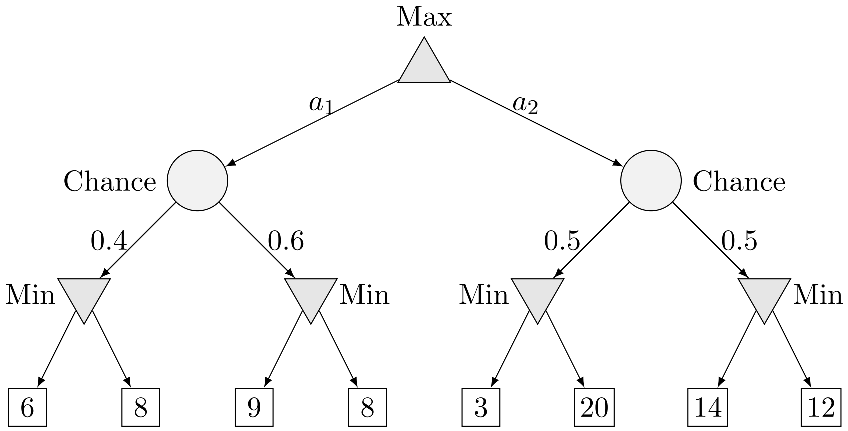
Stochastic Environments Practice

CIS 5210

March 18, 2025

Question 1. Expectimax

What is the expected utility of picking action a_1? What is the expected utility of picking action a_2? Using Expectimax, which action will the Max player select? *(You might need a calculator.)*



Solution:

Expected utility of a_1 is $0.4 * 6 + 0.6 * 8 = 7.2$. Expected utility of a_2 is $0.5 * 3 + 0.5 * 12 = 7.5$. Max will play a_2.

Question 2. Signal in the Noise

Consider a grid-based agent in a Markov Decision Process setting. In these settings, we assume that the agent moves in the direction specified by their action with a certain probability. If the agent does not move in the intended direction, they choose uniformly at random from one of the two directions offset by 90 degrees from the intended direction. (Intending to move LEFT, the agent might move LEFT, UP, or DOWN, but never RIGHT.) The probability that an agent moves in the wrong direction is referred to as the *noise value.*

If an agent has a noise value of 0.2, fill in the blanks in the terms below representing the agent's transition function for taking the action UP from the state represented by coordinate $(x = 5, y = 5)$ with no adjacent obstacles or boundaries. Higher coordinates are to the right and up.

- $T((5, 5), \text{UP}, (5, 5)) =$ _____
- $T((5, 5), \text{UP}, (4, 5)) =$ _____
- $T((5, 5), \text{UP}, (6, 5)) =$ _____
- $T((5, 5), \text{UP}, (5, 4)) =$ _____
- $T((5, 5), \text{UP}, (5, 6)) =$ _____

Solution:

- $T((5,5), \text{UP}, (5,5)) = 0$
- $T((5,5), \text{UP}, (4,5)) = 0.1$
- $T((5,5), \text{UP}, (6,5)) = 0.1$
- $T((5,5), \text{UP}, (5,4)) = 0$
- $T((5,5), \text{UP}, (5,6)) = 0.8$

Question 3.

The Hyperdrive MDP is described by the following transition and reward functions:

$T(s, a, s')$	Cruising	Hyperspace	Crashed
Cruising, Maintain	1.0	0	0
Cruising, Punch It	0.5	0.5	0
Hyperspace, Maintain	0.5	0.5	0
Hyperspace, Punch It	0	0	1.0
Crashed, $*$	0	0	1.0

$R(s, a, s')$	Cruising	Hyperspace	Crashed
Cruising, Maintain	1	1	0
Cruising, Punch It	2	2	0
Hyperspace, Maintain	1	1	0
Hyperspace, Punch It	0	0	-10
Crashed, $*$	0	0	0

1. Assume that the start state is always “Cruising”. Come up with two policies π_1 and π_2 that lead to guaranteed infinite rewards in a setting of the game where the horizon is infinite and $\gamma = 1.0$.
2. If the setting is changed so that $0 < \gamma < 1$, the values of each state will converge to a finite value after a certain number of actions. Can you modify the game in a different way, without changing γ, so that the values would also converge to a finite value? As a hint, consider the role of “Crashed”” as an *absorbing state*.

Solution:

1. $\pi_1(s) =$ Maintain for all values of s. $\pi_2(\text{Cruising}) =$ Punch It, $\pi_2(\text{Hyperspace}) =$ Maintain
2. Adjust the probabilities so that “Crashed” is reachable from all actions and all states. In the limit, a player will crash, capping expected utility.