DFQNetwork and Learned Discount Function

We will follow the discussion in [1, p.8-10]. Our goal here is to relate the weights of **DQFNetwork** to a discount function $\Gamma(t)$. Notice the output of the agent is:

$$Q_\Gamma(s,a) = \sum \alpha_i Q_{\gamma_i}(s,a) \tag{1}$$

where α_i are the weights of **DQFNetwork**, and γ_i are chosen to be 0.99^i for $1 \leq i \leq 10$, which is approximately 0.9 to 0.99 with 0.01 increments. The above equation can be thought of as a Riemann sum approximation of the integral:

$$Q_\Gamma(s,a) = \sum (\gamma_{i+1} - \gamma_i)\omega(\gamma_i) Q_{\gamma_i}(s,a) \sim \int_0^1 \omega(\gamma) Q_\gamma(s,a) \mathrm{d}\gamma \tag{2}$$

where $\omega(\gamma)$ is some (interpolating) function which at $\gamma = \gamma_i$ is $\frac{\alpha_i}{\gamma_{i+1}-\gamma_i}$. Now following equations similar to that in [1, Eq.(5-9)],

$$Q_\pi^\Gamma(s,a) \sim \int_{\gamma=0}^1 \omega(\gamma) Q_\pi^\gamma(s,a) d\gamma = \tag{3}$$

$$\int_{\gamma=0}^1 \mathbb{E}_\pi \left[\sum_t R(s_t,a_t)\gamma^t \omega(\gamma) \middle| s,a \right] d\gamma = \tag{4}$$

$$\mathbb{E}_\pi \left[\sum_t \left(\int_{\gamma=0}^1 \gamma^t \omega(\gamma) d\gamma \right) R(s_t,a_t) \middle| s,a \right] \tag{5}$$

where the first equality comes from Bellman expectation definition of Q_π^γ and the second is just interchanging the expectation E_π operator and integral $\int_0^1 \mathrm{d}\gamma$. Therefore, Q_Γ, the output of **DFQNetwork**, is an approximation of a Q-function with a discount $\Gamma(t)$, given by

$$\Gamma(t) = \int_0^1 w(\gamma)\gamma^t d\gamma \tag{6}$$

Going back to the Riemann sum approximation, we have an approximation of $\Gamma(t)$ as follows:

$$\Gamma(t) \sim \sum (\gamma_{i+1} - \gamma_i)\omega(\gamma_i)\gamma_i^t = \sum \alpha_i \gamma_i^t. \tag{7}$$

Hence, to get $\Gamma(t)$, all we need to do is to implement the code shown in the report.

References

[1] Fedus, W., Gelada, C., Bengio, Y., Bellemare, M. G., and Larochelle, H. (2019). Hyperbolic discounting and learning over multiple horizons. *arXiv preprint arXiv:1902.06865.*