Reinforcement Learning—HW 10

April 20, 2020

1. If $f(x) = 3x^2$, what is $\frac{df}{dx}$?

2. if $g(x, y, z) = 3x^2 + 2xy + y^2$, what are:

   (a) $\frac{\partial g}{\partial x}$?

   (b) $\frac{\partial g}{\partial y}$?

   (c) $\frac{\partial g}{\partial z}$?

3. Let $h(\boldsymbol{w}, \boldsymbol{x}) = 3\boldsymbol{w}^T\boldsymbol{x}$ (where $\boldsymbol{x}$ and $\boldsymbol{w}$ are column vectors of length d). What are:

   (a) $\frac{\partial h}{\partial w_1}$?

   (b) $\frac{\partial h}{\partial w_d}$?

   (c) ∇h?

4. You are trying to optimize a state-value approximation function of the form:
   $\hat{v}(s, \boldsymbol{w}) = w_2 s^2 + w_1 s + w_0$.

   The state loss function is:
   $\overline{VE}_s(\boldsymbol{w}, s) = [v_\pi(s) - \hat{v}(s, \boldsymbol{w})]^2$.

   The overall loss function is: $\overline{VE}(\boldsymbol{w}) = \sum_{s \in S} \frac{1}{|S|} \overline{VE}_s(\boldsymbol{w}, s)$.

   The learning rate, α, is .01. The current value of the column vector $\boldsymbol{w}$ is $[1, 3, 3]$.

   You are given a state $t = 2$ for which $v_\pi(t) = 7$.

   (a) What is the current value of the state loss function for t?

   (b) What is the new value of $\boldsymbol{w}$ after one step of stochastic gradient descent? Show your work.

(c) What is the new value of the state loss function for t? Did the loss go down?

(d) Could the overall loss have gone up? Why or why not?