Fully Connected Neural Networks

Roberto Souza
Assistant Professor
Electrical and Computer Engineering
Schulich School of Engineering

January 2022

Outline

- Learning Goals
- Fully Connected Neural Networks
  - Single-layer Network
  - Multi-layer Network
- Avoiding Overfitting with Dropout
- Summary

Learning Goals

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

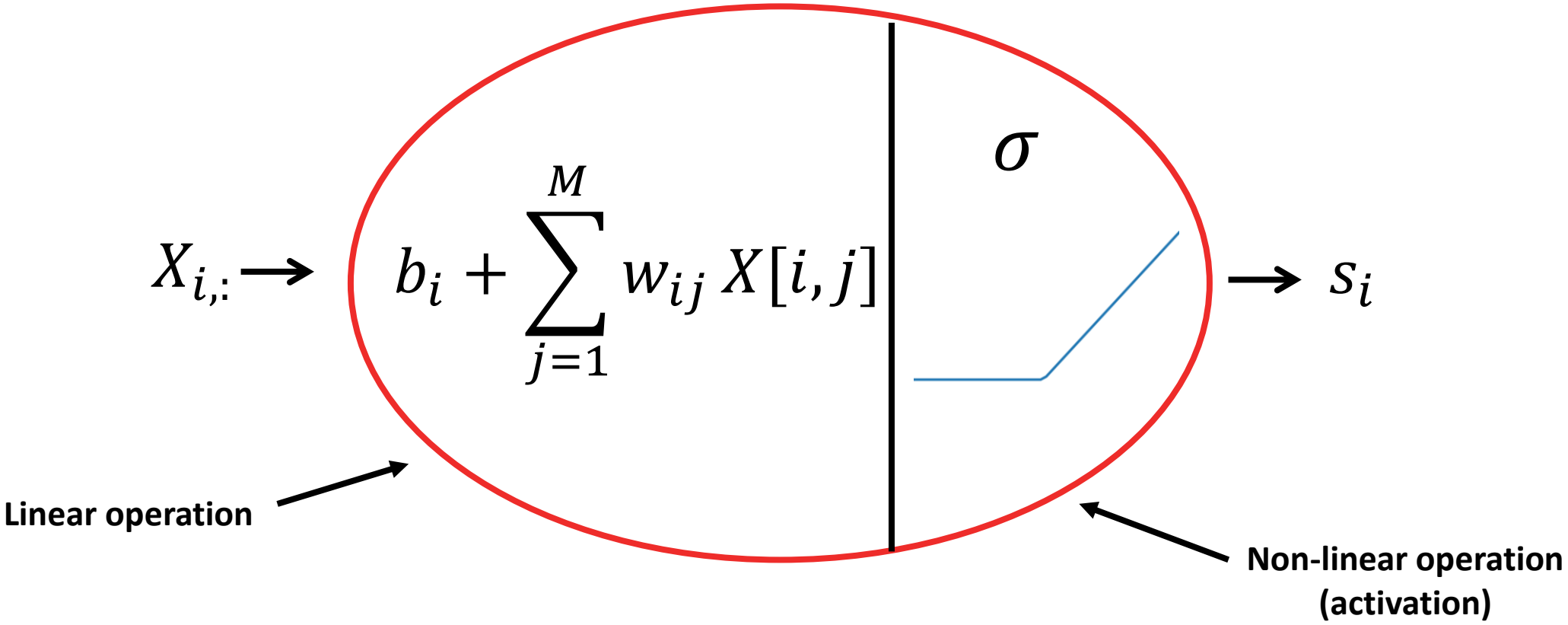
Notation

$$X = \underbrace{\begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1M} \\ x_{21} & x_{22} & \cdots & x_{2M} \\ x_{31} & x_{32} & \cdots & x_{3M} \\ \cdots & \cdots & \cdots & \cdots \\ x_{N1} & x_{N2} & \cdots & x_{NM} \end{bmatrix}}_{\text{N samples with M features}}$$

$$Y = \underbrace{\begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ \cdots \\ y_5 \end{bmatrix}}_{\text{True Labels}}$$

$$\hat{Y} = \underbrace{\begin{bmatrix} \widehat{y_1} \\ \widehat{y_2} \\ \widehat{y_3} \\ \cdots \\ \widehat{y_N} \end{bmatrix}}_{\text{Predicted Labels}}$$

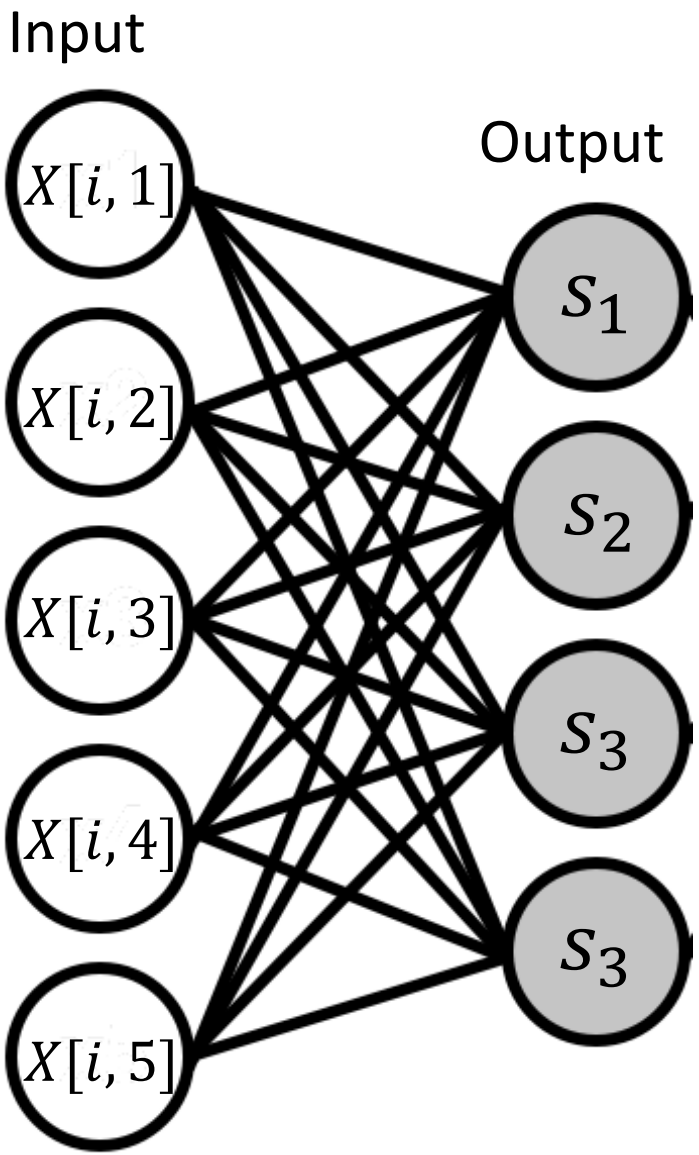
The Neuron Model



- b_i is the bias
- w_{ij} are the weights
- s_i is the output of the neuron
- **σ** is the activation function

UNIVERSITY OF CALGARY

Single-layer Fully Connected Neural Network



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

Matrix formulation

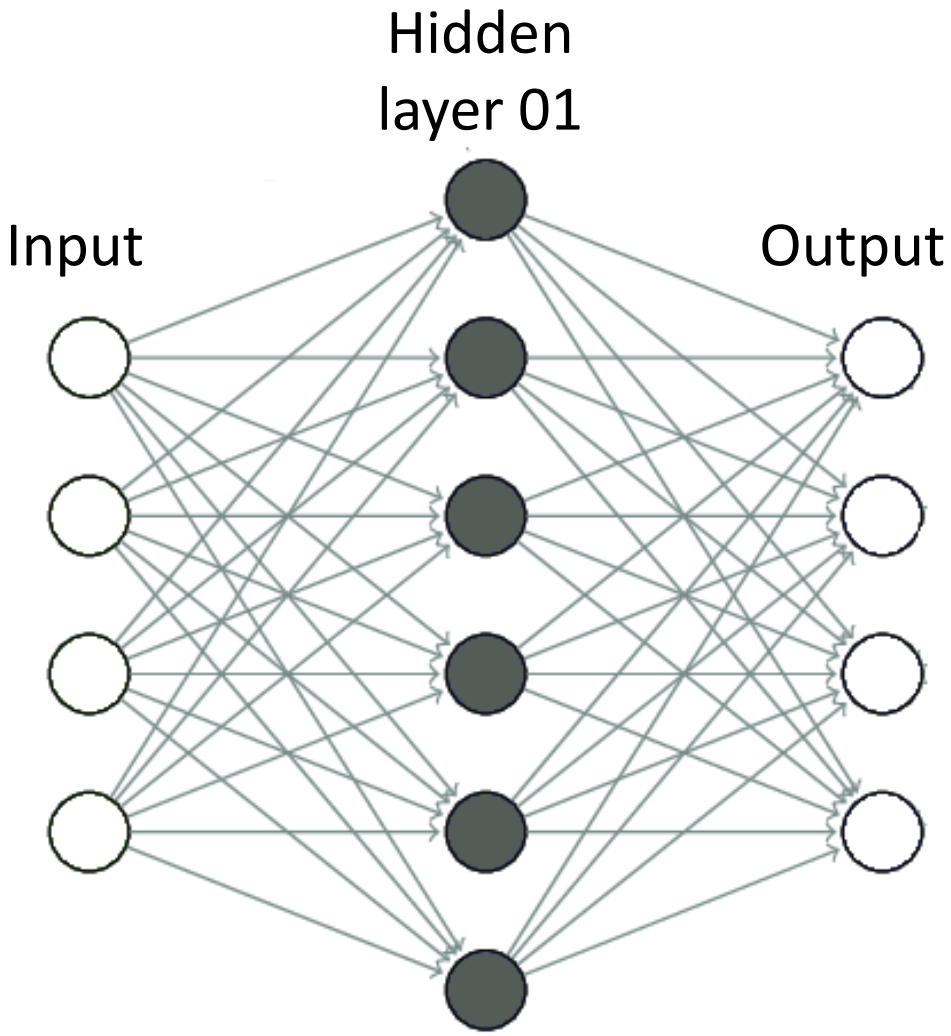
If the activation **σ** is the softmax function, then:

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

- **Number of parameters: C x (M+1) = 4 x 6 = 24**

UNIVERSITY OF CALGARY

Multi-layer Fully Connected Neural Network



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

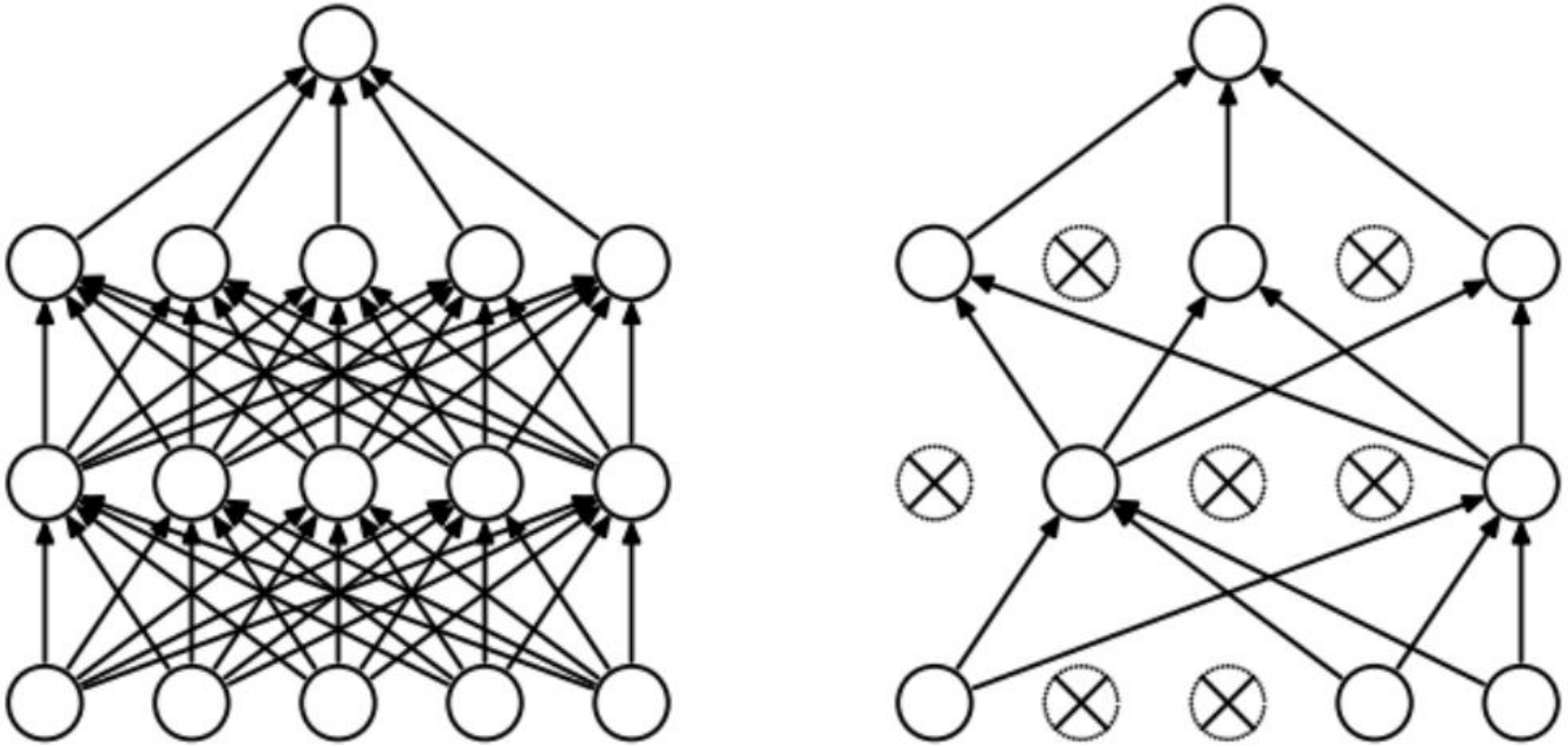
$$\left[S^{(2)}\right] = \sigma_2\left(\left[W^{(2)}\right] S^{(1)} + \left[B^{(2)}\right]\right)$$

- **Number of parameters:**
  - **First layer: (4 x 6) + 6 = 30**
  - **Second layer: (6x4) + 4 = 28**

UNIVERSITY OF CALGARY

Dropout

- Technique to avoid overfitting
- Learn redundant paths -> gain robustness



(a) Standard Neural Net

(b) After applying dropout.

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
- Dropout is a technique to avoid overfitting that makes the neural network learn redundant paths to reach the same decision

Thank you!

UNIVERSITY OF CALGARY