Industrial Engineering Journal
ISSN: 0970-2555
Volume : 52, Issue 10, No. 4, October : 2023

EDGE AND SNAG OF ACTIVATION FUNCTIONS IN NEURAL NETWORKS

Dr. Varalakshmi Dandu, Assistant Professor, JNTUA School of Management Studies, JNTU Ananthapuramu, 515002. **:** **varalakshmi.jntua@gmail.com**

Abstract:
Activation functions play a crucial role in artificial neural networks, introducing nonlinearity and enabling the modeling of complex relationships between inputs and outputs. This paper provides a comprehensive overview of various activation functions commonly used in deep learning, including sigmoid, hyperbolic tangent (tanh), rectified linear unit (ReLU), exponential linear unit (ELU), Swish, and others. The advantages, disadvantages, and considerations related to each activation function are discussed, highlighting their impact on network performance, convergence, and generalization. Additionally, the applications of activation functions across different domains, such as image recognition, natural language processing, robotics, and drug discovery, are explored. The paper also discusses potential future trends in activation function research, including the development of task-specific functions and advancements in neural architecture search. Overall, this paper aims to provide researchers and practitioners with a comprehensive understanding of activation functions and their significance in designing efficient and effective neural networks.

Keywords: Activation functions, neural networks, deep learning, sigmoid, tanh, ReLU, ELU, Swish, neural architecture search.

Introduction:
An activation function is a crucial component of artificial neural networks, serving as a nonlinear transformation that introduces nonlinearity into the network's outputs. The activation function takes the weighted sum of inputs from the previous layer and produces the output for the current neuron.

The importance of activation functions lies in their ability to introduce nonlinearity, enabling neural networks to model complex relationships between inputs and outputs. Without activation functions, the entire neural network would behave like a linear model, limiting its capacity to learn and represent intricate patterns and mappings. Activation functions play a crucial role in artificial neural networks by introducing non-linearity to the model. They help neural networks learn complex relationships in the data and make the model more expressive.

The choice of activation function depends on the specific problem, architecture, and data at hand. While ReLU and its variants are the most commonly used activation functions in deep learning, newer advancements and research may lead to the adoption of different functions for specific scenarios. It's essential to experiment and tune activation functions to find the best fit for a particular neural network architecture and task.

1. **Sigmoid Function (Logistic Function)**
   The sigmoid function, also known as the logistic function, is a common activation function used in artificial neural networks. It takes an input (usually the weighted sum of inputs from the previous layer) and transforms it into an output within the range of (0, 1).
   Formula: f(x) = 1 / (1 + exp(-x))
   Range: (0, 1)
   Advantages:

Smooth and bounded output in the range (0, 1), which can be interpreted as probabilities.
Historically used in older neural networks and binary classification problems.

Disadvantages:

i. Suffers from vanishing gradient problem, which slows down learning in deep networks.
ii. Outputs saturate, leading to the vanishing gradient problem and killing the gradients.

2. **Hyperbolic Tangent Function (tanh):**
The hyperbolic tangent function, often referred to as tanh, is an activation function commonly used in artificial neural networks. It is a variation of the sigmoid function, but with a different range and centering.

Formula: f(x) = (exp(x) - exp(-x)) / (exp(x) + exp(-x))
Range: (-1, 1)
Advantages:
i.Smooth and bounded output in the range (-1, 1).
ii.Centered around zero, which reduces the issues of vanishing gradients compared to the sigmoid function.
Disadvantages:
i.Still suffers from vanishing gradient problem, especially in deep networks.
ii.Outputs saturate and may lead to vanishing gradients.

3.Rectified Linear Unit (ReLU):
The Rectified Linear Unit (ReLU) is a popular activation function used in artificial neural networks.It is one of the simplest and most widely used activation functions due to its effectiveness in deep learning architectures.
Formula: f(x) = max(0, x)
Range: [0, +∞)
Advantages:

i. Simple and computationally efficient.
ii. Addresses vanishing gradient problem by allowing non-zero gradients for positive inputs. Promotes sparse activations, making the network more efficient during training and inference.

Disadvantages:

i. Can suffer from the "dying ReLU" problem, where neurons can get stuck in a state of inactivity during training, causing them to never activate again.
ii. Prone to the "exploding gradient" problem as unbounded positive values can lead to large gradients during training.
iii.

4.Leaky ReLU and Parametric ReLU (PReLU):
Leaky ReLU is an activation function that introduces a small, non-zero slope for negative inputs, preventing the "dying ReLU" problem. Parametric ReLU (PReLU) is a variation of Leaky ReLU with a learnable slope parameter, allowing it to adapt and achieve better performance.
Formula: f(x) = max(ax, x), where 'a' is a small positive constant (typically 0.01).
Range: (-∞, +∞)

Advantages:

i. Address the "dying ReLU" problem by introducing a small slope for negative inputs, preventing neurons from being inactive.
ii. Can improve convergence and generalization in some cases.

Disadvantages:

i. PReLU introduces additional parameters, making the model more complex.

5. Exponential Linear Unit (ELU)

The Exponential Linear Unit (ELU) is an activation function used in artificial neural networks. It is a smooth and continuously differentiable function that addresses some of the limitations of the Rectified Linear Unit (ReLU) activation function, especially for negative inputs.

Formula: $f(x) = x$ if $x > 0$, and $f(x) = a * (\exp(x) - 1)$ if $x \leq 0$, where 'a' is a small positive constant (typically 1).

Range: $(-\infty, +\infty)$

Advantages:

i. Smooth, non-zero gradients for positive and negative inputs, addressing the vanishing gradient problem.
ii. Helps prevent the "dying ReLU" problem.
iii. Captures negative values without resulting in large positive outputs. Disadvantages: Slightly more computationally expensive compared to ReLU.

Disadvantages:

i. Slightly more computationally expensive compared to ReLU.

6.Scaled Exponential Linear Unit (SELU):

The Scaled Exponential Linear Unit (SELU) is an activation function designed to ensure self-normalization in artificial neural networks. It is an extension of the Exponential Linear Unit (ELU) and introduces a scaling factor to maintain a constant mean and variance of activations across layers, promoting more stable and efficient training of deep networks.

Formula: f(x) = scale * (x if x > 0 else alpha * (exp(x) - 1))

In this formula, 'x' represents the input to the function, 'exp' is the exponential function

Range: $(-\infty, +\infty)$

Advantages:

i. Self-normalizing activation function that maintains mean and variance of activations across layers, helping stabilize training in deep networks.
ii. Can lead to improved results in specific architectures (e.g., deep residual networks).

Disadvantages:

i. Requires certain conditions on the weight initialization for self-normalization, limiting its use in arbitrary architectures.

7. Swish:

The Swish activation function is defined as the element-wise multiplication of the input 'x' with the output of the sigmoid function applied to 'x'. It combines linearity for positive inputs and nonlinearity for negative inputs, showing promise in certain deep learning tasks.

Formula: f(x) = x * sigmoid(x)

Range: $(-\infty, +\infty)$

Advantages:

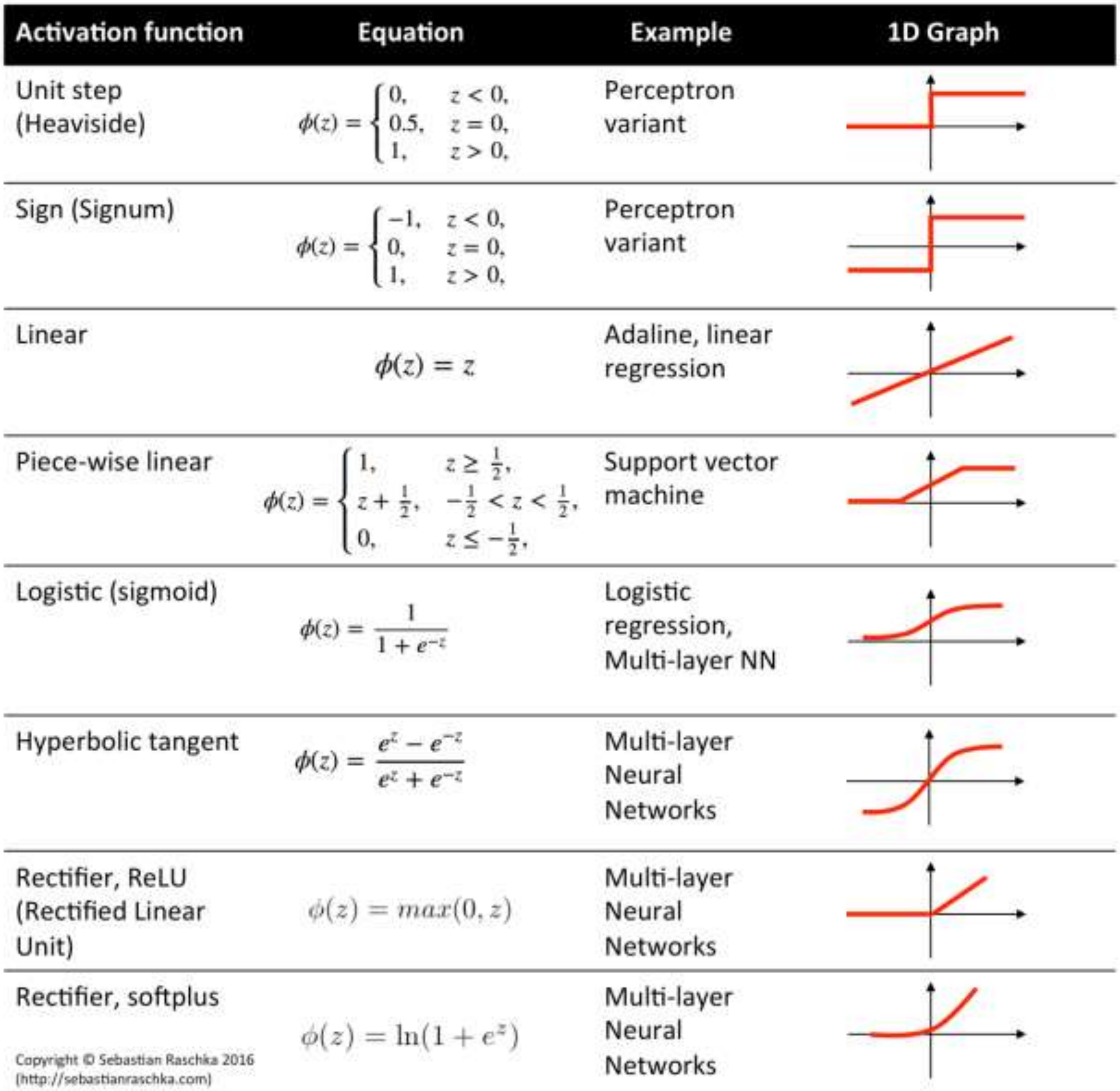
i. Continuously differentiable, allowing for smooth optimization.
ii. Empirically shown to perform well in certain cases, providing faster convergence and better generalization.

Disadvantages:

i. Slightly more computationally expensive compared to ReLU.

Fig no 1: overview of activation functions

Industrial Engineering Journal
ISSN: 0970-2555
Volume : 52, Issue 10, No. 4, October : 2023



Activation function	Equation	Example	1D Graph
Unit step (Heaviside)	$\phi(z) = \begin{cases} 0, & z < 0, \\ 0.5, & z = 0, \\ 1, & z > 0, \end{cases}$	Perceptron variant	
Sign (Signum)	$\phi(z) = \begin{cases} -1, & z < 0, \\ 0, & z = 0, \\ 1, & z > 0, \end{cases}$	Perceptron variant	
Linear	$\phi(z) = z$	Adaline, linear regression	
Piece-wise linear	$\phi(z) = \begin{cases} 1, & z \geq \frac{1}{2}, \\ z + \frac{1}{2}, & -\frac{1}{2} < z < \frac{1}{2}, \\ 0, & z \leq -\frac{1}{2}, \end{cases}$	Support vector machine	
Logistic (sigmoid)	$\phi(z) = \frac{1}{1 + e^{-z}}$	Logistic regression, Multi-layer NN	
Hyperbolic tangent	$\phi(z) = \frac{e^{z} - e^{-z}}{e^{z} + e^{-z}}$	Multi-layer Neural Networks	
Rectifier, ReLU (Rectified Linear Unit)	$\phi(z) = max(0, z)$	Multi-layer Neural Networks	
Rectifier, softplus 	$\phi(z) = \ln(1 + e^{z})$	Multi-layer Neural Networks	

CONSIDERATIONS RELATED TO ACTIVATION FUNCTIONS:
Empirical Evaluation of Activation Functions:

- Several studies have focused on empirically evaluating the performance of different activation functions across various architectures and datasets. Researchers have compared traditional activation functions like sigmoid and tanh with more modern ones like ReLU, Leaky ReLU,

ELU, and Swish. These empirical studies have provided valuable insights into the strengths and weaknesses of each activation function under different scenarios.

Addressing Vanishing Gradient Problem:

- Early activation functions like sigmoid and tanh suffered from the vanishing gradient problem, which hinders learning in deep networks. Research has explored alternative activation functions, such as ReLU and its variants, that alleviate this problem and enable deeper networks to be trained more effectively.

Smooth Activation Functions:

- Smooth activation functions, like ELU and Swish, have gained attention due to their continuous differentiability, which can facilitate optimization in some cases. These functions aim to provide a balance between performance and smoothness in the derivatives.

Activation Functions for Specific Architectures:

- Some research has focused on finding the most suitable activation functions for specific neural network architectures. For instance, SELU has been proposed as a self-normalizing activation function that aims to stabilize training in deep residual networks.

Activation Functions in Recurrent Neural Networks (RNNs) and LSTMs:

- Activation functions are crucial in recurrent neural networks and long short-term memory networks (LSTMs) due to their recurrent nature. Research has investigated how different activation functions impact the performance and convergence of these sequential models.

Activation Functions in Generative Models:

- The choice of activation functions can significantly affect the quality of generated samples in generative models like Generative Adversarial Networks (GANs) and Variational Autoencoders (VAEs). Studies have explored the impact of activation functions on the stability and realism of generated samples.

Activation Functions in Quantized Neural Networks:

- With the growing interest in deploying neural networks on resource-constrained devices, research has examined the behavior of various activation functions in quantized neural networks. Understanding how activation functions perform under low precision is crucial for efficient hardware implementation.

Neural Architecture Search (NAS) for Activation Functions:

- Neural Architecture Search techniques have been employed to automatically discover optimal activation functions for specific tasks or architectures. These studies aim to find novel activation functions that might outperform traditional choices.

Theoretical Analysis of Activation Functions:

- Researchers have explored the theoretical properties of different activation functions, including their convergence properties, expressiveness, and capacity for universal approximation.

APPLICATIONS OF ACTIVATION FUNCTIONS

- The applications of activation functions highlight the diverse and significant impact of activation functions in various domains, where neural networks have shown great potential in solving complex problems and making intelligent decisions.
- Image Recognition and Computer Vision: Activation functions are used in Convolutional Neural Networks (CNNs) to detect and extract features from images. They allow the network to identify edges, shapes, textures, and patterns in the input images, enabling applications like image classification, object detection, and facial recognition.
- Natural Language Processing (NLP): In NLP tasks such as text classification, sentiment analysis, and language translation, activation functions help process and analyze textual data. Recurrent Neural Networks (RNNs) and Long Short-Term Memory (LSTM) networks, which are commonly used in NLP, depend on activation functions to model sequential data.
- Speech Recognition: Activation functions play a role in speech recognition tasks, where the network needs to process audio data and convert speech signals into text or commands.
- Autonomous Vehicles: Activation functions are used in deep learning models that control self-driving cars and autonomous vehicles. These models process sensor data (such as camera inputs, LiDAR, and radar) to make real-time decisions for safe navigation.
- Anomaly Detection: Activation functions are employed in anomaly detection applications, where the network is trained to recognize deviations from normal patterns, helping in identifying fraud, cyberattacks, or any unusual behavior in various systems.
- Recommender Systems: Activation functions are used in collaborative filtering models, which recommend products, movies, or content to users based on their preferences and behavior.
- Financial Forecasting: In finance, activation functions are applied to predict stock prices, market trends, or to model complex financial data for risk assessment and investment strategies.
- Gaming and Reinforcement Learning: Activation functions are utilized in training agents for gaming applications and reinforcement learning tasks. These models learn to make decisions and take actions in an environment to maximize rewards.
- Drug Discovery: Activation functions play a role in the field of computational drug discovery, where neural networks are used to predict molecular properties and identify potential drug candidates.
- Robotics: Activation functions are used in neural networks that control robotic systems, enabling them to perform tasks like object manipulation, grasping, and navigation.
- Applications of activation functions mainly demonstrate the wide-ranging impact and versatility of activation functions in various fields, where neural networks have proven to be effective in solving complex problems and making intelligent decisions.

Future scope of each activation function:

- While it is challenging to predict the exact future use of each activation function, we can make some educated speculations based on current trends and ongoing research. Here are potential future uses for each activation function:
- Sigmoid Activation Function: The sigmoid activation function might continue to find applications in binary classification problems and as an activation function for the output layer in specific cases where the output needs to be in the (0, 1) range, such as probability estimation tasks.

- Hyperbolic Tangent (Tanh) Activation Function: The tanh activation function might see continued use in multi-class classification problems and recurrent neural networks (RNNs) as it provides a range from (-1 to 1), which can be helpful in certain applications.
- Rectified Linear Unit (ReLU) Activation Function: ReLU is likely to remain a popular choice for activation functions in many applications due to its simplicity and effectiveness in preventing the vanishing gradient problem. Its variants, such as Leaky ReLU and Parametric ReLU, might continue to be used in specific cases where the original ReLU has limitations.
- Exponential Linear Unit (ELU) Activation Function: ELU has shown promise in addressing some of the limitations of ReLU, particularly for negative inputs, by introducing a smooth function. Its use might increase, especially in applications where preventing dead neurons and reducing overfitting are crucial.
- Swish Activation Function: Swish has gained attention as a potential alternative to ReLU, and its future use might increase, particularly as researchers and practitioners explore its performance in various applications.
- Softmax Activation Function: The softmax activation function will continue to be widely used in multi-class classification tasks, especially for applications involving natural language processing (NLP) and computer vision, where predicting probabilities for multiple classes is essential.

In the future, as deep learning research advances and new activation functions are proposed, we may see a greater diversity of activation functions tailored to specific tasks. Researchers might design task-specific activation functions that optimize certain properties, such as improved convergence, reduced vanishing/exploding gradients, and better generalization.
Additionally, neural architecture search and autoML techniques might help discover novel activation functions or adapt existing ones to specific problems, leading to further customization and optimization.
It's important to note that the future use of activation functions will largely depend on their performance and suitability for different tasks. As the field of deep learning evolves, we can expect new innovations and improvements in activation functions and their applications.

It's essential to note that the adoption of SELU, PReLU, or any other activation function in future applications will depend on their effectiveness in specific tasks, empirical evaluations, and advancements in deep learning research. New activation functions or improved variants might also emerge in the future, providing alternatives or even better performance in different scenarios. As the field of deep learning continues to evolve, researchers and practitioners will likely explore and optimize activation functions to enhance the performance of neural networks in various domains.

CONCLUSION

From the overall study it is concluded that activation functions are vital for the success of neural networks, introducing nonlinearity and enabling complex pattern recognition. Each function has its strengths and weaknesses, with ReLU and its variants being widely used due to their simplicity and effectiveness in avoiding vanishing gradients. Newer functions like ELU and Swish offer promising alternatives. The choice of activation function depends on the task and architecture. Future research may focus on task-specific functions and advancements in neural architecture search. Activation functions will continue to drive progress in artificial intelligence and machine learning.

In conclusion, activation functions serve as essential building blocks in artificial neural networks, enabling the introduction of nonlinearity and empowering networks to learn complex relationships within data. Throughout this research paper, we have examined various commonly used activation functions, including sigmoid, tanh, ReLU, ELU, Swish, and others, and analyzed their strengths and weaknesses.

Throughout the paper, we have highlighted considerations related to activation functions, such as empirical evaluation, smoothness, suitability for specific architectures, and their impact in various domains. Activation functions play a vital role in tasks ranging from image recognition and natural language processing to robotics and drug discovery. Looking ahead, future research might focus on developing task-specific activation functions optimized for particular problems and exploring advancements in neural architecture search to discover novel activation functions.