Exercise 05

May 25, 2020

1 Temporal-difference methods

a)

One way would be m to solve it mathematically. We know that the probability to end up in either the right (R_{end}) or the left(L_{end}) end starting from the middle (C) is 0.5 for each end, as the problem is symmetrical. For that write

$$P_C(R_{end}) = P_C(L_{end}) = 0.5$$

If we now start from the left side:

$$\begin{aligned} P_A(L_{end}) &= 1 - P_A(R_{end}) \\ &= 1 - P_A(B) \cdot P_B(R_{end}) \\ &= 1 - P_A(B) \cdot (P_B(C) \cdot P_C(R_{end}) + P_B(C) \cdot P_A(R_{end})) \\ &= 1 - 0.5 \cdot (0.5 \cdot 0.5 + 0.5 \ P_A(R_{end}) \\ &\rightarrow P_A(R_{end}) = 0.5 \cdot (0.5 \cdot 0.5 + 0.5 \cdot P_A(R_{end})) \qquad \rightarrow P_A(R_{end}) = \frac{1}{6} \end{aligned}$$

Solving it for starting at B is analogous (using the results from this). Solutions for E and then B are the same as the problem is the same.

Another way to solve this would be via Dynamic programming and just compute every path. Seems more likely that the first solution was used, as that is way easier.

b)

Because it is not using the currently implemented policy to improve its current policy.

2 Batch MC and TD

a)

$$V_{MC}(A) = (0 + 3 + 2 + 2 + 3)/5 = 10/5 = 2$$