Übung 03

Sonntag, 9. Mai 2021 16:54

1 Proofs (5P)

a) Show that the Bellman optimality operator $\mathcal{T}$ is a γ-contraction. Be able to explain all the steps! (2P)

$$(\mathcal{T}v)(s) = \max_a \sum_{s',r} p(s', r|s, a)[r + \gamma v(s')] \tag{1}$$

$$|\mathcal{T}(V(s)) - \mathcal{T}(V'(s))| = \max_s \left| \max_a \sum_{s'r} p(s', r|s, a)[r + \gamma V(s')] - \max_a \sum_{s'r} p(s', r|s, a)[r + \gamma V'(s')] \right|$$

$$= \max_s \left| \max_a \sum_{s'r} p(s', r|s, a)[r + \gamma V(s) - (r + \gamma V'(s'))] \right|$$

$$= \gamma \max_s \left| \max_a \sum_{s'r} p(s', r|s, a)[V(s) - V'(s')] \right|$$

$$\leq \gamma \max_s \left| \max_a \sum_{s'r} p(s', r|s, a) \max_s (V(s) - V'(s)) \right|$$

- s' is included in s
- $\max_s$ → it has to be greater (s' not in s) or equal (s' is in s)
- $\sum_{s'r} p(s', r|s, a) = 1$
- s' is not argument of $V(s)$

$$= \gamma \max_s |\max_a (V(s) - V'(s))|$$

$$\Rightarrow \gamma \max_s |(V(s) - V'(s))| = \gamma \|V - V'\|$$

b) Asuming a general finite MDP (S, A, R, p, γ) where rewards are bounded: $r \in [r_{\min}, r_{\max}]$ for all $r \in R$. Prove the following equations. (3P)

$$\frac{r_{\min}}{1-\gamma} \leq v(s) \leq \frac{r_{\max}}{1-\gamma} \tag{2}$$

$$|v(s) - v(s')| \leq \frac{r_{\max} - r_{\min}}{1-\gamma} \tag{3}$$

a) Implement the value iteration algorithm (see lecture 3 slide 27) in the function *value_iteration*. Use the values for γ and θ that are given in the code. Initialize the value function $V(s)$ to 0 for all states. How many steps does it need to converge? What is the optimal value function? (3P)

Optimal Valuefunctio after 43 steps:

```
[0.01543432 0.01559069 0.02744009 0.01568004 0.02685371 0.
 0.05978021 0.         0.0584134  0.13378315 0.1967357  0.
 0.         0.2465377  0.54419553 0.         ]
```

```
←[41mS←[0mFFF
FHFH
FFFH
HFFG
```

Terminals:

State 5	
State 7	
State 11	
State 12	
State 15	

```
Computed policy:
[1. 3. 2. 3. 0. 0. 0. 0. 3. 1. 0. 0. 0. 2. 1. 0.]
```