Advanced AI Techniques

Prof. Dr. Burgard, Prof. Dr. Nebel, Dr. Kersting
M. Ragni, A. Rottmann
WS 2006/2007

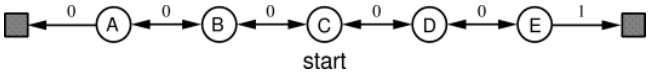
University of Freiburg
Department of Computer Science

Exercise Sheet 8

Due: Tuesday, 9. January 2007

Exercise 8.1

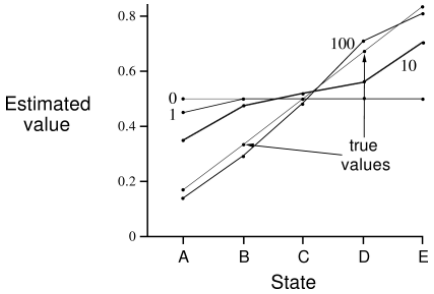
Consider the following random walk example for TD(0) learning with five states A, B, C, D, and E.



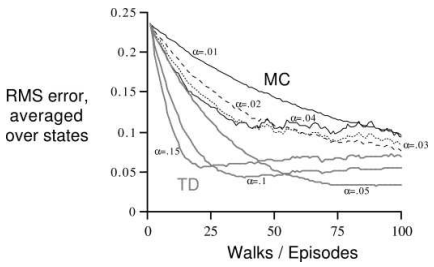
- actions: $\{left, right\}$
- equiprobable random policy
- $\gamma = 1.0$, $\alpha = 0.1$
- initialized value $V(s) = 0.5$, for all s.

(a) Figure (a) shows the values learned by TD(0) after $0, 1, 10$, and 100 episodes. From this figure, it appears that the first episode results in a change in $V(A)$ only. What does this tell you about what happened on the first episode? Why was only the estimate for this state changed? By exactly how much was it changed?

(b) Figure (b) shows the error of the TD method and it seems the error goes down and then up again, particularly for high α's. What could have caused this? Is this a general characteristic of the error for TD(0) or does it depend on how the approximate value function was initialized?



(a) Values learned by TD(0)



(b) Learning curve for TD(0)

Exercise 8.2

Implement for the five state random walk example from exercise 8.1 a:

(a) Every-Visit Monte Carlo Algorithm.

(b) TD(0) algorithm.

Compare your results with Figure (b) from exercise 8.1.

Exercise 8.3

What are the differences (prerequisites, advantages, disadvantages) of Monte Carlo and TD learning? Give an example for a situation when you would rather use Monte Carlo, and when you would prefer TD(0).