

# Research Paper Analysis CS545: Gaussian Error Linear Units (GELUs)

Alex Shrestha

## **1 Introduction**

For this research paper analysis, I decided to focus on activation functions due to their substantial importance in neural networks. Activation functions introduce the much needed non-linearity and thus are fundamental to neural networks. Without non-linear activations, a very deep layered neural network could be represented by a single layer. The paper I will analyze is Gaussian Error Linear Units (GELUs) by Dan Hendrycks and Kevin Gimpel. The paper's goal is to introduce a new activation function called the Gaussian Error Linear Unit or GELU for short. The paper empirically evaluates how the GELU performs against standard non-linear activations ReLU and ELU by running experiments on the MNIST dataset, Twitter POS tagging dataset, TIMIT dataset, and the CIFAR-10/100 dataset.

## **2 Analysis**

## **3 Conclusion**

## **4 References**