Corrections for The Art of Reinforcement Learning (1st edition)

Last updated: April 9, 2024

Page 5, last paragraph, last line: "one of the greatest Go player" should be "one of the greatest Go players" (Ercan Atam)

Page 9, 7th paragraph, 1st line: "the agent may also has its internal state" should be "the agent may also have its internal state" (Ercan Atam)

Page 10, 5th paragraph, 2nd line: "probabilities of chose different actions" should be "probabilities of choosing different actions" (Ercan Atam)

Page 17, 5th paragraph, 5th line: "as the agent may learn to exploit a loophole by simply bouncing the ball back and forth on the same side of the screen without actually clearing any bricks." should be "as the agent may not learn to clear bricks efficiently or prioritize efficient ball movements." (Ercan Atam)

Page 36, 1st paragraph, line 3-4: "so a more accurate estimate that only includes legal actions is $3^5 = 243$." should be "so the actually number of valid deterministic policies may be much smaller." (Ercan Atam)

Page 37, Eq. 2.15:

$$R(s,a) + \gamma \mathbb{E}_\pi \left[Q_\pi(s',a') \;\middle|\; S_{t+1} = s', A_{t+1} = a' \right]$$

should be

$$R(s,a) + \gamma \mathbb{E}_\pi \left[Q_\pi(S_{t+1},A_{t+1}) \;\middle|\; S_t = s, A_t = a \right]$$

(Ercan Atam)

Page 53, last paragraph, 1st line: "Eq. (3.3)" should be "Eq. (3.4)" (Ercan Atam)

Page 57, 3rd paragraph, 6th line: "maximizes the expected value of the next state" should be "maximizes the expected value of immediate reward plus the value of its next state" (Ercan Atam)

Page 57, 3rd paragraph, 7th line: "repeating this process until the end of the episode" should be "repeating this process for all states in the state space" (Ercan Atam)

Page 58, 7th paragraph, 2nd line: "we select the action that yields the highest value of its successor state" should be "we select the action that yields the highest value of immediate reward plus the value of its successor state" (Ercan Atam)

Page 58, 7th paragraph, 4th line and 8th paragraph, 5th line: "highest expected reward" should be "highest expected return" (Ercan Atam)

Page 59, 2nd paragraph, 2nd line: "However, for larger problems," should be "However, for problems where the true model of the MDPs are unknown," (Ercan Atam)

Page 73, last paragraph, 3rd line: "the agent has never visited the state that often during the learning process." should be "the agent has barely visited the state during the learning process." (Ercan Atam)

Page 86, 4th paragraph, 1st line: "There is a special case when the denominator" should be "There is a special case when the numerator" (Ercan Atam)

Page 86, last paragraph, 3rd line: "Specifically, the TD target is weighted" should be "Specifically, the TD error is weighted" (Ercan Atam)

Page 89, 3rd paragraph, 7th line: "the importance sampling ratio becomes zero as well. This means that the value of the sequence will also become zero and be discarded" should be "the importance sampling ratio becomes zero or undefined as well. This means that the value of the sequence will also become zero or undefined and be discarded" (Ercan Atam)