Attention is All You Need

overview of the transformer architecture,
applications and improvements

Felix Karg

March 17, 2023

Institute for Theoretical Informatics:
Artificial Intelligence for Materials Science

Background

Background

Background

- **Multi-Layer Perceptron**
- Activation Functions
- Dropout
- Residual Connections

Multi-Layer Perceptron

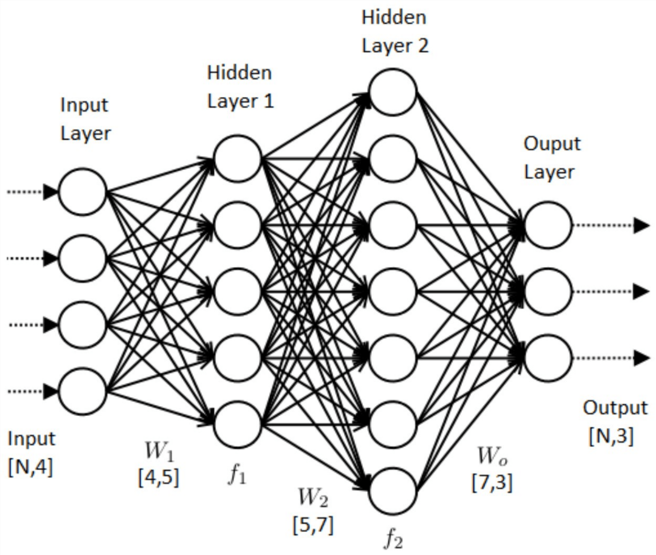
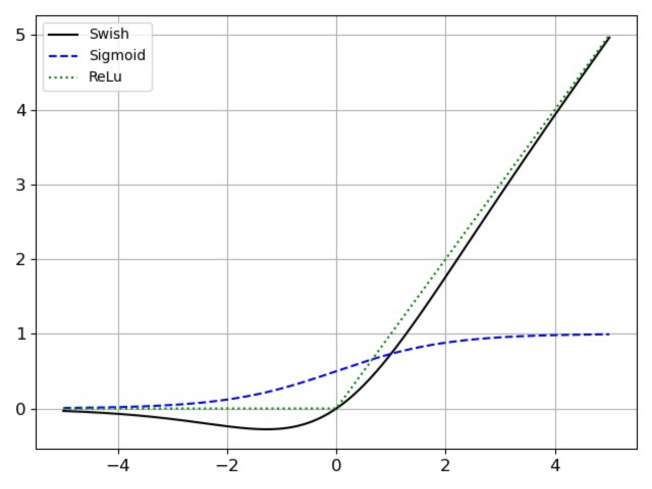


Image Source: Public Domain

Background

- Multi-Layer Perceptron
- **Activation Functions**
- Dropout
- Residual Connections

Common Activation Functions



$swish(x) := x * sigmoid(x)$

Image Source: [1]   SwiGLU introduced by [2]

Background

Multi-Layer Perceptron

Activation Functions

Dropout

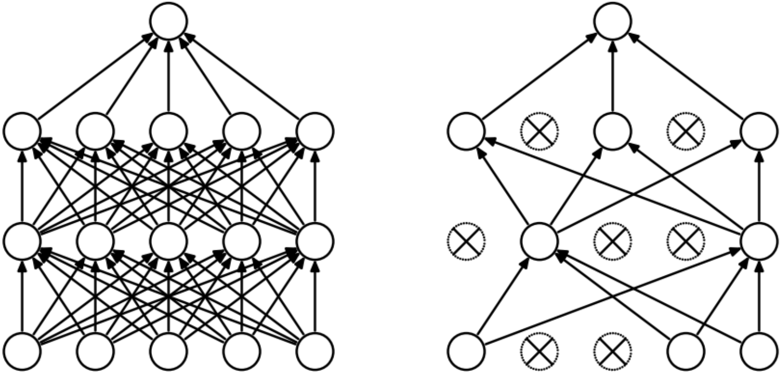
Residual Connections

Dropout I

Problem: neural network training results in highly specialized feature adaptations

”Complex co-adaptations can be trained to work well on a training set, but on novel test data they are far more likely to fail than multiple simpler co-adaptations that achieve the same thing.” [3]

Dropout II



(a) Standard Neural Net

(b) After applying dropout.

Image Source: [3]

Background

Multi-Layer Perceptron

Activation Functions

Dropout

Residual Connections

Residual Connections

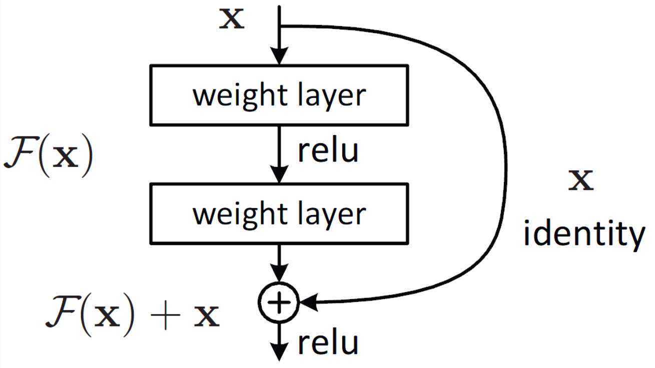


Image Source: [4]

Regularized Residual Connections

Self-Regulated Network [5]

Sources i

[1] H. Chen, A. Didisheim, and S. Scheidegger, “Deep structural estimation: With an application to option pricing,” *arXiv preprint arXiv:2102.09209*, 2021.

[2] N. Shazeer, “Glu variants improve transformer,” *arXiv preprint arXiv:2002.05202*, 2020.

[3] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, “Dropout: A simple way to prevent neural networks from overfitting,” *The journal of machine learning research*, vol. 15, no. 1, pp. 1929–1958, 2014.

Sources ii

[4] K. He, X. Zhang, S. Ren, and J. Sun, “Deep residual learning for image recognition,” in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 770–778, 2016.

[5] J. Xu, Y. Pan, X. Pan, S. Hoi, Z. Yi, and Z. Xu, “RegNet: Self-Regulated Network for Image Classification,” Jan. 2021.