TD(λ)

christian.schroeder

April 2018

1

2 n-step return

(Sutton 2018 p. 290 eq. 12.1)

$$G_{t:t+n} = R_{t+1} + \ldots + \gamma^{n-1} R_{t+n} + \gamma^n V(s_{t+n}),\ 0 \leq t \leq T - n$$

with R_t being rewards.

3 λ-return

(Sutton 2018 p. 312 eq. 12.3)

$$G_t^\lambda = (1-\lambda) \sum_{n=1}^{T-t-1} \lambda^{n-1} G_{t:t+n} + \lambda^{T-t-1} G_t$$

4 n-step truncated λ-return

(Sutton 2018 eq. 12.9)

$$G_{t:h}^\lambda = (1-\lambda) \sum_{n=1}^{h-t-1} \lambda^{n-1} G_{t:t+n} + \lambda^{h-t-1} G_{t:h},\ 0 \leq t < h \leq T$$

Efficient implementation: (eq. 12.10)

$$G_{t:t+k}^\lambda = V(s_t) + \sum_{i=t}^{t+k-1} (\gamma\lambda)^{i-t} \delta_i'$$

where

$$\delta_i' = R_{t+1} + \gamma V(s_{t+1}) - V(s_t)$$

5 Efficient implementation if we need to calculate consecutive $G^\lambda_{t:h}, G^\lambda_{t+1:h}$

(new by author)

$$G^\lambda_{t:h} = V_t + \sum_{i=t}^{h-1} (\gamma\lambda)^{i-t}\delta'_i = V_t + \delta_t + \sum_{i=t+1}^{h-1} (\gamma\lambda)^{i-t}\delta'_i = V_t + \delta_t + \gamma\lambda \sum_{i=t+1}^{h-1} (\gamma\lambda)^{i-(t+1)}\delta'_i$$

at the same time,

$$G^\lambda_{t+1:h} = V_{t+1} + \sum_{i=t+1}^{h-1} (\gamma\lambda)^{i-(t+1)}\delta'_i$$

therefore:

$$G^\lambda_{t+1:h} - \frac{1}{\gamma\lambda} G^\lambda_{t:h} = V_{t+1} - \frac{1}{\gamma\lambda}(V_t + \delta_t)$$

so

$$G^\lambda_{t+1:h} = V_{t+1} - \frac{1}{\gamma\lambda}(V_t + \delta_t - G^\lambda_{t:h})$$

6 Note:

We calculate in t-forward direction (when t-backward would be arguably more efficient as $G^\lambda_{h-1:h}$ is much easier to calculate than $G^\lambda_{0:h}$) as when vectorizing, we want to iterate over sequences of different lengths - each of these does admit $G^\lambda_{0:h}$, but $G^\lambda_{h-1:h}$ could well be NaN if that sequence finished early.

Alternative method Alternatively, we can align all sequence ends at the right hand side and then go backwards:

$$G^\lambda_{t-1:h} = \gamma\lambda(G^\lambda_{t:h} - V_t) + (V_{t-1} + \delta_{t-1}) = \gamma\lambda(G^\lambda_{t:h} - V_t) + (R_t + \gamma V(s_t))$$

,

starting with

$$G^\lambda_{h-1:h} = G_{h-1:h} = R_h + \gamma V_{s_h}$$

After the backward pass, we re-align the sequences with beginnings at left-hand side.