

Tabular Implementation of MaxMin-Q-Learning Algorithm

In this project work your task is to implement the MaxMin-Q-learning algorithm and solve a very simple example problem by using the algorithm. For comparison purposes, try to solve a problem also by using normal Q-learning, i.e. you have two single-agent Q-learners in the same environment.

The environment (Fig. 1) where two learning agents are located is a simple grid world. The agents are initially located in the cells marked with the numbers 1 and 2. There is only one goal cell (marked with a cross) on the grid and the goal of the agents is to reach this cell as fast as possible. After the agents have made their action selections, actions are executed in random order. Therefore only one agent can reach the goal cell implicating the zero-sum property.

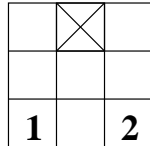


Figure 1: Grid world used for testing MaxMin-Q-Learning

Compare policies generated by the single-agent Q-learning algorithm and MaxMin-Q-learning algorithm. Compare also convergence speeds of the algorithms.