School of Applied Sciences and Engineering
EAFIT - Medellín

From Error to Learning: How Backpropagation Trains Neural Networks

Jorge Iván Padilla-Buriticá

March 9, 2025

Class Topics

Class Overview

Basic Concepts

Computational Graphs and Propagation

Examples

Exercises and activities

Class Objective

To explain the fundamental concepts, mathematical formulation, and practical considerations of the Backpropagation algorithm in neural networks.

Learning Outcomes

- **Explain** the fundamental principles of Backpropagation.

Learning Outcomes

- **Explain** the fundamental principles of Backpropagation.
- **Derive** and analyze the mathematical foundations of Backpropagation.

Learning Outcomes

- **Explain** the fundamental principles of Backpropagation.
- **Derive** and analyze the mathematical foundations of Backpropagation.
- **Apply** the Backpropagation algorithm to practical neural network problems.

Target Audience

Undergraduate, Master's, and PhD students, as well as postgraduate students in applied sciences and engineering, who have basic knowledge of neural networks, calculus, and proficiency in Python or similar.

Class Structure

- Conceptual Explanation
- Mathematical Derivation
- Examples
- Exercises and activities

Basic Concepts - Optimization

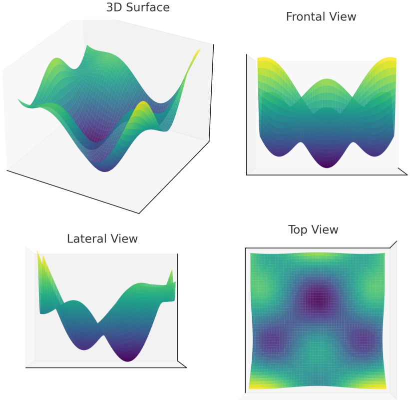


Figure: Multivariate function

Figure: General concept

Derivative: Measures the rate of change of a function.

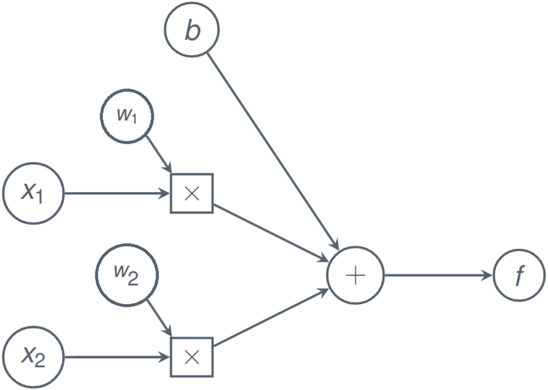
$$f'(x) = \lim_{\Delta x \to 0} \frac{f(x + \Delta x) - f(x)}{\Delta x} \tag{1}$$

Gradient (Generalized Derivative): Used in multivariable functions.

$$\nabla f(x) = \left(\frac{\partial f}{\partial x_1}, \frac{\partial f}{\partial x_2}, \ldots, \frac{\partial f}{\partial x_n} \right)^T \tag{2}$$

Definition: A computational graph represents a mathematical function in a structured form. Here, we compute:

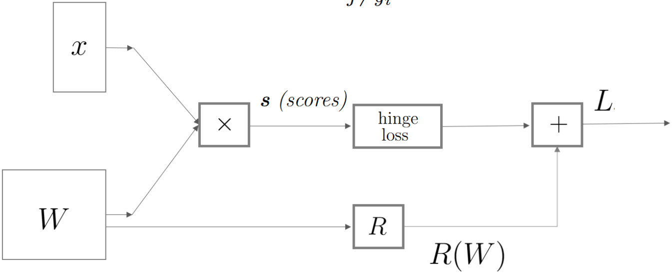
$$f = (x_1 w_1 + x_2 w_2) + b$$



Computational Graph

$$f = Wx$$

$$L_i = \sum_{j \neq y_i} max(0, s_j - s_{y_i} + 1)$$

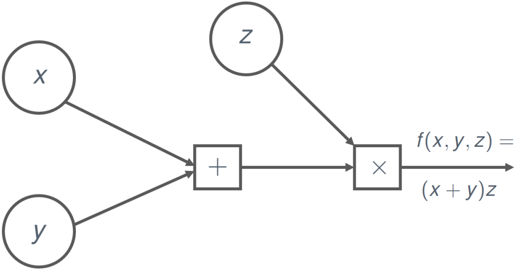


Example 01:

Example: We define the function:

$$f(x, y, z) = (x + y)z$$

Computational Graph:



Example 01:

The function can be rewritten in steps:

$$f(x,y,z) = (x+y)z$$

$$q = x+y, \quad f = q \cdot z$$

Partial derivatives:

$$\frac{\partial f}{\partial q} = z, \quad \frac{\partial f}{\partial z} = q, \quad \frac{\partial q}{\partial x} = 1, \quad \frac{\partial q}{\partial y} = 1$$

Using the **chain rule**, we compute:

$$\frac{\partial f}{\partial x} = \frac{\partial f}{\partial q} \cdot \frac{\partial q}{\partial x} = z$$

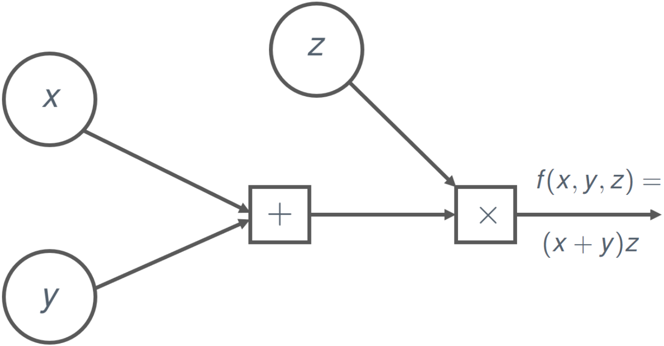
$$\frac{\partial f}{\partial y} = \frac{\partial f}{\partial q} \cdot \frac{\partial q}{\partial y} = z$$

$$\frac{\partial f}{\partial z} = q$$

Example 01:

Forward Propagation :

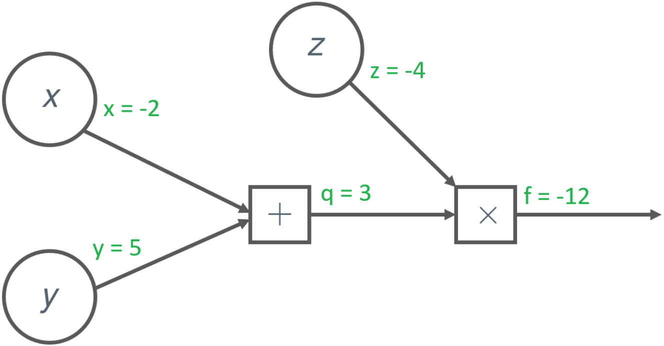
$$f(x, y, x) = (x + y)z = qz$$



Example 01:

Forward Propagation :

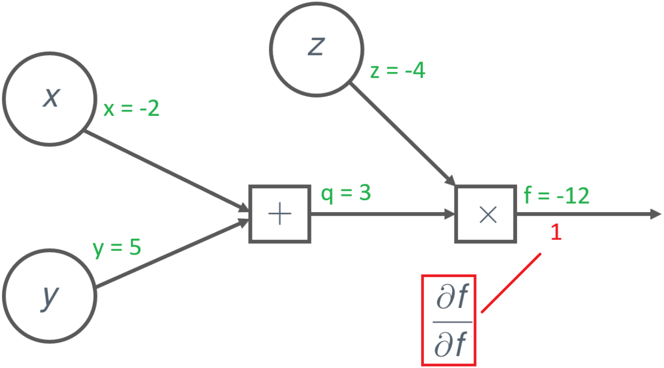
$$f(x, y, x) = (x + y)z = qz$$



Example 01:

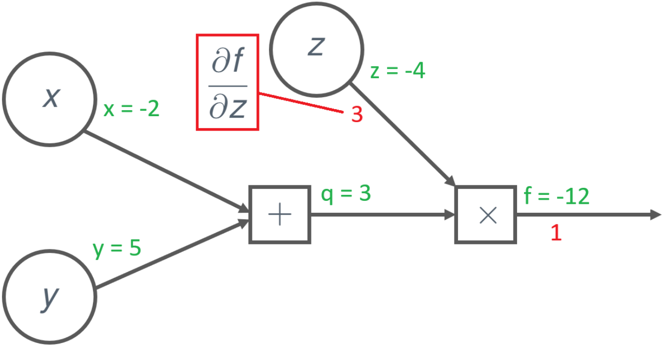
Backward Propagation :

$$f(x, y, x) = (x + y)z = qz$$



Backward Propagation :

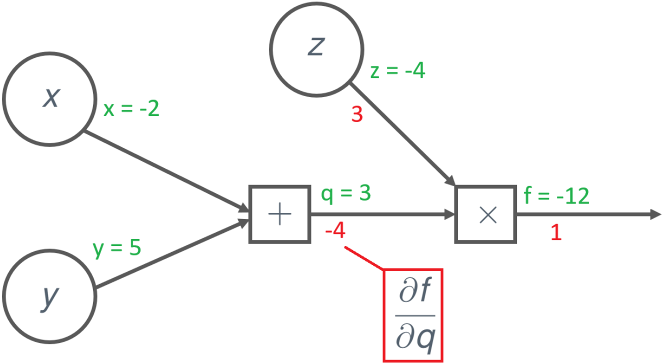
$$f(x, y, x) = (x + y)z = qz$$



Example 01:

Backward Propagation :

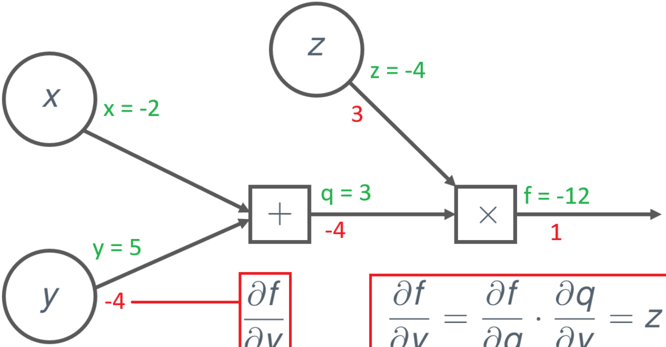
$$f(x, y, x) = (x + y)z = qz$$



Example 01:

Backward Propagation :

$$f(x, y, x) = (x + y)z = qz$$



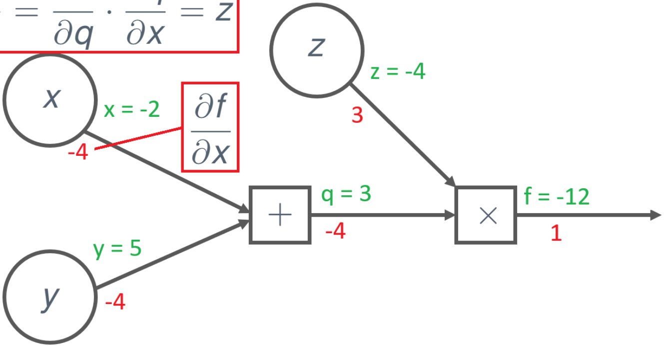
$$\frac{\partial f}{\partial y}$$

$$\frac{\partial f}{\partial y} = \frac{\partial f}{\partial q} \cdot \frac{\partial q}{\partial y} = z$$

Backward Propagation :

$$f(x, y, x) = (x + y)z = qz$$

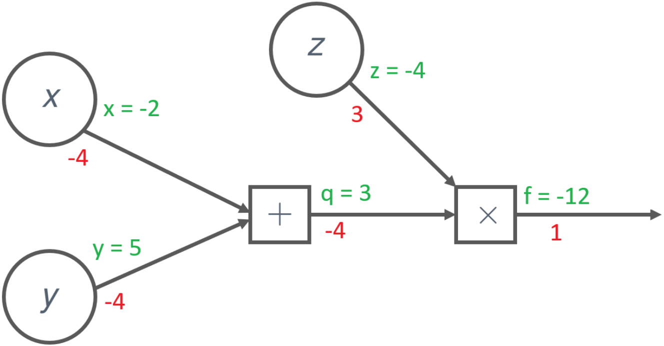
$$\boxed{\frac{\partial f}{\partial x} = \frac{\partial f}{\partial q} \cdot \frac{\partial q}{\partial x} = z}$$



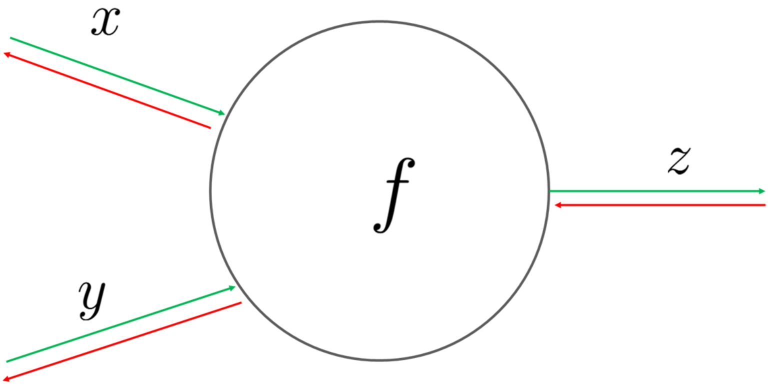
Example 01:

Backward Propagation :

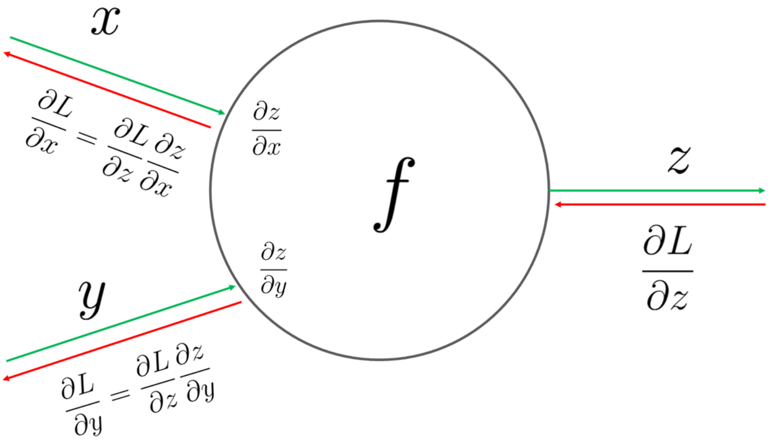
$$f(x, y, x) = (x + y)z = qz$$



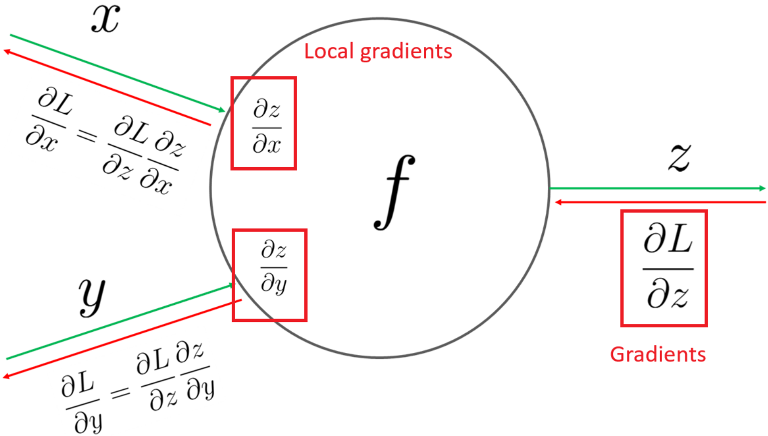
Activations:



Activations:

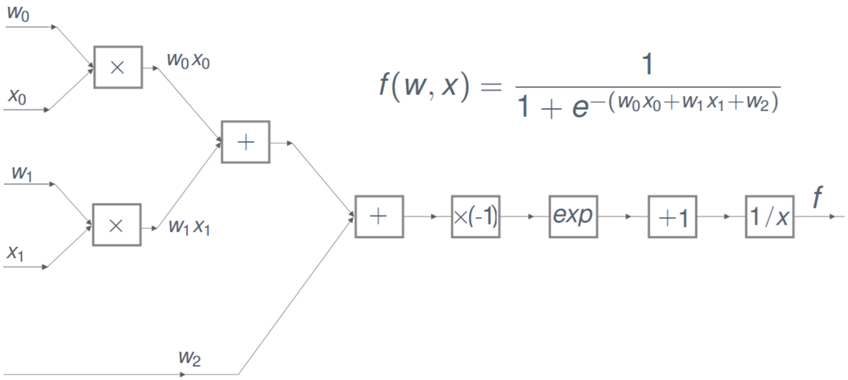


Activations:



Example 02:

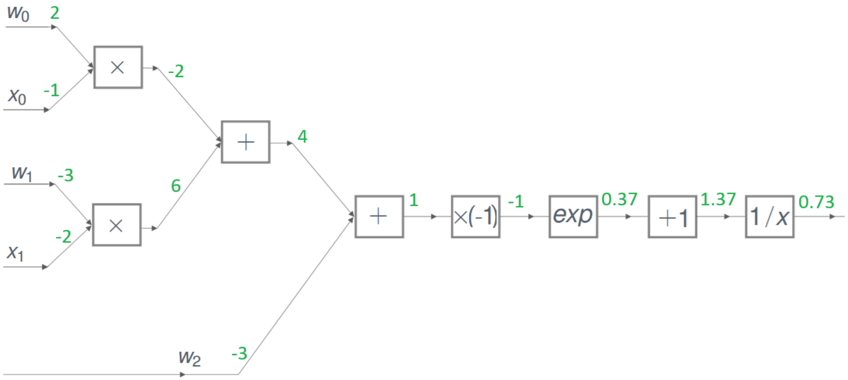
Forward Propagation :



$$f(w, x) = \frac{1}{1 + e^{-(w_0 x_0 + w_1 x_1 + w_2)}}$$

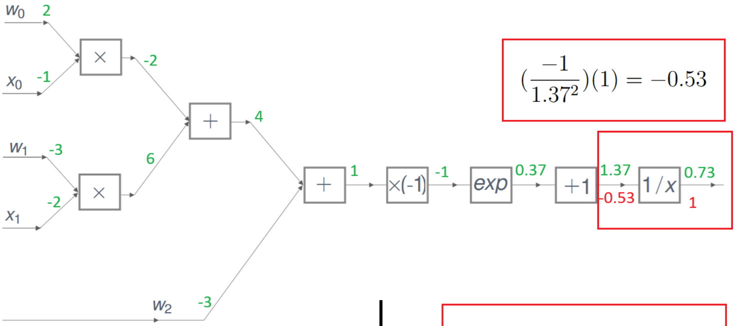
Example 02:

Forward Propagation :



Example 02:

Backward Propagation :



$$\left(\frac{-1}{1.37^2}\right)(1) = -0.53$$

$$f(x) = e^x \rightarrow \frac{df}{dx} = e^x$$

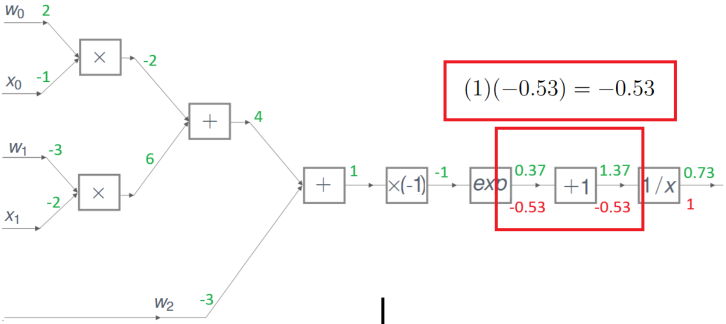
$$f_a(x) = ax \rightarrow \frac{df}{dx} = a$$

$$f(x) = \frac{1}{x} \rightarrow \frac{df}{dx} = -\frac{1}{x^2}$$

$$f_c(x) = c + x \rightarrow \frac{df}{dx} = 1$$

Example 02:

Backward Propagation :



$$\boxed{(1)(-0.53) = -0.53}$$

$$f(x) = e^x \rightarrow \frac{df}{dx} = e^x$$

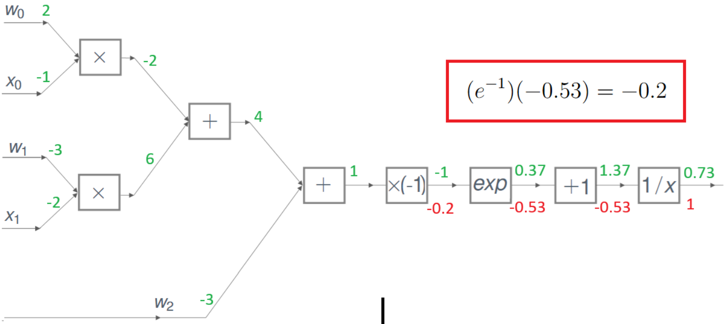
$$f_a(x) = ax \rightarrow \frac{df}{dx} = a$$

$$f(x) = \frac{1}{x} \rightarrow \frac{df}{dx} = -\frac{1}{x^2}$$

$$\boxed{f_c(x) = c + x \rightarrow \frac{df}{dx} = 1}$$

Example 02:

Backward Propagation :



$$\boxed{(e^{-1})(-0.53) = -0.2}$$

$$\boxed{f(x) = e^x \rightarrow \frac{df}{dx} = e^x}$$

$$f_a(x) = ax \rightarrow \frac{df}{dx} = a$$

$$f(x) = \frac{1}{x} \rightarrow \frac{df}{dx} = -\frac{1}{x^2}$$

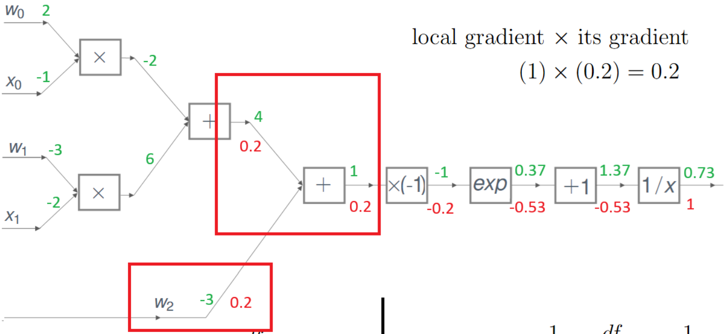
$$f_c(x) = c + x \rightarrow \frac{df}{dx} = 1$$

Example 02:

Backward Propagation :

$$\text{local gradient} \times \text{its gradient}$$

$$(1) \times (0.2) = 0.2$$



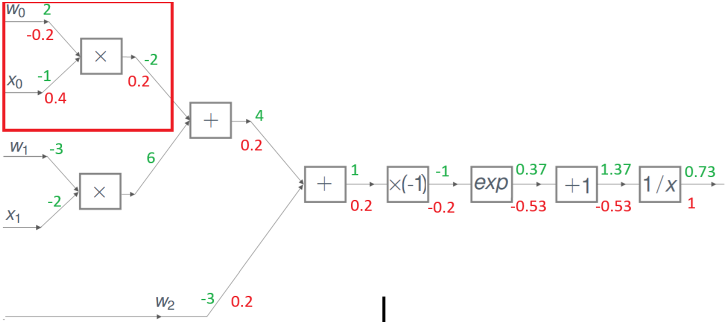
$$f(x) = e^x \rightarrow \frac{df}{dx} = e^x$$

$$f_a(x) = ax \rightarrow \frac{df}{dx} = a$$

$$f(x) = \frac{1}{x} \rightarrow \frac{df}{dx} = -\frac{1}{x^2}$$

$$f_c(x) = c + x \rightarrow \frac{df}{dx} = 1$$

Backward Propagation :



$$f(x) = e^x \rightarrow \frac{df}{dx} = e^x$$

$$f_a(x) = ax \rightarrow \frac{df}{dx} = a$$

$$f(x) = \frac{1}{x} \rightarrow \frac{df}{dx} = -\frac{1}{x^2}$$

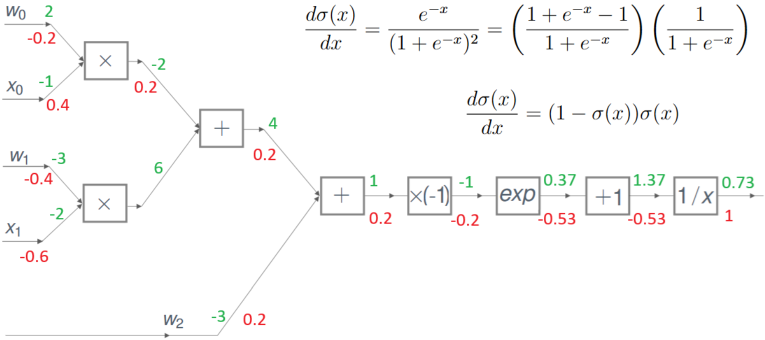
$$f_c(x) = c + x \rightarrow \frac{df}{dx} = 1$$

Another perspective:

$$f(w,x) = \frac{1}{1+e^{-(w_0x_0+w_1x_1+w_2)}} \qquad \sigma(x) = \frac{1}{1+e^{-(x)}}$$

$$\frac{d\sigma(x)}{dx} = \frac{e^{-x}}{(1+e^{-x})^2} = \left(\frac{1+e^{-x}-1}{1+e^{-x}}\right)\left(\frac{1}{1+e^{-x}}\right)$$

$$\frac{d\sigma(x)}{dx} = (1-\sigma(x))\sigma(x)$$



Example 02:

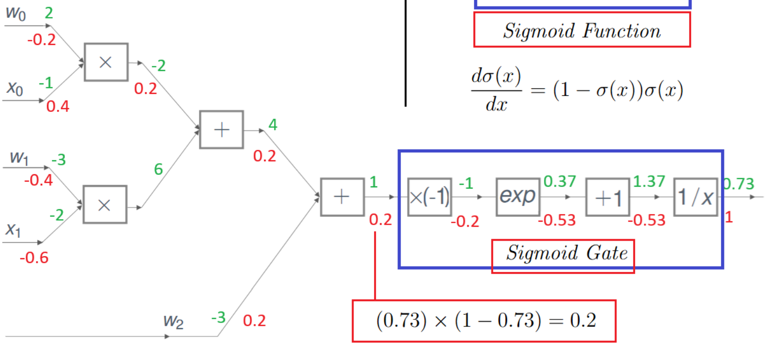
Another perspective:

$$f(w,x) = \frac{1}{1+e^{-(w_0x_0+w_1x_1+w_2)}}$$

$$\boxed{\sigma(x) = \frac{1}{1+e^{-(x)}}}$$

Sigmoid Function

$$\frac{d\sigma(x)}{dx} = (1-\sigma(x))\sigma(x)$$



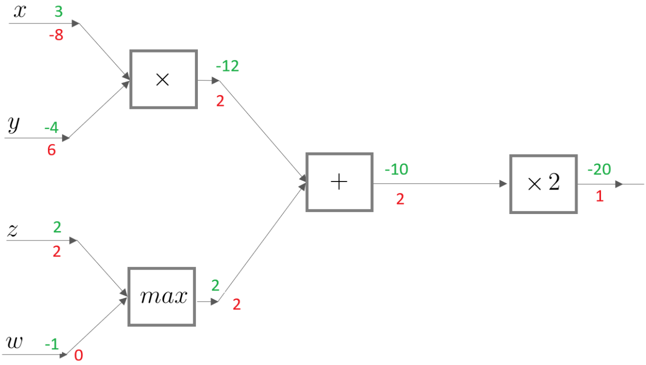
$$(0.73) \times (1-0.73) = 0.2$$

Patterns in backward flow:

Add gate: Gradient distributor

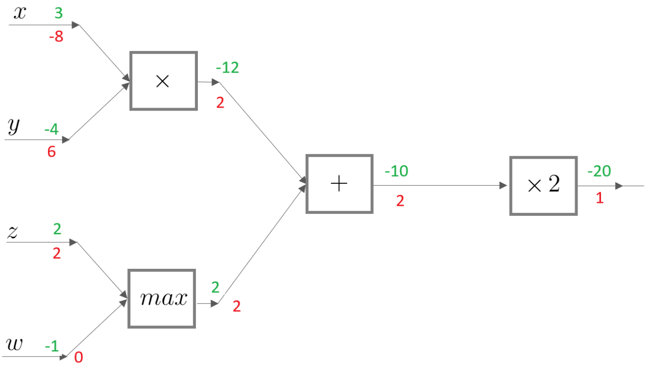
Max gate: Gradient Router

Mul gate: Gradient Switcher



Patterns in backward flow:

Gradients add at branches

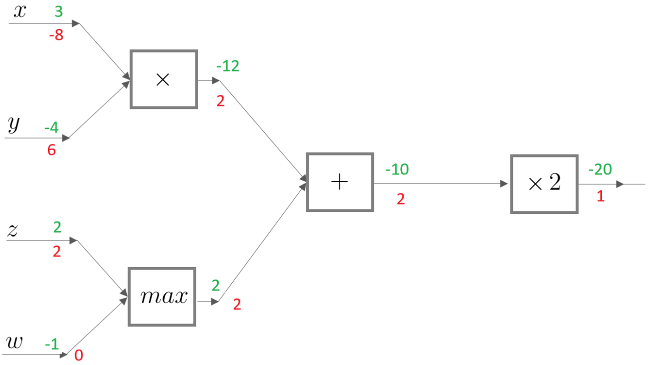


Patterns in backward flow:

Add gate: Gradient distributor

Max gate: Gradient Router

Mul gate: Gradient Switcher



Evaluation Methodology

Task: Please complete Exercises by accessing the following Colab notebook link:

Exercises Notebook (80%)

Questionnaire (20%)

Evaluation Methodology

Assessment Components:

- **Practical Exercises (80%)**
  - Individual solutions to notebook exercises.
  - Submission via email.
  - Evaluated on clarity, mathematical depth, and conceptual accuracy.
- **Multiple-choice Questionnaire (20%)**
  - Focused on backpropagation concepts.
  - Submission via email.
- **Real-Time Polling (15 min)**
  - Conducted through Mentimeter/Kahoot.
  - Anonymous participation.

Timing: 2 hours total

Note: Prior reading of the Google notebook containing exercises, examples, and instructions is required.

Evaluation Methodology

Feedback and Results:

- Immediate formative verbal feedback during session.
- Summary of exercise and polling results shared with students.
- Clarification of common misunderstandings.

Questions?