

Exercise Sheet 9

Submission by Saturday 18th 17:00 via Moodle.

Corrections due by Wednesday 22nd 16:00 via Moodle.

Format for all questions: individual Jupyter notebooks .ipynb.

1 Gridworld (4 points)

- For a 5x5 Gridworld with no pitfalls and no exit, assume that we go up, down, left, right with equal probability. If we try to move over the boundary we remain at the same position. Write down the dynamics/transition probabilities explicitly.
- Using these probabilities generate 10 sample episodes.
- Calculate the reward G_t for each of these samples (assuming that you get a reward of 5 in the central pixel and 1 in each corner).
- Calculate $V(s)$ for all states.

2 Gridworld 2 (3 points)

Based on the gridworld example code from the lectures, try to address the following questions:

- Visualise the time evolution of the state value function Q .
- Augment the gridworld to 25x25 and implement the other updating function Q-learning instead of SARSA.
- Compare the time evolution of SARSA and Q-learning.