2. Q-Value Iteration

Bookmark this page

Homework due May 3, 2023 08:59 -03 Completed

Consider an Markov Decision Process with 6 states $s \in \{0,1,2,3,4,5\}$ and 2 actions by the following transition probability functions

For states 1, 2, and 3:

$$T(s, M, s-1) = 1$$

$$T(s, C, s+2) = 0.7$$

$$T(s, C, s) = 0.3$$

For state 0:

$$T(s, M, s) = 1$$

$$T(s, C, s) = 1$$

For states 4 and 5:

$$T(s, M, s-1) = 1$$

$$T(s, C, s) = 1$$

Note that all transition probabilities not defined by the above are equal to 0.

The rewards R are defined by:

$R(s, a, s') = |s' - s|^{\frac{1}{3}}\ \forall s \neq s'$,

and $R(s, a, s) = (s+4)^{\frac{-1}{2}}, \forall s \neq 0$.

$R(0, M, 0) = R(0, C, 0) = 0$. Also, the discount factor $\gamma = 0.6$.

We initialize $Q_0(s, a) = 0\ \forall s \in \{0,1,2,3,4,5\}$ and $\forall a \in \{C, M\}$

2

6.0/6.0 points (graded)

Input the Q-values **correct to 3 decimal places** after one Q-value iteration

0 ✓

0 ✓

1 ✓

1.016 ✓

1 ✓

1.004 ✓

1 ✓

0.995 ✓

1 ✓

0.354 ✓

1 ✓

0.333 ✓

Submit

You have used 1 of 4 attempts

1 ✓

Submit You have used 1 of 2 attempts

4

5/5 points (graded)

What are the optimal policies we get from ?

● C

○ M

✓

● C

○ M

✓

‹ Previous

Next ›

edX

About

Affiliates

edX for Business

Open edX

Careers

News

Show all posts

? Reward clarification

For the case where both s != s' and s != 0 hold true, which reward function should we use? The first, the sec

Connect

Blog

Contact Us

Help Center

Security

Media Kit

GET IT ON Google play

Download on the App Store