COMP 767 - Assignment 3

Mathieu Nassif

February 17, 2017

The MATLAB files prefixed with A3 contains the necessary code to compare $Q(\sigma)$ with *SARSA* and tree backup.

`A3_model` contains the MDP on which the algorithms are tested. It consists of a walk over the integers from 1 to 9, starting at 5, with the two possible actions *left* and *right*. Going left from 1 or right from 19 result in the end of the episode. Exiting from 19 gives a reward of 1.5, all other transitions give a reward of -1. To change the model in the file, simply implement the necessary logic. When called with no arguments, the function returns a quadruple consisting of (number of integers supporting the MDP, number of actions, starting point, ending point).

`A3_Q` contains the TD algorithms. It implements the algorithm presented in Sutton & Barto for $Q(\sigma)$, using a ϵ-greedy behavior policy and greedy target policy. One can use SARSA or tree backup using this algorithm by passing as the third argument a function handle that always return 1 or 0. The function returns the reward collected after having evaluated the optimal action value function and using a greedy policy (or -1000 if the greedy policy enters an infinite loop), the total number of steps taken, and the estimation of the optimal action value function.

`A3_sigma` contains a function to evaluate the σ_t values. It can be modified to be deterministic based on state and/or action, completely random, or a mix of both. The current implementation returns either 1 or 0, each with probability 0.5, independent of the state or action.

`A3_main` contains a short driver program. It runs the three algorithms multiple times on the model in `A3_model`, varying the number of episodes allowed.

By running the algorithms for 5, 10 and 25 episodes, 10 times each, with the parameters presented in Table 1, we can compare the proportion of results that are able to find their way out and the time taken to do so.

Table 1 Parameters Used in the Evaluation

Parameter	Value
γ	1
α	0.5
ϵ	0.2
n	3

The values for the performance of the algorithms are shown in Table 2 and 3.

Table 2 Proportion of Runs That Found the Way Out

Algorithm	5 episodes per run	10 episodes per run	25 episodes per run
$Q(\sigma)$	30%	30%	60%
SARSA	20%	10%	10%
Tree backup	20%	40%	40%

Table 3 Average Number of Iterations to Complete Each Episode

Algorithm	Nb. of Iterations
$Q(\sigma)$	52
SARSA	124
Tree backup	26