ENGR 4350:Applied Deep Learning

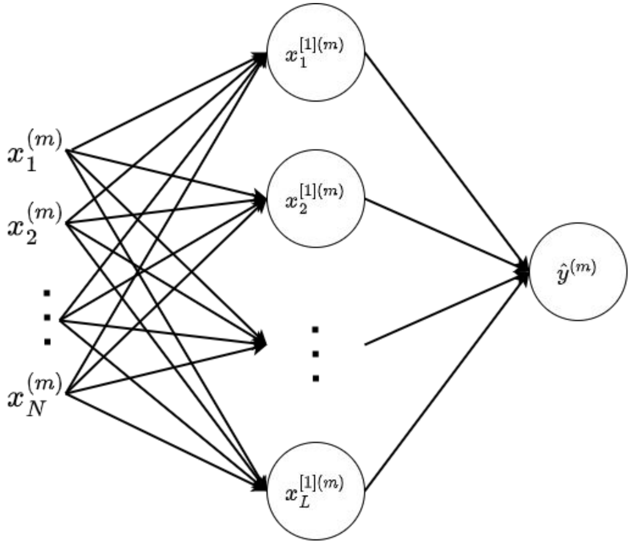
Neural Network: Part 2

09/21/2022

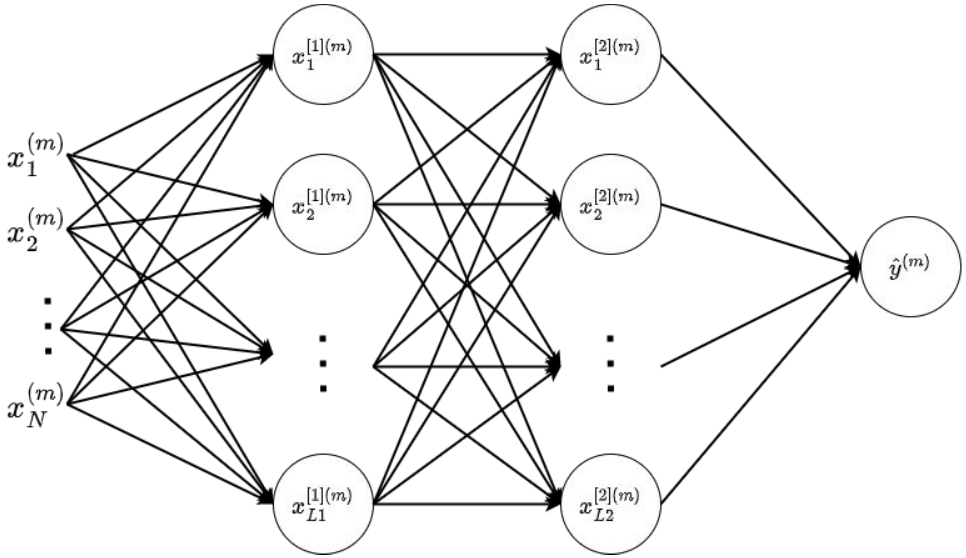
Outline

- Multi-layer Neural Network
- Forward & Backward Propagation

Multi-Layer Neural Network

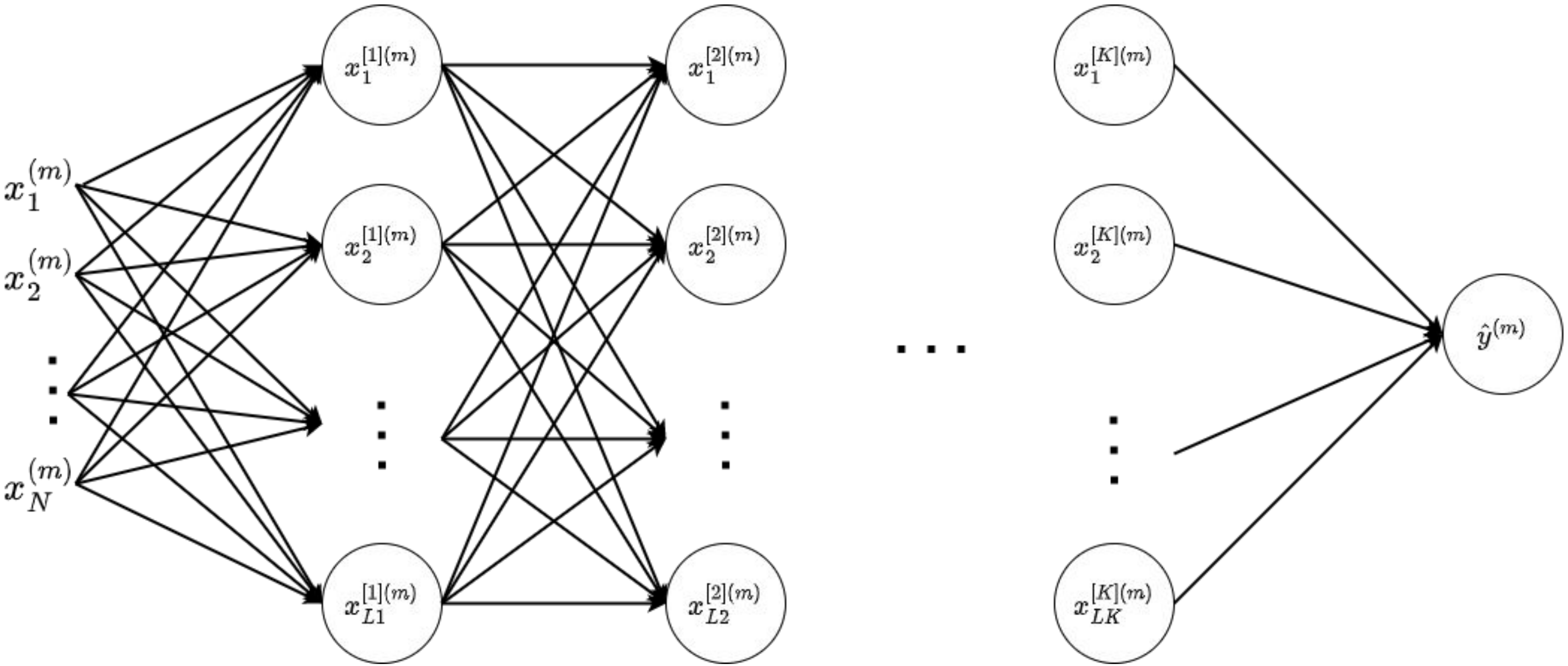


2-layer Neural Network



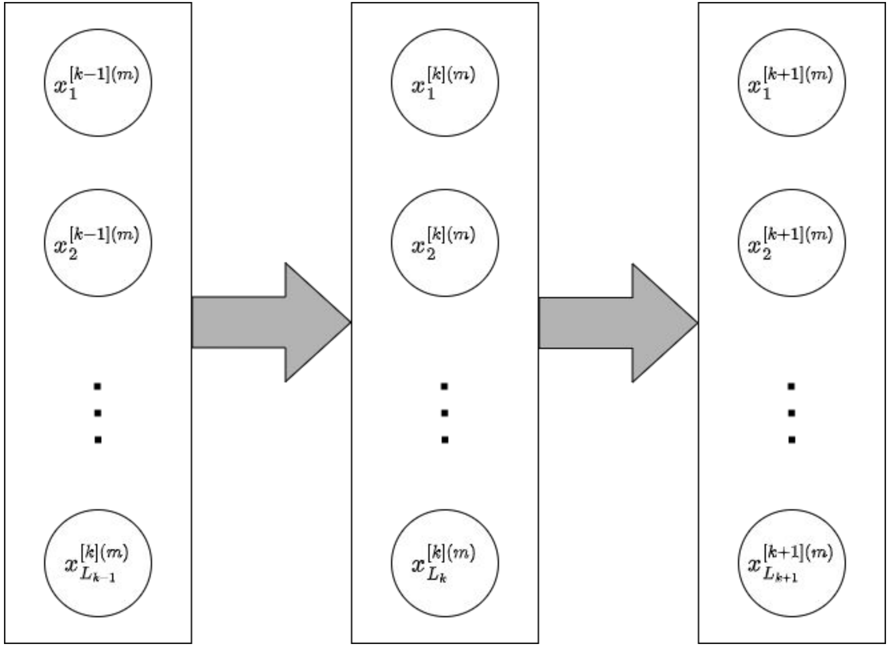
3-layer Neural Network

Multi-Layer Neural Network



K-layer Neural Network

View of a General Layer



Forward Propagation

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

Backward Propagation

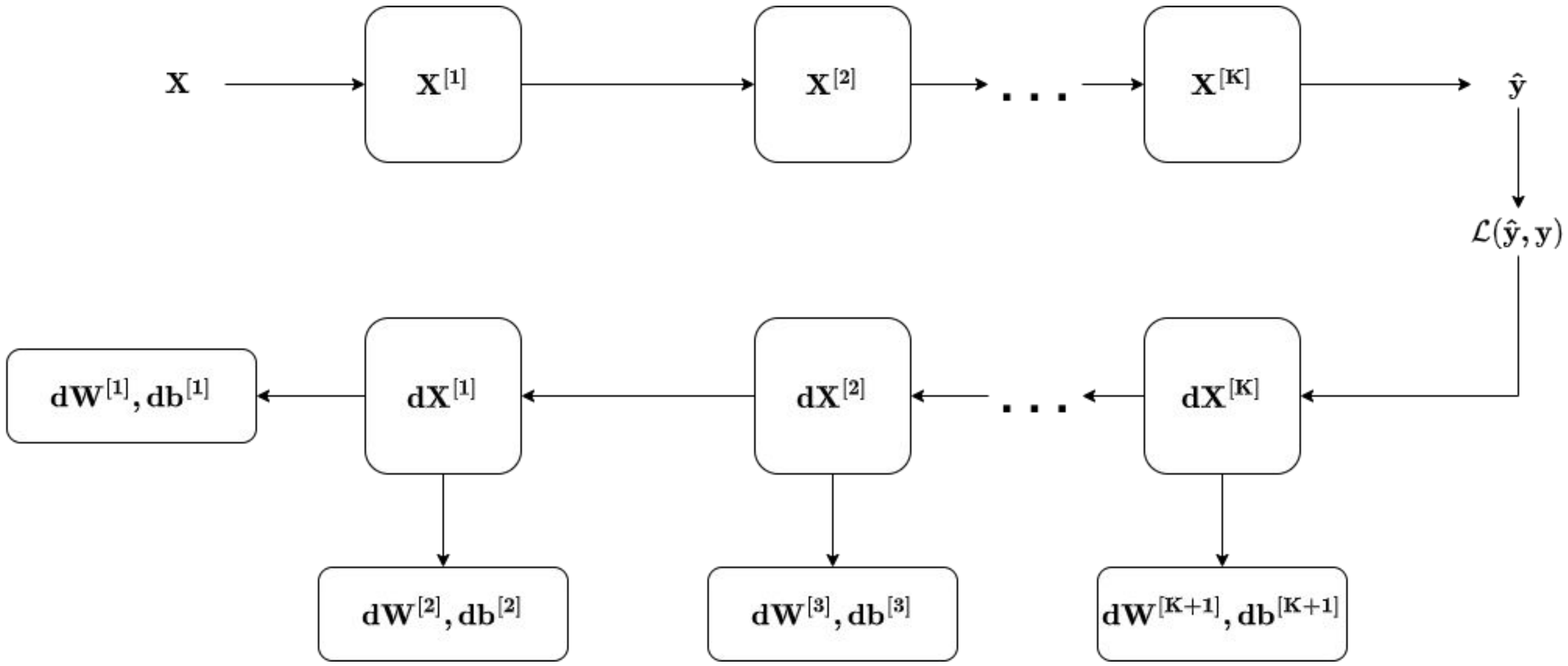
$$d\mathbf{Z}^{[\mathrm{k}]} = d\mathbf{X}^{[\mathrm{k}]} * \mathbf{g}'\left(\mathbf{Z}^{[\mathrm{k}]}\right)$$

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

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

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

Forward/Backward Propagation



Forward Propagation

For layer k=1 to K

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

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

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

Backward Propagation

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

For layer k=K to 1

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

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

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

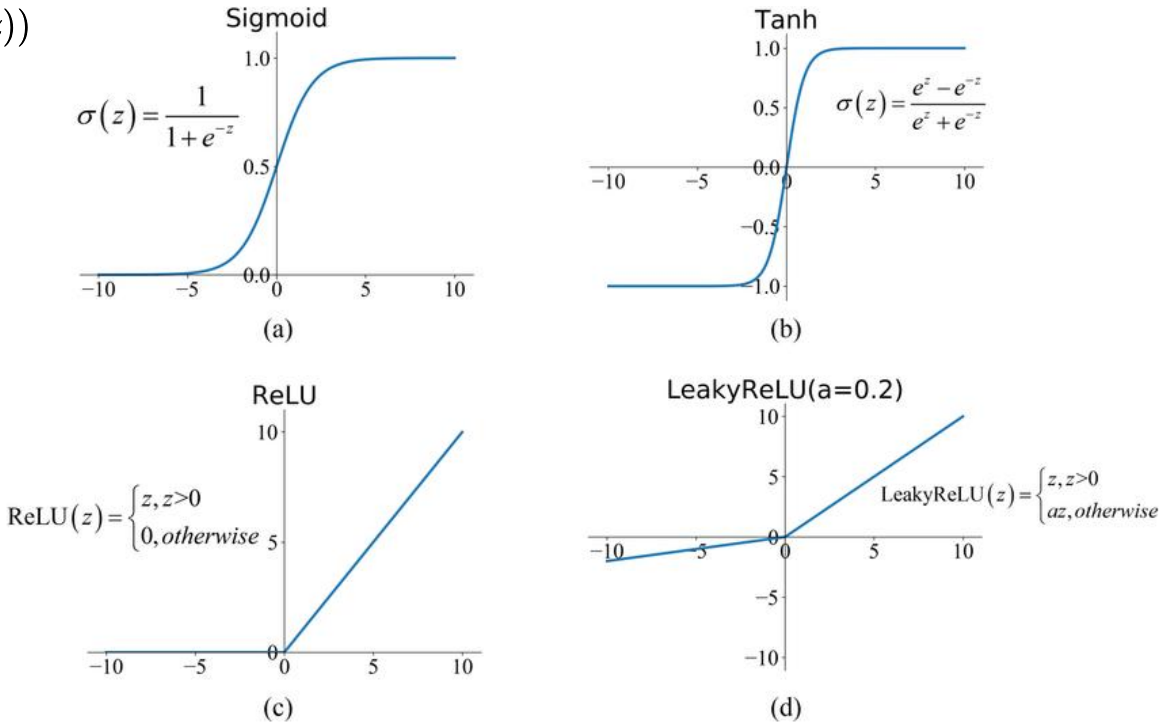
$$d\mathbf{X}^{[\mathrm{k}-1]} = \frac{\partial J}{\partial \mathbf{Z}^{[\mathrm{k}]}} \cdot \frac{\partial \mathbf{Z}^{[\mathrm{k}]}}{\partial \mathbf{X}^{[\mathrm{k}-1]}} = d\mathbf{Z}^{[\mathrm{k}]} \cdot \mathbf{W}^{[\mathrm{k}]\mathbf{T}}$$

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

Derivatives of Activation Functions

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

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



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

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