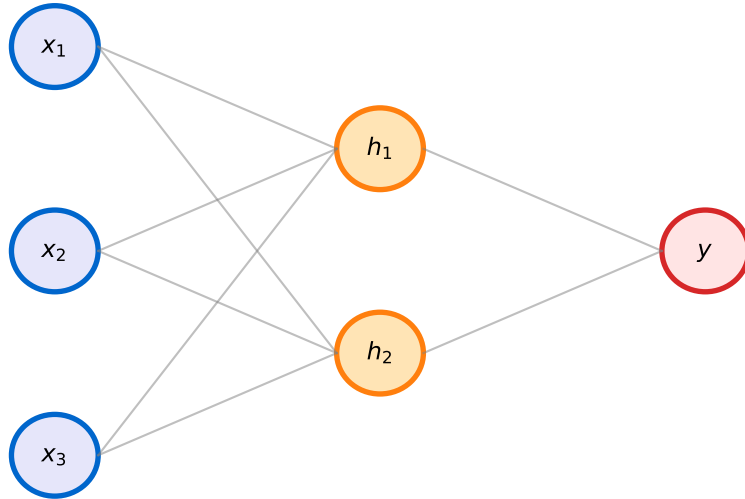# Why Non-Linear Activations Are Essential

Without Non-Linear Activation
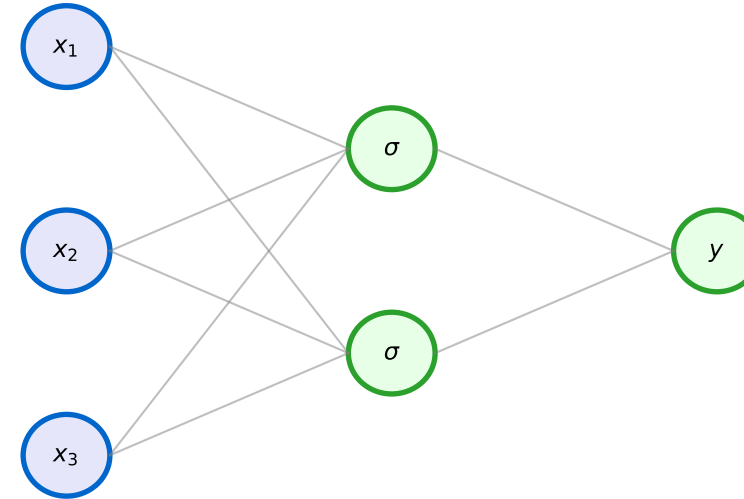
Math:

$$h = W_1 x$$

$$y = W_2 h$$

$$y = W_2(W_1 x)$$

$$y = (W_2 W_1)x$$

$$y = W_{eff} x$$

Still just a
linear transform!

Multiple linear layers = One linear layer

With Non-Linear Activation

Math:

$$h = \sigma(W_1 x)$$

$$y = W_2 h$$

$$y = W_2 \sigma(W_1 x)$$

Cannot simplify!

Non-linearity
breaks collapse

Each layer adds
expressiveness

Activation functions enable deep learning