ENGR 4350: Applied Deep Learning

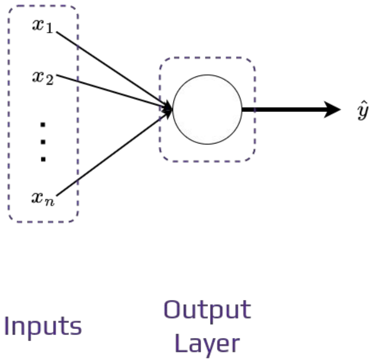
Neural Network: Part 1

09/14/2022

Outline

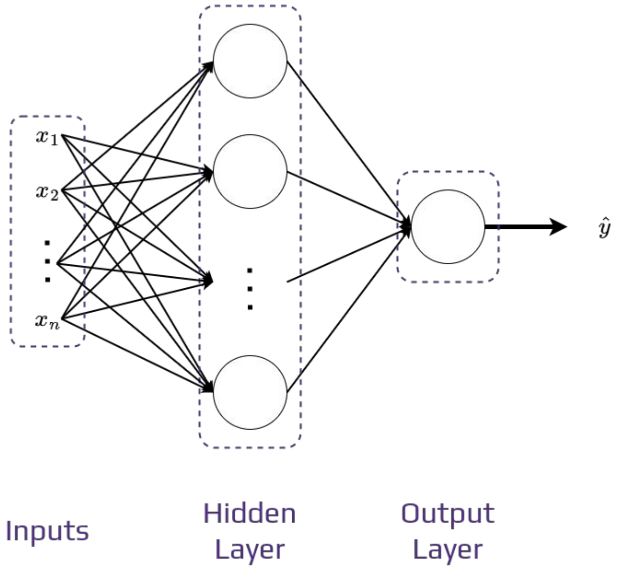
- Neural Network Representation
- Forward & Backward Propagation
- Activation Functions

Neural Network Representation



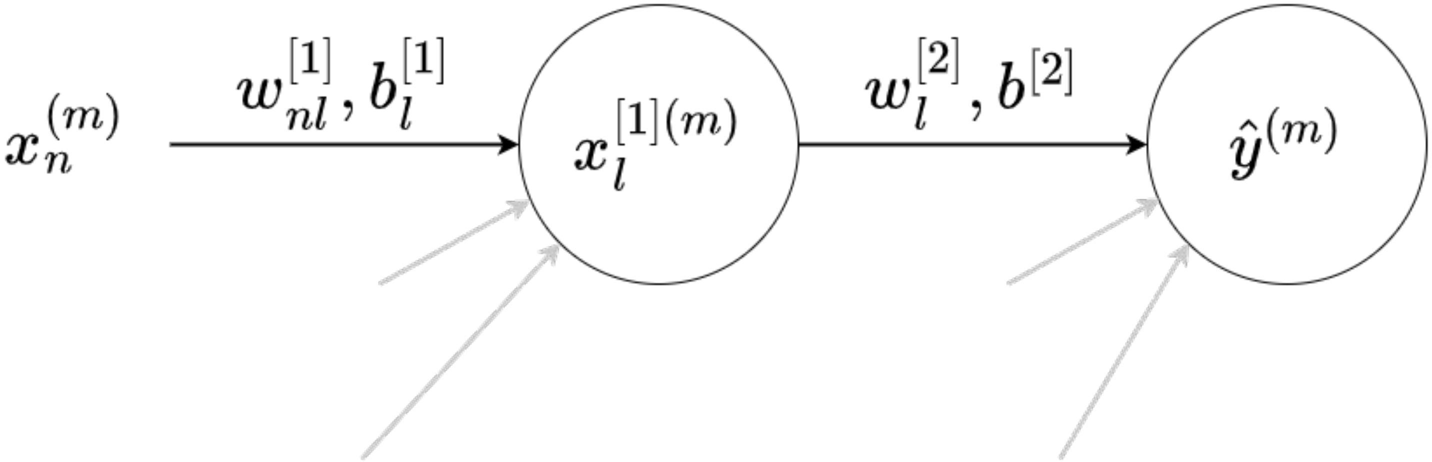
1-Layer neural network, w/o hidden layer

(Logistic regression model)



2-Layer neural network w/ 1 hidden layer

Neural Network Representation



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

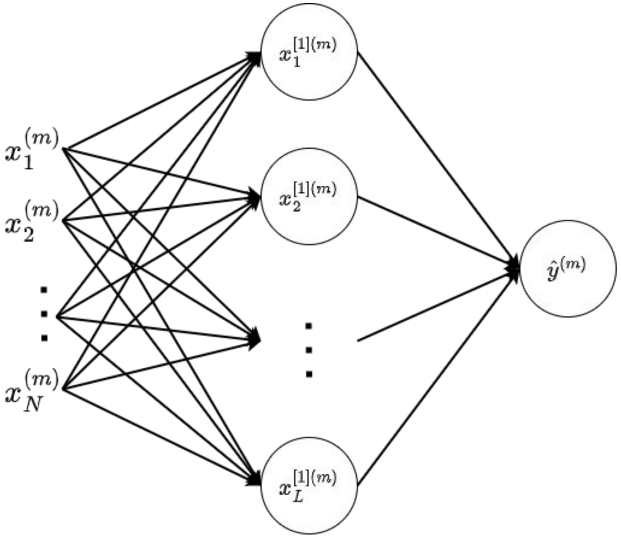
$m \in \{1, \ldots, M\}$: example index

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

$[1]$: 1st layer

$[2]$: 2nd layer

Neural Network Representation



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

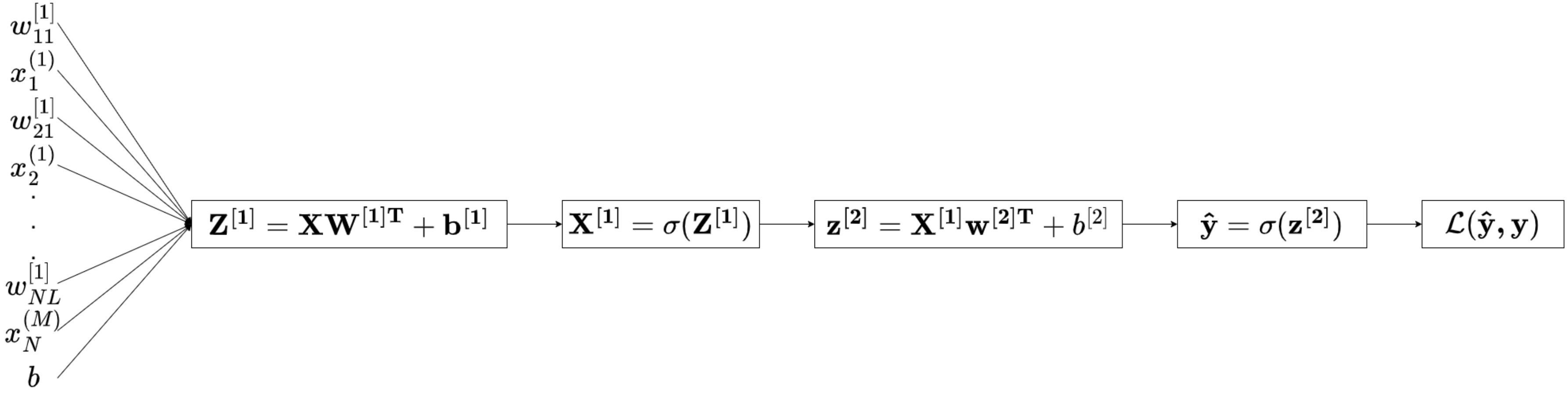
$$x_2^{[1](m)} = \sigma\left(w_{12}^{[1]}x_1^{(m)} + w_{22}^{[1]}x_2^{(m)} + \ldots + w_{N2}^{[1]}x_N^{(m)} + b_2^{[1]}\right)$$

$$\vdots$$

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

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

Neural Network Representation



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

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

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

Forward Propagation

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

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

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

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

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

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

Backward Propagation

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

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

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

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

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

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

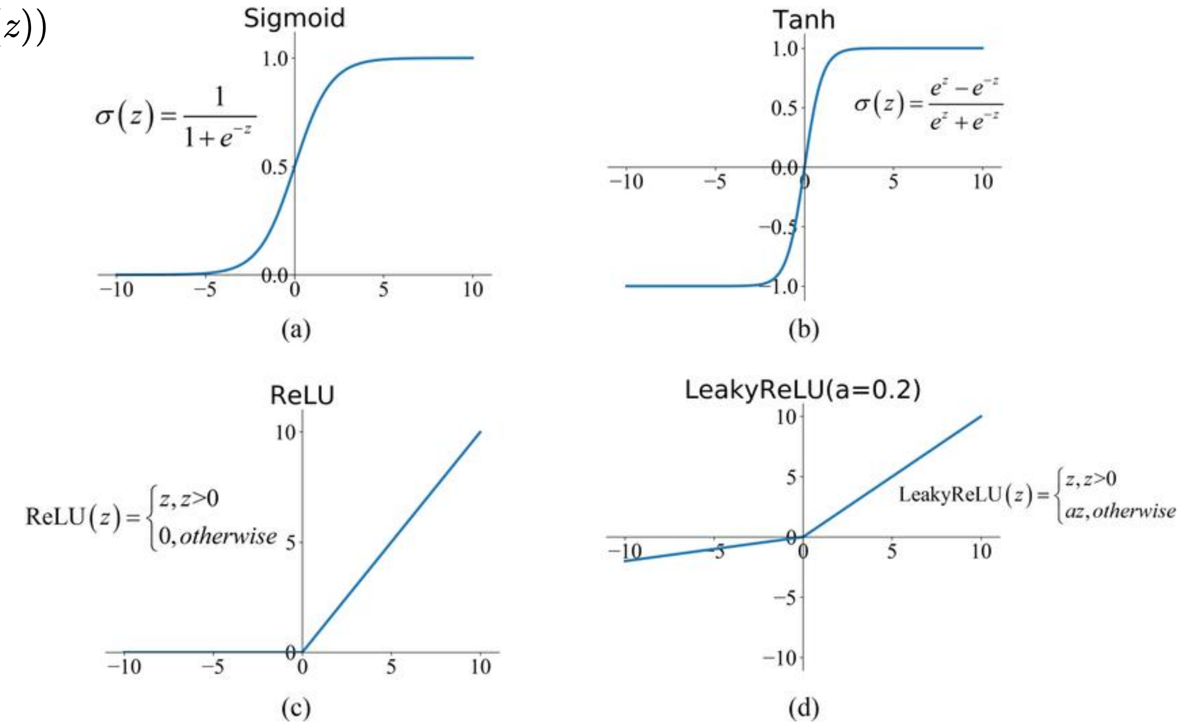
$$d\mathbf{b}^{[1]} = d\mathbf{Z}^{[1]} \cdot \frac{\partial \mathbf{Z}^{[1]}}{\partial \mathbf{b}^{[1]}} = \frac{1}{M} \sum_{m=1}^{M} \left[\left(\hat{\mathbf{y}} - \mathbf{y}\right) \cdot \mathbf{w}^{[\mathbf{2}]} * \mathbf{X}^{[1]} * \left(\mathbf{1} - \mathbf{X}^{[1]}\right)\right]$$

$*$: element-wise multiplication

Activation Functions

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

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



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

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

Backward Propagation

w/ any activation function

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

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

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

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

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

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

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