VDL1

Erik Schwede
Abdalla Arafa
Ehsan Attar
Sai Leela Poduru
Prateek Rathod
Shruti Shrivastava

November 2023

1

1.1

1.2

1.2.1

Due to the chain rule that is used in the back propargation. Every derivitive is multiplied by each other. If there are some realy smal values, the overall derivitive is also going to be close to 0. Wich will limit the learning effect of the neural network.

1.2.2

$\sigma(x) = \frac{1}{1+e^{-x}}$

$\sigma'(x) = \frac{e^x}{(e^x+1)^2}$

$\sigma''(x) = \frac{(e^x-1)e^x}{(e^x+1)^3}$

$\sigma''(x) = 0$

$0 = e^x$

$\lim_{x \to -\infty} e^x = 0$

$\lim_{x \to -\infty} \sigma'(x) = 0$

$0 = e^x - 1$

$1 = e^x$

$x = 0$

$\sigma'(0) = \frac{e^0}{(e^0+1)^2} = 0.25$

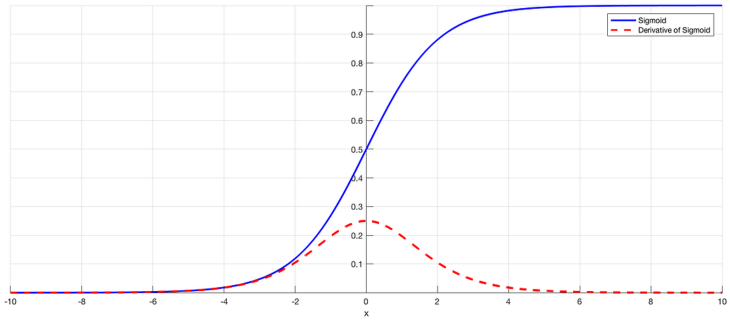

Figure 1: $\sigma(x)$ and $\sigma'(x)$ functions

According to these calculations the upperbound of the derivitive of the sigmoid function is 0.25 and the lowerbound is close to 0 but never reaches 0 due to the caracteristics of the e-function.

1.2.3

$a_1 = \sigma(w_1 x)$
$a_2 = \sigma(w_2 a_1)$
$a_3 = \sigma(w_3 a_2)$
$a_4 = \sigma(w_4 a_3)$
$y = \sigma(w_5 a_4)$

$$\frac{\partial y}{\partial x} = \frac{\partial y}{\partial a_4} \cdot \frac{\partial a_4}{\partial a_3} \cdot \frac{\partial a_3}{\partial a_2} \cdot \frac{\partial a_2}{\partial a_1} \cdot \frac{\partial a_1}{\partial x} = w_5\sigma'(a_4) \cdot w_4\sigma'(a_3) \cdot w_3\sigma'(a_2) \cdot w_2\sigma'(a_1) \cdot w_1\sigma'(x)$$

$$\frac{\partial y}{\partial x} = w_5 w_4 w_3 w_2 w_1 \cdot \sigma'(x)\sigma'(a_1)\sigma'(a_2)\sigma'(a_3)\sigma'(a_4)$$

1.2.4

In the context of the sigmoid function, its derivative σ' has an upper bound of 0.25. When this derivative is used in the chain rule during back propagation, it is multiplied with the weights for each layer. If the weights are less than 1, the gradients will diminish as they are back propagated through the layers. This means that the farther back you go in the network, the smaller the gradients become, potentially approaching zero. As a result, the weights in the early

layers receive very small updates, and these layers may not learn effectively during training.

1.2.5

Exploding gradients refer to a situation in deep learning where the gradients during back propagation become extremely large. This can lead to numerical instability and cause the weights of the neural network to update excessively, potentially resulting in the model's failure to converge during training. Exploding gradient problem in sigmoid activation can happen when the initial weights of the network are very large along with large number of layers. This would happen because sigmoid transform an input space into a space that is between [0,1]. So for large input values the gradients could collect during an update, resulting in very big gradients, which eventually results in huge modifications to network weights, resulting in an unstable network and causing exploding gradients problem.