Neural Network Basics

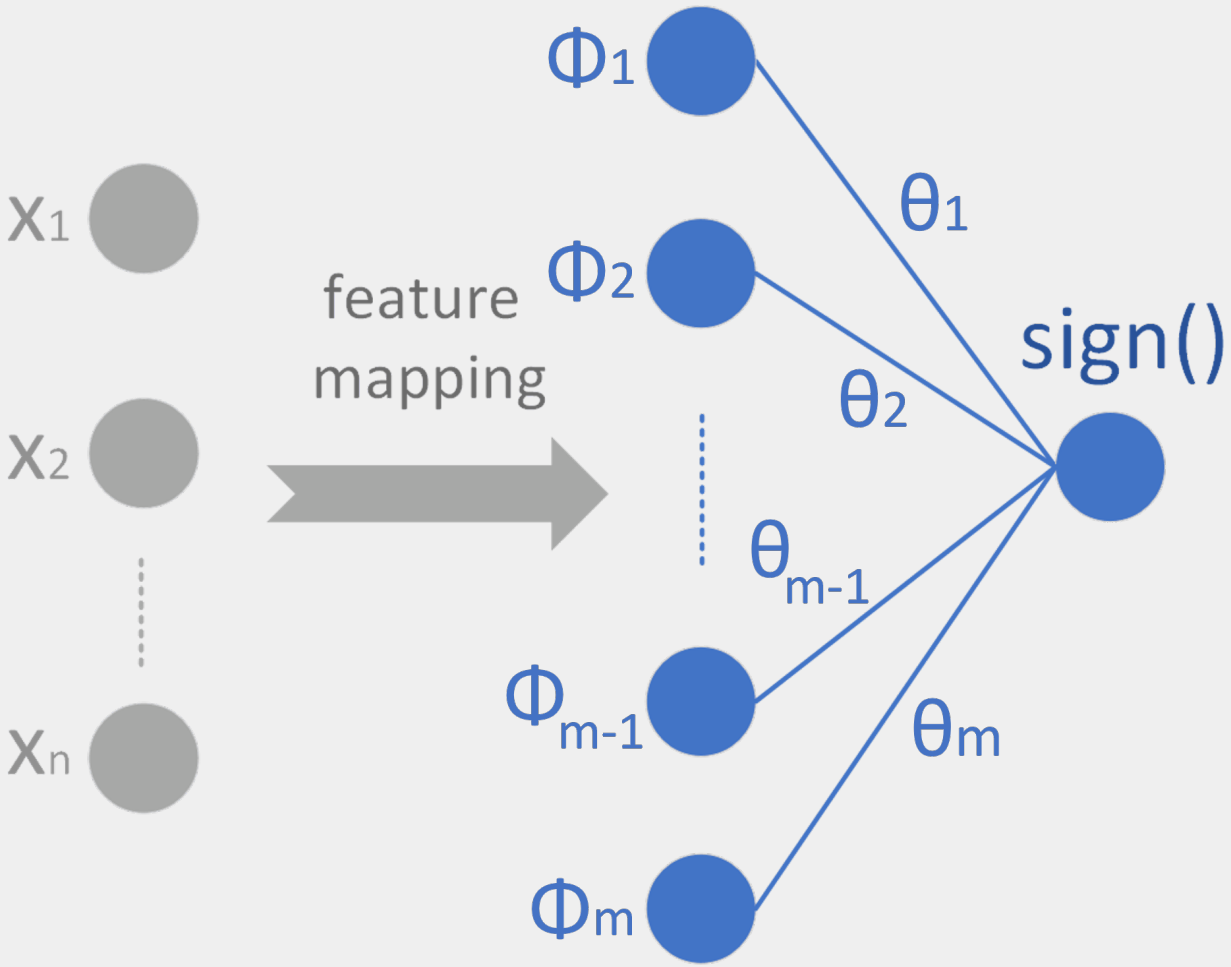
Weizi Li

Department of Computer Science
University of Memphis

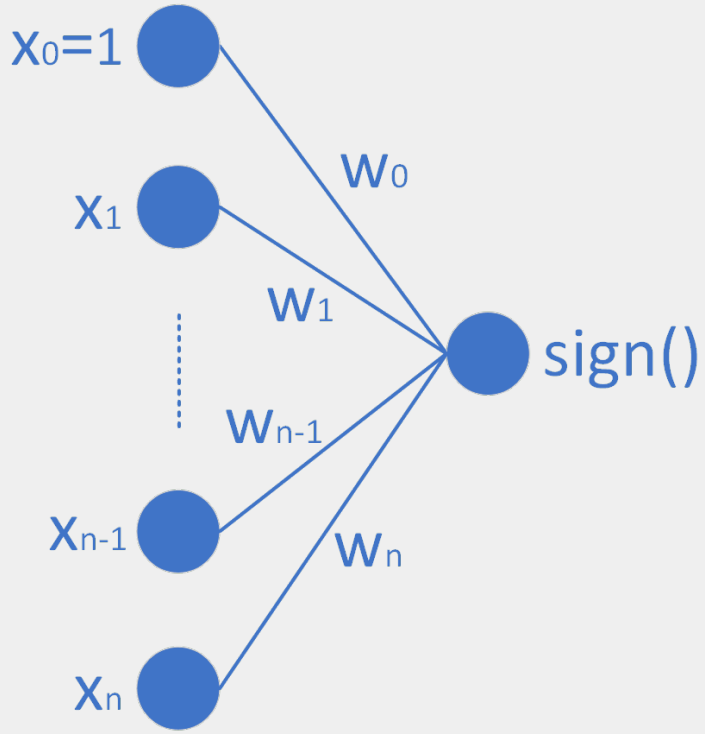
- Recall non-linear classifier via feature mapping:

$$h(x; \theta, \theta_0) = sign(\theta \phi(x))$$

- $x = [x_1, x_2, \ldots, x_n]$
- $\phi(x) = [\phi_1(x), \phi_2(x), \ldots, \phi_m(x)]$
- $\theta = [\theta_1, \theta_2, \ldots, \theta_m]$


x_1
x_2
x_n
feature mapping
Φ_1
Φ_2
Φ_{m-1}
Φ_m
θ_1
θ_2
θ_{m-1}
θ_m
sign()

- x can be raw data but more commonly, features extracted from raw data



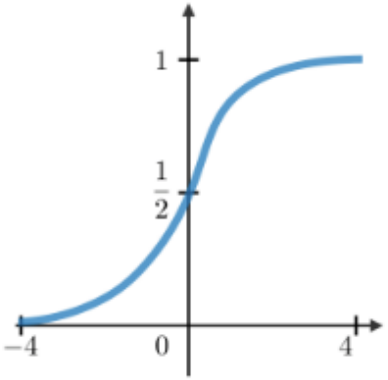
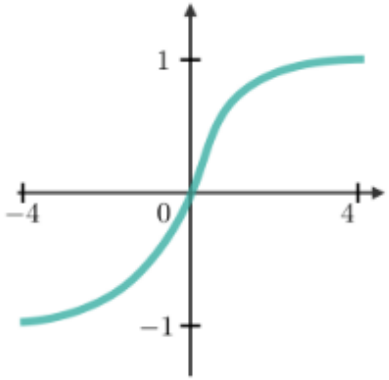
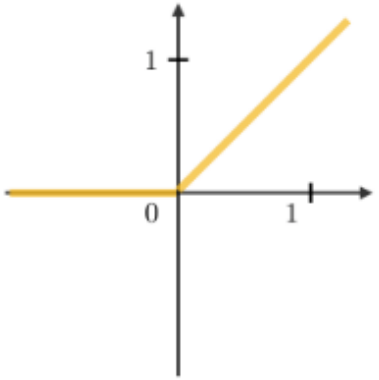
- If $\sum_{i=0}^{n} x_i w_i > 0$, x is $+$, otherwise x is $-$

- A simple neural network that delivers a linear classifier for binary classification
- Guarantee to solve any binary classification task if data are linearly separable

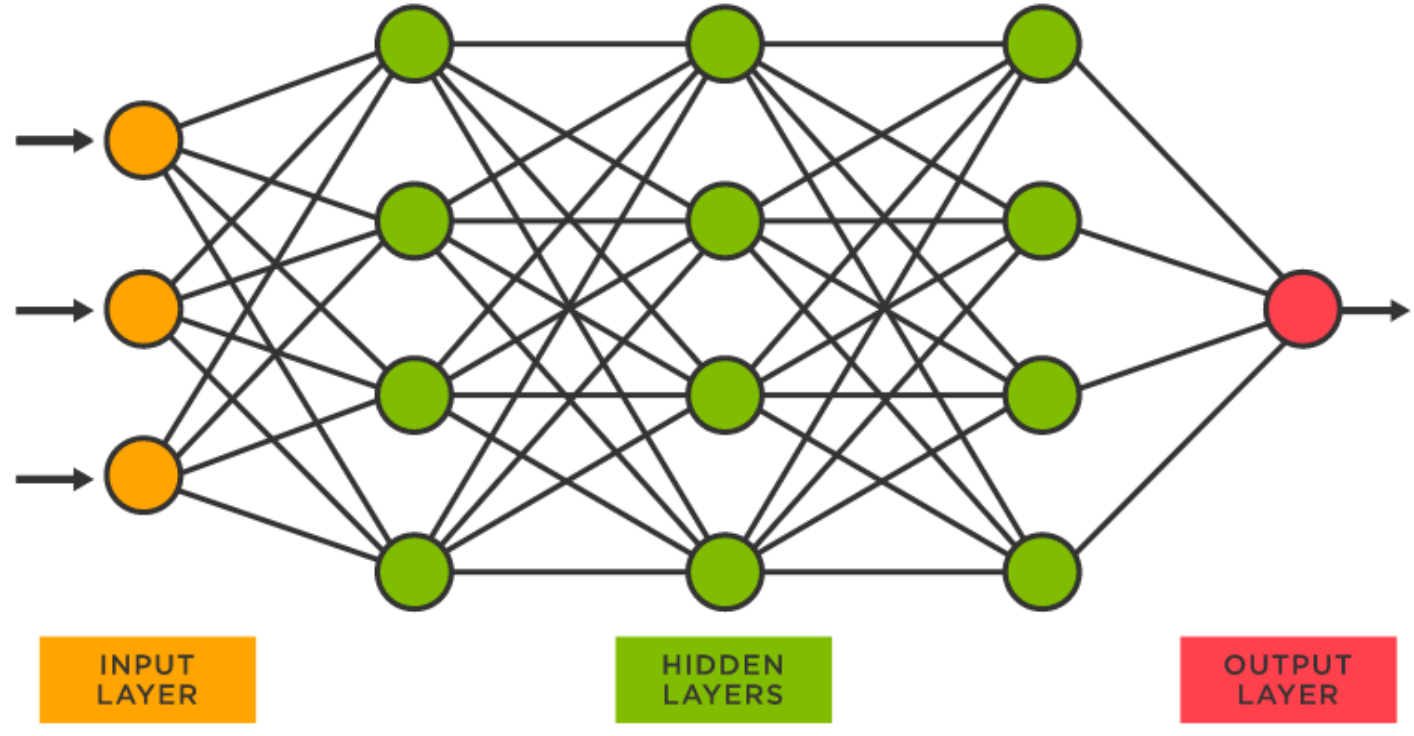
Algorithm 1 Perceptron Learning Algorithm

Require: $\{(X^{(i)}, y^{(i)})\}_{i=1,\ldots,n}$; T
  for $t = 1, \ldots, T$ **do**
    for $i = 1, \ldots, n$ **do**
      if $y^{(i)}(wX^{(i)} + w_0) \leq 0$ **then**
        $w_0 = w_0 + y^{(i)}$
        $w = w + y^{(i)} X^{(i)}$
      end if
    end for
  end for
  return w_0, w

- Use a non-linear activation function instead of simple $sign()$
- $g(z = \sum_{i=0}^{n} x_i w_i)$: Sigmoid, Tanh, ReLU, etc.

Sigmoid	Tanh	RELU
$g(z) = \frac{1}{1+e^{-z}}$	$g(z) = \frac{e^z - e^{-z}}{e^z + e^{-z}}$	$g(z) = \max(0, z)$
		

- Neural networks with multiple hidden layers; backbone for "deep learning"



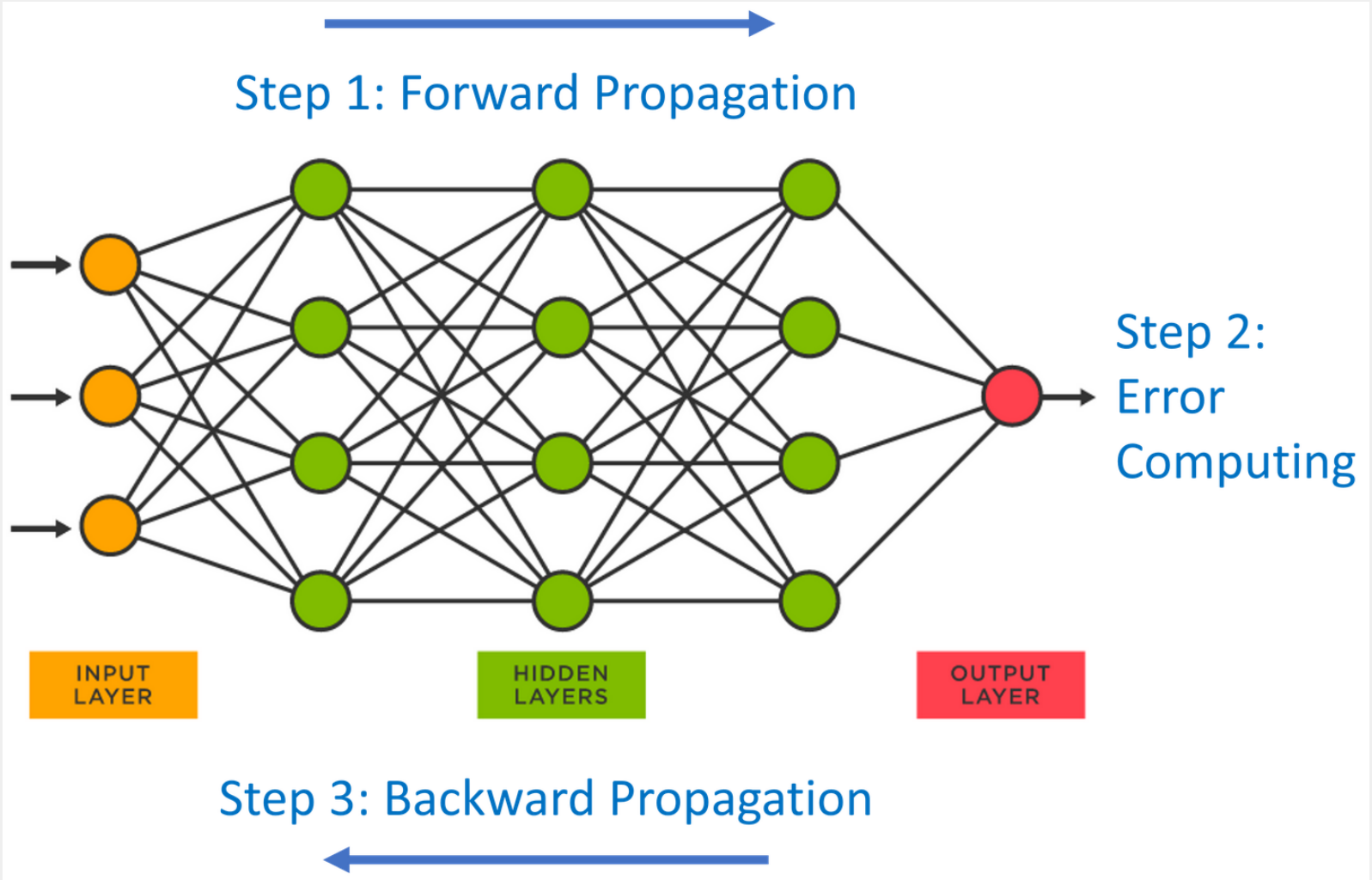
- Number of hidden layers
- Number of units of each layer
- Activation function in the units (usually the same function is used for the entire network)
- Ideal architecture is guided by monitoring validation error

- Playground (link)

- Fully-connected: all units of a layer connect to all units of the next layer
- Convolutional neural networks are not fully connected

- Feedforward: connections among units do not form a cycle
  - Fully-connected, feedforward neural networks are also called multilayer perceptron (MLP)
- Recurrent: connections among units form a cycle

- The concepts of “multiple layers”, “connected”, and “activation function” are loosely from neuroscience
- Modern neural networks are guided by mathematical and engineering principles
- Better to think modern neural networks as function approximators to achieve statistical generalizability


Step 1: Forward Propagation
Step 2:
Error
Computing
INPUT
LAYER
HIDDEN
LAYERS
OUTPUT
LAYER
Step 3: Backward Propagation