580 Final Study Guide

Connor A.

Spring 2020

Chapter 17

1 What's a MDP?

- With an MDP there's **uncertainty** in the result of an action — **in what state we'll end up in after we transition from state s to s'**
- The possible outcomes/states of an action depends only on the state we're currently in.
- Formally, it's defined by — set of states s $\in$ S, set of actions a $\in$ A, and transition function T(s, a, s').
- A transition function T() returns probability of getting to next state we want as well as the next state we actually ended up in.

2 Example MDP *(Grid World)*

- Each cell is a state, 11 states because the "12th" is unreachable
- 5 actions, (N—S—E—W) and 'exiting'
- Reward function R() — Cost of each "step", negative in this case to maximize utility

3 Policy

- We solve a MVP by coming up with a policy — **a reward system**, R(s)
- Given all states, we ask what is the **optimal action** to take at each individual state
- We want something that will **maximize utility**
- Works best if state space is **small**