Understanding $V_\pi(s)$ with Gridworld

$$v_\pi(s) \doteq \sum_a \pi(a|s) \sum_{s',r} p(s',r|s,a)\Big[r + \gamma v_\pi(s')\Big]$$

Verify $V_\pi(s)$ using Bellman equation for this state with γ = 0.9, and equiprobable random policy

3.3	8.8	4.4	5.3	1.5
1.5	3.0	2.3	1.9	0.5
0.1	0.7	0.7	0.4	-0.4
-1.0	-0.4	-0.4	-0.6	-1.2
-1.9	-1.3	-1.2	-1.4	-2.0

Understanding V_{π}(s) with Gridworld

$$v_\pi(s) \doteq \sum_a \pi(a|s) \sum_{s',r} p(s',r|s,a)\Big[r + \gamma v_\pi(s')\Big]$$

$$V^\pi(s) = \sum_{a \in A} \pi(a \mid s) \sum_{s' \in \{s_1,s_2,s_3,s_4\}} P(s' \mid s,a)[0 + \gamma V^\pi(s')]$$

$$V^\pi(s) = 0.25 \sum_{s' \in \{s_1,s_2,s_3,s_4\}} [0 + 0.9 V^\pi(s')]$$

$$V^\pi(s) = 0.25\,[0.9\,(2.3 + 0.4 - 0.4 + 0.7)]$$

$$= 0.25 \cdot [0.9 \cdot 3.0] = 0.675 \approx 0.7$$

3.3	8.8	4.4	5.3	1.5
1.5	3.0	2.3	1.9	0.5
0.1	0.7	0.7	0.4	-0.4
-1.0	-0.4	-0.4	-0.6	-1.2
-1.9	-1.3	-1.2	-1.4	-2.0

Goal: Verify the value of the circled state (which is shown as `0.7`) using the **Bellman expectation equation**, with:

- Discount factor $\gamma = 0.9$
- **Equiprobable random policy**: meaning the agent picks any of the 4 actions with **equal probability** (0.25)
- Each action leads to **1 of 4 neighboring states** (assuming deterministic transitions)

3.3	8.8	4.4	5.3	1.5
1.5	3.0	2.3	1.9	0.5
0.1	0.7	0.7	0.4	-0.4
-1.0	-0.4	-0.4	-0.6	-1.2
-1.9	-1.3	-1.2	-1.4	-2.0

Equation used:

$$v_\pi(s) = \sum_a \pi(a|s) \sum_{s'} p(s'|s,a) \left[r + \gamma v_\pi(s')\right]$$

But in this case:

- All rewards $r = 0$
- Only value of successor states matters
- $\pi(a|s) = 0.25$ (equal for each action)

So the simplified version becomes:

$$v_\pi(s) = 0.25 \cdot \sum_{s'} \left[\gamma \cdot v_\pi(s')\right]$$

$$v_\pi(s) = 0.25 \cdot \gamma \cdot (v(s_1) + v(s_2) + v(s_3) + v(s_4))$$

$$v_\pi(s) = 0.25 \cdot 0.9 \cdot (2.3 + 0.4 + (-0.4) + 0.7)$$

Simplify the inner sum:

$$= 0.25 \cdot 0.9 \cdot 3.0 = 0.25 \cdot 2.7 = \boxed{0.675 \approx 0.7}$$

- Transitions are deterministic
- So each action deterministically leads to one s'
- Therefore $p(s' \mid s, a) = 1$
- So it becomes:

$$v_\pi(s) = \sum_a \pi(a \mid s)[r + \gamma v_\pi(s')]$$

And with:

- $r = 0$
- $\pi(a \mid s) = 0.25$

  It simplifies to:

$$v_\pi(s) = 0.25 \cdot \sum_{s'} \gamma v_\pi(s') = 0.25 \cdot \gamma \cdot \sum_{s'} v_\pi(s')$$