1-1, 1-2

1 input layer with 2 nodes.

2 hidden layers with 3 nodes and 2 nodes respectively. total 5 nodes.

1 output layer with 3 nodes.

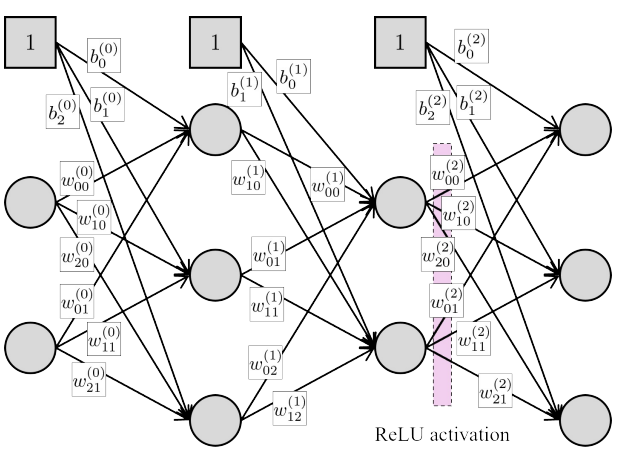
1-3 ReLU activation is used:

$$\text{ReLU}(x) = \begin{cases} x & x \geq 0 \\ 0 & x < 0 \end{cases}$$

1-4 18 weights, 8 biases

26 parameters in total.

1-5



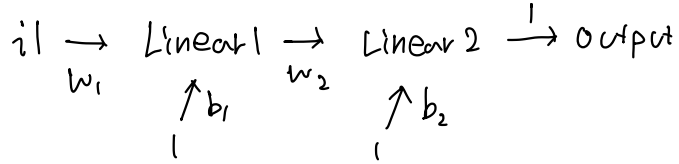
2-1

$$A * B = \begin{bmatrix} 1 & 0 & 0 & 0 & -1 \\ 2 & 1 & 0 & -1 & -2 \\ 2 & 2 & 0 & -2 & -2 \\ 1 & 3 & 0 & -3 & -1 \\ 0 & 1 & 0 & -1 & 0 \end{bmatrix}$$

2-2

$$A * C = \begin{bmatrix} -2 & -3 & -3 & -3 & -2 \\ -1 & -2 & -3 & -2 & -1 \\ 1 & 2 & 2 & 2 & 1 \\ 1 & 1 & 2 & 1 & 1 \\ 0 & 1 & 1 & 1 & 0 \end{bmatrix}$$

3



$$\text{output} = 1 \cdot \text{Linear2}$$
$$\text{Linear2} = w_2 \cdot \text{Linear1} + b_2 \cdot 1$$
$$\text{Linear1} = w_1 \cdot il + b_1 \cdot 1$$

$w_1 = 0.1$ $\quad b_1 = -0.2$ $\quad$ input $= 1$

$w_2 = -0.4$ $\quad b_2 = 0.5$ $\quad$ target $= 0.1$

output : $\quad 1 \cdot (w_2 \cdot (w_1 \cdot 1 + b_1 \cdot 1) + b_2 \cdot 1) = 0.54$

3-1

$$\text{loss} = \frac{1}{2}(\text{output}(1) - 0.1)^2 = \frac{1}{2}(0.54 - 0.1)^2$$
$$= (0.44)^2 = (0.01 \times 4 \times 11)^2$$
$$= 0.1936$$

3-2

$$\frac{\partial loss}{\partial output} = output - 0.1 = 0.44, \quad \frac{\partial output}{\partial linear2} = 1$$

$$\frac{\partial linear2}{\partial linear1} = w_2 = -0.4$$

$$\frac{\partial linear2}{\partial w_2} = Linear1 = -0.1, \quad \frac{\partial linear2}{\partial b_2} = 1$$

$$\frac{\partial linear1}{\partial w_1} = i1 = 1, \quad \frac{\partial linear1}{\partial b_1} = 1$$

cached values in the backpropagation algorithm:

● : 0.44 ● : 0.44 ● : -0.1176

$$\frac{\partial loss}{\partial w_1} = \frac{\partial loss}{\partial output}\frac{\partial output}{\partial linear2}\frac{\partial linear2}{\partial linear1}\frac{\partial linear1}{\partial w_1} = -0.1176$$

$$\frac{\partial loss}{\partial b_1} = \frac{\partial loss}{\partial output}\frac{\partial output}{\partial linear2}\frac{\partial linear2}{\partial linear1}\frac{\partial linear1}{\partial b_1} = -0.1176$$

$$\frac{\partial loss}{\partial w_2} = \frac{\partial loss}{\partial output}\frac{\partial output}{\partial linear2}\frac{\partial linear2}{\partial w_2} = -0.044$$

$$\frac{\partial loss}{\partial b_2} = \frac{\partial loss}{\partial output}\frac{\partial output}{\partial linear2}\frac{\partial linear2}{\partial b_2} = 0.44$$

3-3 $(w_1', b_1', w_2', b_2') = (w_1, b_1, w_2, b_2) - 0.1 \cdot \left(\frac{\partial loss}{\partial w_1}, \frac{\partial loss}{\partial b_1}, \frac{\partial loss}{\partial w_2}, \frac{\partial loss}{\partial b_2}\right)$

$= (0.1176, -0.1824, -0.3956, 0.456)$

3-4 $linear1 = -0.0648$ $linear2 = output = 0.48163488$

$loss = 0.007282259081630721$