**ML\_强化学习**

**填空题**

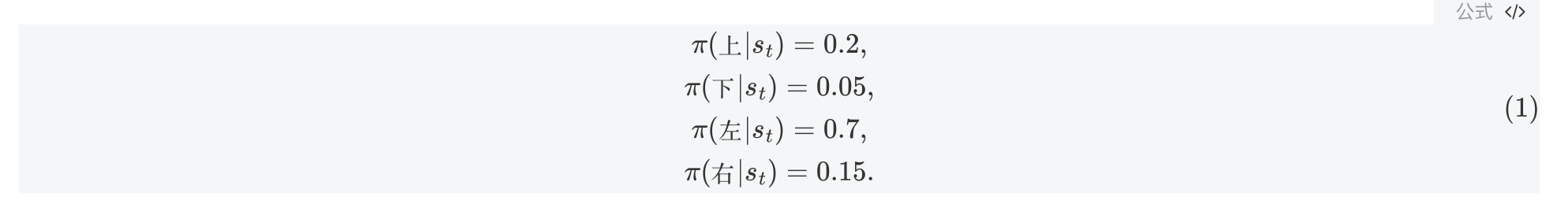
1，最优动作价值函数依赖于**\_\_\_最优策略\_**\_\_。

2，DQN是对**\_\_\_最优动作价值函数\_**\_\_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   =**\_\_\_18\_**\_\_小时，TD绝对误差值||**\_\_\_2\_**\_\_小时；

**选择题**

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



请问，哪个动作会成为?

A，下

B，左

C，4种动作都有可能

2，设随机变量为t时刻的回报，请问依赖于哪些变量？

A，t时刻的状态

B，t时刻的动作

C，和

D，和,

3，动作价值函数是什么的期望？

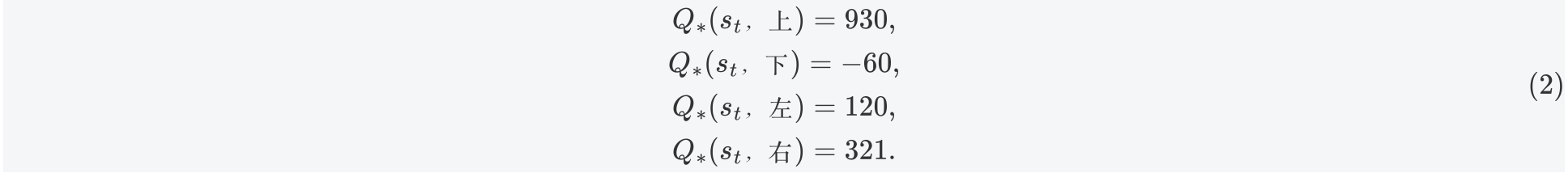
A，奖励

B，回报

C，状态

D，动作

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



请问，哪个动作会成为?

A，上

B，下

C，4种动作都有可能

5，DQN的输出层用于什么激活函数？

A，不需要激活函数，因为Q值可正可负，没有取值范围

B，用sigmoid激活函数，因为Q值介于0和1之间

C，用ReLU激活函数，因为Q值非负

D，用SoftMax激活函数，因为DQN的输出是一个概率分布

6，多臂赌博机是单步强化学习的经典范例，𝜺贪心算法和SoftMax算法用于处理什么问题？

A， 探索-利用问题

B， 奖励函数优化问题

C， 动作选择问题

D， 状态空间问题

7，DQN（深度 Q 网络）是基于什么的强化学习方法？

A， 基于值的方法

B， 基于策略的方法

C， 基于模型的方法

D， 基于探索的方法

8，TD gradient 是与哪种强化学习方法相关的概念？

A， 基于值的方法

B， 基于策略的方法

C， 基于模型的方法

D， 基于探索的方法

9，在强化学习中，基于策略的方法主要关注什么？

A， 最大化奖励

B， 最小化损失

C， 直接学习值函数

D， 直接学习策略函数