

1. 强化学习的基本目标是什么？

- A. 最大化短期奖励
- B. 减少环境的复杂性
- C. 最大化累积奖励的总值
- D. 完全避免随机性

你的答案：

2. 强化学习中，智能体通过什么与环境进行交互？

- A. 网络模型
- B. 状态、动作和奖励
- C. 数据库
- D. 优化算法

你的答案：

3. 强化学习问题通常建模为哪种数学框架？

- A. 线性规划
- B. 动态规划
- C. 马尔可夫决策过程
- D. 贝叶斯网络

你的答案：

4. 在马尔可夫决策过程中，状态的转移和奖励取决于什么？

- A. 环境的复杂度
- B. 当前状态和智能体的动作
- C. 时间步的长度
- D. 算法的类型

你的答案：

5. 什么是确定性策略？

- A. 根据环境随机选择动作
- B. 以概率分布选择动作
- C. 为每个状态指定唯一的动作
- D. 根据外部因素改变策略

你的答案：

6. 值迭代方法的核心思想是什么？

- A. 独立评估值函数
- B. 改进策略后再评估值函数
- C. 将策略评估和策略改进过程合并
- D. 采用监督学习方法

你的答案：

7. Q 学习算法主要用于什么类型的强化学习？

- A. 基于模型的强化学习
- B. 模型无关的强化学习
- C. 基于深度学习的强化学习
- D. 监督学习

你的答案：

8. 在 SARSA 算法中，智能体的学习主要依赖于哪五个元素？

- A. 状态、动作、奖励、环境、目标
- B. 策略、值函数、环境、奖励、目标
- C. 当前状态、动作、奖励、下一状态、下一动作
- D. 状态、价值、奖励、目标、策略

你的答案：

9. 深度 Q 学习 (DQN) 的一个显著特点是什么？

- A. 不需要值函数
- B. 只能处理线性问题
- C. 使用神经网络近似值函数
- D. 无法应用于复杂环境

你的答案：

10. AlphaGo 使用了哪种强化学习方法来学习围棋策略？

- A. 强监督学习
- B. 结合监督学习和强化学习
- C. 随机搜索
- D. 动态规划

你的答案：

11. 在强化学习中，维度灾难是指什么？

- A. 状态空间随着维度增加呈指数增长
- B. 动作数量过多
- C. 环境变化速度过快
- D. 奖励延迟

你的答案：

12. 探索与利用的困境主要体现在以下哪个方面？

- A. 智能体是否应优先探索新信息还是利用已知信息
- B. 状态转移的随机性
- C. 智能体的复杂性
- D. 环境反馈的延迟性

你的答案：

13. 强化学习的一个主要挑战是：

- A. 过于依赖动态规划
- B. 数据需求低
- C. 非平稳环境
- D. 可扩展性差

你的答案：

14. AlphaGo Zero 通过强化学习解决了什么问题？

- A. 如何评估围棋局势的复杂性
- B. 优化搜索时间
- C. 训练深度学习模型
- D. 动态规划的应用

你的答案：

15. 部分可观测性环境对强化学习的影响是：

- A. 提升了智能体的性能
- B. 限制了智能体对环境的全面了解
- C. 增加了数据收集难度
- D. 完全破坏了学习过程

你的答案：

16. 时间延迟的奖励会导致什么问题？

- A. 奖励过于稀疏
- B. 环境不可观测
- C. 对远离的先前状态的影响非常弱
- D. 无法更新值函数

你的答案：

17. 在强化学习中，目标函数通常是什么？

- A. 累积惩罚的最小值
- B. 奖励函数的平均值
- C. 累积奖励的期望值
- D. 智能体的准确度

你的答案：

18. 哪个不是深度强化学习的挑战？

- A. 维度灾难
- B. 非平稳环境
- C. 时间延迟奖励
- D. 线性问题建模

你的答案：

19. 深度强化学习在 Atari 游戏中的典型

应用是基于哪种算法？

- A. SARSA
- B. 蒙特卡洛方法
- C. 深度 Q 学习 (DQN)
- D. 随机搜索

你的答案：

20. 在深度强化学习中，蒙特卡洛树搜索的主要作用是什么？

- A. 增加随机性
- B. 减少数据需求
- C. 通过预测对手动作减少搜索空间
- D. 简化奖励函数

你的答案：