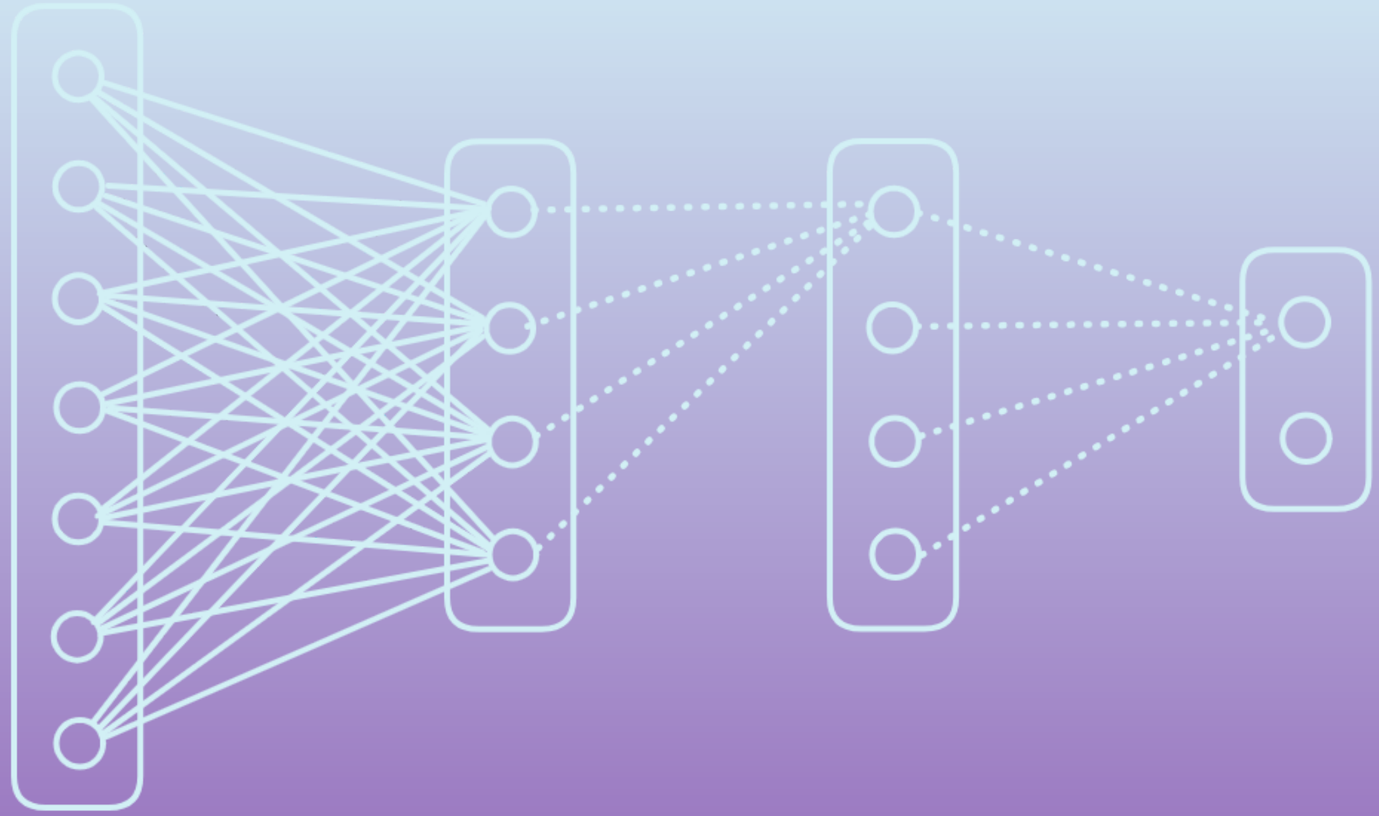
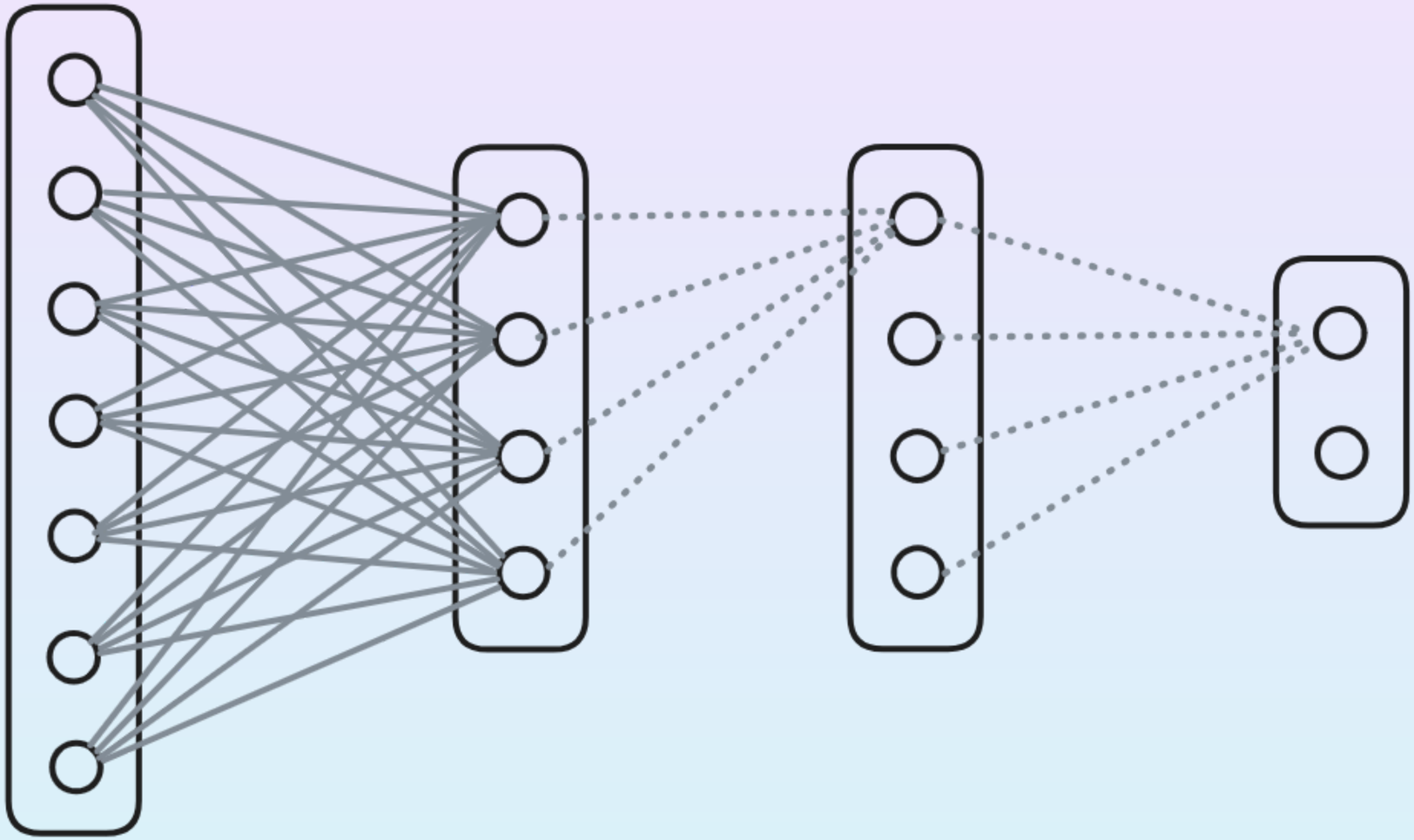
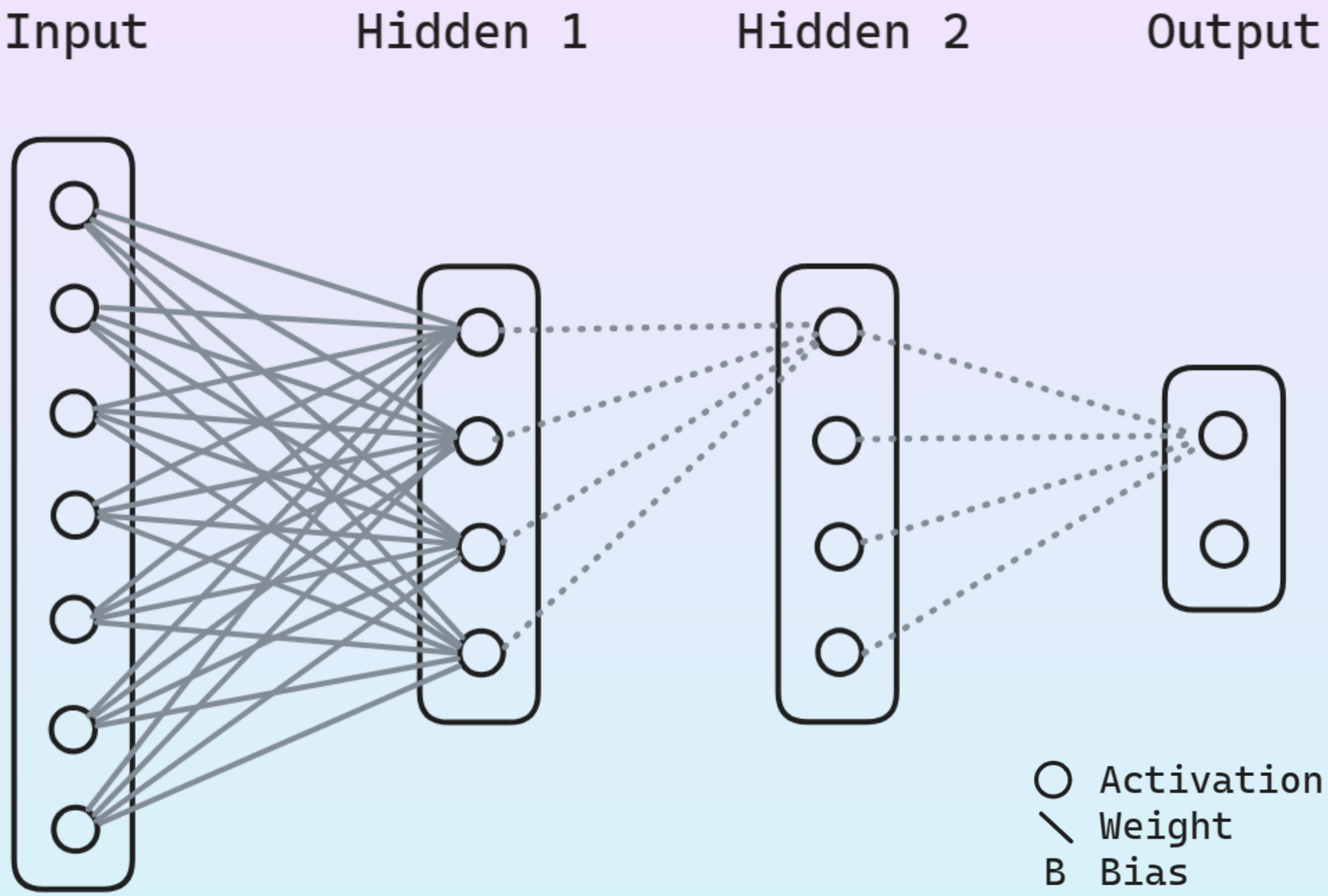
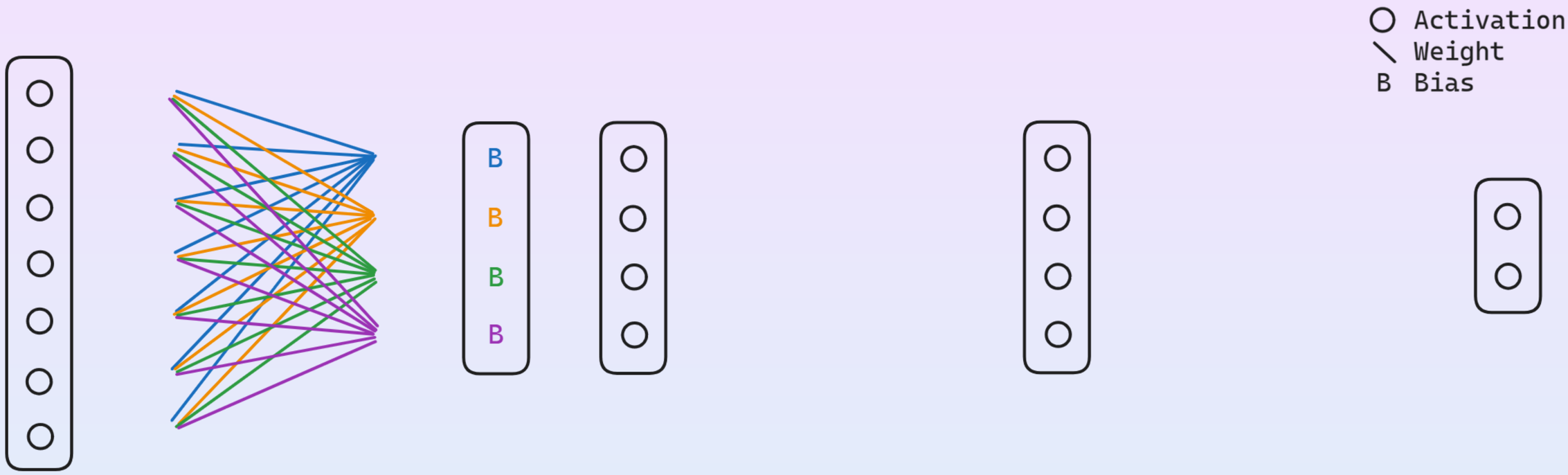


peek in the dark

sebastiaan indesteege




Input
Hidden 1
Hidden 2
Output
Activation
Weight
B Bias



LAYER 1 28 Weights + 4 Biases

A11	W1	W2	W3	W4	W5	W6	W7	B1
A12	W1	W2	W3	W4	W5	W6	W7	B2
A13	W1	W2	W3	W4	W5	W6	W7	B3
A14	W1	W2	W3	W4	W5	W6	W7	B4
A15								
A16								
A17								

LAYER 2 16 W + 4 B

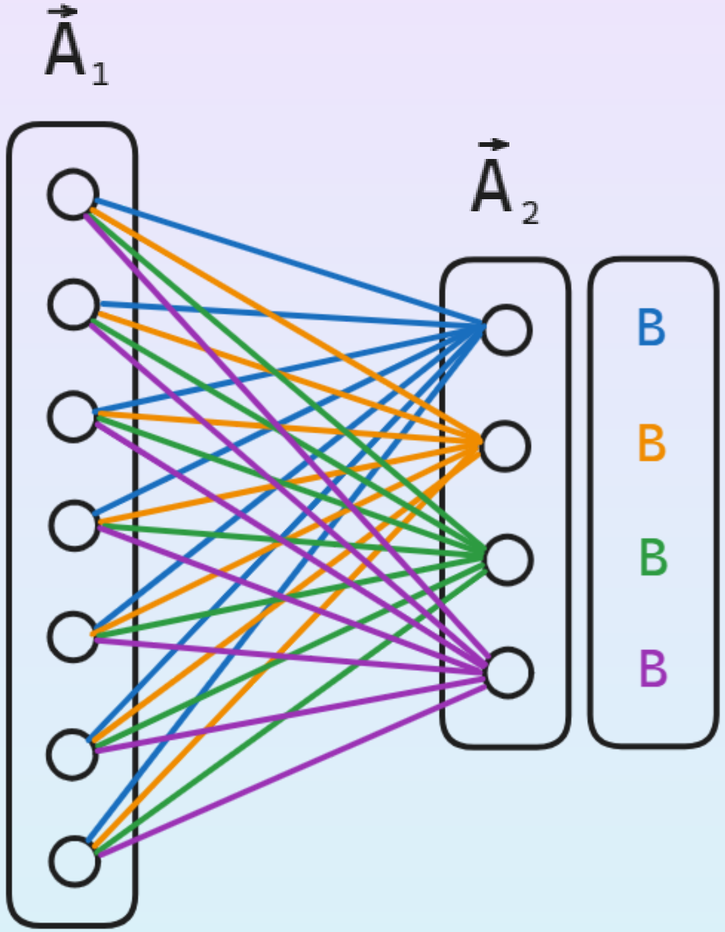
A21	W1	W2	W3	W4	B1
A22	W1	W2	W3	W4	B2
A23	W1	W2	W3	W4	B3
A24	W1	W2	W3	W4	B4

LAYER 3 8 W + 2 B

A31					
A32	W1	W2	W3	W4	B1
A33	W1	W2	W3	W4	B2
A34					

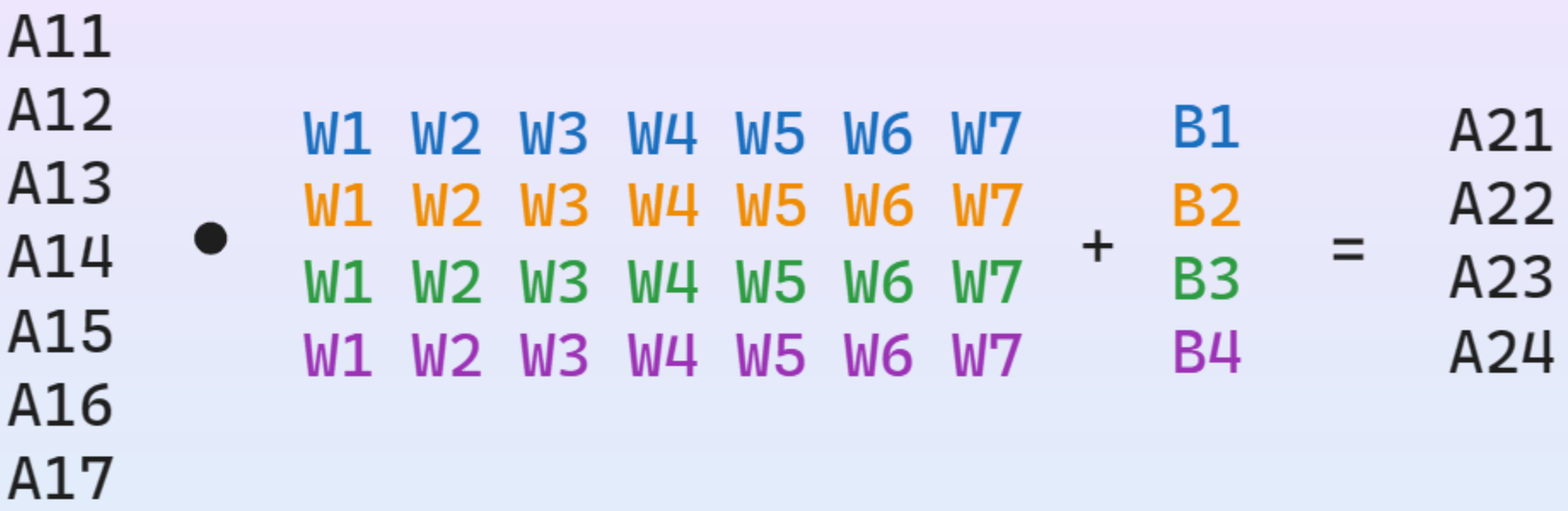
LAYER 4

A41
A42



$$\vec{A}_2 = \vec{A}_1 \bullet [W]_1 + \vec{B}_1$$

$$\vec{A}_2 = \text{ActivationFunction}(\vec{A}_2)$$



A21 = ActFc(W1*A11 + W2*A12 + W3*A13 + W4*A14 + W5*A15 + W6*A16 + W7*A17 + B1)
A22 = ActFc(W1*A11 + W2*A12 + W3*A13 + W4*A14 + W5*A15 + W6*A16 + W7*A17 + B2)
A23 = ActFc(W1*A11 + W2*A12 + W3*A13 + W4*A14 + W5*A15 + W6*A16 + W7*A17 + B3)
A23 = ActFc(W1*A11 + W2*A12 + W3*A13 + W4*A14 + W5*A15 + W6*A16 + W7*A17 + B4)

Initialization?

[-1 : 1]
gaussian distribution

“xavier” or “glorot”
for non linear activation functions

Some Activation Functions