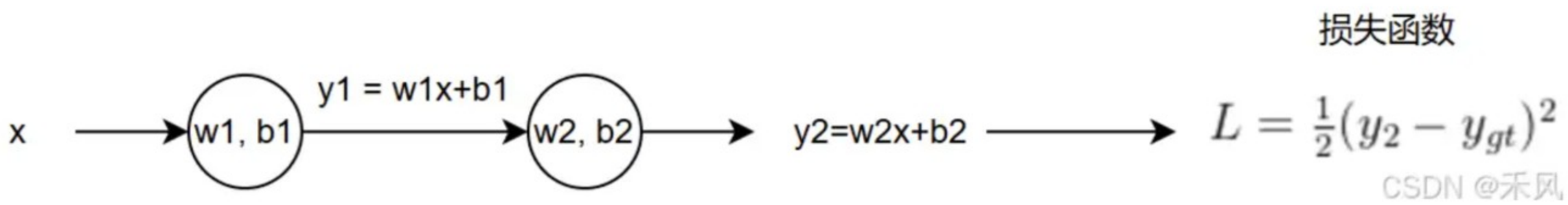
神经网络的训练过程中，正向传播（Forward Propagation）和反向传播（Backward Propagation）是两个核心步骤，分别涉及到信息的传递和误差的调整。

正向传播 (Forward Propagation)

正向传播是神经网络计算输出的过程。在每一层神经元中，输入信号经过加权和（权重w和偏置b）后通过激活函数（如ReLU、Sigmoid等）产生输出，并传递到下一层。输入层接收输入数据之后，隐藏层✦根据输入信号和权重进行计算。每个神经元接收到输入后，先通过加权求和，并加上偏置，然后通过激活函数✦得到输出，最后输出层生成最终的网络输出，通常经过某种形式的激活函数（如softmax或sigmoid✦），得到预测值。公式以及图示如下：

$$\mathbf{z}_l = W_l \mathbf{a}_{l-1} + b_l \quad (加权和)$$

$$\mathbf{a}_l = f(\mathbf{z}_l) \quad (激活函数)$$



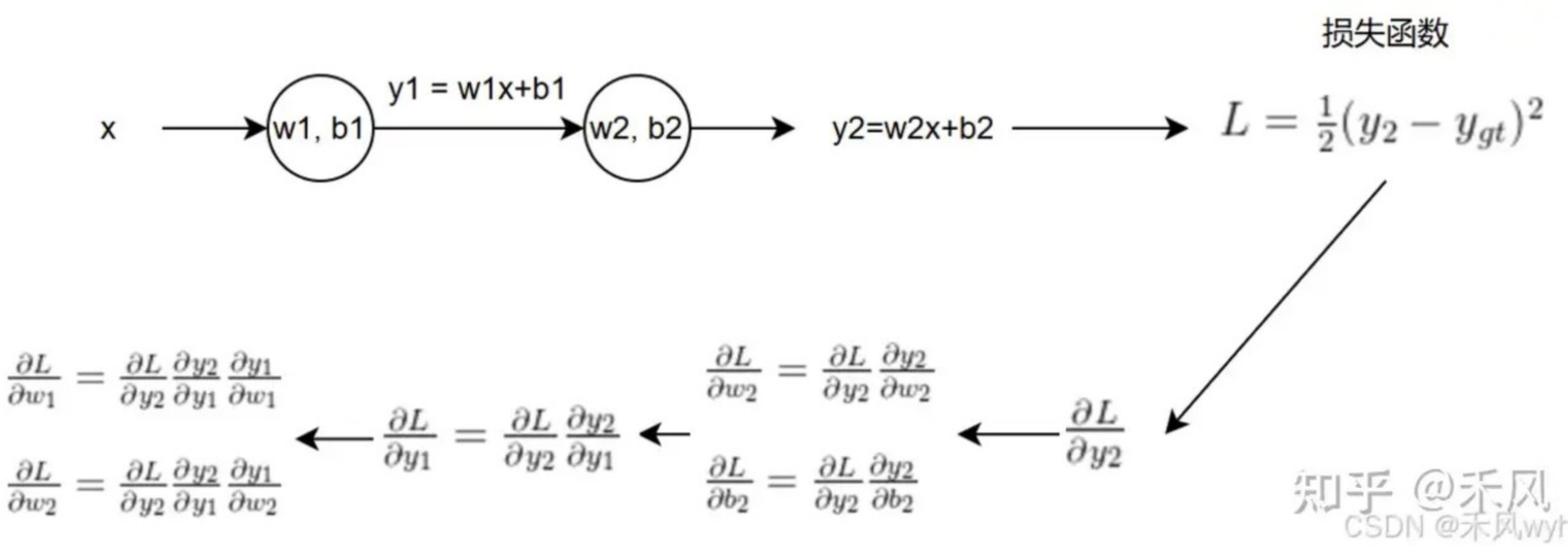
反向传播 (Backward Propagation)

反向传播是神经网络训练✦中用于优化权重的过程，它根据正向传播计算的输出与实际标签之间的误差（通常是损失函数的值）来调整每一层的权重。首先计算损失，通过损失函数（如均方误差、交叉熵损失✦等）计算预测值与真实标签之间的差距；然后计算梯度，反向传播算法✦通过链式法则计算每一层参数的梯度，即损失函数相对于每个参数的偏导数✦，最后进行更新权重，使用梯度下降（或其他优化算法，如Adam）根据计算出的梯度更新权重和偏置。更新公式为：

$$W = W - \eta \frac{\partial L}{\partial W} \quad \text{(梯度下降)}$$

反向传播的关键在于通过逐层计算梯度，从输出层到输入层依次反向传播误差，调整每一层的权重，使得损失最小化。

正向传播是神经网络从输入到输出的过程，主要用于计算输出结果。反向传播是根据输出和目标值之间的差距调整网络参数的过程，主要用于优化网络的权重和偏置。这两个过程交替进行，直到网络的损失函数最小化，达到训练目标。



卷积层✦中的反向传播

之前，我们讨论过在线性层✦中反向传播是如何进行操作的，现在我们讨论一下在卷积层中如何进行反向传播的计算。卷积层的操作本质上就是卷积核对图片进行卷积（对应相乘之后再相加），下图展示了一个3x3的卷积核，对5x5的图片进行卷积操作得到2x2的特征图的过程，其中卷积过程中的stride步长为2。

a1	a2	a3	a4	a5
a6	a7	a8	a9	a10
a11	a12	a13	a14	a15
a16	a17	a18	a19	a20
a21	a22	a23	a24	a25

*

w1	w2	w3
w4	w5	w6
w7	w8	w9

=

z1	z2
z3	z4

特征图上的像素点z1、z2、z3、z4的计算公式如下：

$$z_1 = a_1w_1 + a_2w_2 + a_3w_3 + a_6w_4 + a_7w_5 + a_8w_6 + a_{11}w_7 + a_{12}w_8 + a_{13}w_9$$

$$z_2 = a_3w_1 + a_4w_2 + a_5w_3 + a_8w_4 + a_9w_5 + a_{10}w_6 + a_{13}w_7 + a_{14}w_8 + a_{15}w_9$$

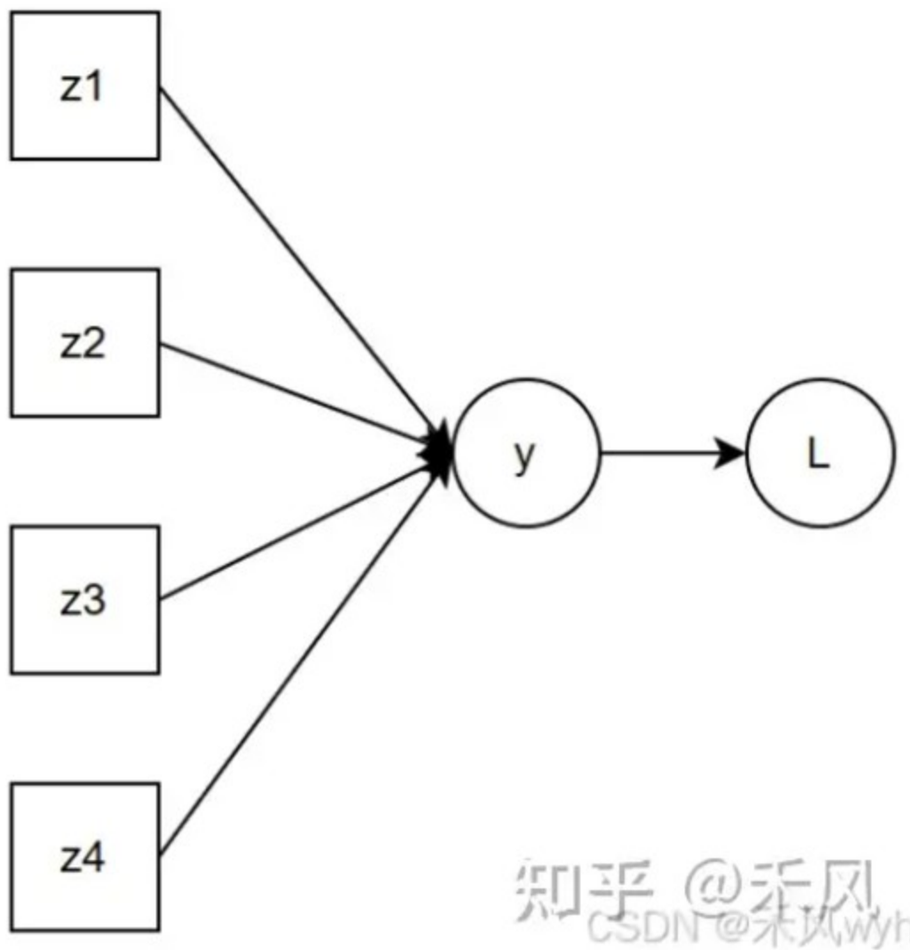
$$z_3 = a_{11}w_1 + a_{12}w_2 + a_{13}w_3 + a_{16}w_4 + a_{17}w_5 + a_{18}w_6 + a_{21}w_7 + a_{22}w_8 + a_{23}w_9$$

$$z_4 = a_{13}w_1 + a_{14}w_2 + a_{15}w_3 + a_{18}w_4 + a_{19}w_5 + a_{20}w_6 + a_{23}w_7 + a_{24}w_8 + a_{25}w_9$$

而通过对得到的特征图（z1、z2、z3、z4）进行操作，可以得到预测值y，最终计算除损失函数L。因为我们最终想要训练的参数值是卷积核上的数值，所以需要对w1...w9进行求偏导，根据链式法则，L先对z1...z4求偏导，然后再进一步分别对w1...w9求偏导。

w1	w2	w3
w4	w5	w6
w7	w8	w9

z1	z2
z3	z4



损失函数L对卷积核上的权重值求偏导的公式如下所示：

$$\frac{\partial L}{\partial w_1}=\frac{\partial L}{\partial z_1}\frac{\partial z_1}{\partial w_1}+\frac{\partial L}{\partial z_2}\frac{\partial z_2}{\partial w_1}+\frac{\partial L}{\partial z_3}\frac{\partial z_3}{\partial w_1}+\frac{\partial L}{\partial z_4}\frac{\partial z_4}{\partial w_1}=a_1\frac{\partial L}{\partial z_1}+a_3\frac{\partial L}{\partial z_2}+a_{11}\frac{\partial L}{\partial z_3}+a_{13}\frac{\partial L}{\partial z_4}$$

$$\frac{\partial L}{\partial w_2}=\frac{\partial L}{\partial z_1}\frac{\partial z_1}{\partial w_2}+\frac{\partial L}{\partial z_2}\frac{\partial z_2}{\partial w_2}+\frac{\partial L}{\partial z_3}\frac{\partial z_3}{\partial w_2}+\frac{\partial L}{\partial z_4}\frac{\partial z_4}{\partial w_2}=a_2\frac{\partial L}{\partial z_1}+a_4\frac{\partial L}{\partial z_2}+a_{12}\frac{\partial L}{\partial z_3}+a_{14}\frac{\partial L}{\partial z_4}$$

$$\frac{\partial L}{\partial w_3}=\frac{\partial L}{\partial z_1}\frac{\partial z_1}{\partial w_3}+\frac{\partial L}{\partial z_2}\frac{\partial z_2}{\partial w_3}+\frac{\partial L}{\partial z_3}\frac{\partial z_3}{\partial w_3}+\frac{\partial L}{\partial z_4}\frac{\partial z_4}{\partial w_3}=a_3\frac{\partial L}{\partial z_1}+a_5\frac{\partial L}{\partial z_2}+a_{13}\frac{\partial L}{\partial z_3}+a_{15}\frac{\partial L}{\partial z_4}$$

$$\frac{\partial L}{\partial w_4}=\frac{\partial L}{\partial z_1}\frac{\partial z_1}{\partial w_4}+\frac{\partial L}{\partial z_2}\frac{\partial z_2}{\partial w_4}+\frac{\partial L}{\partial z_3}\frac{\partial z_3}{\partial w_4}+\frac{\partial L}{\partial z_4}\frac{\partial z_4}{\partial w_4}=a_6\frac{\partial L}{\partial z_1}+a_8\frac{\partial L}{\partial z_2}+a_{16}\frac{\partial L}{\partial z_3}+a_{18}\frac{\partial L}{\partial z_4}$$

$$\frac{\partial L}{\partial w_5}=\frac{\partial L}{\partial z_1}\frac{\partial z_1}{\partial w_5}+\frac{\partial L}{\partial z_2}\frac{\partial z_2}{\partial w_5}+\frac{\partial L}{\partial z_3}\frac{\partial z_3}{\partial w_5}+\frac{\partial L}{\partial z_4}\frac{\partial z_4}{\partial w_5}=a_7\frac{\partial L}{\partial z_1}+a_9\frac{\partial L}{\partial z_2}+a_{17}\frac{\partial L}{\partial z_3}+a_{19}\frac{\partial L}{\partial z_4}$$

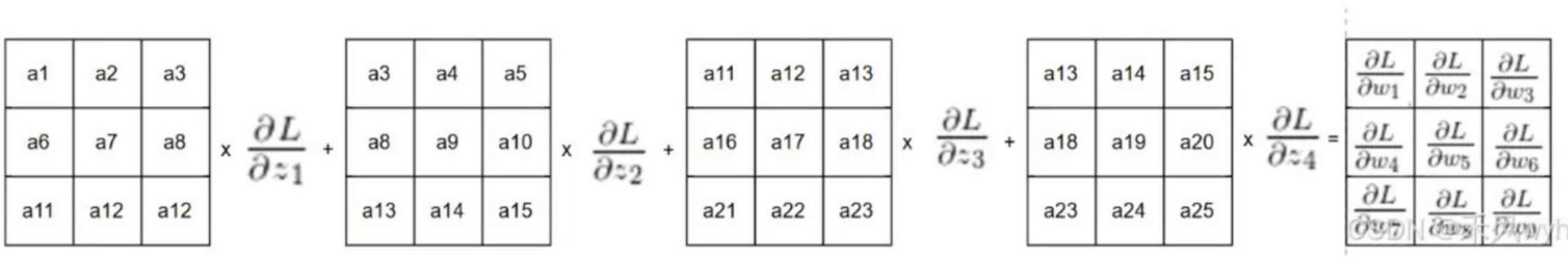
$$\frac{\partial L}{\partial w_6}=\frac{\partial L}{\partial z_1}\frac{\partial z_1}{\partial w_6}+\frac{\partial L}{\partial z_2}\frac{\partial z_2}{\partial w_6}+\frac{\partial L}{\partial z_3}\frac{\partial z_3}{\partial w_6}+\frac{\partial L}{\partial z_4}\frac{\partial z_4}{\partial w_6}=a_8\frac{\partial L}{\partial z_1}+a_{10}\frac{\partial L}{\partial z_2}+a_{18}\frac{\partial L}{\partial z_3}+a_{20}\frac{\partial L}{\partial z_4}$$

$$\frac{\partial L}{\partial w_7}=\frac{\partial L}{\partial z_1}\frac{\partial z_1}{\partial w_7}+\frac{\partial L}{\partial z_2}\frac{\partial z_2}{\partial w_7}+\frac{\partial L}{\partial z_3}\frac{\partial z_3}{\partial w_7}+\frac{\partial L}{\partial z_4}\frac{\partial z_4}{\partial w_7}=a_{11}\frac{\partial L}{\partial z_1}+a_{13}\frac{\partial L}{\partial z_2}+a_{21}\frac{\partial L}{\partial z_3}+a_{23}\frac{\partial L}{\partial z_4}$$

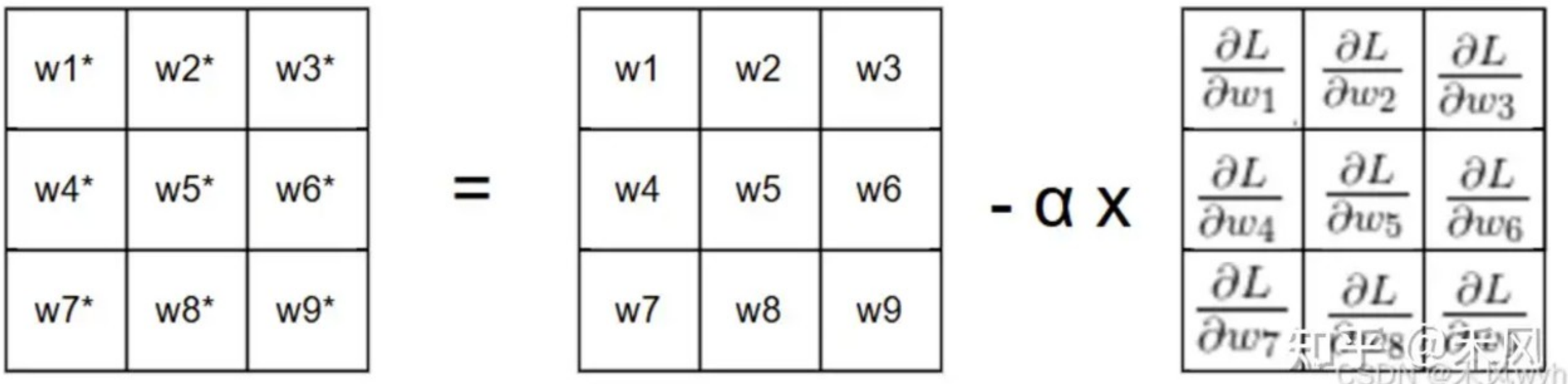
$$\frac{\partial L}{\partial w_8}=\frac{\partial L}{\partial z_1}\frac{\partial z_1}{\partial w_8}+\frac{\partial L}{\partial z_2}\frac{\partial z_2}{\partial w_8}+\frac{\partial L}{\partial z_3}\frac{\partial z_3}{\partial w_8}+\frac{\partial L}{\partial z_4}\frac{\partial z_4}{\partial w_8}=a_{12}\frac{\partial L}{\partial z_1}+a_{14}\frac{\partial L}{\partial z_2}+a_{22}\frac{\partial L}{\partial z_3}+a_{24}\frac{\partial L}{\partial z_4}$$

$$\frac{\partial L}{\partial w_9}=\frac{\partial L}{\partial z_1}\frac{\partial z_1}{\partial w_9}+\frac{\partial L}{\partial z_2}\frac{\partial z_2}{\partial w_9}+\frac{\partial L}{\partial z_3}\frac{\partial z_3}{\partial w_9}+\frac{\partial L}{\partial z_4}\frac{\partial z_4}{\partial w_9}=a_{13}\frac{\partial L}{\partial z_1}+a_{15}\frac{\partial L}{\partial z_2}+a_{23}\frac{\partial L}{\partial z_3}+a_{25}\frac{\partial L}{\partial z_4}$$

将公式列出后，我们可以找到规律，L对z1求偏导的系数对应对应图片上相乘再相加后得到z1的元素部分，L对z2、z3、z4求偏导的系数也是一样的。



得到L对w1...w9的偏导（梯度）之后，我们就可以进行权重更新。



简化之后，可以得到与线性层相似的梯度更新公式：

$$W = W - \eta \frac{\partial L}{\partial W} \quad \text{(梯度下降)}$$

这个过程通过不断地进行正向传播和反向传播，逐步调整卷积核的权重和偏置项，使得网络输出更接近实际目标。