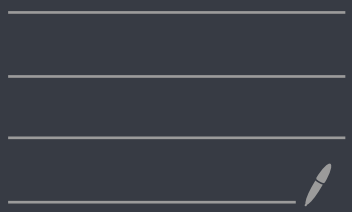
Problem set #2 (due 12/08)



P1、调试并运行助教提供的QL代码，然后：

(i) 比较不同参数设置下的收敛速度；

(ii) 比较状态价值迭代与Q-learning迭代之间的差异；

(iii) 试分析状态价值迭代与Q-learning迭代各有哪些优点。

P2. 设 (X,d) 是一个完备的度量空间且 $T: X \to X$ 满足

$$d(Tx, Ty) \leq r\, d(x, y), \quad \forall x, y \in X,$$

其中，$r \in [0,1)$，令 $x_n = T^n x_0$，$\forall x_0 \in X$.

(i) 试证存在唯一的 $x_* \in X$ 使得 $Tx_* = x_*$;

(ii) 试证 $d(x_{n+1}, x_*) \leq r\, d(x_n, x_*)$;

(iii) 试证 $d(x_{n+1}, x_*) \leq \frac{r}{1-r} d(x_{n+1}, x_n)$.

P3、令{s}为状态空间，{a}为动作空间，$P(s'|s,a)$为从s到s'的转移概率模型，$R(s)$为回报函数。已知状态价值迭代

$$V_{i+1}(s) = R(s) + \gamma \max_{a} \sum_{s'} P(s'|s,a) V_i(s'), \gamma \in [0,1)$$

试证该迭代时存在一个压缩系数为γ的不动点迭代法。