B(E)4M36SMU

Reinforcement Learning 3 - Convergence of Q-learning and SARSA

Monday 15th April, 2019

GLIE

- Off-policy Q-learning needs to satisfy the infinite exploration condition;[1]
- SARSA needs to satisfy both conditions of GLIE in order to converge.
- For further details, please visit: `https://doi.org/10.1023/A:1007678930559`
- If you use ε-greedy policy, it is more practical to use different ε for each state.

[1] This is true if we talk about convergence of Q-values. To obtain agent that behaves optimally (i.e., for policy convergence), you need to satisfy both GLIE conditions.

Learning rate

- In TD-learning and Q-learning, learning rate should decrease so that sum of α_k diverges and sum of α_k^2 converges.
- For further details, please visit: `https://doi.org/10.1023/A:1022676722315`
- Again, it is more practical to use different learning rate for each $Q(s, a)$ value.