A blue and white logo

AI-generated content may be incorrect.

**NIMRA SAFA   
SP22-BCS-049  
ASSIGNMENT NO 04   
TOPIC: ACTIVATION FUNCTION  
SUBMITTED TO : DR.FAISAL AZAM**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Sr No.** | **Activation Function** | **Feature/Target** | **Input Domain** | **Limitations** | **Explanation** |
| 1 | Sigmoid (logistic)[1] | Binary outputs [0,1] | x∈R | Non-zero centered, saturation → vanishing gradient; slow learning | Smooth S-curve: σ(x)=1/(1+e^-x).Good for probability outputs, but gradients vanish at extremes. |
| 2 | Tanh [2] | Zero-centered outputs [–1,1] | x∈R | Still suffers from gradient saturation | Scaled sigmoid: tanh(x)=(e^x − e^−x)/(e^x+e^−x). Better centered, still vanishes at extremes. |
| 3 | ReLU [3] | Sparse activations | x>0⇒x; x≤0⇒0 | Dead neurons, non-zero centered, unbounded output | Fast, simple: introduces linearity for positives, zeros negatives. Standard for hidden layers. |
| 4 | Leaky ReLU  [4] | Address dead ReLU | x>0⇒x; x≤0⇒αx | Small negative slope doesn’t fully solve centering; α hyper Param needed | Leaky variant allows negative gradients to avoid “dying” neurons. |
| 5 | Softplus [7] | Smooth ReLU approximation | ℝ | Slower, softer saturation | f(x)=ln(1+eˣ); smooth and differentiable. |
| 6 | Swish [8] | Smooth, non-monotonic; improved accuracy | ℝ | Slower, β parameter | f(x)=x·sigmoid(x); outperforms ReLU in deep nets. |
| 7 | GELU [9] | Smooth gating; used in transformers | ℝ | Gaussian CDF needed; heavy | f(x)=x·Φ(x); smooth alternative to ReLU. |
| 8 | Linear (Identity) [10] | ℝ → ℝ; regression outputs | ℝ | No non-linearity; must combine with others | g(x)=x; used in regression or bypass layers. |
| 9 | Softmax [6] | ℝⁿ → probability distribution | ℝⁿ | Only for output layers; class-size cost | e^{zᵢ}/Σe^{zⱼ}; standard for classification outputs. |
| 10 | ELU (Exponential LU)[5] | Smooth negative outputs | Same as ReLU domain | More complex, slower computation than ReLU | Linear for positives; exponential approach to −α for negatives—helps mean stay near zero. |

**Explanation :**

* Sigmoid is intuitive for binary outputs but leads to vanishing gradients in deep nets.
* Tanh improves centering (range [–1,1]) but still saturates.
* ReLU is efficient and avoids saturation for positives but can “kill” neurons on the negative side.
* Leaky ReLU and ELU attempt to correct the “dying neuron” and centering issues.
* Softmax is essential for probabilistic multi-class output layers.
* Swish and GELU are modern alternatives that offer smoother gradients and better learning in deep architectures.
* Softplus is a smooth approximation of ReLU but less efficient.
* Linear is best reserved for regression tasks.

**References:**

1. H. Pratiwi *et al.*, “Sigmoid Activation Function in Selecting the Best Model of Artificial Neural Networks,” *Journal of Physics: Conference Series*, vol. 1471, p. 012010, Feb. 2020, doi: <https://doi.org/10.1088/1742-6596/1471/1/012010>.
2. Mian Mian Lau and King Hann Lim, “Review of Adaptive Activation Function in Deep Neural Network,” Dec. 2018, doi: <https://doi.org/10.1109/iecbes.2018.8626714>.
3. C. Banerjee, T. Mukherjee, and E. Pasiliao, “An Empirical Study on Generalizations of the ReLU Activation Function,” *Proceedings of the 2019 ACM Southeast Conference*, Apr. 2019, doi: <https://doi.org/10.1145/3299815.3314450>.
4. J. Xu, Z. Li, B. Du, M. Zhang, and J. Liu, “Reluplex made more practical: Leaky ReLU,” *2020 IEEE Symposium on Computers and Communications (ISCC)*, Jul. 2020, doi: <https://doi.org/10.1109/iscc50000.2020.9219587>.
5. Q. Zheng, D. Tan, and F. Wang, “Improved Convolutional Neural Network Based on Fast Exponentially Linear Unit Activation Function,” *IEEE Access*, vol. 7, pp. 151359–151367, Jan. 2019, doi: <https://doi.org/10.1109/access.2019.2948112>.
6. I. Kouretas and V. Paliouras, “Simplified Hardware Implementation of the Softmax Activation Function,” *IEEE Xplore*, May 01, 2019. <https://ieeexplore.ieee.org/document/8741677>.
7. K. Sun, J. Yu, L. Zhang, and Z. Dong, “A Convolutional Neural Network Model Based on Improved Softplus Activation Function,” *Advances in Intelligent Systems and Computing*, pp. 1326–1335, Jul. 2019, doi: <https://doi.org/10.1007/978-3-030-25128-4_164>.
8. Marina Adriana Mercioni and S. Holban, “P-Swish: Activation Function with Learnable Parameters Based on Swish Activation Function in Deep Learning,” Nov. 2020, doi: <https://doi.org/10.1109/isetc50328.2020.9301059>.
9. M. Lee, “Mathematical Analysis and Performance Evaluation of the GELU Activation Function in Deep Learning,” *Journal of Mathematics*, vol. 2023, pp. 1–13, Aug. 2023, doi: <https://doi.org/10.1155/2023/4229924>.
10. B. Yuen, M. T. Hoang, X. Dong, and T. Lu, “Universal activation function for machine learning,” *Scientific Reports*, vol. 11, no. 1, p. 18757, Sep. 2021, doi: <https://doi.org/10.1038/s41598-021-96723-8>.