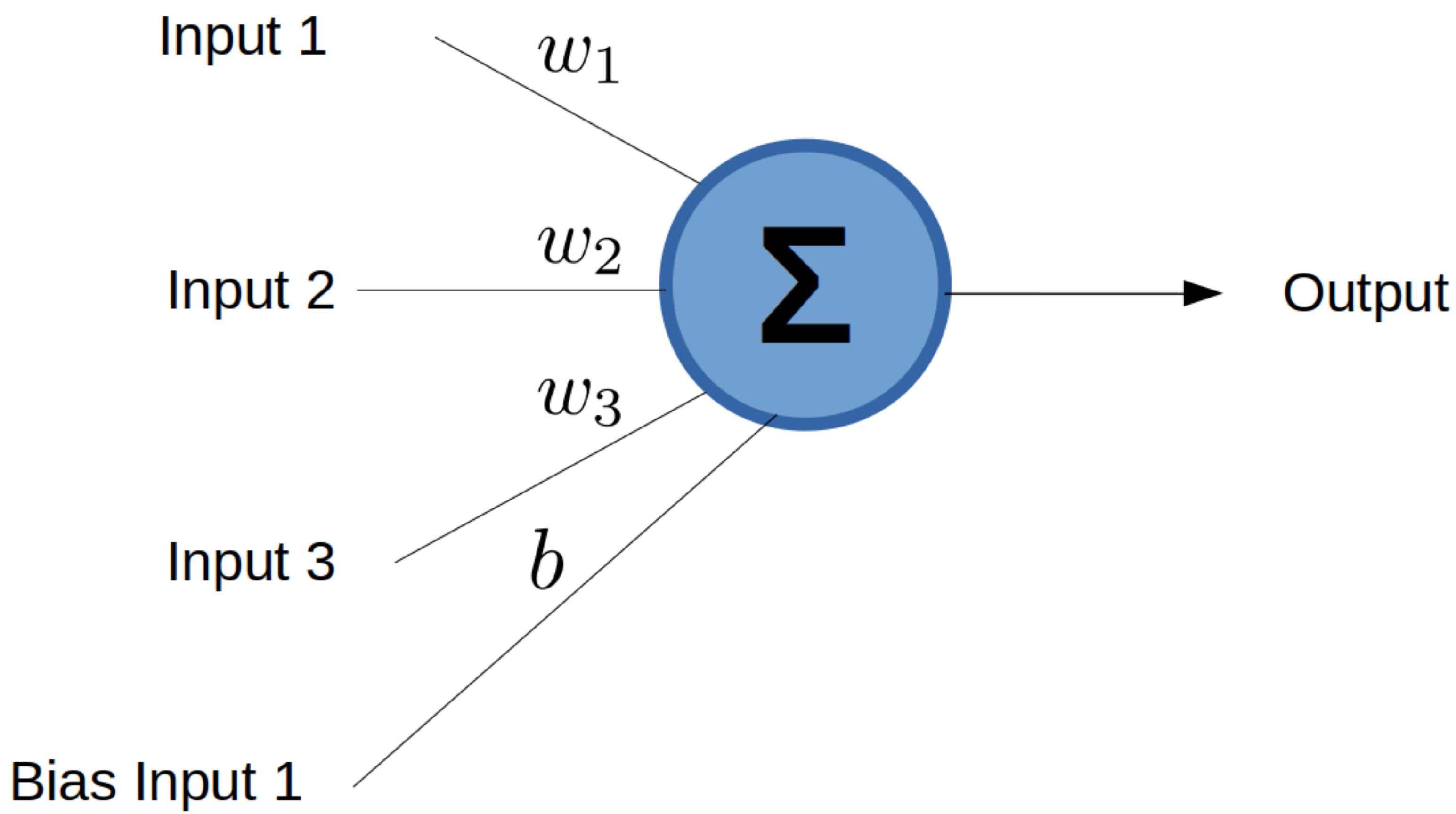
Multilayer Perceptron

Heni Ben Amor, Ph.D.
Assistant Professor
Arizona State University

Ira A. Fulton Schools of Engineering
Arizona State University

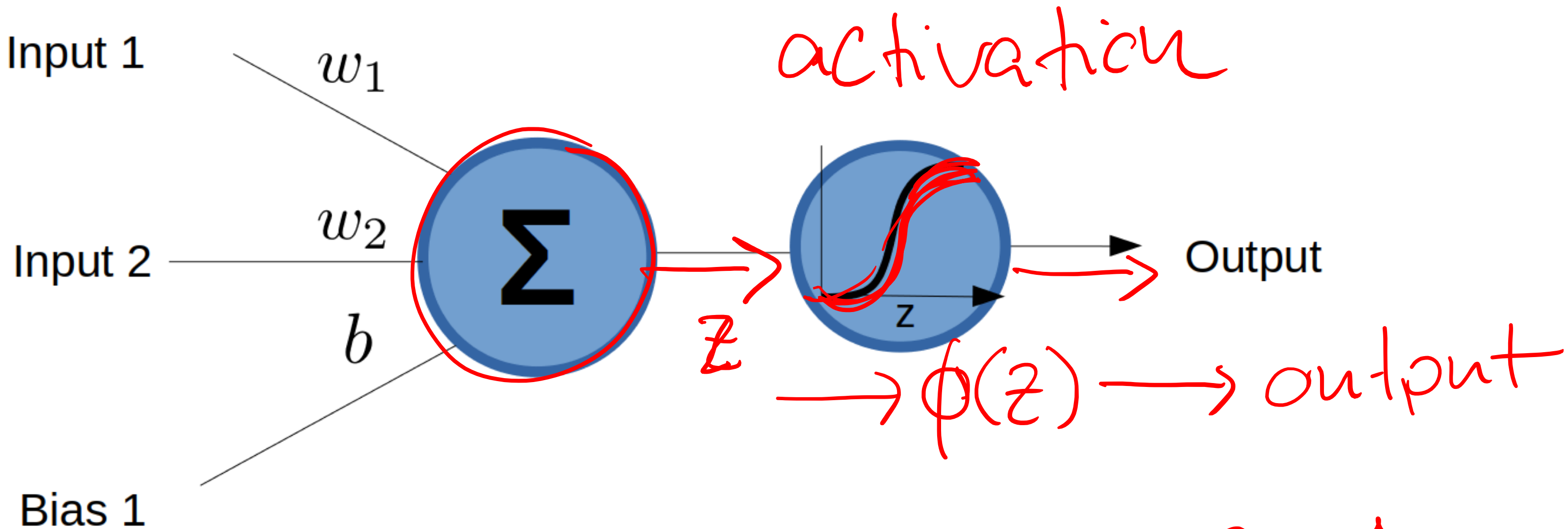
Linear Perceptron with Bias

- The input to the bias term is always 1



Nonlinear Perceptron

- Add nonlinear activation function

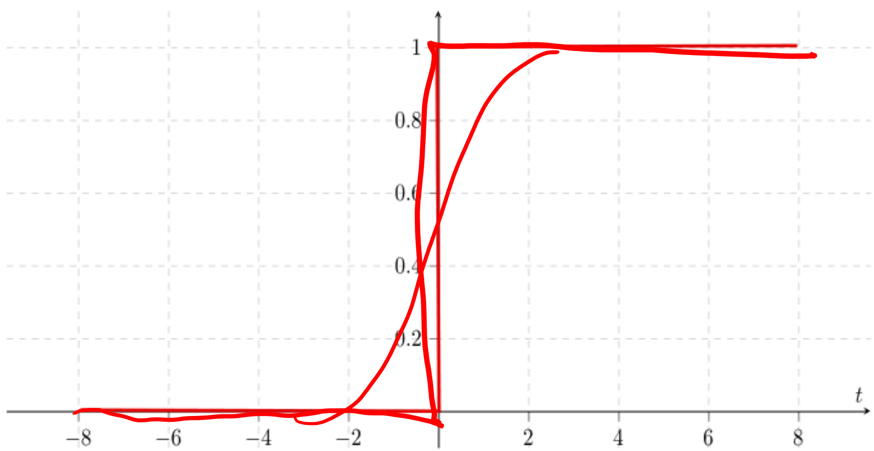


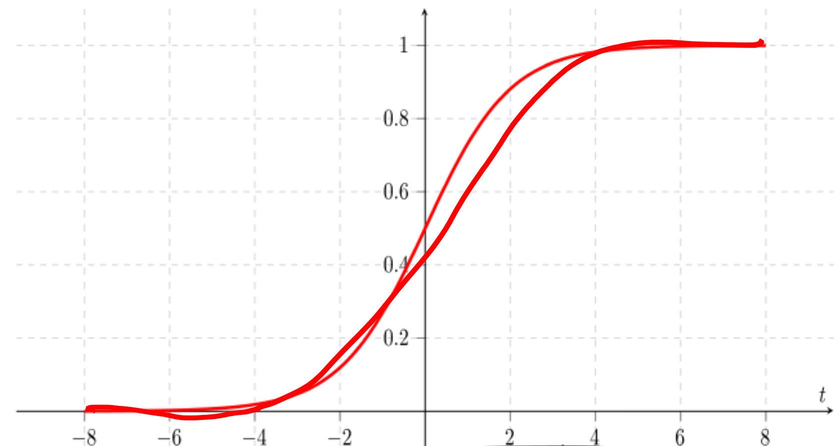
- Output is calculated via: $a = \boxed{\phi}(\overbrace{\mathbf{w}^T\mathbf{x} + b})$
- Possible activation function:

Sigmoid $\phi(\boxed{z}) = \frac{1}{1 + \exp(\boxed{-z})}$

Sigmoid Activation Function

- A **soft** version of a threshold unit





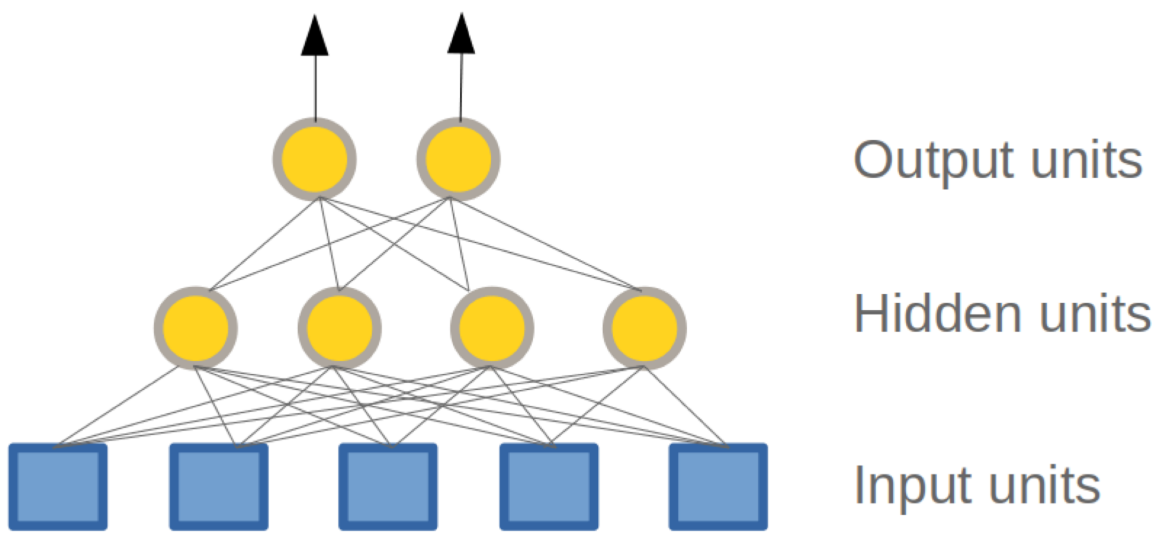
$$f(z) \begin{cases} 1 \text{ if } z > 0 \\ 0 \text{ if } z \leq 0 \end{cases}$$

$$\phi(z) = \boxed{\frac{1}{1+\exp(-z)}}$$

Nice property $\boxed{\dfrac{\partial \phi(z)}{\partial z}} = \boxed{\phi(z)}\boxed{(1-\phi(z))}$

Multi-Layer Perceptron

- Artificial Neural Network (ANN)
- Hierarchy of neurons
- Input layer, hidden layers, output layer
- 2 Layers = all continuous functions
- 3+ Layers = all functions



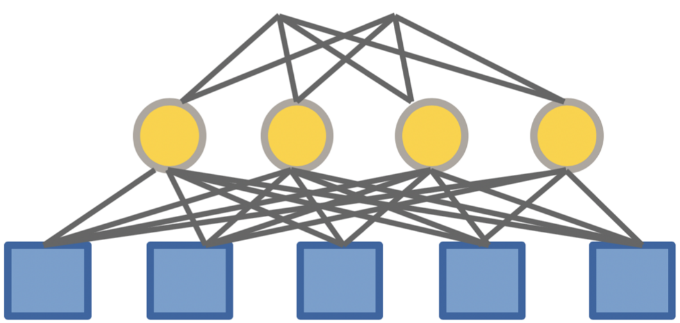
Matrix Representation for Layer

- Each layer can have multiple nodes
- Our output equation only calculated one node

$$a = \phi(\mathbf{w}^T\mathbf{x} + b)$$

- Change equation to matrix notation

$$A = \phi(W X + \mathbf{b})$$



A = activations
W = weights
X = inputs
b = vector of biases

Summary

- Nonlinear activation functions
- ANN: multiple layers of neurons
- Popular activation function: sigmoid
- Sigmoid has a simple derivative
- We can represent each layer in matrix notation