Project

Project

You consider the mini maze environment made up of 9 discrete states and 4 discrete actions. Once you are in the top right, you get a reward of 1 and end up in a terminal state for any subsequent action (max return is 1, see illustration below).

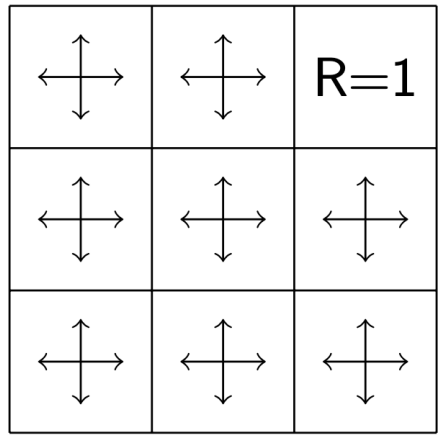

Figure: The mini maze environment ($\gamma = 0.9$). Initial state is bottom left.

In part 1, you work in the tabular context:

- Solve using tabular Q-learning and ϵ-greedy. Provide the optimal Q-values, discuss the learning rate α and ϵ (5 points).

Project

If you don't have much previous experience with function approximators such as deep learning, you can go for the following part 2 (only 4points for part 2, max grade 9/10):

- Increase the size of the grid to 5*5 and add a reward of 0.5 and terminal states for any subsequent action from the top left. What happens with the convergence to the optimal solution? What happens at some point to the optimal policy if you decrease the discount factor?

If you aim for the highest grade (5points for part 2), you can solve the mini maze problem using function approximators for $\gamma = 0.9$ and 5*5 grid instead of 3*3.

- Provide illustrations of the solutions of your optimal Q-values (2 points)
- Discuss the hyper-parameters and the convergence (2 points)

Project

- Provide a report of max 2 pages (+appendix allowed)
- Provide the source code in python