

ENGR 4350: Applied Deep Learning

Neural Network: Part 2

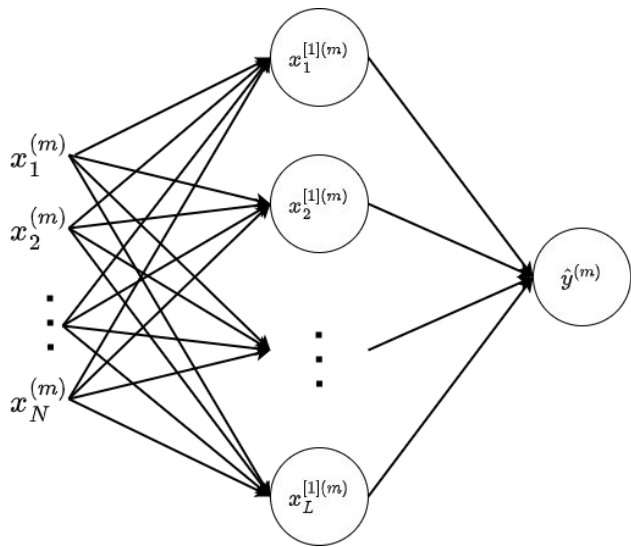
09/21/2022



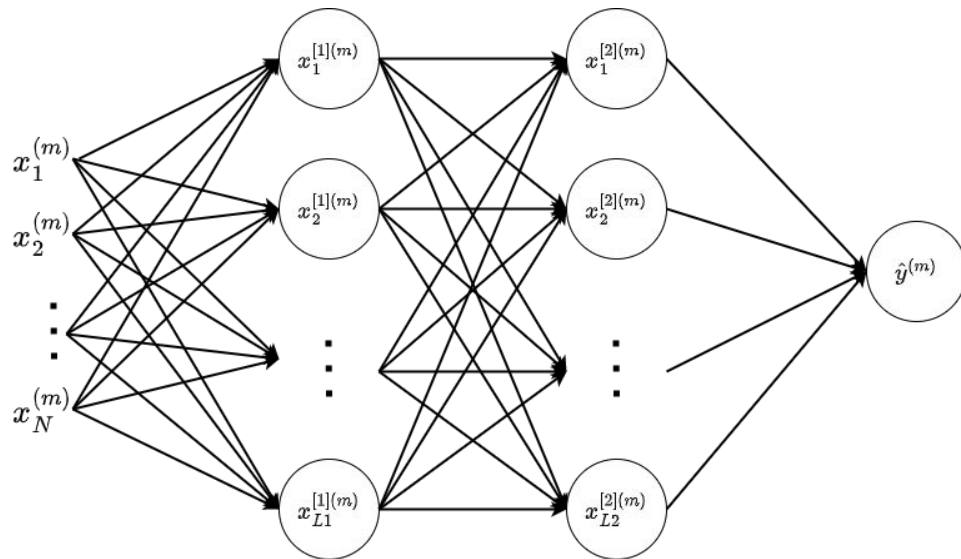
Outline

- Multi-layer Neural Network
- Forward & Backward Propagation

Multi-Layer Neural Network

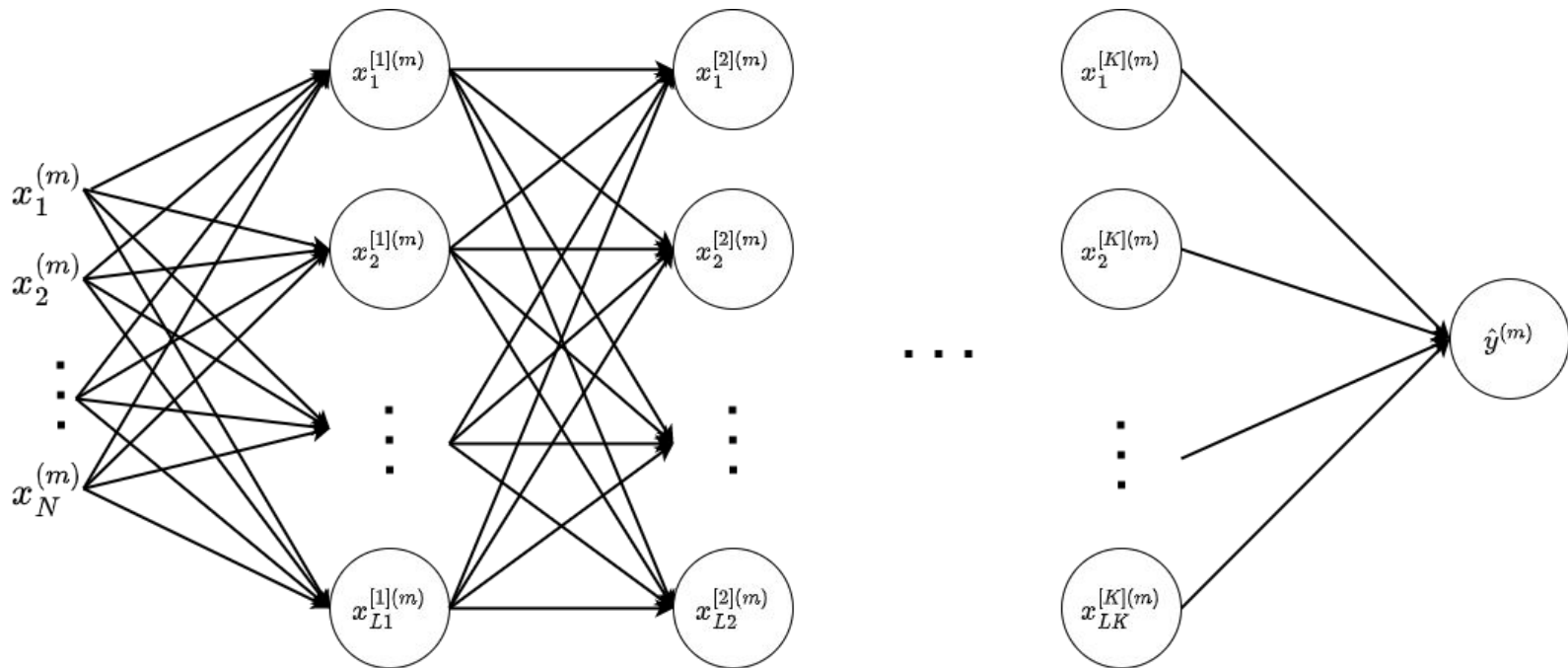


2-layer Neural Network



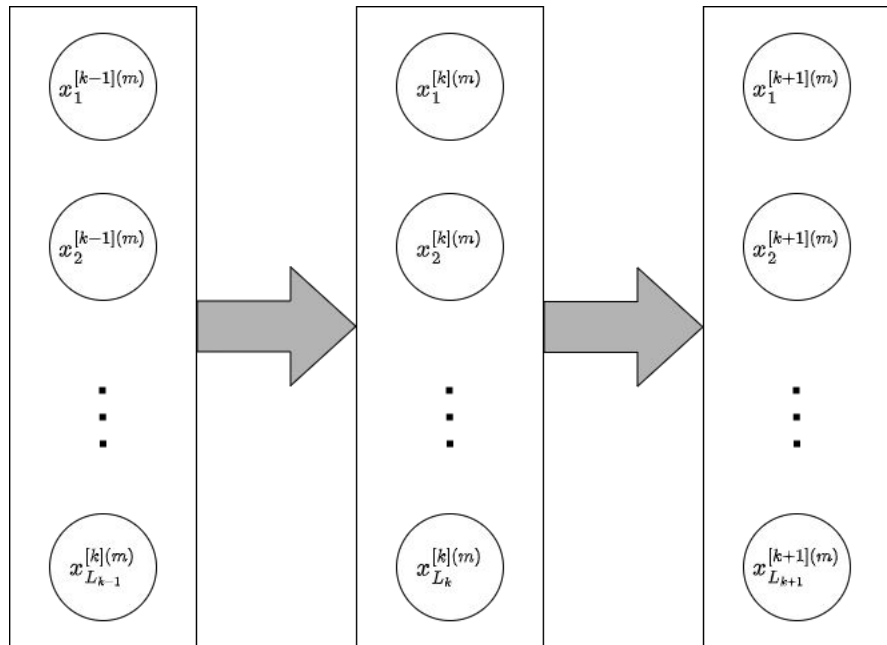
3-layer Neural Network

Multi-Layer Neural Network



K-layer Neural Network

View of a General Layer



Forward Propagation

$$\mathbf{X}^{[k]} = g\left(\mathbf{X}^{[k-1]} \cdot \mathbf{W}^{[k]} + \mathbf{b}^{[k]}\right)$$

Backward Propagation

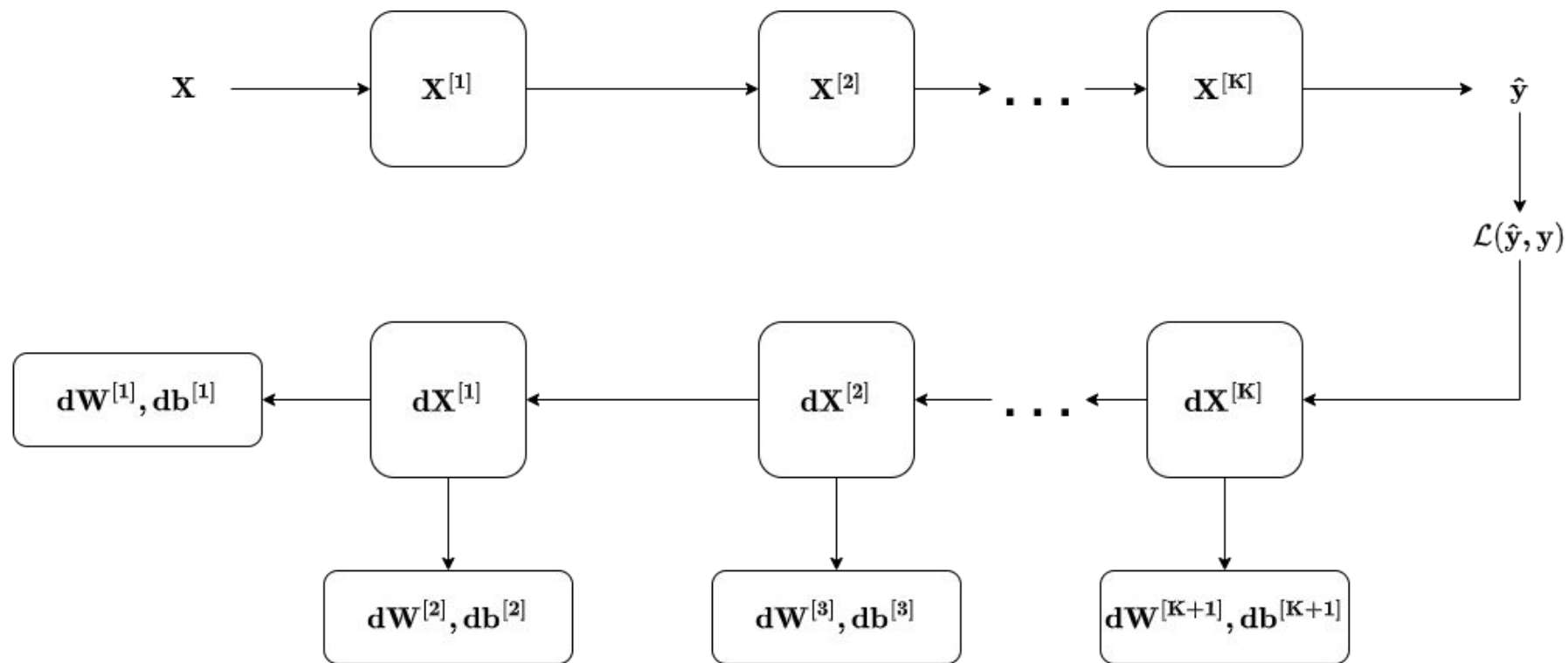
$$d\mathbf{Z}^{[k]} = d\mathbf{X}^{[k]} * \mathbf{g}'(\mathbf{Z}^{[k]})$$

$$d\mathbf{W}^{[k]} = \frac{1}{M} \mathbf{X}^{[k-1]\mathbf{T}} \cdot d\mathbf{Z}^{[k]}$$

$$d\mathbf{b}^{[k]} = \frac{1}{M} \sum_{m=1}^M d\mathbf{Z}^{[k]}$$

$$d\mathbf{X}^{[k-1]} = d\mathbf{Z}^{[k]} \cdot \mathbf{W}^{[k]\mathbf{T}}$$

Forward/Backward Propagation



Forward Propagation

For layer $k=1$ to K

$$\mathbf{X}^{[k]} = g_k\left(\mathbf{X}^{[k-1]} \cdot \mathbf{W}^{[k]} + \mathbf{b}^{[k]}\right)$$

where, input: $\mathbf{X} = \mathbf{X}^{[0]}$, output: $\hat{\mathbf{y}} = \mathbf{X}^{[K]}$

$$J\left(\mathbf{W}^{[1]}, \dots, \mathbf{W}^{[K]}, \mathbf{b}^{[1]}, \dots, \mathbf{b}^{[K]}\right) = \frac{1}{M} \sum_{m=1}^M \mathcal{L}(\hat{\mathbf{y}}, \mathbf{y})$$

Backward Propagation

Set $d\mathbf{X}^{[K]} = \frac{\partial J}{\partial \hat{\mathbf{y}}}$

For layer $k=K$ to 1

$$d\mathbf{Z}^{[k]} = \frac{\partial J}{\partial \mathbf{X}^{[k]}} \cdot \frac{\partial \mathbf{X}^{[k]}}{\partial \mathbf{Z}^{[k]}} = d\mathbf{X}^{[k]} * \mathbf{g}'(\mathbf{Z}^{[k]})$$

$$d\mathbf{W}^{[k]} = \frac{\partial J}{\partial \mathbf{Z}^{[k]}} \cdot \frac{\partial \mathbf{Z}^{[k]}}{\partial \mathbf{W}^{[k]}} = \frac{1}{M} \mathbf{X}^{[k-1]\mathbf{T}} \cdot d\mathbf{Z}^{[k]}$$

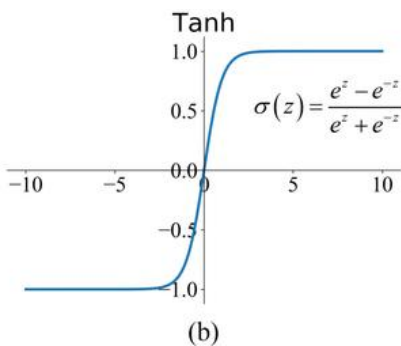
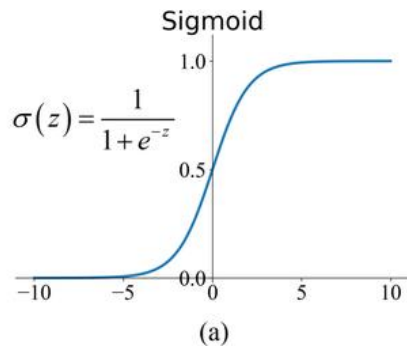
$$d\mathbf{b}^{[k]} = \frac{\partial J}{\partial \mathbf{Z}^{[k]}} \cdot \frac{\partial \mathbf{Z}^{[k]}}{\partial \mathbf{b}^{[k]}} = \frac{1}{M} \sum_1^M d\mathbf{Z}^{[k]\mathbf{T}}, \text{ axis}=0, \text{ keepdims}=\text{True}$$

$$d\mathbf{X}^{[k-1]} = \frac{\partial J}{\partial \mathbf{Z}^{[k]}} \cdot \frac{\partial \mathbf{Z}^{[k]}}{\partial \mathbf{X}^{[k-1]}} = d\mathbf{Z}^{[k]} \cdot \mathbf{W}^{[k]\mathbf{T}}$$

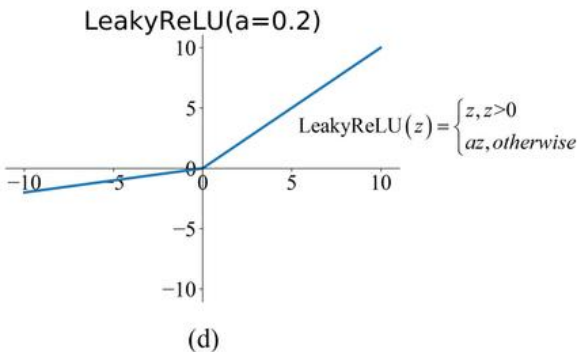
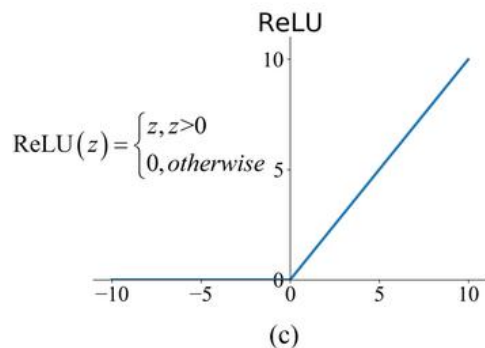
$$d\mathbf{X}^{[k]} \leftarrow d\mathbf{X}^{[k-1]}$$

Derivatives of Activation Functions

$$\sigma'(z) = \sigma(z)(1 - \sigma(z))$$



$$\sigma'(z) = 1 - \sigma^2(z)$$



$$\text{ReLU}'(z) = \begin{cases} 1, & z > 0 \\ 0, & \text{otherwise} \end{cases}$$

$$\text{LeakyReLU}'(z) = \begin{cases} 1, & z > 0 \\ a, & \text{otherwise} \end{cases}$$