Week3 Notes

===========

Lesson 1: Policies and Value Functions

Recognize that a policy is a distribution over actions for each possible state

Describe the similarities and differences between stochastic and deterministic policies

deterministic policy maps a state to action.

stochastic policy - multiple actions can be selected for each state.

Identify the characteristics of a well-defined policy

Generate examples of valid policies for a given MDP

Policies must depend on current state. Not time or previous state.

the state value function is the expected return from a given state

An action value describes what happens when the agent first selects a particular action. More formally, the action value of a state is the expected return if the agent selects action A and then follows policy Pi.

Describe the roles of state-value and action-value functions in reinforcement learning

short term vs long term gain.

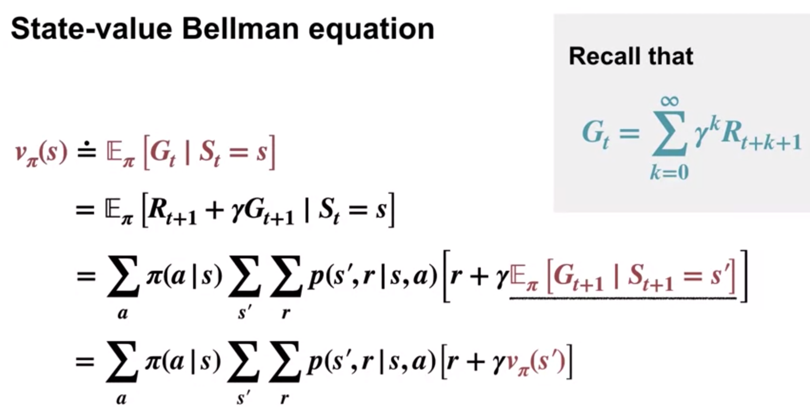
Describe the relationship between value functions and policies

a state value function refers to the expected return from a given state under a specific policy, and an action value function refers to the expected return from a given state after selecting a particular action and then following a given policy.

Create examples of valid value functions for a given MDP

Lesson 2: Bellman Equations

Derive the Bellman equation for state-value functions



Derive the Bellman equation for action-value functions

Understand how Bellman equations relate current and future values

The important thing to note is that the Bellman equation reduced an unmanageable infinite sum over possible futures, to a simple linear algebra problem. Perhaps for this small problem, you can come up with other ways to work out the values of each of these states. However the Bellman equation provides a powerful general relationship for MDPs. In this case, we used the Bellman equation to directly write down a system of equations for the state values, and then some the system to find the values

Use the Bellman equations to compute value functions

Lesson 3: Optimality (Optimal Policies & Value Functions)

Define an optimal policy

the goal of reinforcement learning is not just to evaluate specific policies. Ultimately, we want to find a policy that obtains as much reward as possible in the long run

Understand how a policy can be at least as good as every other policy in every state

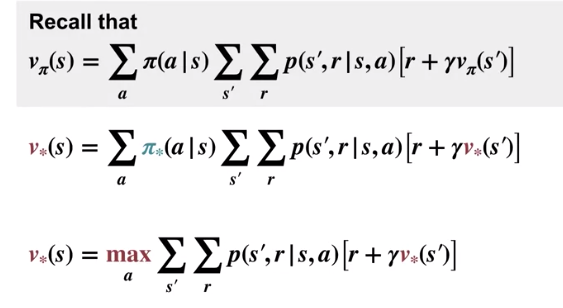
Identify an optimal policy for given MDPs

The most important things to take away are an optimal policy is defined as the policy with

the highest value in all states. At least one optimal policy always exists but there may be more than one. The exponential number of possible policies, makes searching for the optimal policy by brute force intractable.

Derive the Bellman optimality equation for state-value functions

Derive the Bellman optimality equation for action-value functions



Understand how the Bellman optimality equations relate to the previously introduced Bellman equations

Understand the connection between the optimal value function and optimal policies

Verify the optimal value function for given MDPs