# Reinforcement Learning Notes

## The Agent-Environment Interface

Markov Decision Process is solving the problem of learning from interaction to achieve a goal. The learner and decision maker is called an agent.

## Policies and Value Functions

Almost all reinforcement learning algorithms involve estimating **value functions** – functions of state (or state-action pairs) that estimate how good it is for the agent to be in a given state (or how good it is to perform a given action in a given state). The notion of “how good” is defined in terms of future rewards that can be expected, or to be precise, in terms of expected return. Of course, the reward the agent can expect to receive in the future depend on what actions it will take. Accordingly, value functions are defined with respect to particular ways of acting, called **policies**.

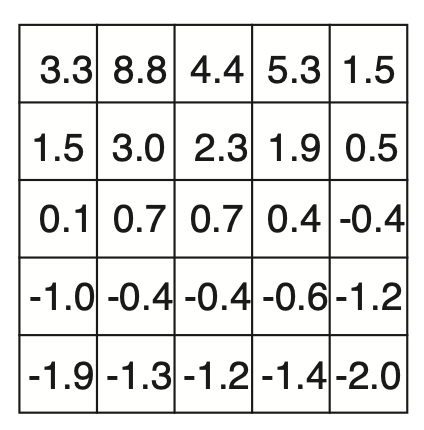
Formally, a **policy** is a mapping from states to probabilities of selecting each possible action. If an agent is following policy at time , then is the probability that if . Like , is an ordinary function which defines a probability distribution.

***Gridworld***

Gridworld is an example illustrating the use of a finite Markov Decision Process.

The cells of the grid correspond to the states of the environment. At each cell, four actions are possible: ***north***, ***south***, ***east***, and ***west***, which deterministically cause the agent to move one cell in the respective direction on the grid. Actions that would take the agent off the grid leave its location unchanged, but also result in a reward of -1.

Other actions result in a reward of 0, except those that move the agent out of the special states and . From state , all four actions yield a reward of +10 and take the agent to . From state B



A picture containing symbol, font, white, graphics

Description automatically generatedA picture containing diagram, line

Description automatically generated

//TODO: finish the discussion on Markov decision processes