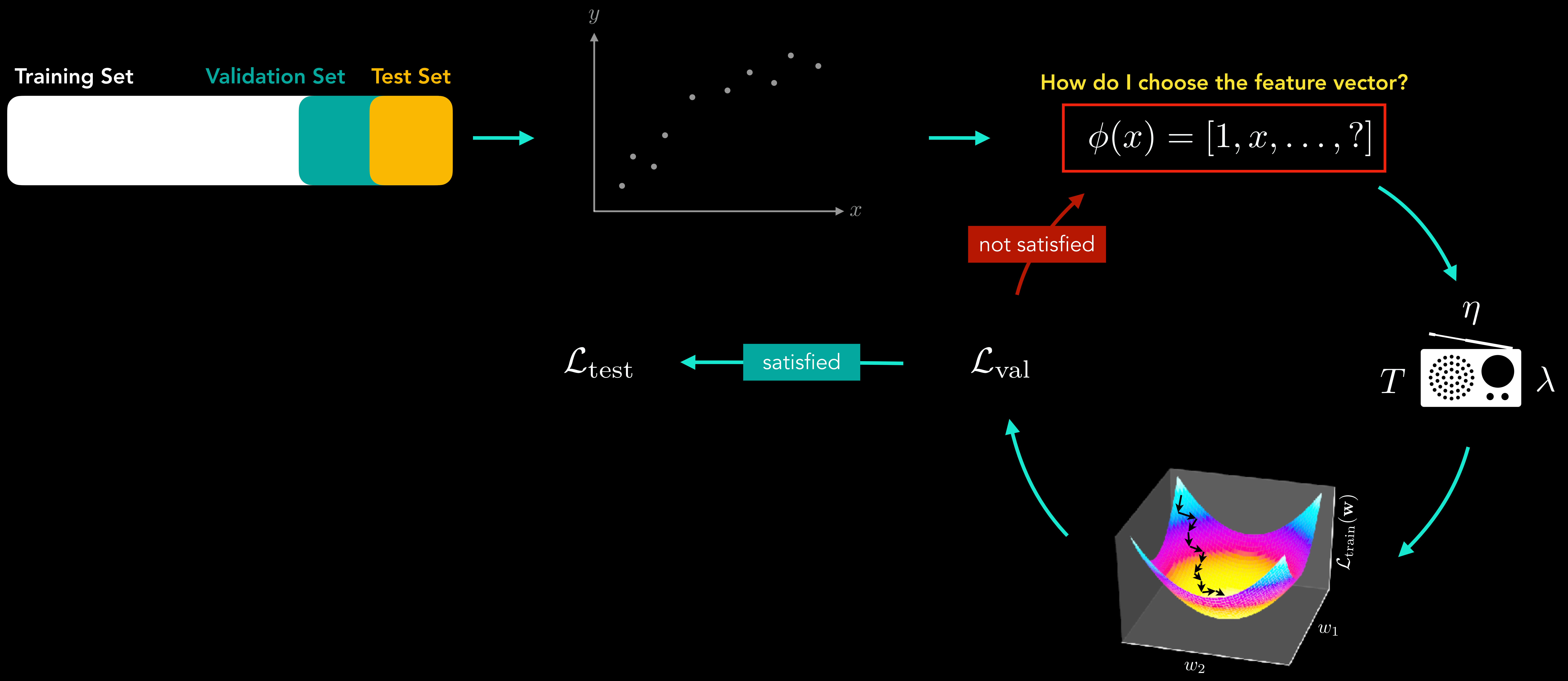
Deep Learning

The ML workflow



How do I choose the feature vector?

$$\phi(x) = [1, x, \ldots, ?]$$

$$f_{\mathbf{w}}(x) = \mathbf{w} \cdot \boxed{\phi(x)}$$

$\mathbf{x}$

$$\phi(x) = [1,\ x]$$

$$\phi(x) = [1, x, x^2, x^3]$$

$$\phi(x) = [1,\ \ x,\ \sin(3x)]$$

??????????????

How do I choose the feature vector?

$$\phi(x) = [1, x, \ldots, ?]$$

Decision Boundary

$$\phi(x) \cdot \mathbf{w} = 0$$

→ **Boat**

Linear Predictor

$$f_{\mathbf{w}}(x) = \mathbf{x} \cdot \mathbf{w}$$

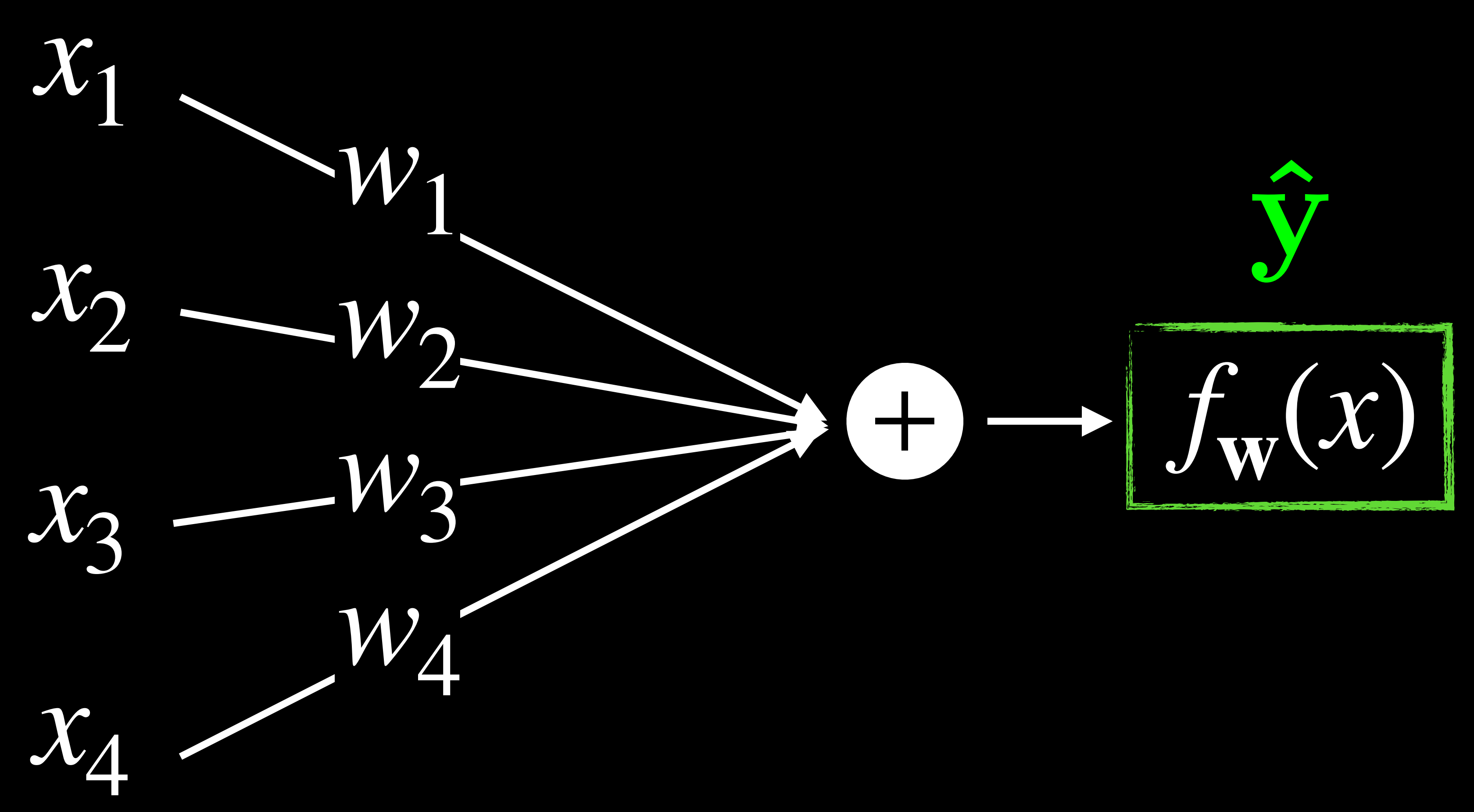
$$\mathbf{w} = [w_1, w_2, w_3, w_4]$$

$$\mathbf{x} = [x_1, x_2, x_3, x_4]$$

$$f_{\mathbf{w}}(x) = w_1x_1 + w_2x_2 + w_3x_3 + w_4x_4$$

→

Network Representation



Linear Predictor

2 outputs?

$$\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} \longrightarrow \begin{bmatrix} \hat{y}_1 \\ \hat{y}_2 \end{bmatrix}$$

$$\mathbf{x} \qquad\qquad \hat{\mathbf{y}}$$

3 * 2 fitting parameters

$$\hat{y}_1 = \mathbf{w}_1 \cdot \mathbf{x} = w_{11}x_1 + w_{12}x_2 + w_{13}x_3$$

$$\hat{y}_2 = \mathbf{w}_2 \cdot \mathbf{x} = w_{21}x_1 + w_{22}x_2 + w_{23}x_3$$

Matrix form

$$\begin{bmatrix} \hat{y}_1 \\ \hat{y}_2 \end{bmatrix} = \begin{bmatrix} w_{11} & w_{12} & w_{13} \\ w_{21} & w_{22} & w_{23} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}$$

$$\hat{\mathbf{y}} = \mathbf{W} \mathbf{x}$$

From Matrix to Network

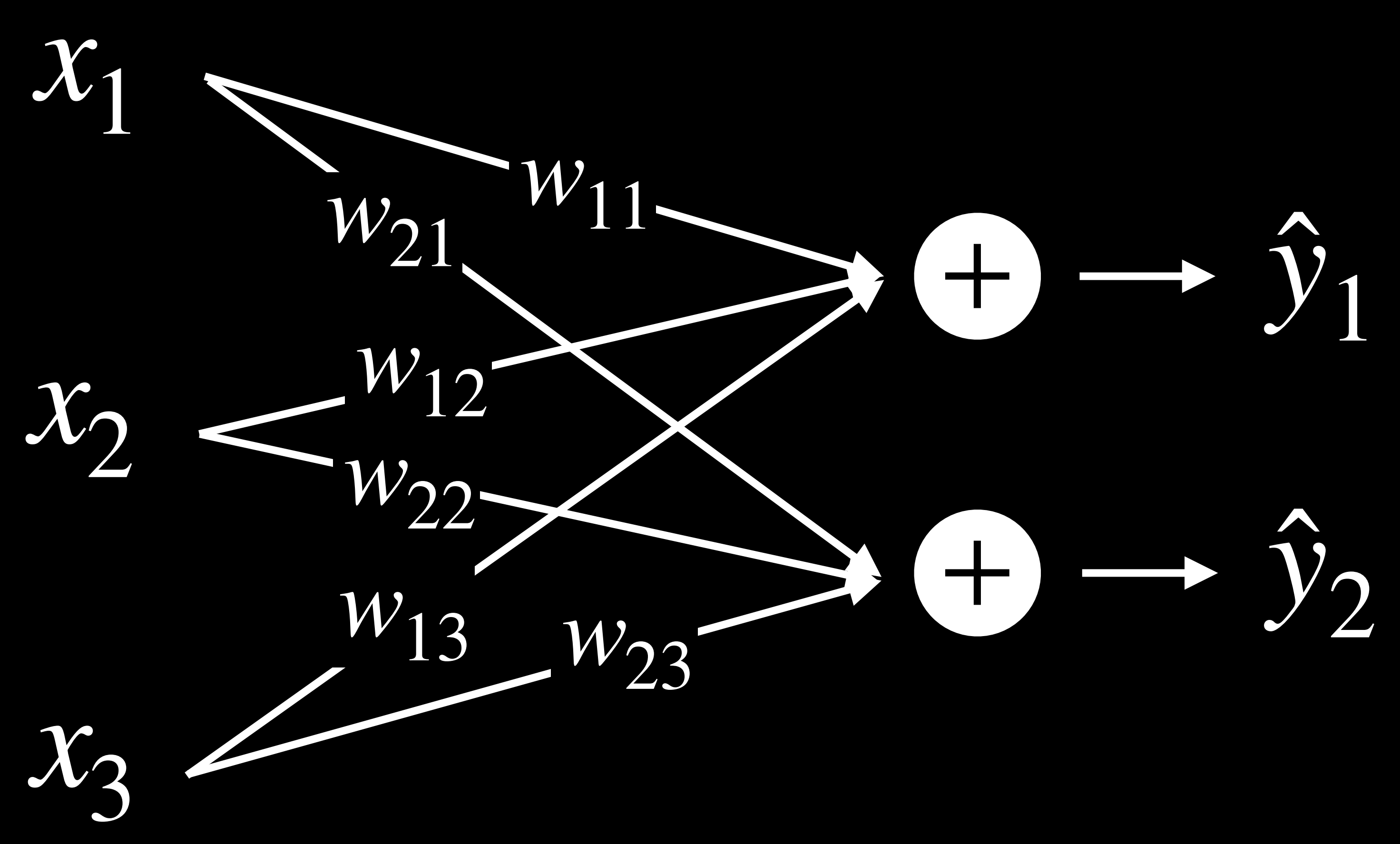
Matrix Representation

$$\begin{bmatrix}\hat{y}_1\\ \hat{y}_2\end{bmatrix} = \begin{bmatrix} w_{11} & w_{12} & w_{13} \\ w_{21} & w_{22} & w_{23}\end{bmatrix}\begin{bmatrix}x_1\\ x_2\\ x_3\end{bmatrix}$$

$$\hat{\mathbf{y}} = \mathbf{W}\,\mathbf{x}$$

Network Representation

Account for Bias: $1 = x_1$



Index notation

$$\hat{y}_i = \sum_{j=1}^{n} w_{ij}x_j$$

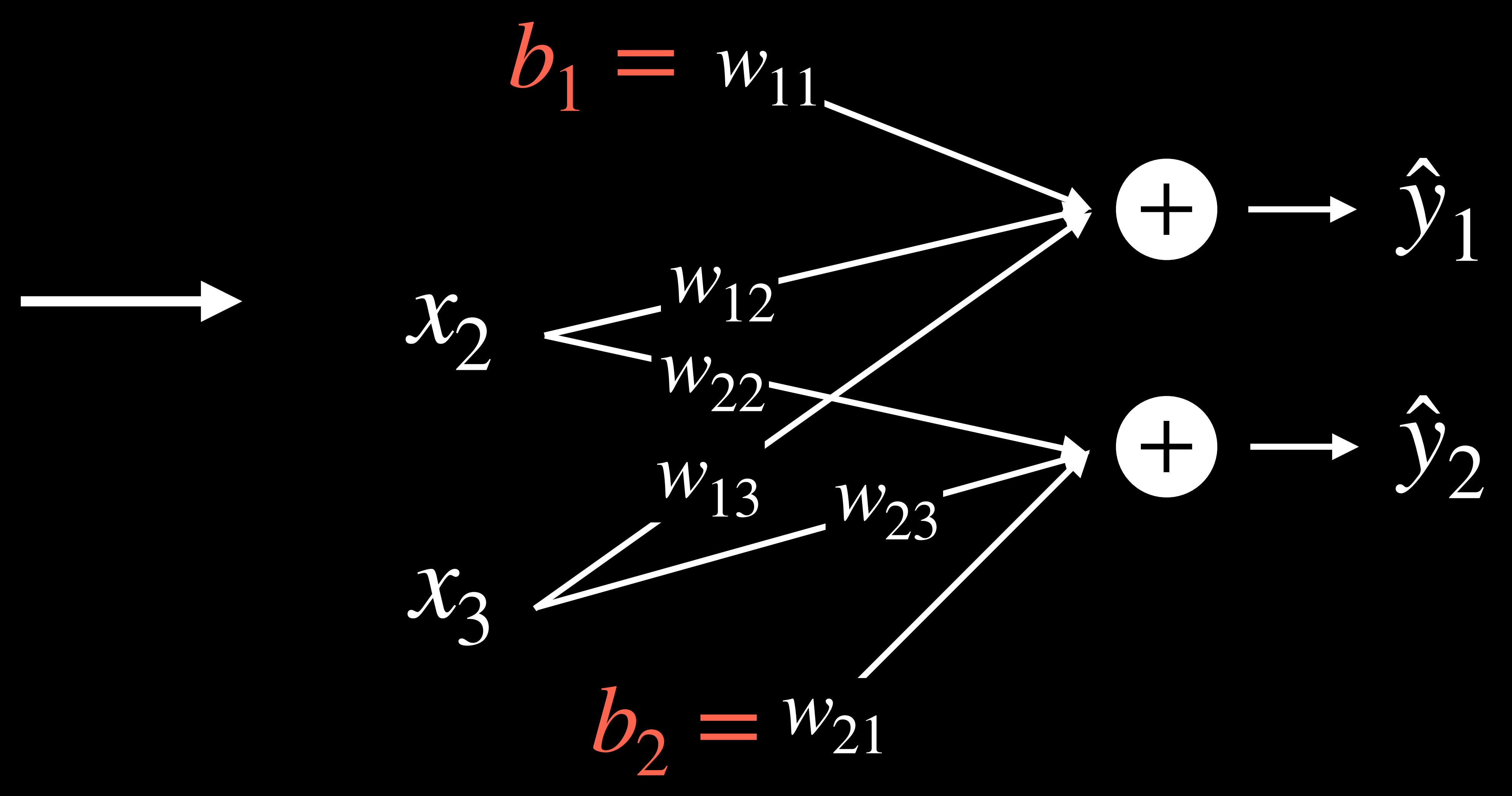
Linear Predictor - Explicit Bias

Matrix Representation

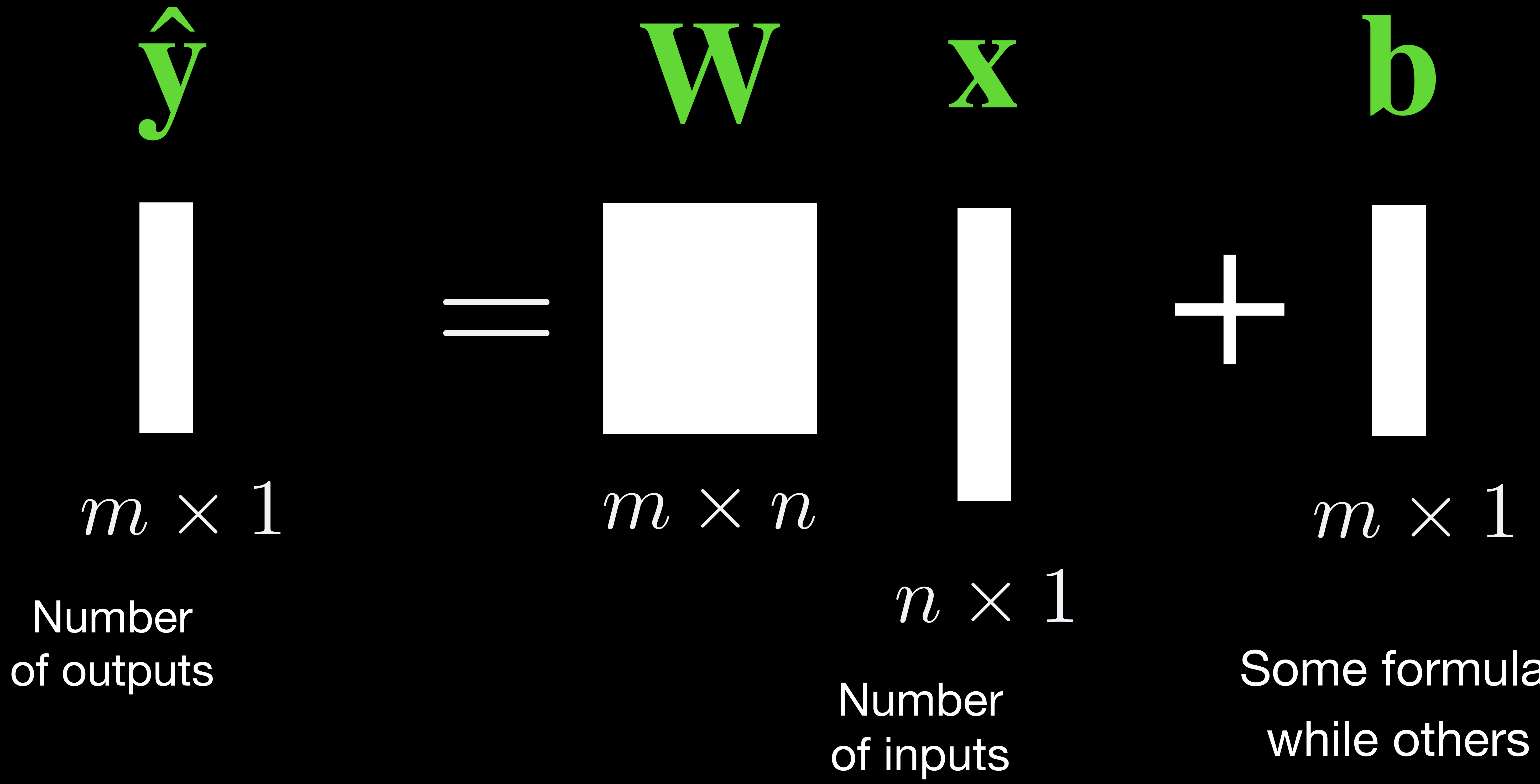
$$\begin{bmatrix}\hat{y}_1 \\ \hat{y}_2\end{bmatrix} = \begin{bmatrix} w_{12} & w_{13} \\ w_{22} & w_{23}\end{bmatrix} \begin{bmatrix} x_2 \\ x_3 \end{bmatrix} + \begin{bmatrix} w_{11} \\ w_{21}\end{bmatrix}$$

$$\hat{\mathbf{y}} = \mathbf{W} \quad \mathbf{x} \quad \mathbf{b}$$

Network Representation



Linear Predictor

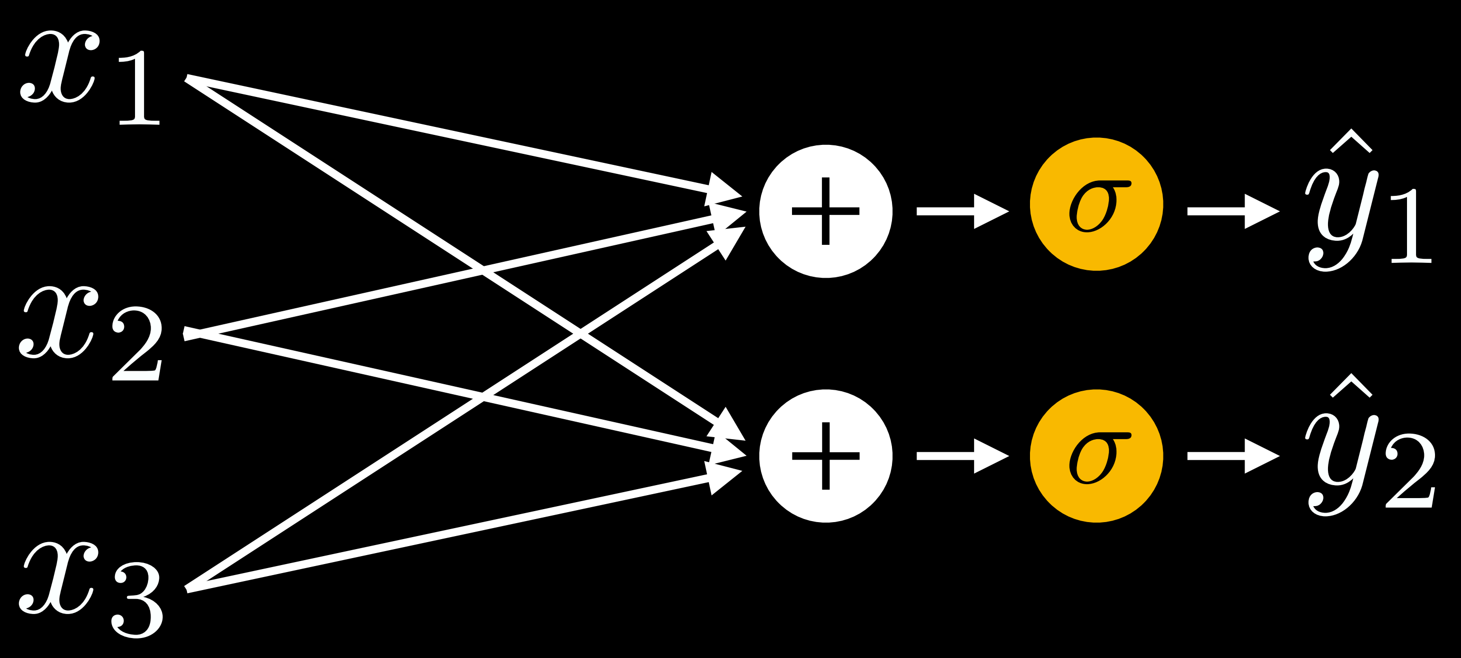


Some formulations explicitly account for $\mathbf{b}$, while others include the bias as part of $\mathbf{x}$

Here we omit $\mathbf{b}$ for simplicity of representation

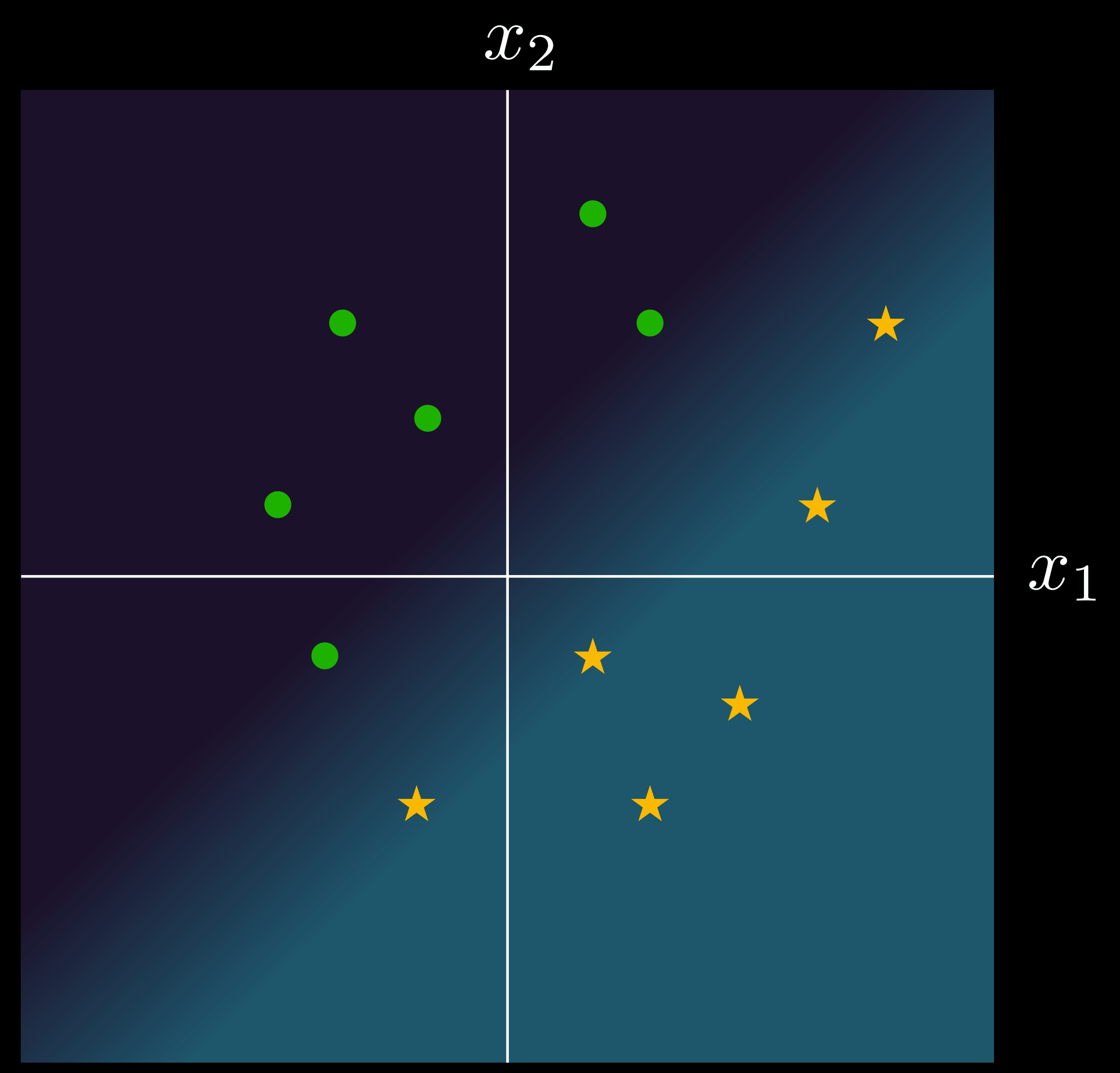
Nonlinear Predictor

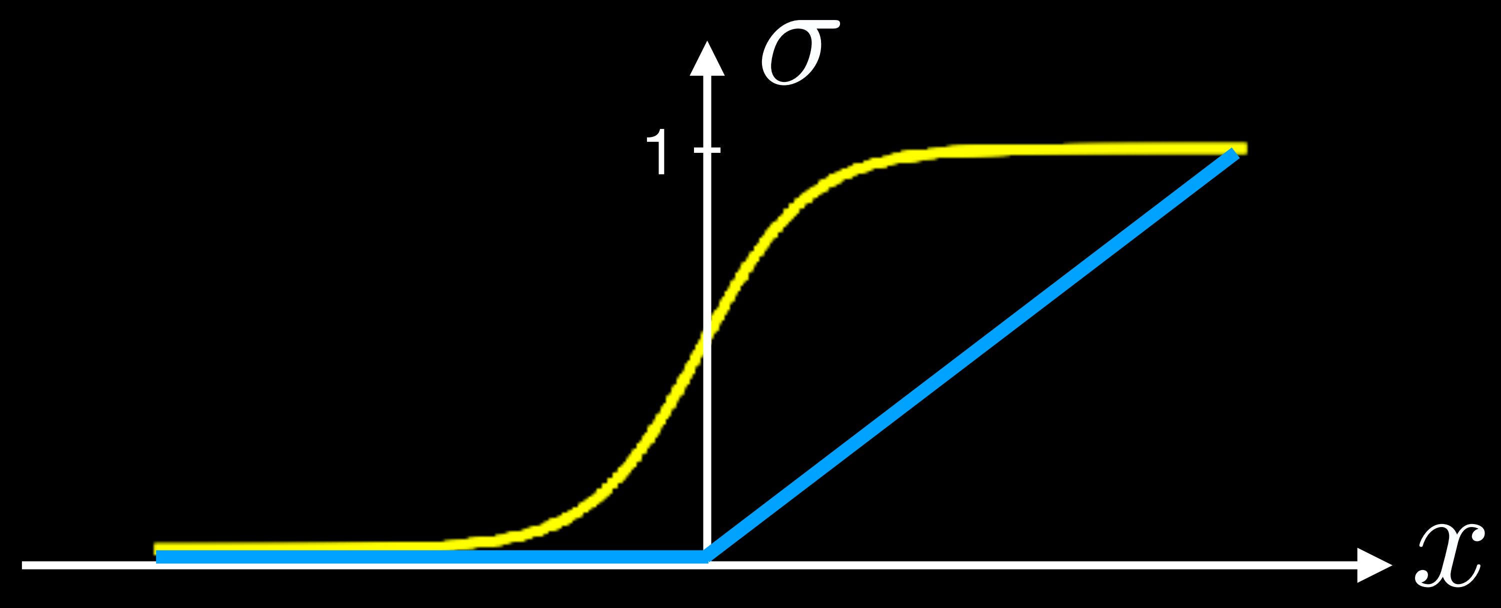
$$\hat{\mathbf{y}} = \sigma(\mathbf{W}\mathbf{x})$$



Logistic function: $\sigma(x) = \frac{1}{1 + e^{-x}}$

ReLU: $\sigma(x) = xH(x)$

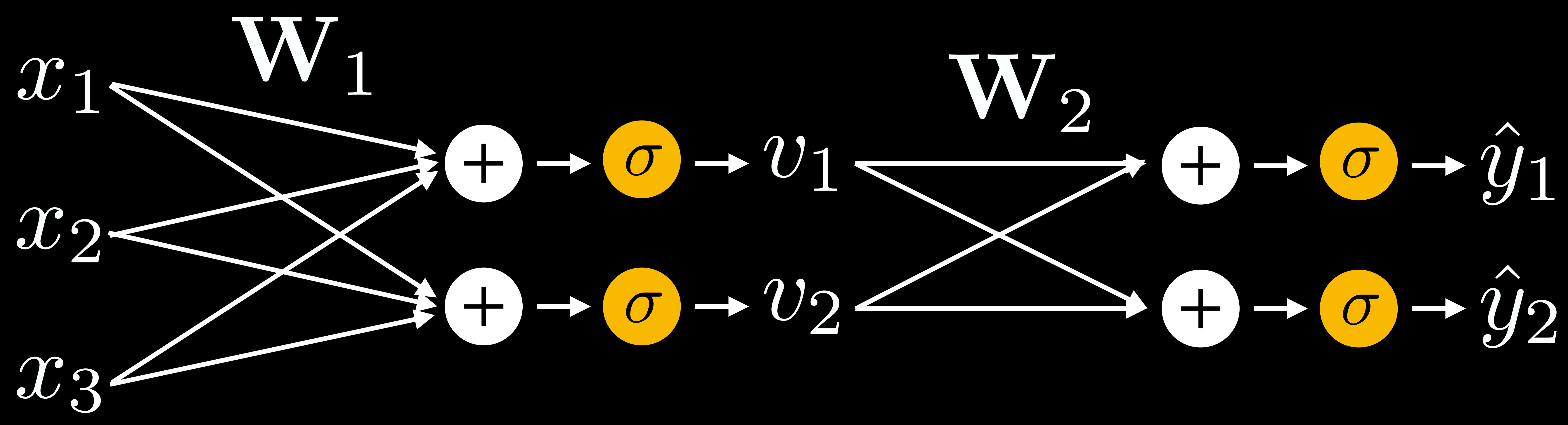




Activation Function

Neural network

$$\hat{\mathbf{y}} = \sigma\left(\mathbf{W}_2\, \sigma\left(\mathbf{W}_1 \mathbf{x}\right)\right)$$



$$\mathbf{x} \rightarrow \boxed{\sigma \circ \mathbf{W}_2 \circ \sigma \circ \mathbf{W}_1} \rightarrow \hat{\mathbf{y}}$$

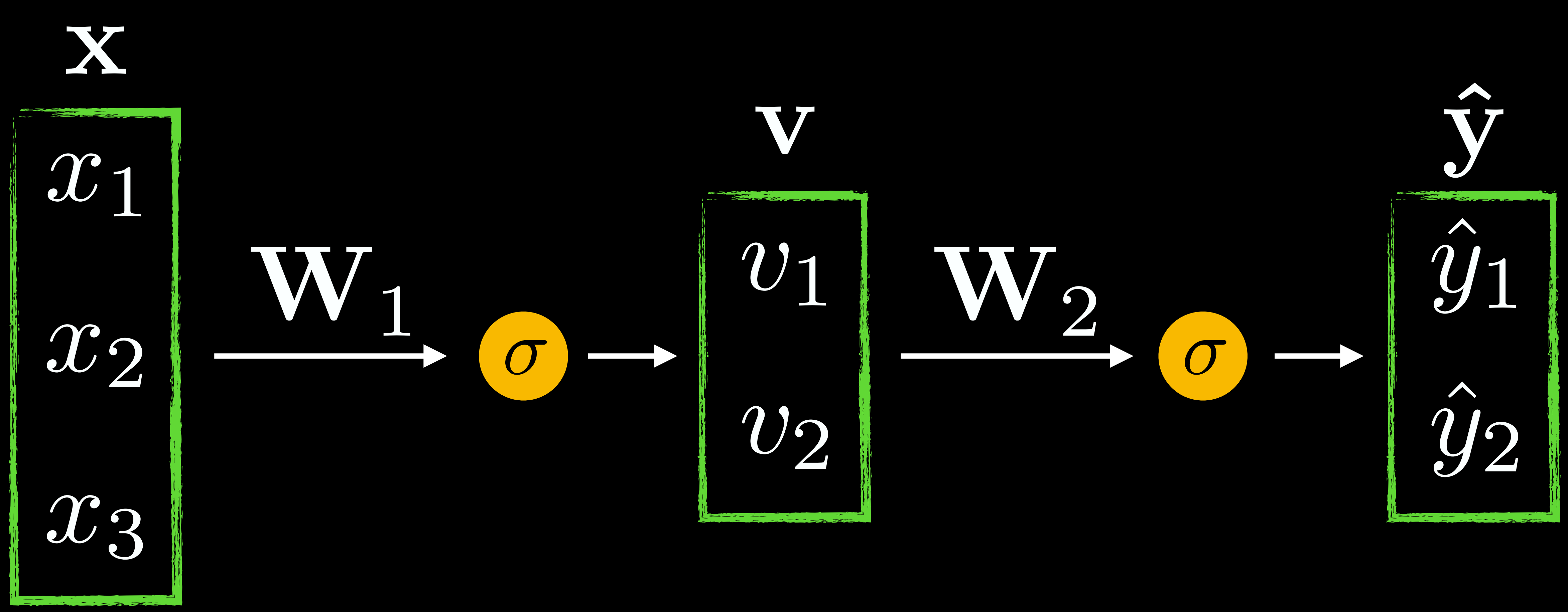
Nonlinear Linear Nonlinear Linear

Neural network

$$\hat{\mathbf{y}} = \sigma\left(\mathbf{W}_2\, \sigma\left(\mathbf{W}_1 \mathbf{x}\right)\right)$$

$$\mathbf{v} = \sigma(\mathbf{W}_1 \mathbf{x})$$

$$\hat{\mathbf{y}} = \sigma(\mathbf{W}_2 \mathbf{v})$$

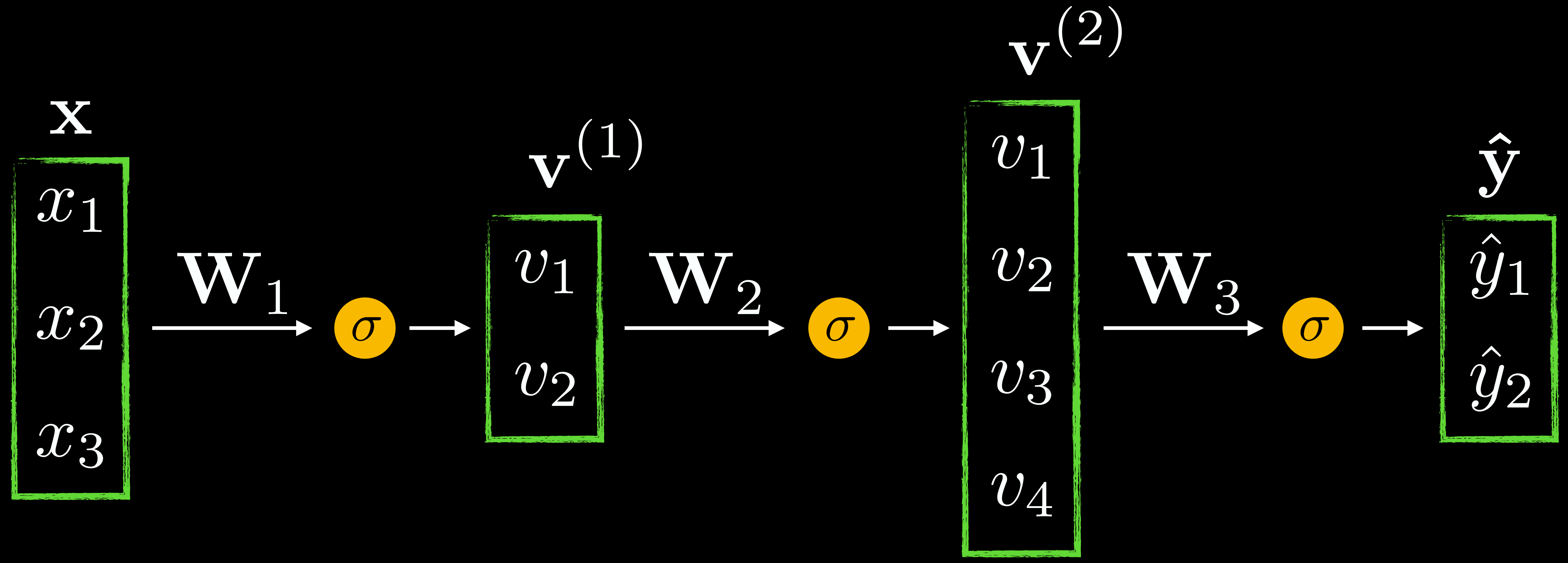


Hidden layer

Can be interpreted as a learned $\phi(\mathbf{x})$

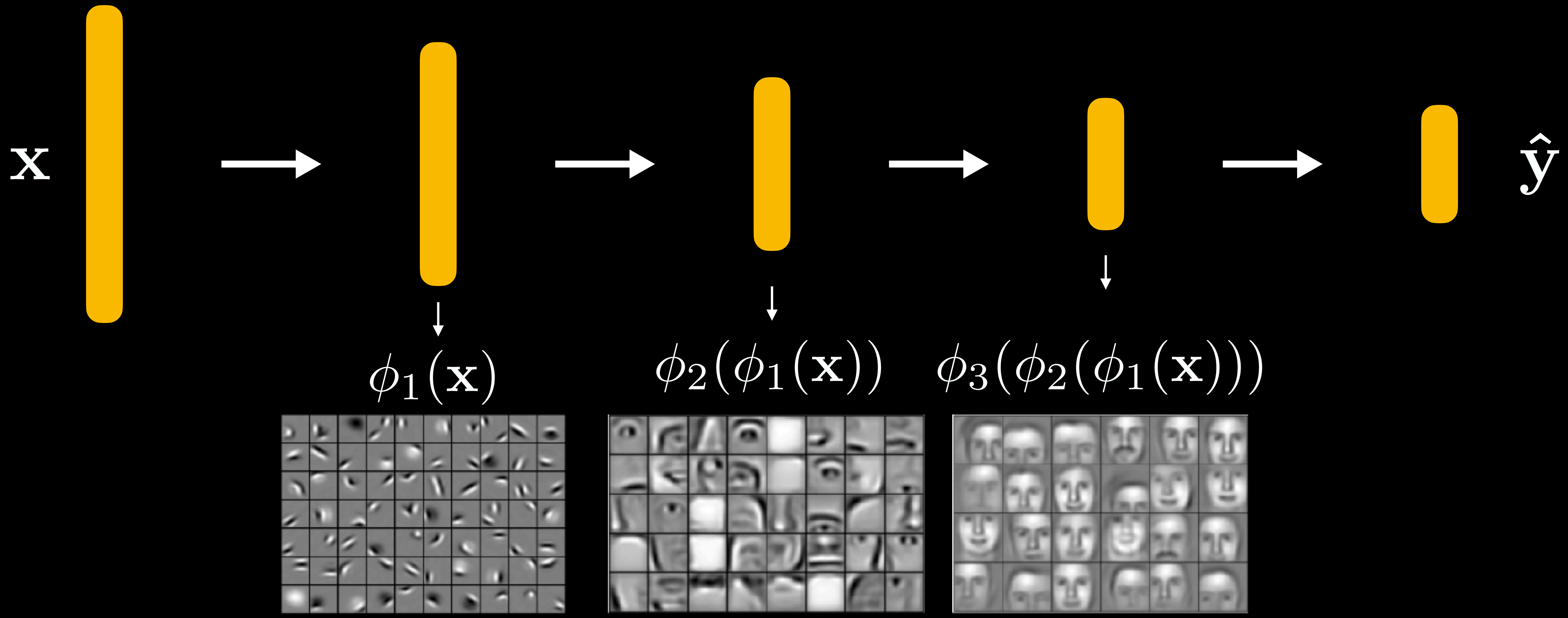
Deep network

$$\hat{\mathbf{y}} = \sigma\left(\mathbf{W}_3\ \sigma\left(\mathbf{W}_2\ \sigma\left(\mathbf{W}_1\mathbf{x}\right)\right)\right)$$



$$\mathbf{v}^{(1)} = \sigma\left(\mathbf{W}_1\mathbf{x}\right) \qquad \mathbf{v}^{(2)} = \sigma\left(\mathbf{W}_2\mathbf{v}^{(1)}\right) \qquad \hat{\mathbf{y}} = \sigma\left(\mathbf{W}_3\mathbf{v}^{(2)}\right)$$

Why *deep* learning?



Feature learning

Lee, Honglak, et al. "Unsupervised learning of hierarchical representations with convolutional deep belief networks." Communications of the ACM 54.10 (2011)