Neural Arithmetic Units

Andreas Madsen[†‡]
amwebdk@gmail.com

Alexander Rosenberg Johansen[†]
alexander@herhjemme.dk

[†]Technical University of Denmark [‡]Computationally Demanding

Abstract

Neural networks often perform extremely when interpolating within the numerical range of the training dataset. However, when extrapolating outside this range they often fail, because the neural network have learned an approximation and not the exact rule. Previous work on neural arithmetic logic unit (NALU), attempts to solve this issue by learning exact arithmetic operations. Those results are unfortunately extremely difficult to reproduce consistently. In this paper we propose two alternative arithmetic units, one for addition and one for multiplication. We show that these units have fewer parameters, converges much more consistently, learns faster that previous arithmetic units, and the weight have meaningful discrete values.

1 Introduction

The ability for neurons to hold numbers and do arithmetic operations have been documented in both humans, non-human primates **?**, as well as newborn chicks **?**, and even bees **?**. These neurons are thus considered to be a vital component of intelligence.

In the context of neural networks, it have been repeatedly shown that neural networks are extremely capable of approximating complex unknown functions from a dataset. However, their ability to predict actually is often limited to interpolation within the data that the network have been trained with. Just slight variations outside of this training dataset domain **?**, can cause critical failures that may be dangerous in some applications. The task of extrapolating accurately, is thus a critical and important problem.

So far, there have not been much research in training exact arithmetic operations for extrapolation. The most important research, is by far, the recently proposed NAC and NALU **?**. However, as we will show both analytically and empirically, their results are highly exaggerated, as the proposed arithmetic units are extremely difficult to make convergence consistently. In fact, even without gating mechanism their multiplication operator can't converge on extremely simple problems.

Motivated by these convergence issue, we focus on the arithmetic units themself, without considering the gating mechanism in NALU. That is, we will assume that the appropriate required operation is already known, or can empirically be found by varying the network architecture which is very common when developing neural networks.

In the interrest of scientific integrity, we have made the code for all experiments, and more, available on GitHub: `https://github.com/AndreasMadsen/stable-nalu`.

Preprint. Under review.

2 Analysis of NAC and NALU

Acknowledgments

Use unnumbered third level headings for the acknowledgments. All acknowledgments go at the end of the paper. Do not include acknowledgments in the anonymized submission, only in the final paper.

References