CMPUT 609 Assignment 5

Shaurya Seth

March 5, 2023

Question 1

The λ-return can be written recursively as:

$$\begin{aligned}
G_t^\lambda &= (1-\lambda)[G_{t:t+1} + \lambda G_{t:t+2} + \lambda^2 G_{t:t+3} + \dots] \\
&= (1-\lambda)[(R_{t+1} + G_{t+1:t+1}) + (R_{t+1} + \lambda G_{t+1:t+2}) + (R_{t+1} + \lambda^2 G_{t+1:t+3}) + \dots] \\
&= (1-\lambda)[(R_{t+1} + \lambda R_{t+1} + \lambda^2 R_{t+1} + \dots) + \gamma(G_{t+1:t+1} + \lambda G_{t+1:t+2} + \lambda^2 G_{t+1:t+3} + \dots)] \\
&= (1-\lambda)\left[\frac{R_{t+1}}{(1-\lambda)} + \gamma G_{t+1:t+1}\right] + \gamma\lambda(1-\lambda)(G_{t+1:t+2} + \lambda G_{t+1:t+3} + \dots) \\
&= R_{t+1} + \gamma(1-\lambda)G_{t+1:t+1} + \gamma\lambda G_{t+1}^\lambda \\
&= R_{t+1} + \gamma(1-\lambda)\hat{v}(S_{t+1}, \mathbf{w}_t) + \gamma\lambda G_{t+1}^\lambda
\end{aligned}$$

Question 2

The half-life of the weighting decay can be given as:

$$\begin{aligned}
(1-\lambda)\lambda^{\tau_\lambda - 1} &= \frac{(1-\lambda)}{2} \\
\lambda^{\tau_\lambda - 1} &= \frac{1}{2} \\
\tau_\lambda - 1 &= \log_\lambda \frac{1}{2} \\
\tau_\lambda &= \log_\lambda \frac{1}{2} + 1
\end{aligned}$$

Question 3

The error term of the λ-return algorithm can be written as the sum of TD errors:

$$\begin{aligned}
G_t^\lambda - \hat{v}(S_t, \mathbf{w}) &= R_{t+1} + \gamma(1-\lambda)\hat{v}(S_{t+1}, \mathbf{w}) + \gamma\lambda G_{t+1}^\lambda - \hat{v}(S_t, \mathbf{w}) \\
&= R_{t+1} + \gamma\hat{v}(S_{t+1}, \mathbf{w}) - \gamma\lambda\hat{v}(S_{t+1}, \mathbf{w}) + \gamma\lambda G_{t+1}^\lambda - \hat{v}(S_t, \mathbf{w}) \\
&= \delta_t + \gamma\lambda(G_{t+1}^\lambda - \hat{v}(S_{t+1}, \mathbf{w})) \\
&= \delta_t + \gamma\lambda\delta_{t+1} + \gamma^2\lambda^2\delta_{t+2} + \dots + \gamma^{T-t-1}\lambda^{T-t-1}\delta_{T-1} + 0 \\
&= \sum_{k=t}^{T-1} \gamma^{k-t}\lambda^{k-t}\delta_k
\end{aligned}$$

Question 4

The sum of the weight updates computed during the episode can be written as:

$$\begin{aligned}
\sum_{t=0}^{T-1} \alpha\delta_t \mathbf{z}_t &= \sum_{t=0}^{T-1} \alpha\delta_t [\gamma\lambda \mathbf{z}_{t-1} + \nabla\hat{v}(S_t, \mathbf{w})] \\
&= \sum_{t=0}^{T-1} \alpha\delta_t [\gamma\lambda(\gamma\lambda \mathbf{z}_{t-2} + \nabla\hat{v}(S_{t-1}, \mathbf{w})) + \nabla\hat{v}(S_t, \mathbf{w})] \\
&= \sum_{t=0}^{T-1} \alpha\delta_t [\gamma^2\lambda^2 \mathbf{z}_{t-2} + \gamma\lambda\nabla\hat{v}(S_{t-1}, \mathbf{w}) + \nabla\hat{v}(S_t, \mathbf{w})] \\
&= \sum_{t=0}^{T-1} \alpha\delta_t [\gamma^t\lambda^t \hat{v}(S_0, \mathbf{w}) + \cdots + \gamma\lambda\nabla\hat{v}(S_{t-1}, \mathbf{w}) + \nabla\hat{v}(S_t, \mathbf{w})] \\
&= \sum_{t=0}^{T-1} \alpha\delta_t \sum_{k=0}^{t} \gamma^{t-k}\lambda^{t-k}\nabla\hat{v}(S_k, \mathbf{w}) \\
&= \sum_{t=0}^{T-1} \alpha \left[\sum_{k=t}^{T-1} \gamma^{k-t}\lambda^{k-t}\delta_k\right] \nabla\hat{v}(S_t, \mathbf{w}) \qquad \text{(using the summation rule)} \\
&= \sum_{t=0}^{T-1} \alpha[G_t^\lambda - \hat{v}(S_t, \mathbf{w})]\nabla\hat{v}(S_t, \mathbf{w})
\end{aligned}$$

Question 5

(12.10) can be derived as:

$$\begin{aligned}
G_{t:t+k}^\lambda &= (1-\lambda)\sum_{n=1}^{k-1} \lambda^{n-1} G_{t:t+n} + \lambda^{k-1} G_{t:t+k} \\
&= \sum_{n=1}^{k-1} \lambda^{n-1} G_{t:t+n} - \sum_{n=1}^{k-1} \lambda^{n} G_{t:t+n} + \lambda^{k-1} G_{t:t+k} \\
&= \sum_{n=1}^{k} \lambda^{n-1} G_{t:t+n} - \sum_{n=1}^{k-1} \lambda^{n} G_{t:t+n} \\
&= \sum_{n=0}^{k-1} \lambda^{n} G_{t:t+n+1} - \sum_{n=1}^{k-1} \lambda^{n} G_{t:t+n} \\
&= G_{t:t+1} + \sum_{n=1}^{k-1} \lambda^{n} G_{t:t+n+1} - \sum_{n=1}^{k-1} \lambda^{n} G_{t:t+n} \\
&= G_{t:t+1} + \sum_{n=1}^{k-1} \lambda^{n} [G_{t:t+n+1} - G_{t:t+n}] \\
&= R_{t+1} + \gamma\hat{v}(S_{t+1}, \mathbf{w}_t) + \sum_{n=1}^{k-1} \lambda^{n} G_{t+n:t+n+1} \\
&= R_{t+1} + \gamma\hat{v}(S_{t+1}, \mathbf{w}_t) + \sum_{n=t+1}^{t+k-1} \lambda^{n-t} G_{n:n+1} \\
&= R_{t+1} + \gamma\hat{v}(S_{t+1}, \mathbf{w}_t) + \sum_{n=t+1}^{t+k-1} (\gamma\lambda)^{n-t} [R_{n+1} + \gamma\hat{v}(S_{n+1}, \mathbf{w}_n) - \hat{v}(S_n, \mathbf{w}_{n-1})] \\
&= \hat{v}(S_t, \mathbf{w}_{t-1}) + \sum_{n=t}^{t+k-1} (\gamma\lambda)^{n-t} \delta_n'
\end{aligned}$$

Question 6

The following modification needs to be made in the $\mathcal{F}(S, A)$ loop:

$z_{\text{sum}} \leftarrow 0$
Loop for i in $\mathcal{F}(S, A)$:
    $z_{\text{sum}} \leftarrow z_{\text{sum}} + z_i$
Loop for i in $\mathcal{F}(S, A)$:
    $\delta \leftarrow \delta - w_i$
    $z_i \leftarrow z_i + 1 - \alpha\gamma\lambda z_{\text{sum}}$

Question 7

The equations can be given as:

$$\begin{aligned}
G^{\lambda s}_{t:h} &= R_{t+1} + \gamma_{t+1}((1 - \lambda_{t+1})\hat{v}(S_{t+1}, \mathbf{w}_t) + \lambda_{t+1}G^{\lambda s}_{t+1:h}) \\
G^{\lambda a}_{t:h} &= R_{t+1} + \gamma_{t+1}((1 - \lambda_{t+1})\hat{q}(S_{t+1}, A_{t+1}, \mathbf{w}_t) + \lambda_{t+1}G^{\lambda a}_{t+1:h}) \\
G^{\lambda a}_{t:h} &= R_{t+1} + \gamma_{t+1}((1 - \lambda_{t+1})\bar{V}_t(S_{t+1}) + \lambda_{t+1}G^{\lambda a}_{t+1:h})
\end{aligned}$$

For the basis step we have:

$$\begin{aligned}
G^{\lambda s}_{t:t} &= R_{t+1} + \gamma_{t+1}(1 - \lambda_{t+1})\hat{v}(S_{t+1}, \mathbf{w}_t) \\
G^{\lambda a}_{t:t} &= R_{t+1} + \gamma_{t+1}(1 - \lambda_{t+1})\hat{q}(S_{t+1}, A_{t+1}, \mathbf{w}_t) \\
G^{\lambda a}_{t:t} &= R_{t+1} + \gamma_{t+1}(1 - \lambda_{t+1})\bar{V}_t(S_{t+1})
\end{aligned}$$

Question 8

(12.24) becomes exact when the value function doesn't change:

$$\begin{aligned}
G^{\lambda s}_0 - V_0 &= \rho_0(R_1 + \gamma_1((1 - \lambda_1)V_1 + \lambda_1 G^{\lambda s}_1)) + (1 - \rho_0)V_0 - V_0 \\
&= \rho_0 R_1 + \rho_0\gamma_1(1 - \lambda_1)V_1 + \rho_0\gamma_1\lambda_1 G^{\lambda s}_1 - \rho_0 V_0 \\
&= \rho_0 R_1 + \rho_0\gamma_1 - \rho_0\gamma_1\lambda_1 V_1 + \rho_0\gamma_1\lambda_1 G^{\lambda s}_1 - \rho_0 V_0 \\
&= \rho_0([R_1 + \gamma_1 V_1 - V_0] - \gamma_1\lambda_1[V_1 - G^{\lambda s}_1]) \\
&= \rho_0(\delta_0 + \gamma_1\lambda_1[G^{\lambda s}_1 - V_1]) \qquad \text{(notice the recursion)} \\
&= \rho_0 \sum_{k=0}^{\infty} \delta^s_k \prod_{i=1}^{k} \gamma_i\lambda_i\rho_i
\end{aligned}$$

Question 9

For the truncated version we can sum till $h - 1$ instead of ∞:

$$G^{\lambda s}_{t:h} \approx \hat{v}(S_t, \mathbf{w}_t) + \rho_t \sum_{k=t}^{h-1} \delta^s_k \prod_{i=t+1}^{k} \gamma_i\lambda_i\rho_i$$

Question 10

The action-based TD error is given by:

$$\delta^a_0 = R_1 + \gamma_1\bar{V}(S_1) - Q_0$$

where

$$\bar{V}(s) = \sum_a \pi(a|s)\hat{q}(s, a, \mathbf{w})$$

Now (12.27) becomes exact when the value function doesn't change:

$$\begin{aligned}
G_0^{\lambda a} - Q_0 &= R_1 + \gamma_1(\bar{V}(S_1) + \lambda_1\rho_1[G_1^{\lambda a} - Q_1]) - Q_0 \\
&= R_1 + \gamma_1\bar{V}(S_1) + \gamma_1\lambda_1\rho_1[G_1^{\lambda a} - Q_1]) - Q_0 \\
&= \delta_0^a + \gamma_1\lambda_1\rho_1[G_1^{\lambda a} - Q_1]) \qquad \text{(notice the recursion)} \\
&= \sum_{k=0}^{\infty} \delta_k^a \prod_{i=1}^{k} \gamma_i\lambda_i\rho_i
\end{aligned}$$

Question 11

Again for the truncated version we can sum till $h-1$ instead of ∞:

$$G_{t:h}^{\lambda a} \approx \hat{q}(S_t, A_t, \mathbf{w}_t) + \sum_{k=t}^{h-1} \delta_k^a \prod_{i=t+1}^{k} \gamma_i\lambda_i\rho_i$$

Question 12

Starting with the update rule:

$$\begin{aligned}
\mathbf{w}_{t+1} &= \mathbf{w}_t + \alpha[G_t^{\lambda a} - \hat{q}(S_t, A_t, \mathbf{w}_t)]\nabla\hat{q}(S_t, A_t, \mathbf{w}_t) \\
&\approx \mathbf{w}_t + \alpha\left(\sum_{k=t}^{\infty} \delta_k^a \prod_{i=t+1}^{k} \gamma_i\lambda_i\rho_i\right)\nabla\hat{q}(S_t, A_t, \mathbf{w}_t) \qquad \text{(using 12.27)}
\end{aligned}$$

After summing over time we get:

$$\begin{aligned}
\sum_{t=0}^{\infty}(\mathbf{w}_{t+1} - \mathbf{w}_t) &\approx \sum_{t=0}^{\infty}\sum_{k=t}^{\infty} \alpha\delta_k^a\nabla\hat{q}(S_t, A_t, \mathbf{w}_t) \prod_{i=t+1}^{k} \gamma_i\lambda_i\rho_i \\
&\approx \sum_{k=0}^{\infty}\sum_{t=0}^{k} \alpha\delta_k^a\nabla\hat{q}(S_t, A_t, \mathbf{w}_t) \prod_{i=t+1}^{k} \gamma_i\lambda_i\rho_i \qquad \text{(using summation rule)} \\
&= \sum_{k=0}^{\infty} \alpha\delta_k^a \sum_{t=0}^{k} \nabla\hat{q}(S_t, A_t, \mathbf{w}_t) \prod_{i=t+1}^{k} \gamma_i\lambda_i\rho_i
\end{aligned}$$

We write the expression from the second sum as an eligibility trace:

$$\begin{aligned}
\mathbf{z}_k &= \sum_{t=0}^{k} \nabla\hat{q}(S_t, A_t, \mathbf{w}_t) \prod_{i=t+1}^{k} \gamma_i\lambda_i\rho_i \\
&= \sum_{t=0}^{k-1} \nabla\hat{q}(S_t, A_t, \mathbf{w}_t) \prod_{i=t+1}^{k} \gamma_i\lambda_i\rho_i + \hat{q}(S_k, A_k, \mathbf{w}_k) \\
&= \gamma_k\lambda_k\rho_k\left[\sum_{t=0}^{k-1} \nabla\hat{q}(S_t, A_t, \mathbf{w}_t) \prod_{i=t+1}^{k-1} \gamma_i\lambda_i\rho_i\right] + \hat{q}(S_k, A_k, \mathbf{w}_k) \\
&= \gamma_k\lambda_k\rho_k\mathbf{z}_{k-1} + \hat{q}(S_k, A_k, \mathbf{w}_k)
\end{aligned}$$

We get (12.29):

$$\mathbf{z}_t = \gamma_t\lambda_t\rho_t\mathbf{z}_{t-1} + \hat{q}(S_t, A_t, \mathbf{w}_t)$$

Question 13

For dutch traces we have:

$$\mathbf{z}_t = \rho_t(\gamma_t\lambda_t\mathbf{z}_{t-1} + (1 - \alpha\gamma_t\lambda_t\mathbf{z}_{t-1}^\top\mathbf{x}_t))\mathbf{x}_t \qquad \text{(state-value methods)}$$

$$\mathbf{z}_t = \gamma_t \lambda_t \rho_t \mathbf{z}_{t-1} + (1 - \alpha \gamma_t \lambda_t \rho_t \mathbf{z}_{t-1}^\top \mathbf{x}_t)\mathbf{x}_t \qquad \text{(action-value methods)}$$

For replacing traces we have:

$$z_{i,t} = \begin{cases} \rho_t & \text{if } x_{i,t} = 1 \\ \gamma_t \lambda_t \rho_t \mathbf{z}_{t-1} & \text{otherwise} \end{cases} \qquad \text{(state-value methods)}$$

$$z_{i,t} = \begin{cases} 1 & \text{if } x_{i,t} = 1 \\ \gamma_t \lambda_t \rho_t \mathbf{z}_{t-1} & \text{otherwise} \end{cases} \qquad \text{(action-value methods)}$$

Question 14

Assuming $\gamma = 1$ we would see the following sequence:

$$\begin{aligned} z_{\text{wrong},0} &= 0 & z_{\text{right},0} &= 0 \\ z_{\text{wrong},1} &= 1 & z_{\text{right},1} &= 0 \\ z_{\text{wrong},2} &= \lambda + 1 & z_{\text{right},2} &= 0 \\ z_{\text{wrong},2} &= \lambda(\lambda + 1) & z_{\text{right},2} &= 1 \end{aligned}$$

By solving $\lambda(\lambda + 1) > 1$ we find that the trace parameter would have to be greater than 0.61.