

隐藏层

维度: d

+

预训练权重

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

$$B = 0$$

$$A = N(0, \sigma)$$

维度: d

隐藏层

