NAME: Andrew Caian ....

CSCI S-89c Deep Reinforcement Learning

Part I of Final

Suppose each state $s \in \mathcal{S}$ of the Markov Decision Process can be represented by a vector of 3 real-valued features: $\mathbf{x}(s) = (x_1(s), x_2(s), x_3(s))^T$.

Given some policy π, suppose we model the state value function $v_\pi(s)$ with a *fully connected feedforward neural network* (please see the table below) which has three inputs ($x_1(s)$, $x_2(s)$, and $x_3(s)$), one hidden layer that consists of two neurons (u_1 and u_2) with Leaky Rectified Linear Unit (Leaky ReLU) activation functions, and one output ($\hat{v}(s, \mathbf{w})$) with the Leaky ReLU activation function.

The explicit representation of this network is

input layer	hidden layer	output layer
x_1 x_2 x_3	$u_1 = f(w_{01}^{(1)} + w_{11}^{(1)} x_1 + w_{21}^{(1)} x_2 + w_{31}^{(1)} x_3)$ $u_2 = f(w_{02}^{(1)} + w_{12}^{(1)} x_1 + w_{22}^{(1)} x_2 + w_{32}^{(1)} x_3)$	$\hat{v} = f(w_0^{(2)} + w_1^{(2)} u_1 + w_2^{(2)} u_2)$

Here, $f(x)$ denotes the following Leaky ReLU:

$$f(x) = \begin{cases} x, & \text{if } x \geq 0, \\ 0.1x, & \text{if } x < 0. \end{cases}$$

Assume that the weights,

$$\mathbf{w} = \big(\underbrace{w_{01}^{(1)}, w_{11}^{(1)}, w_{21}^{(1)}, w_{31}^{(1)}, w_{02}^{(1)}, w_{12}^{(1)}, w_{22}^{(1)}, w_{32}^{(1)}}_{\text{hidden layer}}, \underbrace{w_0^{(2)}, w_1^{(2)}, w_2^{(2)}}_{\text{output layer}} \big)^T,$$

are currently estimated as follows:

hidden layer	output layer
$w_{01}^{(1)} = -0.8, w_{11}^{(1)} = 0.2, w_{21}^{(1)} = 0.3, w_{31}^{(1)} = 0.9$ $w_{02}^{(1)} = 0.3, w_{12}^{(1)} = -0.5, w_{22}^{(1)} = -0.2, w_{32}^{(1)} = -0.4$	$w_0^{(2)} = 0.1, w_1^{(2)} = -0.3, w_2^{(2)} = 1.4$

Assume the agent minimizes the mean squared error loss function,

$$L \doteq \frac{1}{2} \left(\hat{v}(S_t, \mathbf{w}) - v_\pi(S_t) \right)^2,$$

using Stochastic Gradient Descent (SGD), i.e. the Neural Network is trained in mini-batches of size 1.

If for current state S_t, the features are $x_1(S_t) = 1.2$, $x_2(S_t) = 0.4$, and $x_3(S_t) = 0.3$; and the agent "observes" $v_\pi(S_t)$ (this, of course, means the agent uses MC return, 1-step TD return, etc. as a "measurement" of $v_\pi(S_t)$) to be 3.2, please find the next SGD update of the weights using $\alpha = 0.1$:

$$\mathbf{w} - \alpha \nabla L,$$

where $\nabla L \doteq \Big(\underbrace{\frac{\partial L}{\partial w_{01}^{(1)}}, \frac{\partial L}{\partial w_{11}^{(1)}}, \frac{\partial L}{\partial w_{21}^{(1)}}, \frac{\partial L}{\partial w_{31}^{(1)}}, \frac{\partial L}{\partial w_{02}^{(1)}}, \frac{\partial L}{\partial w_{12}^{(1)}}, \frac{\partial L}{\partial w_{22}^{(1)}}, \frac{\partial L}{\partial w_{32}^{(1)}}}_{\text{hidden layer}}, \underbrace{\frac{\partial L}{\partial w_{0}^{(2)}}, \frac{\partial L}{\partial w_{1}^{(2)}}, \frac{\partial L}{\partial w_{2}^{(2)}}}_{\text{output layer}} \Big)^T.$

Please notice that the "measurement" of the state-value $v_\pi(S_t)$ here is considered to be independent of $\mathbf{w}$ (please see, for example, the Semi-gradient 1-step Temporal-Difference (TD) prediction).

SOLUTION:

Andrew Caide
CSCI S-89c Final

$$L \doteq \frac{1}{2}\left(\hat{v}(s,t) - v_\pi(s_t)\right)^2$$

$x_1 = 1.2$

$x_2 = 0.4$

$x_3 = 0.3$

$v_\pi(s_t) = 3.2$

$$\hat{v} = f(w_0^{(2)} + w_1^{(2)} u_1 + w_2^{(2)} u_2)$$

$u_1 = (-0.8 + 0.2(1.2) + 0.3(0.4) + 0.9(0.3))$

$= -0.17 \Rightarrow \boxed{-0.017} = u_1$

$u_2 = (.3 + -0.5(1.2) + -0.2(0.4) + -0.4(0.3))$

$= -0.5 \Rightarrow \boxed{-0.05}$

$\hat{v} = 0.1 - 0.017(-0.3) - 0.05(1.4)$

$= \boxed{0.0351 = \hat{v}}$

a) $\varepsilon^{(2)} \doteq \partial L / \partial \hat{v} = \hat{v} - v_\pi(s_t) = 0.0351 - 3.2 = \boxed{-3.165}$

error of output layer

b) $\varepsilon_h^{(1)} = \frac{\partial L}{\partial u_h} = \frac{\partial L}{\partial \hat{v}} \cdot \partial \hat{v} / \partial u_h = \varepsilon^{(2)} \frac{\partial \hat{v}}{\partial u_h} = \varepsilon^{(2)} f'(z^{(2)}) w_h^{(2)}$

error of hidden layer

$h = 1, 2,$

$$\boxed{\begin{aligned} \varepsilon^{(2)} w_1^{(2)} &= -3.165(-0.3) = 0.9495 \\ \varepsilon^{(2)} w_2^{(2)} &= -3.165(1.4) = -4.431 \end{aligned}}$$

c) $\frac{\partial L}{\partial w_h^{(2)}} = \frac{\partial L}{\partial \hat{v}} \frac{\partial \hat{v}}{\partial w_h^{(2)}} = \varepsilon^{(2)} \frac{\partial}{\partial w_h^{(2)}} \left[f(w_0^{(2)} + w_1^{(2)} u_1 + w_2^{(2)} u_2) \right] = \varepsilon^{(2)} f'(z^{(2)}) u_h$

$$\boxed{\begin{aligned} \frac{\partial L}{\partial w_0^{(2)}} &= -3.165 \\ \frac{\partial L}{\partial w_1^{(2)}} &= -3.165 \cdot 1 \cdot (-0.017) = 0.054 \\ \frac{\partial L}{\partial w_2^{(2)}} &= -3.165 \cdot 1 \cdot (0.0351) = -0.1111 \end{aligned}}$$

d) $$\frac{\partial L}{\partial w_{jh}^{(1)}} = \frac{\partial L}{\partial u_h} \frac{\partial u_h}{\partial w_{jh}^{(1)}} = \varepsilon^{(1)} \frac{\partial}{\partial w_{jh}^{(1)}} \left[f\left(w_{0h}^{(1)} + w_{1h}^{(1)} x_1 + w_{2h}^{(1)} x_2 + w_{3h}^{(1)} x_3\right) \right]$$

$$= \varepsilon^{(1)} f'(z_h^{(1)}) x_j \qquad \text{for } j = 0,1,2,3, \quad h = 1,2,\cancel{3}$$

$$\frac{\partial L}{\partial w_{01}^{(1)}} = \varepsilon_1^{(1)} f'(z_1^{(1)}) x_0 = \varepsilon_1^{(1)} x_0 = 0.9445$$

$$\frac{\partial L}{\partial w_{11}^{(1)}} = \varepsilon_1^{(1)} x_1 = 0.9445(1.2) = \cancel{0.3798}\ 1.1394$$

$$\frac{\partial L}{\partial w_{21}^{(1)}} = \cancel{0.9445}\ 0.9445(0.4) = 0.3798$$

$$\cancel{\frac{\partial L}{\partial w}}\ \frac{\partial L}{\partial w_{31}^{(1)}} = 0.9495(0.3) = 0.2849$$

$$\cancel{\frac{\partial L}{\partial w_{41}^{(1)}} = 0.9495}$$

$$\frac{\partial L}{\partial w_{02}^{(1)}} = \varepsilon_2^{(1)} x_0 = -4.431$$

$$\frac{\partial L}{\partial w_{12}^{(1)}} = \varepsilon_2^{(1)} x_1 = -4.431(1.2) = -5.317$$

$$\frac{\partial L}{\partial w_{22}^{(1)}} = \varepsilon_2^{(1)} x_2 = -4.431(0.4) = -1.7724$$

$$\frac{\partial L}{\partial w_{32}^{(1)}} = \varepsilon_2^{(1)} x_3 = -4.431(0.3) = -1.3293$$

note

$$\alpha = 0.1$$
$$\nabla L = \left(\frac{\partial L}{\partial w_{HL}}, \frac{\partial L}{\partial w_{OL}}\right) \&$$

$$\frac{\partial L}{\partial w_{HL}} = \left(\frac{\partial L}{\partial w_{jh}}\right) \text{etc}$$

$$\frac{\partial L}{\partial w_{OL}} = \left(\frac{\partial L}{\partial w_0}, \frac{\partial L}{\partial w_1}, \frac{\partial L}{\partial w_2}\right)$$

$$w - \alpha \nabla L$$

$$= (-0.8, 0.2, 0.3, 0.9, 0.3, 0.5, -0.2, -0.4, 0.1, -0.3, 1.4) -$$

$$(0.095, 0.14, 0.038, 0.0285, -.4431, -0.5317, -0.1772, -0.133, -0.3165, 0.0064, -0.01111)$$

$$\boxed{= (-0.895, 0.06, 0.262, 0.8715, 0.7431, 0.0317, -0.0228, -0.267, 0.417, 0.295, 1.4111)}$$