

ENGR 4350: Applied Deep Learning

Neural Network: Part 1

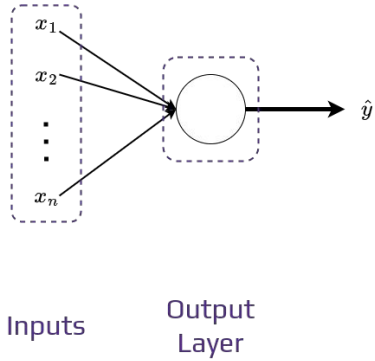
09/14/2022



Outline

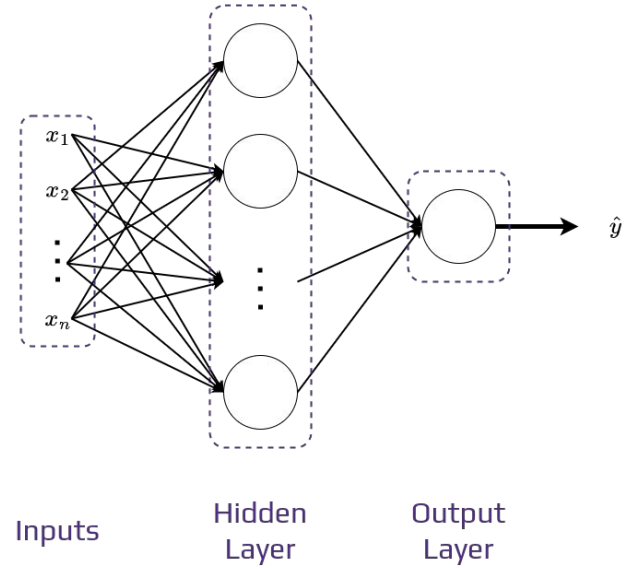
- Neural Network Representation
- Forward & Backward Propagation
- Activation Functions

Neural Network Representation



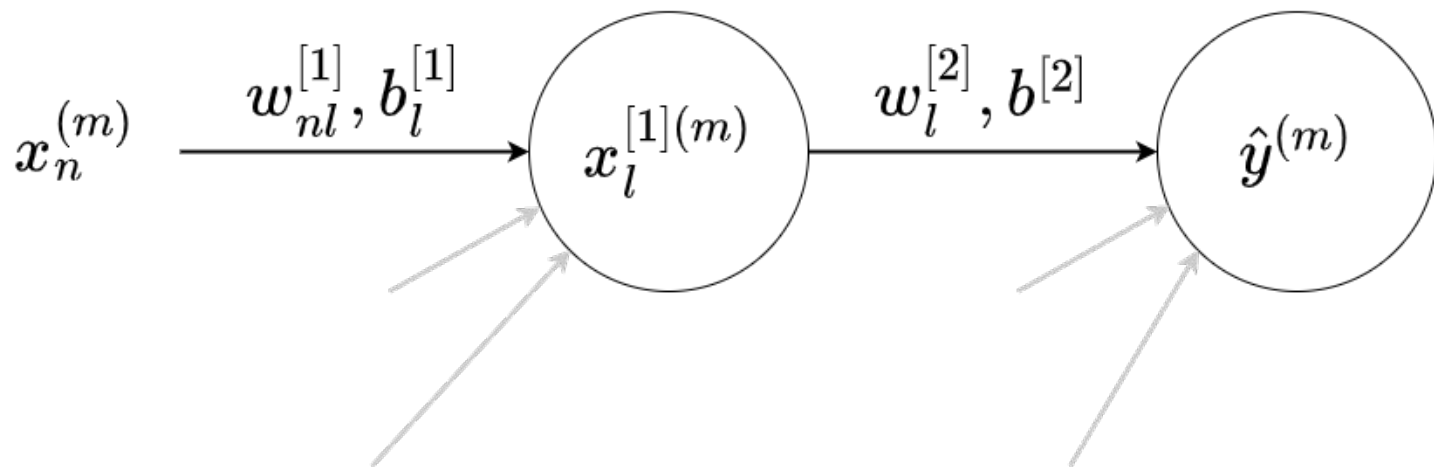
1-Layer neural network, w/o hidden layer

(Logistic regression model)



2-Layer neural network w/ 1 hidden layer

Neural Network Representation



$l \in \{1, \dots, L\}$: neuron index

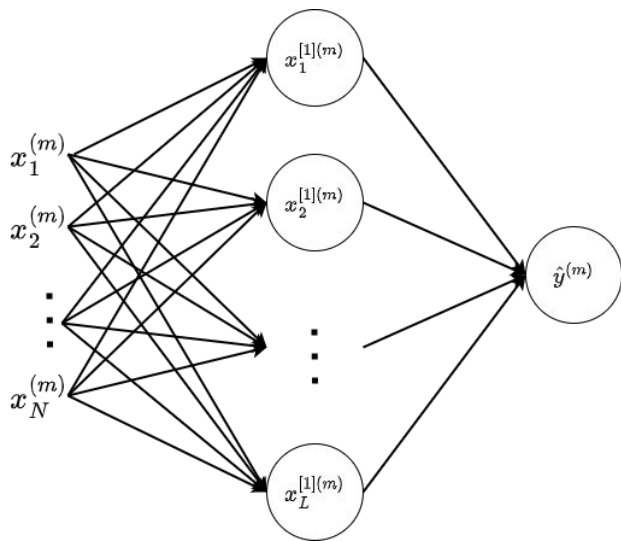
$m \in \{1, \dots, M\}$: example index

$n \in \{1, \dots, N\}$: feature index

[1]: 1st layer

[2]: 2nd layer

Neural Network Representation



$$x_1^{[1](m)} = \sigma\left(w_{11}^{[1]}x_1^{(m)} + w_{21}^{[1]}x_2^{(m)} + \dots + w_{N1}^{[1]}x_N^{(m)} + b_1^{[1]}\right)$$

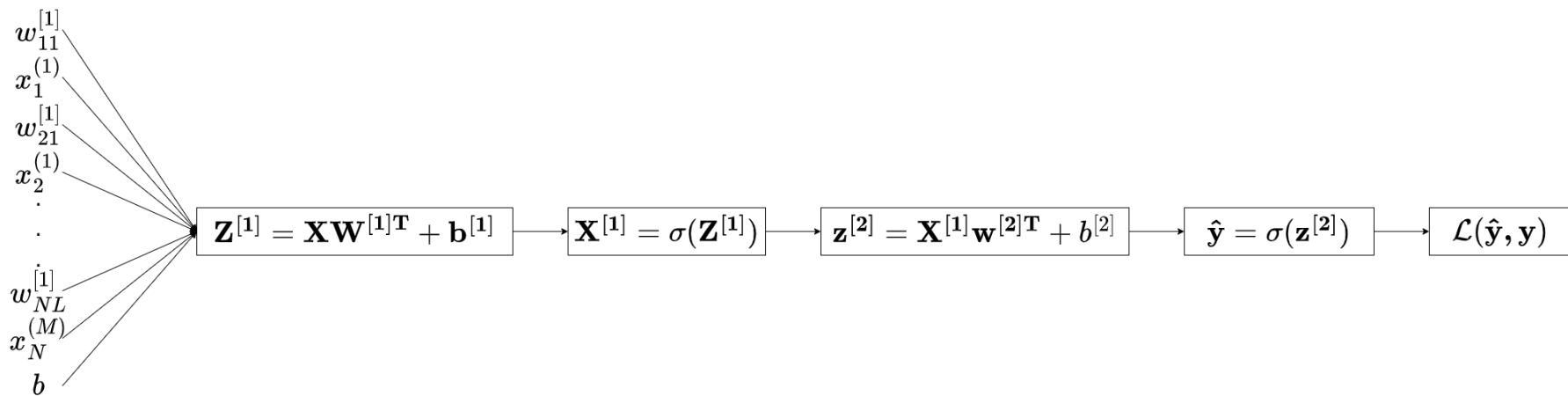
$$x_2^{[1](m)} = \sigma\left(w_{12}^{[1]}x_1^{(m)} + w_{22}^{[1]}x_2^{(m)} + \dots + w_{N2}^{[1]}x_N^{(m)} + b_2^{[1]}\right)$$

⋮

$$x_L^{[1](m)} = \sigma\left(w_{1L}^{[1]}x_1^{(m)} + w_{2L}^{[1]}x_2^{(m)} + \dots + w_{NL}^{[1]}x_N^{(m)} + b_L^{[1]}\right)$$

$$\hat{y}^{(m)} = x^{[2](m)} = \sigma\left(w_1^{[2]}x_1^{[1](m)} + w_2^{[2]}x_2^{[1](m)} + \dots + w_L^{[2]}x_L^{[1](m)} + b^{[2]}\right)$$

Neural Network Representation



$$\mathbf{X} : M \times N \quad \mathbf{W}^{[1]} : L \times N \quad \mathbf{b}^{[1]} : 1 \times L \quad \mathbf{Z}^{[1]} : M \times L$$

$$\mathbf{X}^{[1]} : M \times L \quad \mathbf{w}^{[2]} : 1 \times L \quad \mathbf{b}^{[2]} : 1 \times 1 \quad \mathbf{z}^{[2]} : M \times 1 \quad \hat{y} : M \times 1$$

$$\mathbf{y} : M \times 1$$

Forward Propagation

$$\mathbf{X}^{[1]} = \sigma(\mathbf{X}\mathbf{W}^{[1]} + \mathbf{b}^{[1]})$$

$$\sigma \left(\begin{bmatrix} x_1^{(1)} & x_2^{(1)} & \dots & x_N^{(1)} \\ x_1^{(2)} & x_2^{(2)} & \dots & x_N^{(2)} \\ \vdots & \vdots & \ddots & \vdots \\ x_1^{(M)} & x_2^{(M)} & \dots & x_N^{(M)} \end{bmatrix} \cdot \begin{bmatrix} w_{11}^{[1]} & w_{12}^{[1]} & \dots & w_{1L}^{[1]} \\ w_{21}^{[1]} & w_{22}^{[1]} & \dots & w_{2L}^{[1]} \\ \vdots & \vdots & \ddots & \vdots \\ w_{N1}^{[1]} & w_{N2}^{[1]} & \dots & w_{NL}^{[1]} \end{bmatrix} + \begin{bmatrix} b_1^{[1]} & b_2^{[1]} & \dots & b_L^{[1]} \\ b_1^{[1]} & b_2^{[1]} & \dots & b_L^{[1]} \\ \vdots & \vdots & \ddots & \vdots \\ b_1^{[1]} & b_2^{[1]} & \dots & b_L^{[1]} \end{bmatrix} \right) = \begin{bmatrix} x_1^{1} & x_2^{1} & \dots & x_L^{1} \\ x_1^{[1](2)} & x_2^{[1](2)} & \dots & x_L^{[1](2)} \\ \vdots & \vdots & \ddots & \vdots \\ x_1^{[1](M)} & x_2^{[1](M)} & \dots & x_L^{[1](M)} \end{bmatrix}$$

$$\hat{\mathbf{y}} = \mathbf{x}^{[2]} = \sigma(\mathbf{X}^{[1]}\mathbf{w}^{[2]\mathbf{T}} + \mathbf{b}^{[2]})$$

$$\sigma \left(\begin{bmatrix} x_1^{1} & x_2^{1} & \dots & x_L^{1} \\ x_1^{[1](2)} & x_2^{[1](2)} & \dots & x_L^{[1](2)} \\ \vdots & \vdots & \ddots & \vdots \\ x_1^{[1](M)} & x_2^{[1](M)} & \dots & x_L^{[1](M)} \end{bmatrix} \cdot \begin{bmatrix} w_1^{[2]} \\ w_2^{[2]} \\ \vdots \\ w_L^{[2]} \end{bmatrix} + \begin{bmatrix} b^{[2]} \\ b^{[2]} \\ \vdots \\ b^{[2]} \end{bmatrix} \right) = \begin{bmatrix} \hat{y}^{(1)} \\ \hat{y}^{(2)} \\ \vdots \\ \hat{y}^{(M)} \end{bmatrix}$$

$$\mathcal{L}(\hat{\mathbf{y}}, \mathbf{y}) = -(\mathbf{y} \log \hat{\mathbf{y}} + (1 - \mathbf{y}) \log (1 - \hat{\mathbf{y}}))$$

$$J(\mathbf{W}^{[1]}, \mathbf{b}^{[1]}, \mathbf{w}^{[2]}, \mathbf{b}^{[2]}) = \frac{1}{M} \sum_{m=1}^M \mathcal{L}(\hat{\mathbf{y}}, \mathbf{y})$$

Backward Propagation

$$d\mathbf{z}^{[2]} = \frac{\partial J}{\partial \mathbf{z}^{[2]}} = \hat{\mathbf{y}} - \mathbf{y}$$

$$d\mathbf{w}^{[2]} = d\mathbf{z}^{[2]} \cdot \frac{\partial \mathbf{z}^{[2]}}{\partial \mathbf{w}^{[2]}} = \frac{1}{M} (\hat{\mathbf{y}} - \mathbf{y})^T \cdot \mathbf{X}^{[1]}$$

$$d\mathbf{b}^{[2]} = d\mathbf{z}^{[2]} \cdot \frac{\partial \mathbf{z}^{[2]}}{\partial \mathbf{b}^{[2]}} = \frac{1}{M} \sum_{m=1}^M (\hat{\mathbf{y}} - \mathbf{y})^T$$

$$d\mathbf{X}^{[1]} = d\mathbf{z}^{[2]} \cdot \frac{\partial \mathbf{z}^{[2]}}{\partial \mathbf{X}^{[1]}} = (\hat{\mathbf{y}} - \mathbf{y}) \cdot \mathbf{w}^{[2]}$$

$$d\mathbf{Z}^{[1]} = d\mathbf{X}^{[1]} \cdot \frac{\partial \mathbf{X}^{[1]}}{\partial \mathbf{Z}^{[1]}} = [(\hat{\mathbf{y}} - \mathbf{y}) \cdot \mathbf{w}^{[2]}] * \mathbf{X}^{[1]} * (\mathbf{1} - \mathbf{X}^{[1]})$$

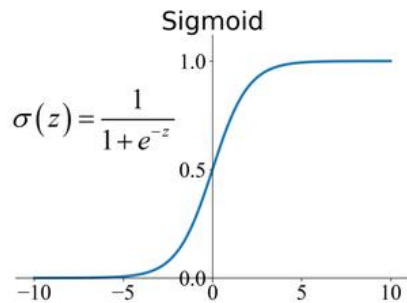
$$d\mathbf{W}^{[1]} = d\mathbf{Z}^{[1]} \cdot \frac{\partial \mathbf{Z}^{[1]}}{\partial \mathbf{W}^{[1]}} = \frac{1}{M} [(\hat{\mathbf{y}} - \mathbf{y}) \cdot \mathbf{w}^{[2]} * \mathbf{X}^{[1]} * (\mathbf{1} - \mathbf{X}^{[1]})]^T \cdot \mathbf{X}$$

$$d\mathbf{b}^{[1]} = d\mathbf{Z}^{[1]} \cdot \frac{\partial \mathbf{Z}^{[1]}}{\partial \mathbf{b}^{[1]}} = \frac{1}{M} \sum_{m=1}^M [(\hat{\mathbf{y}} - \mathbf{y}) \cdot \mathbf{w}^{[2]} * \mathbf{X}^{[1]} * (\mathbf{1} - \mathbf{X}^{[1]})]$$

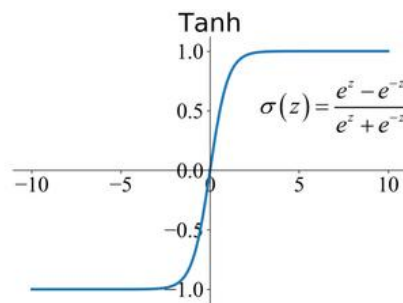
* : element-wise multiplication

Activation Functions

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

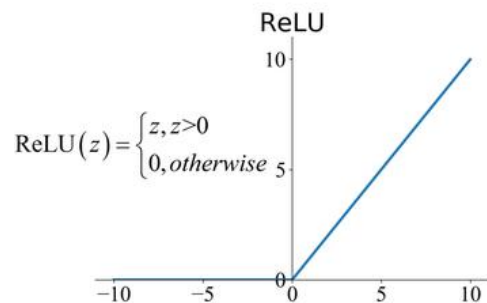


(a)

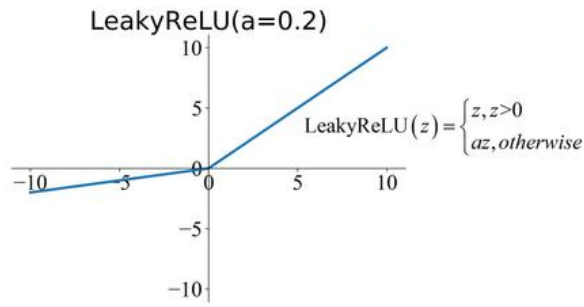


(b)

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



(c)



(d)

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

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

Backward Propagation

w/ any activation function

$$d\mathbf{z}^{[2]} = \frac{\partial J}{\partial \mathbf{z}^{[2]}} = \hat{\mathbf{y}} - \mathbf{y}$$

$$d\mathbf{w}^{[2]} = d\mathbf{z}^{[2]} \cdot \frac{\partial \mathbf{z}^{[2]}}{\partial \mathbf{w}^{[2]}} = \frac{1}{M} (\hat{\mathbf{y}} - \mathbf{y})^T \cdot \mathbf{X}^{[1]}$$

$$d\mathbf{b}^{[2]} = d\mathbf{z}^{[2]} \cdot \frac{\partial \mathbf{z}^{[2]}}{\partial \mathbf{b}^{[2]}} = \frac{1}{M} \sum_{m=1}^M (\hat{\mathbf{y}} - \mathbf{y})$$

$$d\mathbf{X}^{[1]} = d\mathbf{z}^{[2]} \cdot \frac{\partial \mathbf{z}^{[2]}}{\partial \mathbf{X}^{[1]}} = (\hat{\mathbf{y}} - \mathbf{y}) \cdot \mathbf{w}^{[2]}$$

$$d\mathbf{Z}^{[1]} = d\mathbf{X}^{[1]} \cdot \frac{\partial \mathbf{X}^{[1]}}{\partial \mathbf{Z}^{[1]}} = [(\hat{\mathbf{y}} - \mathbf{y}) \cdot \mathbf{w}^{[2]}] * \mathbf{g}'(\mathbf{Z}^{[1]})$$

$$d\mathbf{W}^{[1]} = d\mathbf{Z}^{[1]} \cdot \frac{\partial \mathbf{Z}^{[1]}}{\partial \mathbf{W}^{[1]}} = \frac{1}{M} [(\hat{\mathbf{y}} - \mathbf{y}) \cdot \mathbf{w}^{[2]} * \mathbf{g}'(\mathbf{Z}^{[1]})]^T \cdot \mathbf{X}$$

$$d\mathbf{b}^{[1]} = d\mathbf{Z}^{[1]} \cdot \frac{\partial \mathbf{Z}^{[1]}}{\partial \mathbf{b}^{[1]}} = \frac{1}{M} \sum_{m=1}^M [(\hat{\mathbf{y}} - \mathbf{y}) \cdot \mathbf{w}^{[2]} * \mathbf{g}'(\mathbf{Z}^{[1]})]$$