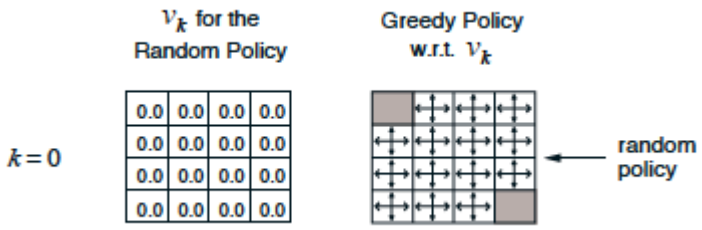
Summary



First step of policy iteration in gridworld example (Sutton and Barto, 2017)

Introduction

- In the **dynamic programming** setting, the agent has full knowledge of the MDP. (This is much easier than the **reinforcement learning** setting, where the agent initially knows nothing about how the environment decides state and reward and must learn entirely from interaction how to select actions.)

An Iterative Method

- In order to obtain the state-value function v_π corresponding to a policy π, we need only solve the system of equations corresponding to the Bellman expectation equation for v_π.
- While it is possible to analytically solve the system, we will focus on an iterative solution approach.

Iterative Policy Evaluation

- **Iterative policy evaluation** is an algorithm used in the dynamic programming setting to estimate the state-value function v_π corresponding to a policy π. In this approach, a Bellman update is applied to the value function estimate until the changes to the estimate are nearly imperceptible.