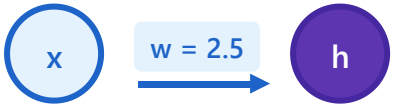
Weights and Biases

Weights (W)

Matrix Form

$$W^{(l)} \in \mathbb{R}^{(m \times n)}$$

$$\begin{bmatrix} w_{11} & w_{12} & \cdots & w_{1n} \\ w_{21} & w_{22} & \cdots & w_{2n} \\ \vdots & \vdots & \vdots & \vdots \end{bmatrix}$$



1. **Connection strength** between neurons
2. **Feature importance:** larger |w| = more influence
3. **Learned during training** via gradient descent

Biases (b)

Vector Form

$$b^{(l)} \in \mathbb{R}^{m}$$

$$\begin{bmatrix} b_1 \\ b_2 \\ \vdots \end{bmatrix}$$

1. **Shifts activation:** controls threshold
2. **Model flexibility:** allows fitting data better
3. **One per neuron** in the layer

Layer Computation

Weighted Sum

$W^{(l)}h^{(l-1)}$

+

Bias Term

$b^{(l)}$

Complete Formula: $z^{(l)} = W^{(l)}h^{(l-1)} + b^{(l)} \rightarrow h^{(l)} = \sigma(z^{(l)})$