B， 最小化损失

C， 直接学习值函数

D， 直接学习策略函数

答案

填空题

1. **最优策略**
2. **Q-learning**
3. **18小时**；**2小时**；

选择题

1. C，4种动作都有可能。
2. D；
3. B；
4. A；
5. A；
6. B；
7. A；
8. A；
9. D；