ML_强化学习

填空题

1，最优动作价值函数Q_*依赖于_____。

2，DQN是对_____的近似。

3，驾车按照“甲，乙，丙”行驶，从甲地出发，模型预计需要行驶20小时，实际行驶6小时到达乙地，模型预计还需12个小时到达丙地，如果我们用TD算法更新模型，那么TD目标$\hat{y}$ =_____小时，TD绝对误差值$|\delta|$_____小时；

选择题

1，设$A = \{上，下，左，右\}$为动作空间，s_t为当前状态，π为策略函数，策略函数的输出：

$$\begin{aligned}\pi(上|s_t) &= 0.2,\\ \pi(下|s_t) &= 0.05,\\ \pi(左|s_t) &= 0.7,\\ \pi(右|s_t) &= 0.15.\end{aligned} \tag{1}$$

请问，哪个动作会成为a_t?

A，下

B，左

C，4种动作都有可能

2，设随机变量U_t为t时刻的回报，请问U_t依赖于哪些变量?

A，t时刻的状态S_t

B，t时刻的动作A_t

C，S_t和A_t

D，$S_t, S_{t+1}, S_{t+2}, \cdots$和$A_t, A_{t+1}, A_{t+2}, \cdots$

3，动作价值函数是什么的期望?

A，奖励

B，回报

C，状态

D，动作

4，设$A = \{上，下，左，右\}$为动作空间，s_t为当前状态，Q_*为最优动作价值函数，策略函数的输出：