This is ungraded examination preparation.

CSCI 4511/6511 - Exam Prep 5

Write your name here:_____

Instructions:
This is exam prep and is not for credit. You should try to complete it on your own, without references.

This is ungraded examination preparation.

1 Markov Decision Processes

The Ignatius Coffee Company in Cleric, AZ roasts coffee for Carl's Diner in South Cleric. Every week, Ignatius has to order coffee beans from his supplier in Honduras. The beans arrive the following week.

Ignatius has noticed a pattern in Carl's ordering:

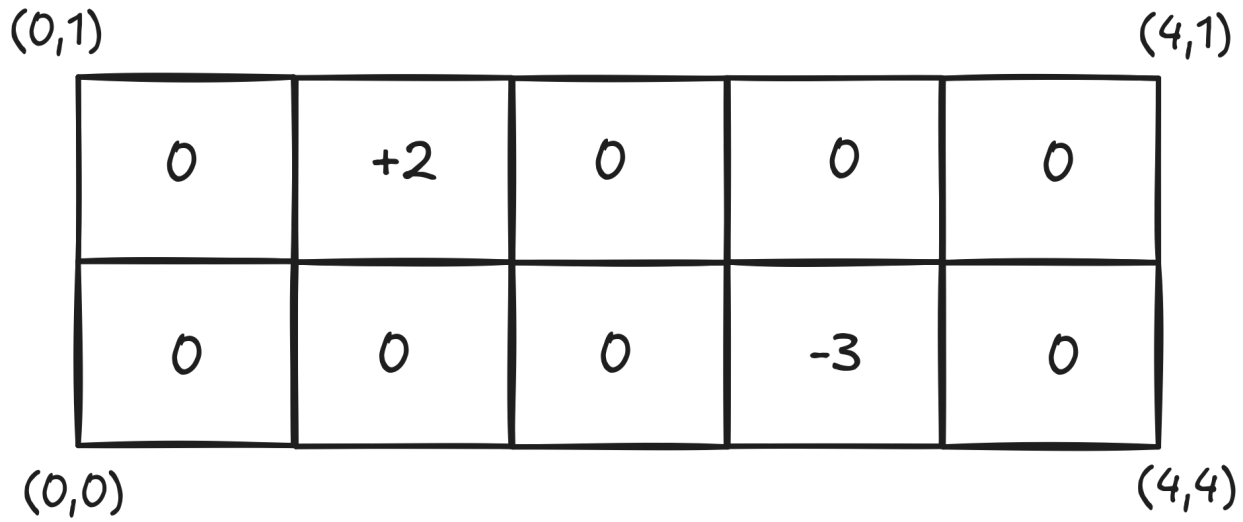
- Carl orders at least 40, and not more than 50 bags of coffee each week.
- Each week's order is within two bags' of the previous week's order, with equal probability within these values:
  - If Carl ordered 45 bags last week, the next week is equally probable to be 43, 44, 45, 46, or 47 bags.
  - If Carl ordered 40 bags last week, the next week is equally probable to be 40, 41, or 42 bags.
- Ignatius can order any number of bags of coffee from his supplier:
  - Bags of unroasted coffee cost Ignatius $2
  - Ignatius sells roasted coffee bags to carl for $10 each
  - Ignatius values every missed order[1] at negative $3.

Model Ignatius's decision for coffee ordering as a Markov Decision Process. You only need to set up the problem: show all states, state transitions, and rewards, and describe mathematically an algorithm to solve the problem. You do not need to provide a solution.

[1] A missed order is a bag of coffee that Carl would like to buy, but can't, because Ignatius is out of stock

2 Reinforcement Learning

For a problem with states arranged on the following grid, with rewards for entering each grid square shown:



- Consider an action space where an agent can deterministically move to any adjacent (shares an edge) state.
- Use a learning rate α of 0.5
- The problem will terminate when reward exceeds 4 or falls below -4; do not discount.
- Start with $Q(s, a) = 0 \quad \forall s, a$

2.1 Sarsa

Show how the Sarsa algorithm updates $Q(s, a)$ for an agent starting at point $(0, 0)$. moving one state to the right, and then moving one state up.

2.2 Sarsa-λ

Show how the Sarsa-λ algorithm updates $Q(s, a)$ for an agent starting at point $(0, 0)$ and moving right three times. Use a λ value of 0.5.

This is the back of ungraded examination preparation.

This is the back of ungraded examination preparation.