recall that,

total return = immediate reward + the discounted expected return from the next state under policy π.

- [x] An agent acts greedily with respect to combination of immediate reward and the expected return under policy π.
- [ ] An agent acts greedily with respect to the immediate reward only and ignores the remaining expected return under policy π.
- [ ] Relying on the estimates of expected return under policy π may lead to deterioration of an agent's performance in some states. This is so because the estimates will no longer valid as soon as policy is changed (improved).

Policy Iteration

测验, 7 个问题

- [ ] Making several policy improvements in a row may increase the performance of a new policy.

1 point

3。
How many different value functions can correspond to any particular policy function?

- ( ) Depends on number of actions.
- (•) One
- ( ) Depends on number of states.
- ( ) Infinite

1 point

4。
Why we don't need the precise solution of a system of Bellman equations?

- [ ] The solution of such system of equations is intractable on any modern supercomputer. Thus we have to approximate.
- [ ] The system of Bellman equations may have no solution at all. Thus we should be satisfied with an approximation.
- [x] After reaching some precision level further refinements of the solution will not change the result of subsequent policy improvement.
- [ ] We want to sacrifice the global optimality for much faster convergence.

2 points

5。
Generalised Policy Iteration (GPI)

- [x]

- [ ] does not depend on initialization.
- [ ] requires to improve policy in each and every state before subsequent policy evaluation.
- [ ] converges to local optimum.
- [x] does not require to perform policy evaluation until its convergence
- [ ] depends on initialization.
- [x] converges to global optimum.
- [ ] requires to perform policy evaluation until convergence at every iteration.
- [x] does not require to improve policy in each and every state as long as policy in any state is improved once in a while

---

1 point

6。
How can we recover the optimal policy solely from q^* function?

- ( ) Sample from a distribution that is proportional to q-values.
- ( ) Find the action that is closest in q-value to average q-value over actions.
- ( ) It is impossible without the knowledge of environment dynamics.
- (•) With argmax operator.
- ( ) With max operator.

---

1 point

7。
What is the difference between Policy Iteration and Value Iteration?

- [ ] Value Iteration updates value function until numerical convergence of all its state values before each policy improvement step.

Policy Iteration

测验, 7 个问题

- [x] Policy Iteration updates value function until numerical convergence of all its state values before each policy improvement step.
- [x] Value Iteration perform only one iteration of policy evaluation before policy improvement step.
- [ ] Policy Iteration perform only one iteration of policy evaluation before policy improvement step.

- [x] 我（**伟臣 沈**）了解提交不是我自己完成的作业 将永远不会通过此课程或导致我的 Coursera 帐号被关闭。

了解荣誉准则的更多信息

Submit Quiz
