人工智能导论

----行为学派例题

1）假设有一个 3×2 的网格世界，起始点为节点 1，目标点为节点 6，智能体可进行上、下、左、右移动的动作。当到达节点 6 时，获得的奖励为+10，其余所有动作的奖励均为-1。

4	5	6（Goal）
1（Start）	2	3

假设有如下策略：

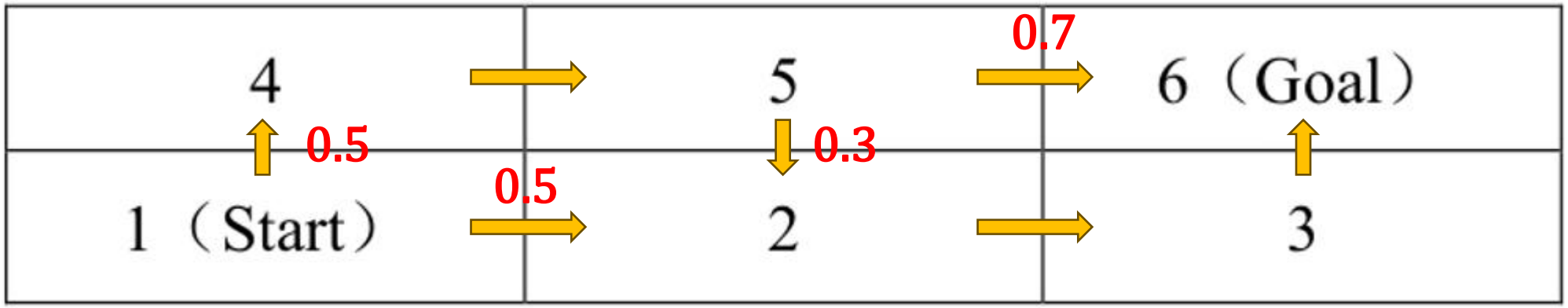
Right	30%Down， 70%Right	6（Goal）
50%Up， 50%Right	Right	Up

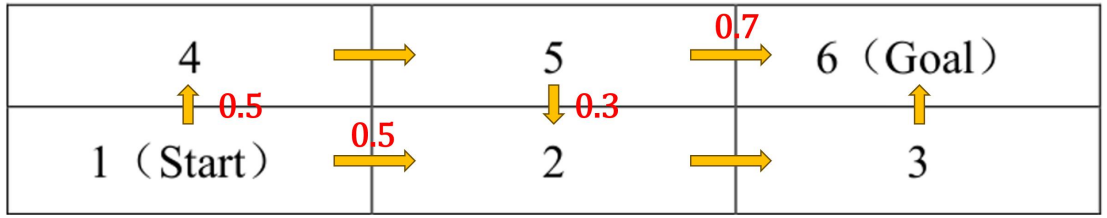
令折扣因子 Gamma=0.9，试计算 1 状态和 5 状态的值函数（3 分）。

$$V_\pi(s) = E(R_1 + \gamma R_2 + \gamma^2 R_3 + \cdots + \gamma^{T-1} R_T)$$

（0.9^2 = 0.81，0.9^3 = 0.729，0.9^4 = 0.6561，0.9^5 = 0.59049）

Right	30%Down， 70%Right	6（Goal）
50%Up， 50%Right	Right	Up





$$V(6) = 0$$

$$V(3) = r(3, \uparrow, 6) + \gamma * V(6) = 10$$

$$V(2) = r(2, \rightarrow, 3) + \gamma * V(3) = -1 + 0.9 * 10 = 8$$

$$V(5) = P(\downarrow|5) * \left(r(5, \downarrow, 2) + \gamma * V(2)\right) + P(\rightarrow|5) * \left(r(5, \rightarrow, 6) + \gamma * V(6)\right)$$
$$= 0.3 * (-1 + 0.9 * 8) + 0.7 * (10 + 0.9 * 0) = 8.86$$

$$V(4) = r(4, \rightarrow, 5) + \gamma * V(5) = -1 + 0.9 * 8.86 = 6.9740$$

$$V(1) = P(\uparrow|1) * \left(r(1, \uparrow, 4) + \gamma * V(4)\right) + P(\rightarrow|1) * \left(r(1, \rightarrow, 2) + \gamma * V(2)\right)$$
$$= 0.5 * (-1 + 0.9 * 6.9740) + 0.5 * (-1 + 0.9 * 8) = 5.7383$$