

准备好后再次尝试

获得的成绩 66.66% 最新提交作业的评分 66.67% 通过条件 80% 或更高

再试

您正在查看此评估的自动翻译版本
您可以根据需要切换回去，以源语言查看这些内容。即使您更改语言，也不会丢失任何进度。[显示英语（English）版本](#)

取消

1.

0 / 1个点

以下哪项能准确描述状态-行动值函数 $Q(s, a)$ ？

- ☐ 如果您从状态 s 开始，采取行动 a （一次），之后的行为达到最佳状态，那么它就是您的回报。
- ☐ 如果您从状态 s 开始，并反复采取行动 a ，这就是回报。
- ☒ 如果从状态 s 开始，并采取最佳行为，收益就是这样。
- ☐ 如果您从 s 开始，并采取行动 a （一次），就会立即得到回报。

错误
回顾一下，我们需要先执行 a 。

2.

1 / 1个点

您正在控制一个有 3 个动作的机器人：←（左）、→（右）和停止。在给定状态 s 下，您计算出 $Q(s, \leftarrow) = -10$, $Q(s, \rightarrow) = -20$, $Q(s, \text{STOP}) = 0$ 。

在 s 状态下的最优操作是什么？

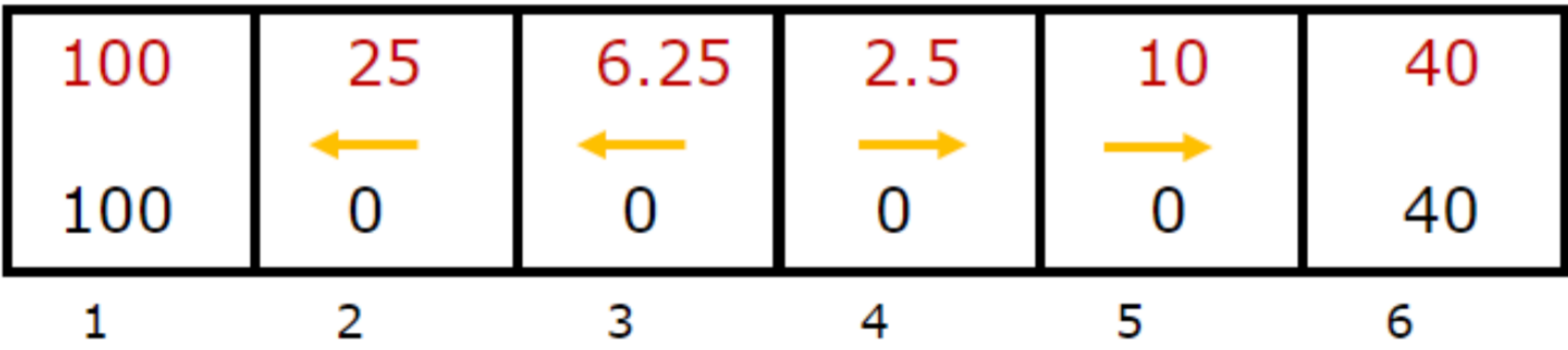
- ☒ 停止
- ☐ ←（左）
- ☐ →（右）
- ☐ 无法判断

正确
是的，因为这具有最大的价值。

3.

1 / 1个点

对于这个问题， $\gamma = 0.25$ 。下图显示了每个状态的收益和最优行动。请计算 $Q(5, \leftarrow)$ 。



← return
← action
← reward

$Q(5, \leftarrow) = ?$

- ☒ 0.625
- ☐ 0.391
- ☐ 1.25
- ☐ 2.5

正确
是的，我们在状态 5 中获得了 0 奖励。然后， $0 * 0.25$ ，因为我们的行动向左移动了，所以在状态 4 中的奖励打了折扣。现在，我们从状态 4 开始的行为是最优的。因此，我们从状态 4 向右移动到状态 5，并获得 $0 * 0.25^2$ 折扣奖励。最后，我们从状态 5 向右移动到状态 6，得到 $40 * 0.25^3$ 的折扣奖励。将这些相加，得到 0.625。