September 25, 2022

7 Exercises

7.1 Question

In Chapter 6 we noted that the Monte Carlo error can be written as the sum of TD errors (6.6) if the value estimates don't change from step to step. Show that the n-step error used in (7.2) can also be written as a sum TD errors (again if the value estimates don't change) generalizing the earlier result.

Answer

Value estimates are assumed not to change thus we can omit value estimate subscripts such that $V_t(S_t) = V_{t+1}(S_t)$.

n-step TD Error used in 7.2 is:

$G_{t:t+n} - V(S_t) = R_{t+1} + \gamma G_{t+1:t+n} - V(S_t)$
$G_{t:t+n} - V(S_t) = R_{t+1} + \gamma V(S_{t+1}) - V(S_t) + \gamma G_{t+1:t+n} - \gamma V(S_{t+1})$
$G_{t:t+n} - V(S_t) = \delta_t + \gamma(G_{t+1:t+n} - V(S_{t+1}))$
$G_{t:t+n} - V(S_t) = \delta_t + \gamma\delta_{t+1} + \gamma^2(G_{t+2:t+n} - V(S_{t+2}))$
$G_{t:t+n} - V(S_t) = \sum_{k=t}^{t+n-1} \gamma^{k-t}\delta_k$

7.2 Question

(programming) With an n-step method, the value estimates do change from step to step, so an algorithm that used the sum of TD errors (see previous exercise) in place of the error in (7.2) would actually be a slightly different algorithm. Would it be a better algorithm or a worse one? Devise and program a small experiment to answer this question empirically.

Answer

The chart above shows regular n-step TD with different n parameters. The chart below shows the same configuration with unchanged value functions.

Value function updates are applied only after an episode terminates.

Using sum of TD errors as in place of the error in 7.2 performs worse in all n and α values.