Assignment 11

Machine Learning, Summer term 2014, Norman Hendrich

To be discussed in exercise groups on July 07-09

Exercise 11.1 (Random-Walk task, 3 points) In the lecture slides (based on the Sutton&Barto book, chapter 7), a short (5 states) random-walk was chosen for the demonstration of TD(0). However, a longer random-walk (19 states) was used for the demonstration of on-line and off-line TD(λ). Why?

Exercise 13.2 (Implement RL-learning with radial-basis functions, 9 points) Document and implement a RL learning task of your own choice that involves function-approximation with radial-basis functions. For example, another grid-world, a game, or a physics model like the mountain-car. Describe the chosen task and your design of the state- and action-spaces, and the performance of your learner.

For example, model the mountain car problem using a grid of $N = n \times n$ radial basis functions $\{\mu_i, \sigma_i\}$ for the 2-dimensional state space $s = \{x_t, \dot{x}_t\}$ (position and velocity of the car). The means μ_i are created by evenly separating each dimension

$$\mu_i = \left\{ x_{min} + j \cdot \frac{x_{max} - x_{min}}{n-1}, \quad \dot{x}_{min} + k \cdot \frac{\dot{x}_{max} - \dot{x}_{min}}{n-1} \right\}$$

For each action $a_t \in \{-1, 0, +1\}$, approximate $Q(a, s_t)$ as

$$Q(a, s_t) = \sum_{i=1}^{N} \theta_N \phi(||s_t - \mu_i||, \sigma_i)$$

Try to implement Q-learning with RBFs as above, and $n = 8$, $\sigma = 0.1$. The dynamics used in the mountain-car problem is

$$x_{t+1} = \text{bound}\big[x_t + \dot{x}_t\big]$$

$$\dot{x}_{t+1} = \text{bound}\big[\dot{x}_t + 0.001 \cdot a - 0.0025 \cdot \cos(3 \cdot x_t)$$

where the bound operator ensures that the car remains within the position limits of $-1.2 \leq x \leq 0.5$, and the velocity limits $|\dot{x}| < 0.07$. The initial parameters are $x = -0.5$ and $\dot{x} = 0$, and the goal position is $x = 0.5$. The reward is -1 on every time-step, and the task finishes when the car reaches the goal position $x = 0.5$.