Chapter 11

Exercise 11.1:

Tabular update:

$$v_{t+n}(S_t) = v_{t+n-1}(S_t) + \alpha\rho_{t:t+n-1}[G_{t:t+n} - v_{t+n-1}(S_t)]$$

The corresponding semi-gradient off policy update will be:

$$w_{t+n} = w_{t+n-1} + \alpha\rho_{t:t+n-1}[G_{t:t+n} - v(S_t, w_{t+n-1})]\partial v(S_t, w_{t+n-1})$$

Where,

For episodic and discounted tasks:

$$G_{t:t+n} = R_{t+1} + \gamma R_{t+2} + \gamma^2 R_{t+3} + ... + \gamma^{n-1} R_{t+n} + \gamma^n v(S_{t+n}, w_{t+n-1})$$

For continuing and undiscounted tasks:

$$G_{t:t+n} = R_{t+1} - \overline{R}_t + R_{t+2} - \overline{R}_{t+1} + ... + R_{t+n} - \overline{R}_{t+n-1} + v(S_{t+n}, w_{t+n-1})$$

Exercise 11.2:

Eq 7.11:

$$Q_{t+n}(S_t, A_t) = Q_{t+n-1}(S_t, A_t) + \alpha\rho_{t+1:t+n-1}[G_{t:t+n} - Q_{t+n-1}(S_t, A_t)]$$

Semi gradient form:

$$w_{t+n} = w_{t+n-1} + \alpha\rho_{t+1:t+n-1}[G_{t:t+n} - Q(S_t, A_t, w_{t+n-1})]\partial Q(S_t, A_t, w_{t+n-1})$$

Eq 7.17:

For episodic and discounted tasks:

$$G_{t:h} = R_{t+1} + \gamma(\sigma_{t+1}\rho_{t+1} + (1 - \sigma_{t+1})\pi(A_{t+1}|S_{t+1}))(G_{t+1:h} - Q_{h-1}(S_h, A_h)) + \gamma\overline{v}_{h-1}(S_{t+1})$$

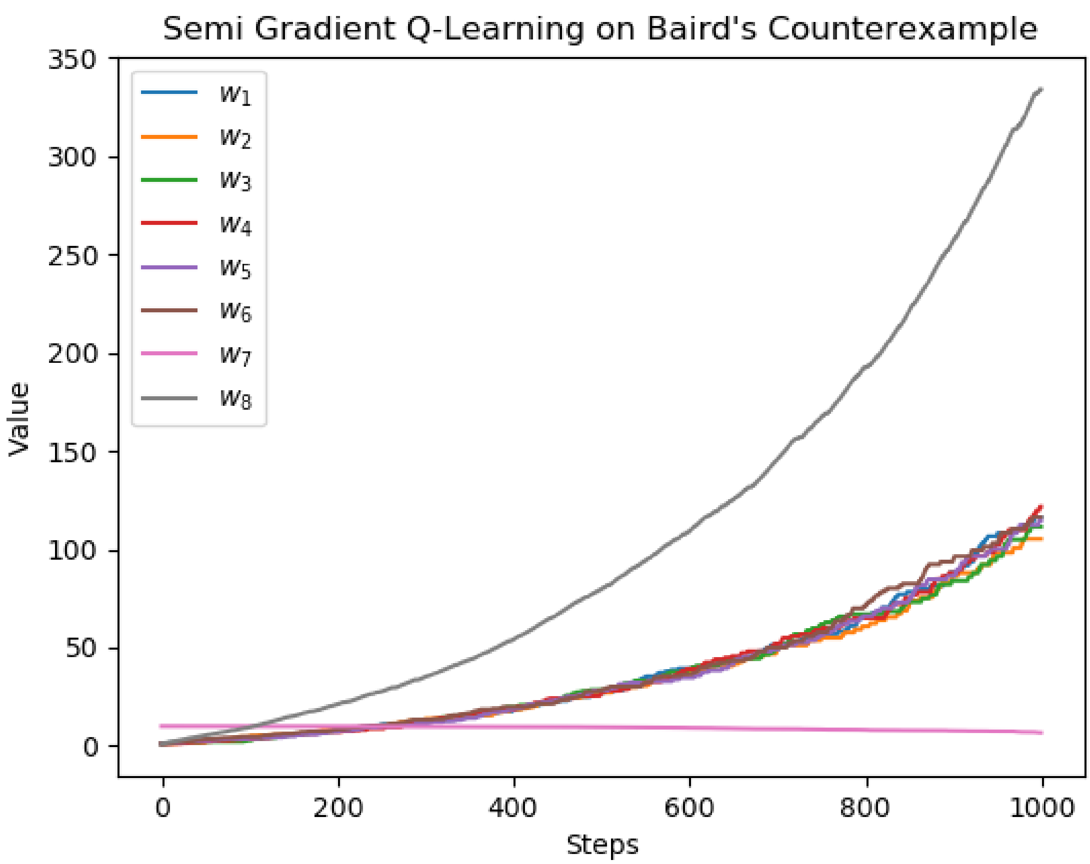
For continuing and undiscounted tasks:

$$G_{t:h} = R_{t+1} - \overline{R}_t + (\sigma_{t+1}\rho_{t+1} + (1 - \sigma_{t+1})\pi(A_{t+1}|S_{t+1}))(G_{t+1:h} - Q_{h-1}(S_h, A_h)) + \overline{v}_{h-1}(S_{t+1})$$

Exercise 11.3:

Code: Check [Github]

Results:



Exercise 11.4:

$$\overline{RE}(w) = E[(G_t - v(S_t, w))^2]$$

$$\overline{RE}(w) = E[(G_t - v(S_t, w) + v_\pi(S_t) - v_\pi(S_t))^2]$$

$$\overline{RE}(w) = E[(G_t - v_\pi(S_t) + v_\pi(S_t) - v(S_t, w))^2]$$

$$\overline{RE}(w) = E[(G_t - v_\pi(S_t))^2 + (v_\pi(S_t) - v(S_t, w))^2 + 2(G_t - v_\pi(S_t))((v_\pi(S_t) - v(S_t, w))]$$

$$\overline{RE}(w) = E[(G_t - v_\pi(S_t))^2] + E[(v_\pi(S_t) - v(S_t, w))^2] + E[2(G_t - v_\pi(S_t))((v_\pi(S_t) - v(S_t, w))]$$

$$\overline{RE}(w) = E[(G_t - v_\pi(S_t))^2] + \sum_s \mu(s)[(v_\pi(s) - v(s, w))^2\ ]$$
$$+ E[2(G_t - v_\pi(S_t))((v_\pi(S_t) - v(S_t, w))]$$

$$\overline{RE}(w) = E[(G_t - v_\pi(S_t))^2] + \overline{VE}(w)$$
$$+ E[2(G_t - v_\pi(S_t))((v_\pi(S_t) - v(S_t, w))]$$

$$\overline{RE}(w) = E[(G_t - v_\pi(S_t))^2] + \overline{VE}(w)$$
$$+ E[2(G_t v_\pi(S_t) - G_t v_\pi(S_t) v(S_t, w)) - v_\pi(S_t)^2 + v_\pi(S_t) v(S_t, w))]$$

$$\overline{RE}(w) = E[(G_t - v_\pi(S_t))^2] + \overline{VE}(w)$$
$$+ E[2G_t v_\pi(S_t)] - E[2G_t v(S_t, w))] - 2v_\pi(S_t)^2 + E[2v_\pi(S_t) v(S_t, w))]$$

$$\overline{RE}(w) = E[(G_t - v_\pi(S_t))^2] + \overline{VE}(w)$$
$$+ E[2G_t] v_\pi(S_t) - E[2G_t v(S_t, w))] - 2v_\pi(S_t)^2 + 2v_\pi(S_t) E[v(S_t, w))]$$

$$\overline{RE}(w) = E[(G_t - v_\pi(S_t))^2] + \overline{VE}(w)$$
$$+ 2v_\pi(S_t)^2 - 2v_\pi(S_t) E[v(S_t, w))] - 2v_\pi(S_t)^2 + 2v_\pi(S_t) E[v(S_t, w))]$$

$$\overline{RE}(w) = E[(G_t - v_\pi(S_t))^2] + \overline{VE}(w)$$