

h



预训练权重

$$W \in \mathbb{R}^{d \times d}$$

$$B = 0$$

r

$$A = \mathcal{N}(0, \sigma^2)$$

d

x

