

Fully Connected Neural Networks

Roberto Souza
Electrical and Software Engineering
Schulich School of Engineering

W2026

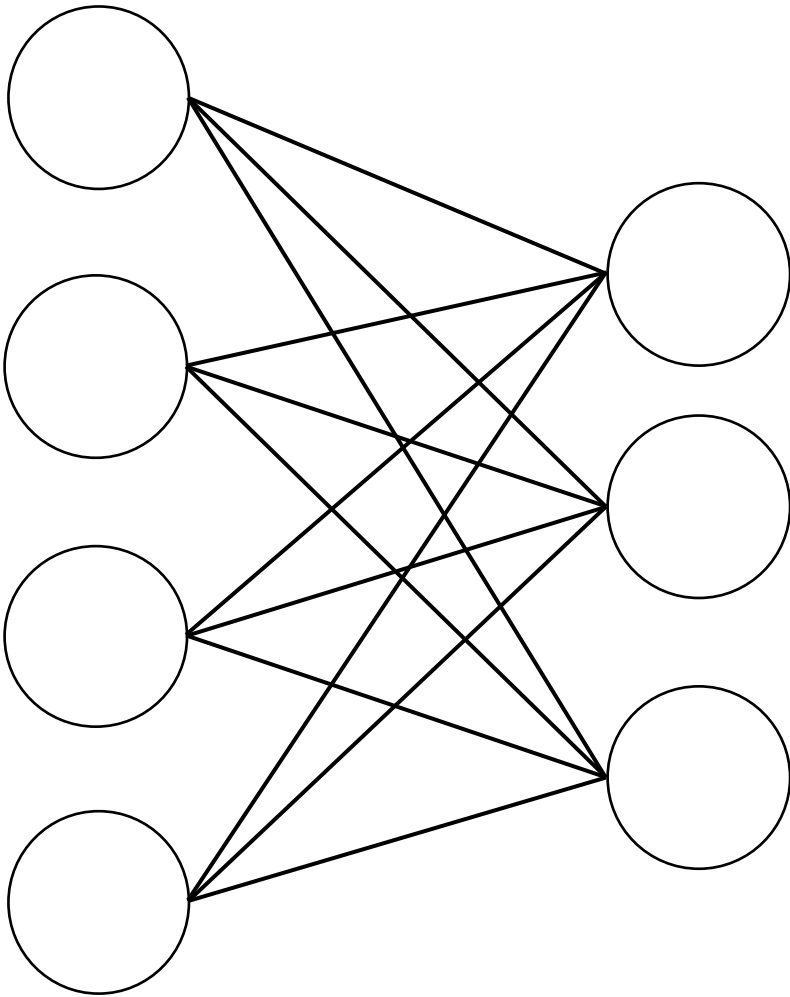
Outline

- Learning Goals
- Fully Connected Neural Networks
 - Single-layer Network
 - Multi-layer Network
- Summary
- Tutorial

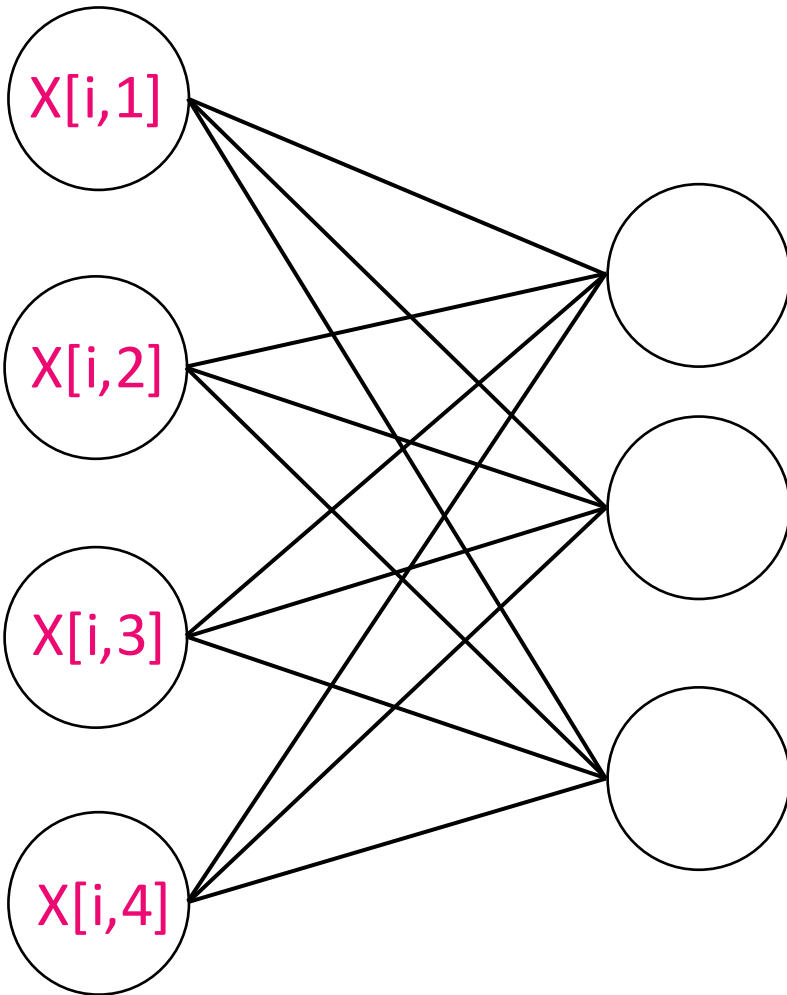
Learning Goals

- Introduce fully connected neural networks
- Learn how to compute the number of parameters of your model

Fully connected neural networks (FCNN)



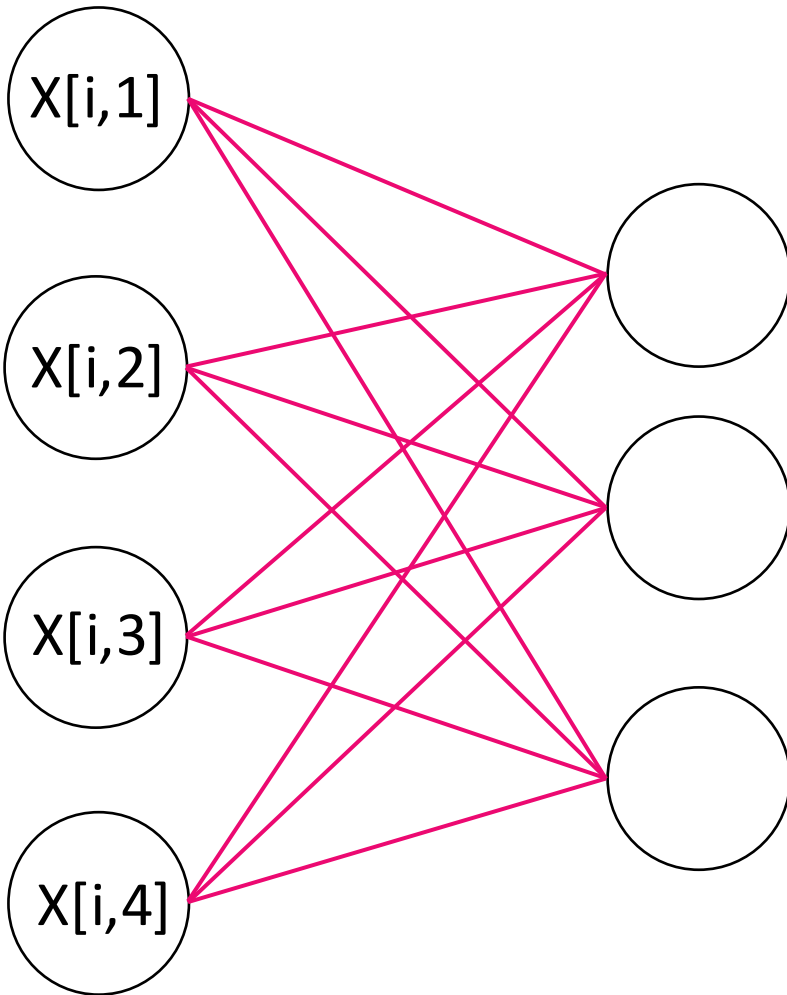
Inputs



$$X = \begin{bmatrix} x_{11} & x_{12} & \dots & x_{1M} \\ x_{21} & x_{22} & \dots & x_{2M} \\ \dots & \dots & \dots & \dots \\ x_{N1} & x_{N2} & \dots & x_{NM} \end{bmatrix}$$

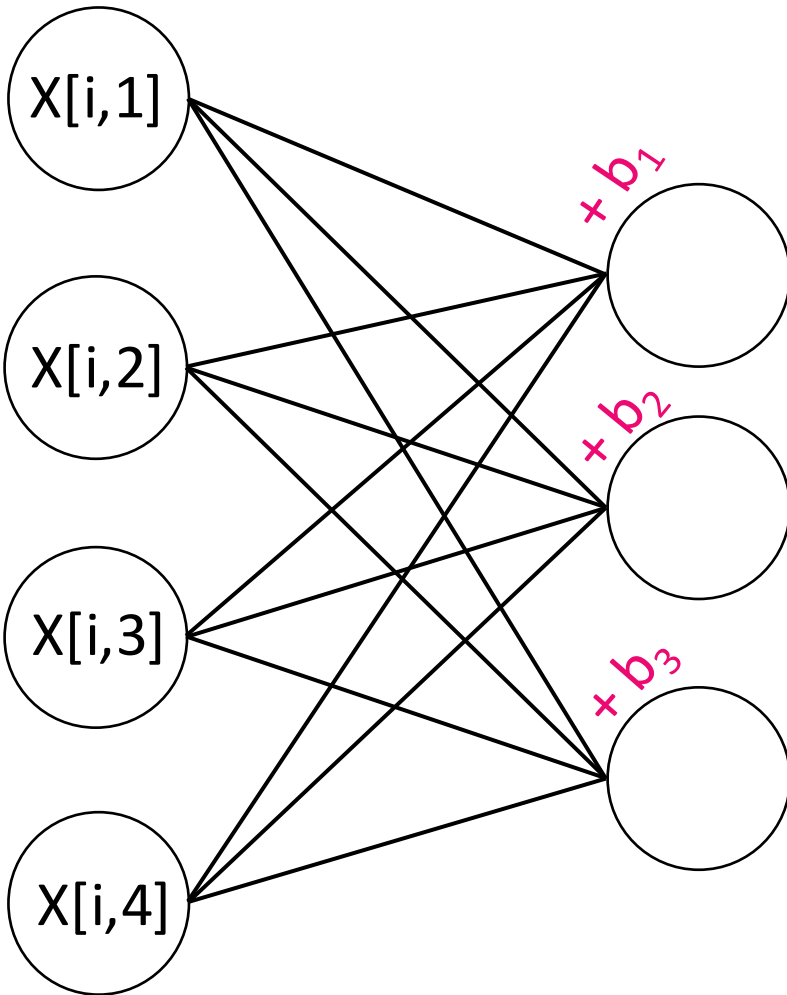
N samples with M
features

Weights



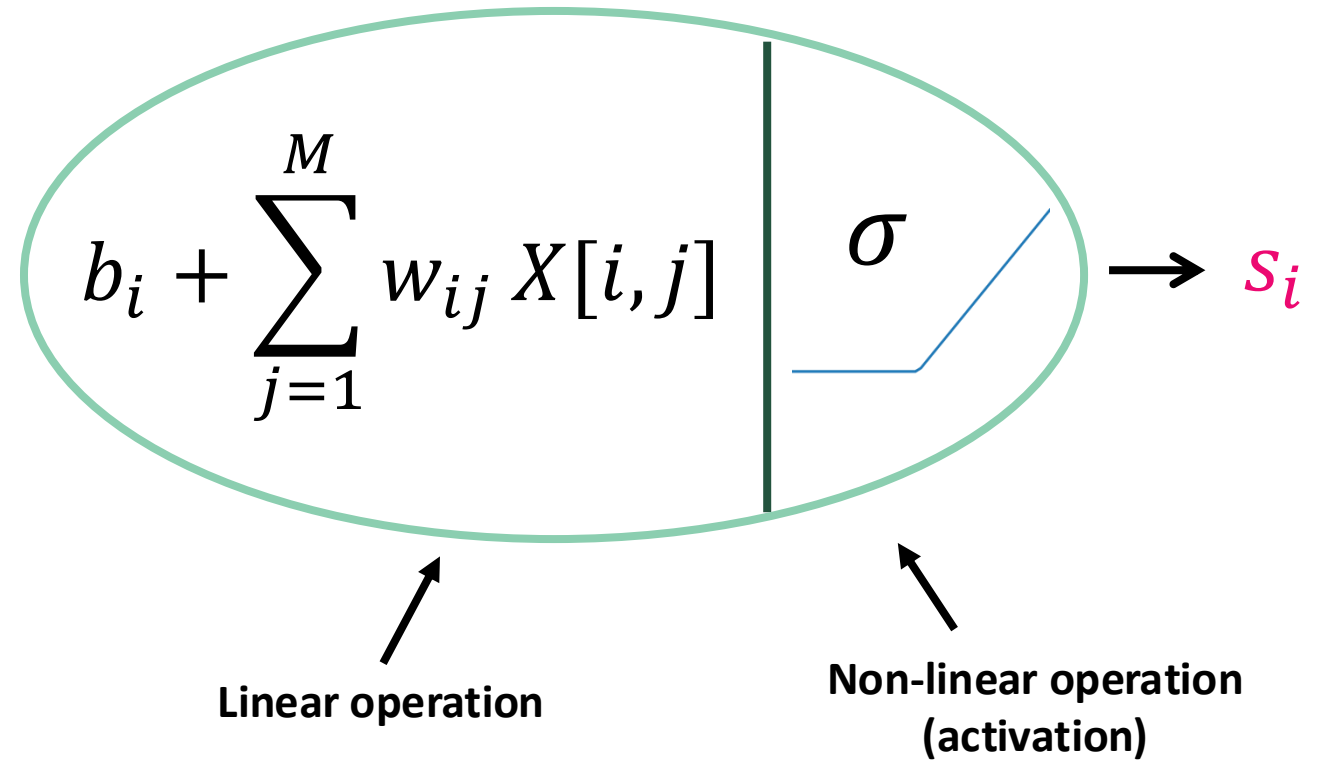
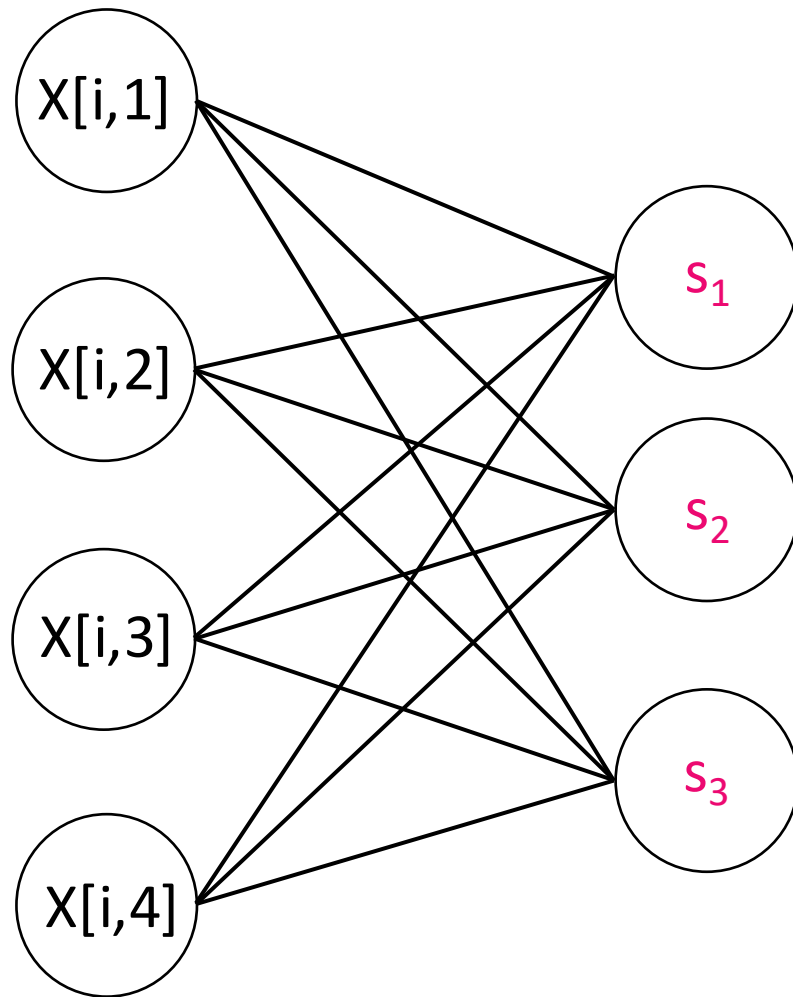
$$W = \underbrace{\begin{bmatrix} w_{11} & w_{12} & \dots & w_{1M} \\ w_{21} & w_{22} & \dots & w_{2M} \\ \dots & \dots & \dots & \dots \\ w_{N1} & w_{N2} & \dots & w_{NM} \end{bmatrix}}_{\text{N outputs M inputs}}$$

Bias

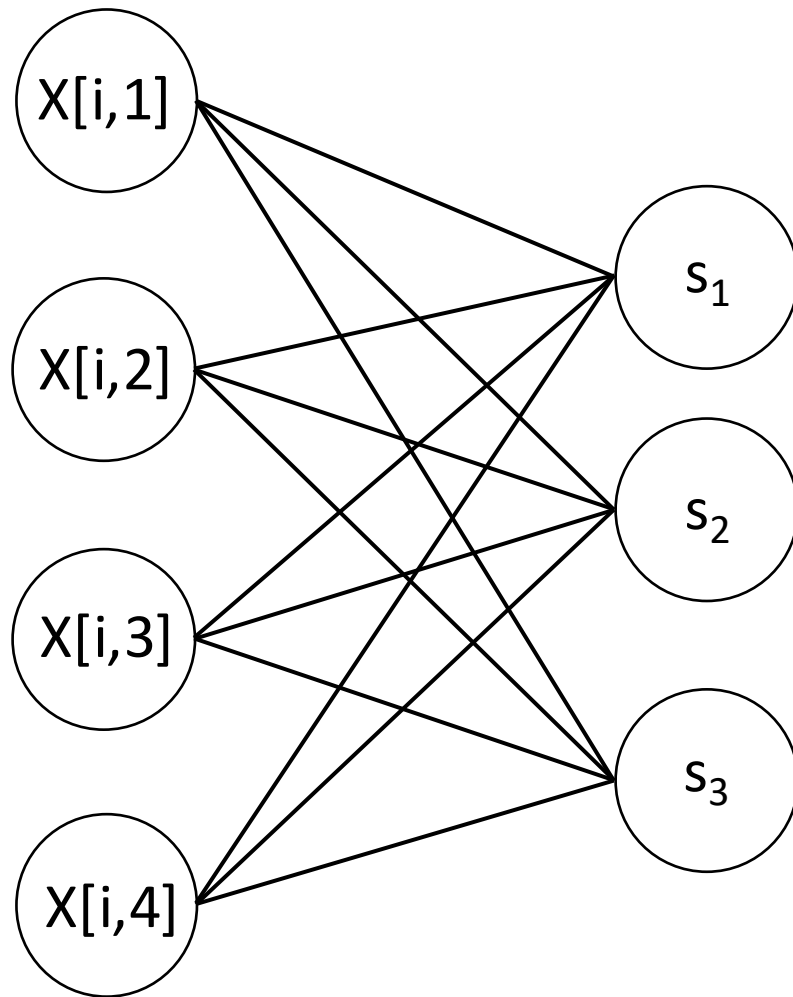


$$b = \underbrace{\begin{bmatrix} b_1 \\ \vdots \\ b_M \end{bmatrix}}_{M \text{ outputs}}$$

The Neuron Model



Outputs



$$Y = \begin{bmatrix} y_1 \\ y_2 \\ \dots \\ y_N \end{bmatrix}$$

True Labels

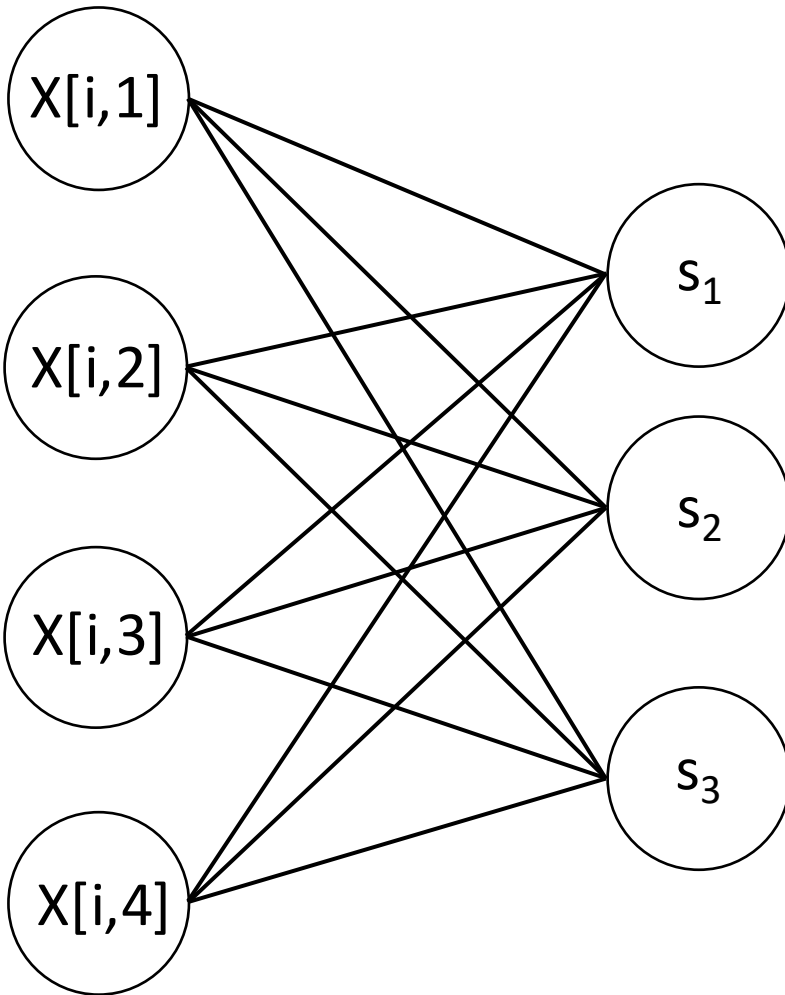
$$\hat{Y} = \begin{bmatrix} \hat{y}_1 \\ \hat{y}_2 \\ \dots \\ \hat{y}_N \end{bmatrix}$$

Predicted Labels

If the activation σ is the softmax function, then:

$$\hat{y} = \underset{\forall i}{\operatorname{argmax}}(s_i)$$

Single-layer FCNN



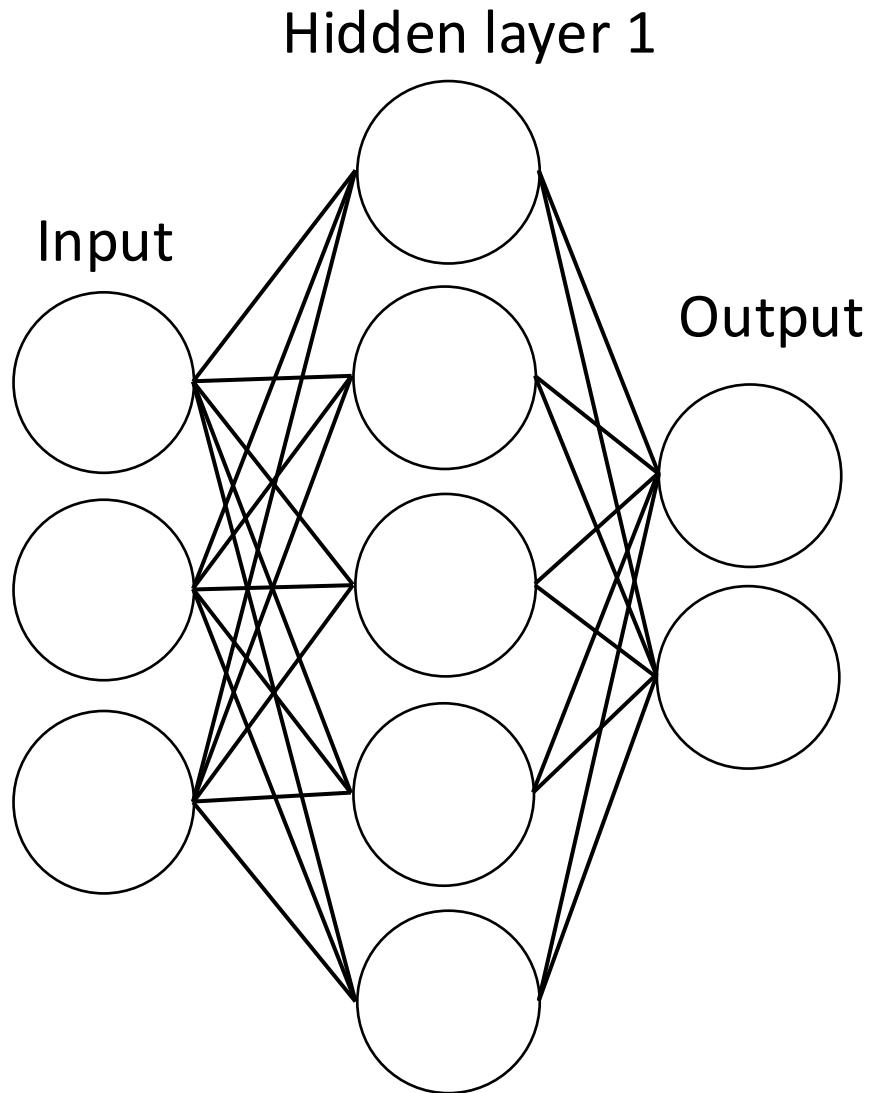
Matrix formulation:

$$[S]_{C \times 1} = \sigma([W]_{C \times M} X_{i,:}^T + [B]_{C \times 1})$$

Number of parameters:

$$(M + 1) \times C = (4 + 1) \times 3 = 15$$

Multi-layer FCNN



$$[S^{(1)}] = \sigma_1([W^{(1)}]X_{i,:}^T + [B^{(1)}])$$

$$[S^{(2)}] = \sigma_2([W^{(2)}]S^{(1)} + [B^{(2)}])$$

Number of parameters:

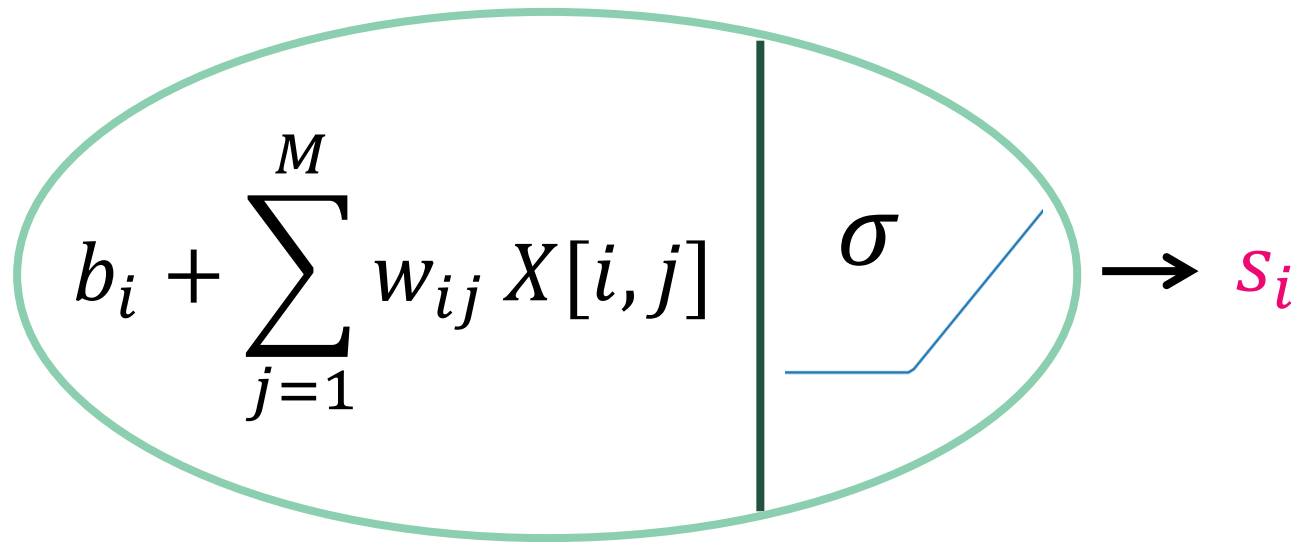
First layer: $(3 + 1) \times 5 = 20$

Second layer: $(5 + 1) \times 2 = 12$

Total: 32

Summary

- Fully connected neural networks alternate linear operations (matrix multiplication + bias term) and non-linear activations



- The number of parameters in each layer is given by the (number of inputs +1) x the number of outputs

Thank you!

