Week2 Notes

=============

Lesson 1: Introduction to Markov Decision Processes

Understand Markov Decision Processes, or MDPs



Describe how the dynamics of an MDP are defined

Understand the graphical representation of a Markov Decision Process

Explain how many diverse processes can be written in terms of the MDP framework

Lesson 2: Goal of Reinforcement Learning

Describe how rewards relate to the goal of an agent

That all of what we mean by goals and purposes can be well thought of as

the maximization of the expected value of the cumulative sum of a received

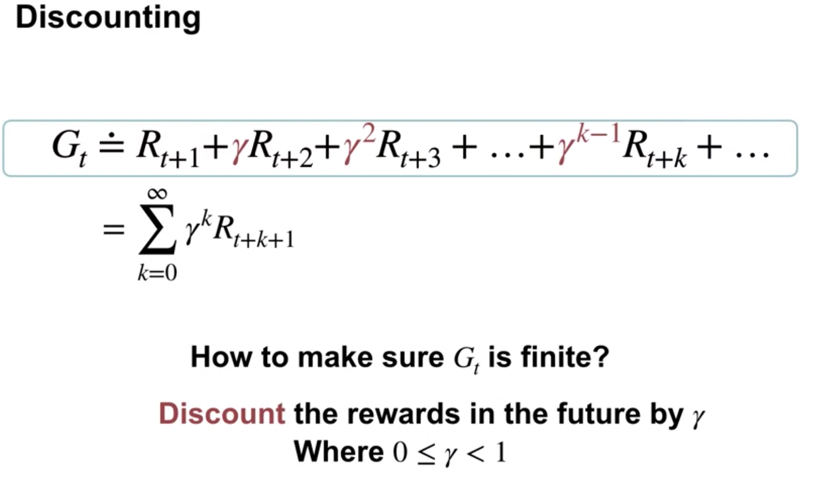
scalar signal (called reward).

Understand episodes and identify episodic tasks

In the simplest case the return is the sum of the rewards: Gt . = Rt+1 + Rt+2 + Rt+3 + ··· + RT , (3.7) where T is a final time step. This approach makes sense in applications in which there is a natural notion of final time step, that is, when the agent–environment interaction breaks naturally into subsequences, which we call episodes, 7 such as plays of a game, trips through a maze, or any sort of repeated interaction. Each episode ends in a special state called the terminal state, followed by a reset to a standard starting state or to a sample from a standard distribution of starting states

Lesson 3: Continuing Tasks

Formulate returns for continuing tasks using discounting



Describe how returns at successive time steps are related to each other

Understand when to formalize a task as episodic or continuing

Returns at successive time steps are related to each other in a way that is important for the theory and algorithms of reinforcement learning: Gt . = Rt+1 + Rt+2 + 2Rt+3 + 3Rt+4 + ··· = Rt+1 + Rt+2 + Rt+3 + 2Rt+4 + ··· = Rt+1 + Gt+1

Suppose \gamma=0.8*γ*=0.8 and we observe the following sequence of rewards: R\_1 = -3*R*1​=−3, R\_2 = 5*R*2​=5, R\_3=2*R*3​=2, R\_4 = 7*R*4​=7, and R\_5 = 1*R*5​=1, with T=5*T*=5. What is G\_0*G*0​? Hint: Work Backwards and recall that G\_t = R\_{t+1} + \gamma G\_{t+1}*Gt*​=*Rt*+1​+*γGt*+1​.

gamma=0.8

R\_1= -3

R\_2 = 5

R\_3 = 2

R\_4 = 7

R\_5 = 1

T = 5

R\_1 + (gamma\*R\_2) + ((gamma\*\*2)\*R\_3) + ((gamma\*\*3)\*R\_4) + ((gamma\*\*4)\*R\_5)

=6.273600000000001