Group G GHC Task (to be presented Mo., March 4, 2024)

Centering on Temporal Difference Learning

Consider the following world called DEF World consisting of states 1, 2 and 3 and operators a, b and for visiting states 2 and 3 the agent receives a reward of 2 and 10, respectively.

a

b/0.9

a

b/0.1

a

b

DEF World

a) Apply temporal difference learning to the DEF World, depicted above, relying on the following assumptions: [6]

* The agent starts in state 3 and applies aaaa (applies action ‘a’ 4 times)
* γ is 1.0 and α is 0.5
* If state 1 is visited a reward of 0 is obtained
* Utilities of the 3 states are initialized with 0

What are the utilities of states 1, 2, 3 after ‘aaaa’ has been applied? Do not only give the final result but also how you derived the final result including the formulas you used!

b) What role does the parameters α and γ play when updating the utilities in temporal difference learning?