机器学习第六次作业

2020年12月

1 第一题

MDP中对于给定的策略π，状态-值函数（state-value function）为$v_\pi(s)$。

（1）证明下面两个对于状态-值函数的定义等价:

$$\begin{aligned} v_\pi(s) &= \mathrm{E}_\pi[G_t|S_t = s] \\ v_\pi(s) &= \mathrm{E}_\pi[G_0|S_0 = s] \end{aligned} \tag{1}$$

（2）对于策略π_1和π_2，如下定义基于状态-值函数的偏序关系$\geq$:

$$\pi_1 \geq \pi_2 \iff \forall s \in \mathcal{S}, v_{\pi_1}(s) \geq v_{\pi_2}(s) \tag{2}$$

请举例说明上述关系不是全序的，即存在π_1和π_2, $\pi_1 \not\geq \pi_2$且$\pi_2 \not\geq \pi_1$。

（3）证明对于任意MDP一定存在一个最优策略π_*，使得$\forall \pi, \pi_* \geq \pi$。

2 第二题

强化学习算法在模拟环境中的实践。

（1）请基于OpenAI Gym Taxi-v3环境实现Sarsa和Sarsa(λ)算法。（可参考助教提供的Q-learning算法）

（2）对算法的学习率、更新步长、Sarsa(λ)中λ等超参数进行调优，分析不同参数下的算法表现。

（3）对比OnPolicy算法Q-learning、OffPolicy算法Sarsa、Sarsa(λ)实验效果，并进行分析（包括最终奖励值，完成任务所需动作数，迭代稳定性等）。

本题需要提交完整代码并形成实验报告。