

# Neural Highways: Unraveling the Role of Nonlinearity Functions

Algotecniq LLC

July 23, 2023

What makes neural nets so powerful?

NONLINEAR FUNCTIONS are indispensable elements in the field of neural networks. Their critical role is underscored by the necessity of representing and modeling complex, nonlinear correlations between inputs and outputs.

Drawing an analogy from geographical terrains, consider the relationship between inputs and outputs as an expedition. In cases where the journey is a direct route from one point to another, akin to a highway drive, a linear function is an adequate representative model. However, when the expedition involves maneuvering around turns, scaling hills, descending into valleys, or circumventing obstacles — a direct path no longer suffices. This necessitates a more sophisticated guide that can handle the complexity of the journey — and this is precisely the role played by a nonlinearity function in a neural network.

TO EXTEND THE METAPHOR, consider a complex Lego construction following intricate instructions. You cannot create simple, straight structures with only linear, straight pieces at your disposal. Introduce a set of unique, curved pieces, however, and the possibilities for the types of structures you can build are vastly expanded. Analogously, nonlinearity functions in a neural network are those ‘special, curved Lego pieces,’ opening up a realm of complexities the network can capture.

## Two Examples

THE SIGMOID FUNCTION, characterized by the equation  $f(x) = \frac{1}{1+e^{-x}}$ , has a confined output range of  $(0,1)$ , making it suitable for mapping arbitrary inputs to probabilities. It was commonly used in early neural networks and still finds continued use in specific applications like binary classification tasks.

THE SWISH FUNCTION, defined as  $f(x) = x \cdot \frac{1}{1+e^{-x}}$ , is an activation function developed by Google researchers. It introduces nonlinearity, allows smooth gradient flow, and helps avoid “neuron death” by retaining the input’s sign. The Swish function has shown enhanced performance in various deep learning models. However, it adds computational complexity, potentially increasing processing times. Despite this, the Swish function holds promising potential, and its role in diverse network architectures is a focus of ongoing research in machine learning.

Sigmoid Function:  $y = \frac{1}{1+e^{-x}}$

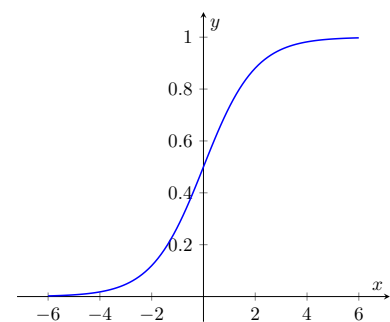


Figure 1: The Sigmoid Function – Showcasing its Characteristic ‘S’ Shape and Saturation Behavior

Swish Function:  $y = x \cdot \frac{1}{1+e^{-x}}$

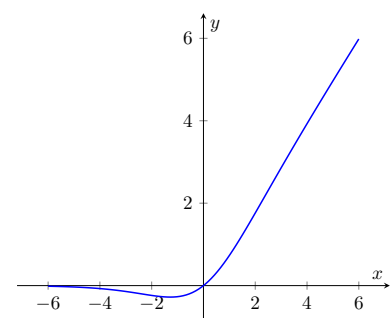


Figure 2: Visualizing the Swish Activation Function: Balancing Non-linearity and Gradient Flow