DL4\_6 TD法

5

状態価値は「移動先の地点の報酬とその地点の状態価値の和」の平均

「移動先の地点の報酬とその地点の状態価値の和」の方向へ価値を修正している

移動先の地点の報酬とその地点の状態価値の和() : TDターゲット

9

方策評価を行うために状態価値関数ではなく、行動価値関数を評価する必要がある。

があれば、それから決定的な方策を求めることができる。

先ほどと同様に、Q関数にもTD法を適用できる。

10

13

が方策に従うとき、

15

Q関数のベルマン最適方程式

これをサンプリング化(特定の行動を選択したと考える)すると、