Reference Solutions to the Quiz 7

Xiangyu Yang

May 21, 2020

1 Lecture 23

1). Please derive the updating rule, if we use ReLU as the activation function.

Sol: Before proceeding, we introduce the indicator function $\mathbb{I}(\cdot)$, meaning if condition $\cdot$ is met, then return 1; otherwise, return 0.

In the backward pass, we consider changes in any w_i, $i = 1, \ldots, n$ affecting the total error E. This is achieved by simply applying the chain rule, i.e.,

$$\begin{aligned} \frac{\partial E}{\partial w_i} &= \sum_d \frac{\partial E}{\partial o_d} \frac{\partial o_d}{\partial \mathrm{net}_d} \frac{\partial \mathrm{net}_d}{\partial w_i} \\ &= \sum_d (o_d - t_d) \mathbb{I}(\mathrm{net}_d \geq 0) x_{d,i}, \end{aligned} \tag{1}$$

where we use the fact that the derivative of ReLU function is the defined indicator function above. We hence update the weights as follows

$$\begin{aligned} w_i &= w_i - \eta \frac{\partial E}{\partial w_i} \\ &= w_i - \eta \sum_d (o_d - t_d) \mathbb{I}(\mathrm{net}_d \geq 0) x_{d,i}. \end{aligned} \tag{2}$$

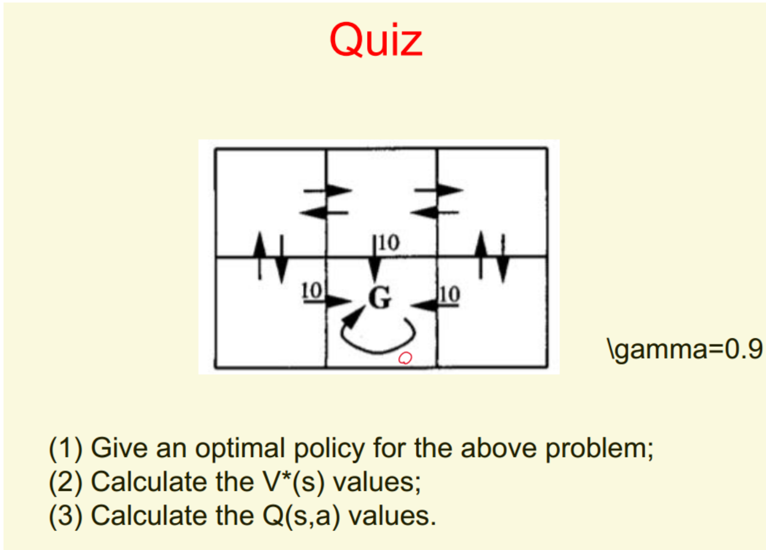
2). Compare the difference between the error gradients of the sigmoid function and the ReLU function.

Sol: The error gradients of the sigmoid function reads

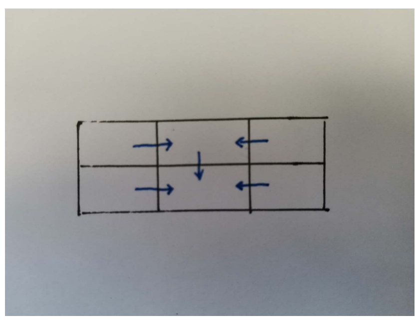
$$\begin{aligned} \frac{\partial E}{\partial w_i} &= \sum_d \frac{\partial E}{\partial o_d} \frac{\partial o_d}{\partial \mathrm{net}_d} \frac{\partial \mathrm{net}_d}{\partial w_i} \\ &= \sum_d (o_d - t_d) o_d (1 - o_d) x_{d,i}. \end{aligned} \tag{3}$$

We first note that the backward updating is a gradient-based learning method. From (3), we observe that the derivative of the sigmoid function is always smaller than 1 (i.e., consider $o_d(1 - o_d)$). Indeed, it is at most 0.25. This would cause significant side effects if you have many layers as the product of many smaller than 1 values goes to zero very quickly. However, RELU activation fixes the vanishing gradients problem because it only saturates in one direction.

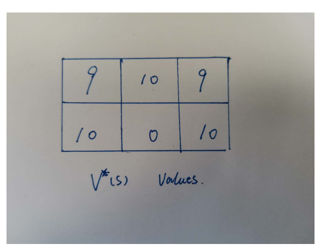
2 Lecture 24



Sol:

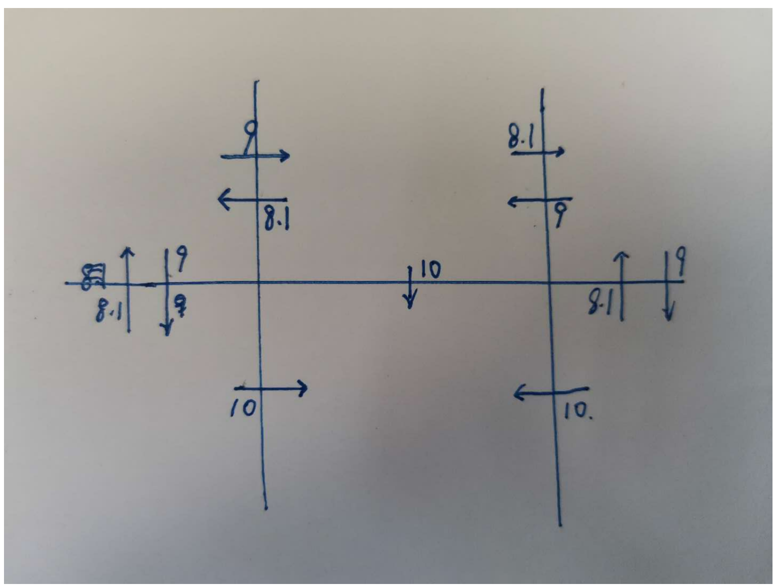


(1)



(2)

(3)


9
8.1
8.1
9
9
8.1
9
10
8.1
9
10
10.