Quiz\_Lecture23

Name：刘承奇 Student ID：2018533249

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





where  is a step function .

So, the updating rule is .

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

For the sigmoid function , when , its error gradients becomes very small (about 0), so it may have the problem of gradient disappearance. For the ReLU function, , when , it is a linear function. So, it will more unlikely have the problem of gradient disappearance. It can also screen better for the data.