

This handout includes space for every question that requires a written response. Please feel free to use it to handwrite your solutions (legibly, please). If you choose to typeset your solutions, the `README.md` for this assignment includes instructions to regenerate this handout with your typeset  $\text{\LaTeX}$  solutions.

---

1.a

## Initialization (Iteration 0)

- $V(-2) = 0$  (Terminal State)
- $V(-1) = 0$
- $V(0) = 0$
- $V(1) = 0$
- $V(2) = 0$  (Terminal State)

## Iteration 1

Using the Bellman equation for value iteration,  $V(s)$  values are calculated as:

For state  $-1$ :

$$V(-1) = \max(0.2 \times (-5 + 0) + 0.8 \times (20 + 0), 0.3 \times (-5 + 0) + 0.7 \times (20 + 0))$$

$$V(-1) = \max(16, 13.5)$$

$$V(-1) = 16$$

For state  $0$ :

$$V(0) = \max(0.2 \times (-5 + 0) + 0.8 \times (-5 + 0), 0.3 \times (-5 + 0) + 0.7 \times (-5 + 0))$$

$$V(0) = \max(-5, -5)$$

$$V(0) = -5$$

For state  $1$ :

$$V(1) = \max(0.2 \times (100 + 0) + 0.8 \times (-5 + 0), 0.3 \times (100 + 0) + 0.7 \times (-5 + 0))$$

$$V(1) = \max(16, 26.5)$$

$$V(1) = 26.5$$

## Iteration 2

For state  $-1$ :

$$V(-1) = \max(0.2 \times (-5 + -5) + 0.8 \times (20 + 0), 0.3 \times (-5 + -5) + 0.7 \times (20 + 0))$$

$$V(-1) = \max(14, 11)$$

$$V(-1) = 14$$

For state  $0$ :

$$V(0) = \max(0.2 \times (-5 + 26.5) + 0.8 \times (-5 + 16), 0.3 \times (-5 + 26.5) + 0.7 \times (-5 + 16))$$

$$V(0) = \max(13.1, 14.15)$$

$$V(0) = 14.15$$

For state 1:

$$V(1) = \max(0.2 \times (100 + 0) + 0.8 \times (-5 + -5), 0.3 \times (100 + 0) + 0.7 \times (-5 + -5))$$

$$V(1) = \max(12, 23)$$

$$V(1) = 23$$

## Summary

### After Iteration 0:

- $V(-2) = 0$
- $V(-1) = 0$
- $V(0) = 0$
- $V(1) = 0$
- $V(2) = 0$

### After Iteration 1:

- $V(-2) = 0$
- $V(-1) = 16$
- $V(0) = -5$
- $V(1) = 26.5$
- $V(2) = 0$

### After Iteration 2:

- $V(-2) = 0$
- $V(-1) = 14$
- $V(0) = 14.15$
- $V(1) = 23$
- $V(2) = 0$

1.b

- $S(-1)$ : the best policy is take  $A(-1)$ , which will have  $V_{\text{opt}}(-1) = 14$
- $S(0)$ : the best policy is take  $A(1)$ , which will have  $V_{\text{opt}}(0) = 14.15$
- $S(1)$ : the best policy is take  $A(1)$ , which will have  $V_{\text{opt}}(0) = 23$

2.a

4.b

4.d

5.a

5.b



5.c

5.d