Problem 2

Linear function approximation with Q learning using target n/w

$$w = \begin{bmatrix} w_0 \\ w_1 \\ w_2 \end{bmatrix} \in R^3$$

S & $a \in \{-1, 0, 1\}$

feature vector of

$$\phi = \begin{bmatrix} 2.5 \\ a \\ 0.5 \end{bmatrix}$$

$$q(S, a; w) = w^T \phi$$

$$= [w_0 \; w_1 \; w_2] \begin{bmatrix} 2.5 \\ a \\ 0.5 \end{bmatrix}$$

$$= w_0 * 2.5 + w_1 * a + w_2 * 0.5$$

2.

$$Q^{Target} = Q(s', a'; w^-)$$

where, $w^- = \begin{bmatrix} w_0^- \\ w_1^- \\ w_2^- \end{bmatrix}$ weight vector for Q^{target} n/w

TD_target

$$y = r + \gamma \max_{a'} q(s', a'; w^-)$$

TD_error = TD_target − $Q(s, a; w)$

$$J(w) = MSE\left(y - q(s, a; w)\right)$$

$$J(w) = \frac{1}{2}\left((r + \gamma \max_{a'} q(s', a'; w^-)) - q(s, a; w)\right)$$

$$J(w) = \frac{1}{2}\left(q(s, a; w) - y\right)^2$$

minimize this loss function

3. $w = \begin{bmatrix} -2 \\ 1 \\ -1 \end{bmatrix}$ $\bar{w} = \begin{bmatrix} -1 \\ 2 \\ 1 \end{bmatrix}$

Sample (s, a, s', a')

$(1, 0, 2, 2)$

(a)

$q(s, a; w) = w^T \phi = \begin{bmatrix} -2 & 1 & -1 \end{bmatrix} \begin{bmatrix} 2.5 \\ a \\ 0.5 \end{bmatrix}$

$= -2 \cdot 2.5 + 1 \cdot a + -1 \cdot 0.5$

$= -2 \cdot 2.1 + 1 \cdot 0 + -1 \cdot 0.5$

$= -4 + 0 - 0.5 = -4.5$

(b) check which is action produces max q

$q(s', a'; \bar{w}) = \bar{w}^T \phi = \begin{bmatrix} -1 & 2 & 1 \end{bmatrix} \begin{bmatrix} 2.5 \\ a \\ 0.5 \end{bmatrix}$

next action

$\alpha = 0.2$ assume $\gamma = 0.9$

$a \in \{-1, 0, 1\}$ $\quad s' = 2$

$a = -1$

$$q(s', a'; w) = w^T \phi = \begin{bmatrix} -1 & 2 & 1 \end{bmatrix} \begin{bmatrix} 2.5' \\ -1 \\ 0.5 \end{bmatrix}$$

$$= -1 * 2 * 2 + 2 * -1 + 1 * 0.5$$

$$= -4 - 2 + 0.5 = \boxed{-5.5}$$

$a = 0$

$$q(s', a'; w) = w^T \phi = \begin{bmatrix} -1 & 2 & 1 \end{bmatrix} \begin{bmatrix} 2.5' \\ 0 \\ 0.5 \end{bmatrix}$$

$$= -1 * 2 * 2 + 0 + 0.5$$

$$= -4 + 0.5 = \boxed{-3.5}$$

$a = 1$

$$q(s', a'; w) = w^T \phi = \begin{bmatrix} -1 & 2 & 1 \end{bmatrix} \begin{bmatrix} 2.5' \\ 1 \\ 0.5 \end{bmatrix}$$

$$= -1 * 2 * 2 + 2 * 1 + 1 * 0.5$$

$$= -4 + 2 + 0.5 = \boxed{-1.5}$$

$$\max_{a'} q(s', a'; w) = -1.5$$

$a = 1$

© TD-error & gradient

$$y = r + \gamma \max_{a'} q(s', a'; w^-)$$

$$= 2 + 0.9 * -1.5$$

$$= \boxed{0.65}$$

$$\delta = 0.65 - (-4.5)$$

$$= 5.15$$

$$J(w) = \frac{1}{2}\left(y - q(s,a;w)\right)^2$$

$$\nabla_w J(w) = \left(q(s,a;w) - y\right) \nabla_w \left(q(s,a;w)\right)$$

$$= \delta * \nabla_w \left(w^T \phi(s,a)\right)$$

$$\nabla_w J(w) = \delta * \phi(s,a) = -5.15 \begin{bmatrix} 2.5 \\ a \\ 0.5 \end{bmatrix}$$

$$= -5.15 \begin{bmatrix} 2 * 1 \\ 0 \\ 0.5 \end{bmatrix} = \begin{bmatrix} -10.3 \\ 0 \\ -2.575 \end{bmatrix}$$

(d) update weight

$$w \leftarrow w - \alpha \nabla_w J(w)$$

$$W_{new} = \begin{bmatrix} -2 \\ 1 \\ -1 \end{bmatrix} - (0.2) \begin{bmatrix} -10.3 \\ 0 \\ -2.575 \end{bmatrix}$$

$$= \begin{bmatrix} -2 \\ 1 \\ -1 \end{bmatrix} - \begin{bmatrix} -2.06 \\ 0 \\ -0.515 \end{bmatrix}$$

$$= \begin{bmatrix} 0.06 \\ 1 \\ -0.485 \end{bmatrix}$$