Reinforcement Learning (SS18) - Exercise 2

Daniel Hennes

11.05.2018 (due 16.05.2018)

1. Show that the Bellman optimality operator $\mathcal{T}$ is a γ–contraction.

$$(\mathcal{T}v)(s) = \max_a \sum_{s',r} p(s', r|s, a)\Big[r + \gamma v(s')\Big]$$

2. Assuming a general finite MDP $(\mathcal{S}, \mathcal{A}, \mathcal{R}, p, \gamma)$ where rewards are bounded: $r \in [r_{min}, r_{max}]$ for all $r \in \mathcal{R}$. Prove that:

(a)

$$\frac{r_{min}}{1-\gamma} \leq v(s) \leq \frac{r_{max}}{1-\gamma} \quad \text{for all } s \in \mathcal{S}$$

(b)

$$|v(s) - v(s')| \leq \frac{r_{max} - r_{min}}{1-\gamma} \quad \text{for all } s, s' \in \mathcal{S}$$

3. Solve the *cleaning robot* problem (see Exercise 2) with value iteration. Programming language is your choice. Bring your laptop to present your code and the solution in class. You may be asked to run your program for different values of γ and θ.