Policy Gradients

Prof. Alfio Ferrara

Reinforcement Learning

Introduction

We discuss how, instead of estimating the value of actions in order to pick up the right one, we can learn a parametrized policy as a tool to select actions. In VFA, the policy is generated from the value function, for example using ϵ-greedy strategy. Now, we will directly parametrize the policy:

$$\pi_\theta(s, a) = \mathbb{P}(a \mid s; \theta) \tag{1}$$

in order to find a policy π with the best V^π. Also in this case, we work on model-free algorithms.