Neural Network Training with Approximate Logarithmic Computations[1]

International Conference on Acoustics, Speech and Signal Processing

Arnab Sanyal, Peter A. Beerel, Keith M. Chugg

Ming Hsieh Department of Electrical & Computer Engineering
University of Southern California
Los Angeles, CA, USA

Virtual Tele-conference Presentation, May 2020

[1] Supported in part by the National Science Foundation (CCF-1763747).

- Enabling Neural Network training on edge-devices.

Goals

- Enabling Neural Network training on edge-devices.
- Computation reduction is of paramount importance.

Goals

Neural Network Training with Approximate Logarithmic Computations

- Enabling Neural Network training on edge-devices.
- Computation reduction is of paramount importance.
- Variety of approaches already exist - sparsity, pruning, quantization.

Goals

Neural Network Training with Approximate Logarithmic Computations

- Enabling Neural Network training on edge-devices.
- Computation reduction is of paramount importance.
- Variety of approaches already exist - sparsity, pruning, quantization.
- Our method – design end-to-end training in a logarithmic number system (LNS).

Goals

Neural Network Training with Approximate Logarithmic Computations

- Enabling Neural Network training on edge-devices.
- Computation reduction is of paramount importance.
- Variety of approaches already exist - sparsity, pruning, quantization.
- Our method – design end-to-end training in a logarithmic number system (LNS).
    - All NN operations needs to be defined in LNS

Goals

Neural Network Training with Approximate Logarithmic Computations

- Enabling Neural Network training on edge-devices.
- Computation reduction is of paramount importance.
- Variety of approaches already exist - sparsity, pruning, quantization.
- Our method – design end-to-end training in a logarithmic number system (LNS).
    - All NN operations needs to be defined in LNS
    - In LNS multiplications are cheap but addition are computationally expensive

Goals

Neural Network Training with Approximate Logarithmic Computations

- Enabling Neural Network training on edge-devices.
- Computation reduction is of paramount importance.
- Variety of approaches already exist - sparsity, pruning, quantization.
- Our method – design end-to-end training in a logarithmic number system (LNS).
  - All NN operations needs to be defined in LNS
  - In LNS multiplications are cheap but addition are computationally expensive
  - Resort to **Approximate Logarithmic Fixed-Point Computations**

- LNS independently invented and published as an alternative to floating-point number system[2]

[2] DOI: 10.1109/TC.1979.1675442

[2] DOI: 10.1109/T-C.1975.224172

Motivations and Previous Research in LNS

Neural Network Training with Approximate Logarithmic Computations

- LNS independently invented and published as an alternative to floating-point number system[2]
- Extensively explored in communications[3], processor design[4], re-configurable architectures[5], and a number of signal processing applications[6]

[2] DOI: 10.1109/TC.1979.1675442
[2] DOI: 10.1109/T-C.1975.224172
[3] **DOI: 10.1109/ICC.1995.524253**
[4] **DOI: 10.1109/ARITH.2011.15**
[5] **DOI: 10.1109/FPT.2006.270342**
[6] **DOI: 10.1049/el:19710039**

Motivations and Previous Research in LNS

Neural Network Training with Approximate Logarithmic Computations

- LNS independently invented and published as an alternative to floating-point number system[2]
- Extensively explored in communications[3], processor design[4], re-configurable architectures[5], and a number of signal processing applications[6]
- Explored in context of back-propagation but prior to resurgence of neural networks[7]

[2] DOI: 10.1109/TC.1979.1675442
[4] DOI: 10.1109/ARITH.2011.15
[2] DOI: 10.1109/T-C.1975.224172
[5] DOI: 10.1109/FPT.2006.270342
[7] **DOI: 10.1109/ICNN.1997.616150**
[3] DOI: 10.1109/ICC.1995.524253
[6] DOI: 10.1049/el:19710039

Motivations and Previous Research in LNS

- LNS encoded weights for inference[8] and extensions to LNS MACs restricted to positive numbers[9]

[8] DOI: 10.1109/ICASSP.2017.7953288

[9] arXiv: 1603.01025

Motivations and Previous Research in LNS

Neural Network Training with Approximate Logarithmic Computations

- LNS encoded weights for inference[8] and extensions to LNS MACs restricted to positive numbers[9]
- Log-encoding on posits relying on conversions to and from linear domain to perform additions[10]

[8] DOI: 10.1109/ICASSP.2017.7953288

[9] arXiv: 1603.01025

[10] **arXiv: 1811.01721**

Motivations and Previous Research in LNS

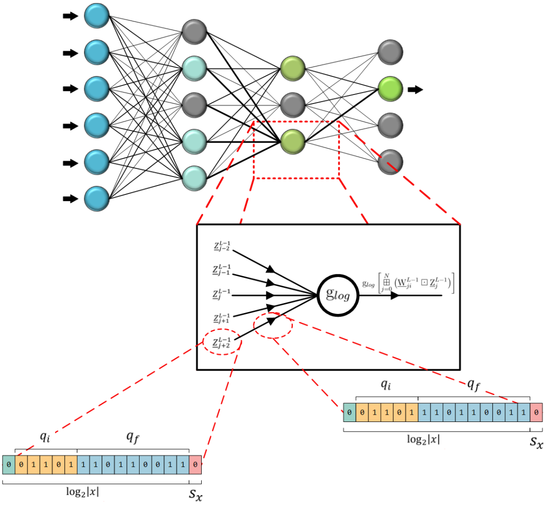
Neural Network Training with Approximate Logarithmic Computations

- LNS encoded weights for inference[8] and extensions to LNS MACs restricted to positive numbers[9]
- Log-encoding on posits relying on conversions to and from linear domain to perform additions[10]
- LNS circuit implementation for inference on pre-trained recurrent neural network[11]

[8] DOI: 10.1109/ICASSP.2017.7953288
[9] arXiv: 1603.01025
[10] arXiv: 1811.01721
[11] **DOI: 10.1109/MOCAST.2018.8376572**

LNS Neural Network Pipeline

Neural Network Training with Approximate Logarithmic Computations



Equivalence

$$
\begin{aligned}
v \longleftrightarrow \underline{V} &= (V, s_v) \\
V &= \log_2(|v|) \\
s_v &= \mathrm{sign}(v)
\end{aligned}
$$

Equivalence

$$
\begin{aligned}
v \longleftrightarrow \underline{V} &= (V, s_v) \\
V &= \log_2(|v|) \\
s_v &= \mathrm{sign}(v)
\end{aligned}
$$

Multiplication

$$
\begin{aligned}
u = xy \longleftrightarrow \underline{U} &= \underline{X} \boxdot \underline{Y} \\
U &= X + Y \\
s_u &= \overline{(s_x \veebar s_y)}
\end{aligned}
$$

Addition

$$z = x + y \longleftrightarrow \underline{Z} = \underline{X} \boxplus \underline{Y}$$

$$Z = \begin{cases} \max(X, Y) + \Delta_+ \left(|X - Y|\right) & s_x = s_y \\ \max(X, Y) + \Delta_- \left(|X - Y|\right) & s_x \neq s_y \end{cases}$$

$$s_z = \begin{cases} s_x & X > Y \\ s_y & X \leq Y \end{cases}$$

Addition

$$z = x + y \longleftrightarrow \underline{Z} = \underline{X} \boxplus \underline{Y}$$

$$Z = \begin{cases} \max(X, Y) + \Delta_{+}\left(|X - Y|\right) & s_x = s_y \\ \max(X, Y) + \Delta_{-}\left(|X - Y|\right) & s_x \neq s_y \end{cases}$$

$$s_z = \begin{cases} s_x & X > Y \\ s_y & X \leq Y \end{cases}$$

$$\Delta_{+}(d) = \log_2\left(1 + 2^{-d}\right) \qquad d \geq 0$$

$$\Delta_{-}(d) = \log_2\left(1 - 2^{-d}\right) \qquad d \geq 0$$

Subtraction

$$t = x - y \longleftrightarrow \underline{T} = \underline{X} \boxminus \underline{Y} = \underline{X} \boxplus (\underline{Y}, \overline{s_y})$$

Mathematical Operations

Neural Network Training with Approximate Logarithmic Computations

Subtraction	$t = x - y \longleftrightarrow \underline{T} = \underline{X} \boxminus \underline{Y} = \underline{X} \boxplus (Y, \overline{s_y})$
Exponentiation	$w = x^y \longleftrightarrow \underline{W} = (yX, 1)$

Mathematical Operations

Neural Network Training with Approximate Logarithmic Computations

Subtraction

$$t = x - y \longleftrightarrow \underline{T} = \underline{X} \boxminus \underline{Y} = \underline{X} \boxplus (Y, \overline{s_y})$$

Exponentiation

$$w = x^y \longleftrightarrow \underline{W} = (yX, 1)$$

Approximate additions

Subtraction	$t = x - y \longleftrightarrow \underline{T} = \underline{X} \boxminus \underline{Y} = \underline{X} \boxplus (Y, \overline{s_y})$
Exponentiation	$w = x^y \longleftrightarrow \underline{W} = (yX, 1)$

Approximate additions

- Approximate Δ to reduce addition computation complexity

Mathematical Operations

Neural Network Training with Approximate Logarithmic Computations

Subtraction	$t = x - y \longleftrightarrow \underline{T} = \underline{X} \boxminus \underline{Y} = \underline{X} \boxplus (Y, \overline{s_y})$
Exponentiation	$w = x^y \longleftrightarrow \underline{W} = (yX, 1)$

Approximate additions

- Approximate Δ to reduce addition computation complexity
- Two different approximations explored

Subtraction $$t = x - y \longleftrightarrow \underline{T} = \underline{X} \boxminus \underline{Y} = \underline{X} \boxplus (Y, \overline{s_y})$$

Exponentiation $$w = x^y \longleftrightarrow \underline{W} = (yX, 1)$$

Approximate additions

- Approximate Δ to reduce addition computation complexity
- Two different approximations explored
  - Look-Up Table (LUT) based approximations

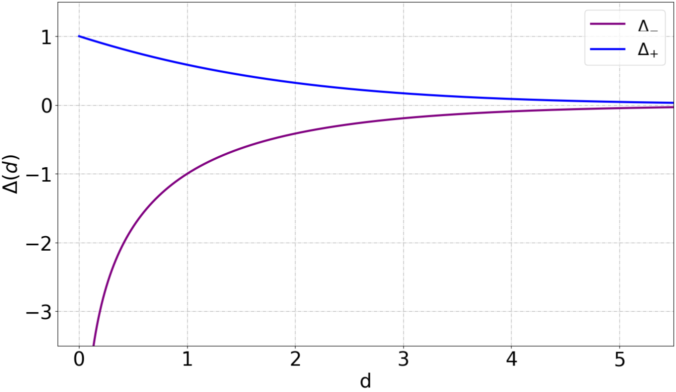
Subtraction	$t = x - y \longleftrightarrow \underline{T} = \underline{X} \boxminus \underline{Y} = \underline{X} \boxplus (Y, \overline{s_y})$
Exponentiation	$w = x^y \longleftrightarrow \underline{W} = (yX, 1)$

Approximate additions

- Approximate Δ to reduce addition computation complexity
- Two different approximations explored
  - Look-Up Table (LUT) based approximations
  - Bit-shift (BS) based approximations

Visualizing Δ from LNS ⊞ Additions

Neural Network Training with Approximate Logarithmic Computations



Look-Up Table Based Approximations

Neural Network Training with Approximate Logarithmic Computations

- Idea - Store the Δ terms as a Look-Up Table (LUT)

Look-Up Table Based Approximations

- Idea - Store the Δ terms as a Look-Up Table (LUT)
- LUT specified by three parameters – fixed point width, dynamic range, resolution

Look-Up Table Based Approximations

Neural Network Training with Approximate Logarithmic Computations

- Idea - Store the Δ terms as a Look-Up Table (LUT)
- LUT specified by three parameters – fixed point width, dynamic range, resolution
  - Dynamic range determined by fixed point width as Δ terms die down to 0

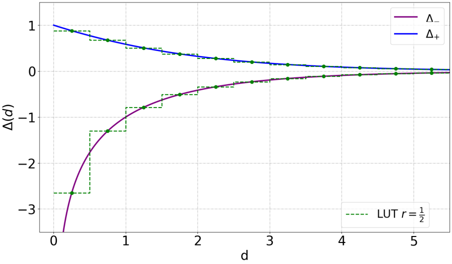
Look-Up Table Based Approximations

- Idea - Store the Δ terms as a Look-Up Table (LUT)
- LUT specified by three parameters – fixed point width, dynamic range, resolution
  - Dynamic range determined by fixed point width as Δ terms die down to 0
  - Resolution is a hyper-parameter

- Idea - Store the Δ terms as a Look-Up Table (LUT)
- LUT specified by three parameters – fixed point width, dynamic range, resolution
  - Dynamic range determined by fixed point width as Δ terms die down to 0
  - Resolution is a hyper-parameter
  - Fixed-point width is a hyper-parameter

Look-Up Table Based Approximations

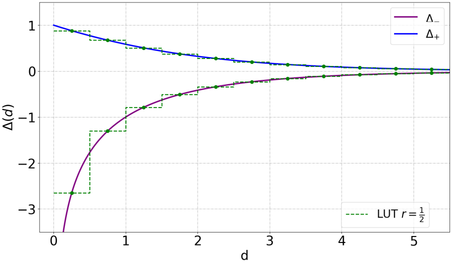
Neural Network Training with Approximate Logarithmic Computations

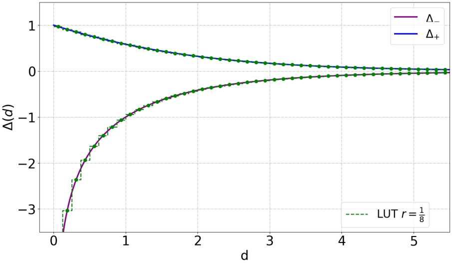


LUT Resolution $r = \frac{1}{2}$

Look-Up Table Based Approximations

Neural Network Training with Approximate Logarithmic Computations





LUT Resolution $r = \frac{1}{2}$

LUT Resolution $r = \frac{1}{8}$

Bit-Shift Based Approximations

Neural Network Training with Approximate Logarithmic Computations

Taylor Series approximation

$$\log_e(1 \pm x) \approx \pm x \qquad 0 \le x \ll 1$$

$$\Delta_{\pm}(d) = \log_2\left(1 \pm 2^{-d}\right) \approx \pm \log_2 e \times 2^{-d}$$

Bit-Shift Based Approximations

Neural Network Training with Approximate Logarithmic Computations

Taylor Series approximation

$$\log_e(1 \pm x) \approx \pm x \qquad 0 \leq x \ll 1$$

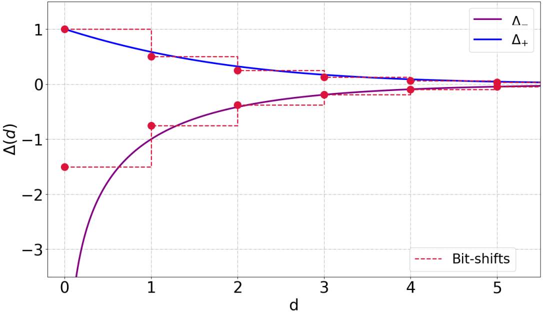
$$\Delta_{\pm}(d) = \log_2\left(1 \pm 2^{-d}\right) \approx \pm \log_2 e \times 2^{-d}$$

Further Simplification

$$\log_2 e = 1.442695\cdots \approx 1.4375 = 2^0 + 2^{-1} - 2^{-4} \approx 1.5 = 2^0 + 2^{-1} \approx 1 = 2^0$$

Bit-Shift Based Approximations

Neural Network Training with Approximate Logarithmic Computations



$$\Delta_+(d) \approx \mathbb{BS}\,(1, -d)$$
$$\Delta_-(d) \approx -\mathbb{BS}\,(1.5, -d)$$

Can be thought of as a LUT with $r = 1$

Multiply accumulate
(MAC)

The LNS Neuron

Neural Network Training with Approximate Logarithmic Computations

Multiply accumulate (MAC)

$$z_i = \sum_j w_{ij} x_j + b_i \longleftrightarrow \underline{Z}_i = \boxplus_j \underline{W}_{ij} \boxdot \underline{X}_j \boxplus \underline{B}_i$$

The LNS Neuron

Neural Network Training with Approximate Logarithmic Computations

Multiply accumulate (MAC)

$$z_i = \sum_j w_{ij} x_j + b_i \longleftrightarrow \underline{Z}_i = \boxplus_j \underline{W}_{ij} \boxdot \underline{X}_j \boxplus \underline{B}_i$$

LNS bit-width Constraint

The LNS Neuron

Multiply accumulate (MAC)

$$z_i = \sum_j w_{ij} x_j + b_i \longleftrightarrow \underline{Z}_i = \boxplus_j \underline{W}_{ij} \boxdot \underline{X}_j \boxplus \underline{B}_i$$

LNS bit-width Constraint

$$W_{\mathrm{log}} \geq 1 + \max\left(\left\lceil \log_2\left(b_i + 1\right)\right\rceil, \left\lceil \log_2 b_f \right\rceil\right) + W_{\mathrm{lin}}$$

Experiments suggest that $W_{\mathrm{log}} \approx W_{\mathrm{lin}}$ suffices in practice. In this work, set $W_{\mathrm{log}} = W_{\mathrm{lin}}$

LNS Weight Initialization

- To avoid hundreds of thousands of parameter initialization on prior distribution and taking logarithm of them, use standard change of measure approaches from probability to derive desired distribution

LNS Weight Initialization

- To avoid hundreds of thousands of parameter initialization on prior distribution and taking logarithm of them, use standard change of measure approaches from probability to derive desired distribution
- Weights generally initialized from symmetric distributions. Hence the sign parameter can be initialized randomly and independently of the magnitude from a $Bernoulli(\frac{1}{2})$ distribution.

LNS Weight Initialization

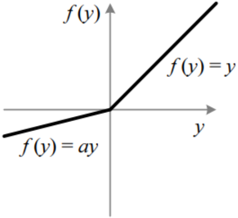
Neural Network Training with Approximate Logarithmic Computations

- To avoid hundreds of thousands of parameter initialization on prior distribution and taking logarithm of them, use standard change of measure approaches from probability to derive desired distribution
- Weights generally initialized from symmetric distributions. Hence the sign parameter can be initialized randomly and independently of the magnitude from a $Bernoulli(\frac{1}{2})$ distribution.
- The magnitude distribution for weights reduce to,

$$f_W(y) = 2^{y+1} \times \log_e 2 \times f_w(2^y)$$

The Log-Leaky ReLU Activation

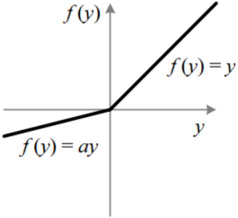
Neural Network Training with Approximate Logarithmic Computations



Parametric ReLU

The Log-Leaky ReLU Activation

Neural Network Training with Approximate Logarithmic Computations



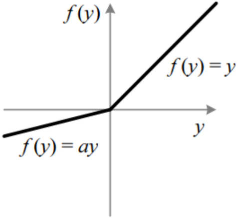
Parametric ReLU

- First proposed in ICCV 2015[12], fixes the *Dying ReLU* problem

[12] DOI: 10.1109/ICCV.2015.123

The Log-Leaky ReLU Activation

Neural Network Training with Approximate Logarithmic Computations



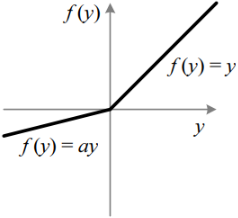
Parametric ReLU

- First proposed in ICCV 2015[12], fixes the *Dying ReLU* problem
- Attractive to this research as Dying ReLU could make activations ∞ in LNS

[12] DOI: 10.1109/ICCV.2015.123

The Log-Leaky ReLU Activation

Neural Network Training with Approximate Logarithmic Computations



Parametric ReLU

- First proposed in ICCV 2015[12], fixes the *Dying ReLU* problem
- Attractive to this research as Dying ReLU could make activations ∞ in LNS
- This brings us to *Log-Leaky ReLU*

$$g_{\text{llReLU}}\left((X, s_x)\,\middle|\,\beta\right) = \begin{cases} (X, s_x) & s_x = 1 \\ (X + \beta, s_x) & s_x = 0 \end{cases}$$

[12] DOI: 10.1109/ICCV.2015.123

LNS Soft-Max

Gradient calculation

LNS Soft-Max

Neural Network Training with Approximate Logarithmic Computations

Gradient calculation

$$p_{ij} = \frac{e^{a_{ij}}}{\sum_{j=1}^{N} e^{a_{ij}}}$$

$$\delta_{ij} = p_{ij} - y_{ij}$$

LNS Soft-Max

Gradient calculation

$$p_{ij} = \frac{e^{a_{ij}}}{\sum_{j=1}^{N} e^{a_{ij}}}$$

$$\delta_{ij} = p_{ij} - y_{ij}$$

Log-probabilities

LNS Soft-Max

Neural Network Training with Approximate Logarithmic Computations

Gradient calculation

$$p_{ij} = \frac{e^{a_{ij}}}{\sum_{j=1}^{N} e^{a_{ij}}}$$

$$\delta_{ij} = p_{ij} - y_{ij}$$

Log-probabilities

$$\log_2 p_{ij} = (a_{ij} \log_2 e) - \mathop{\boxplus}_{j=1}^{N} (a_{ij} \log_2 e, 1)$$

LNS Soft-Max

Neural Network Training with Approximate Logarithmic Computations

Gradient calculation

$$p_{ij} = \frac{e^{a_{ij}}}{\sum_{j=1}^{N} e^{a_{ij}}}$$

$$\delta_{ij} = p_{ij} - y_{ij}$$

Log-probabilities

$$\log_2 p_{ij} = (a_{ij} \log_2 e) - \boxplus_{j=1}^{N} (a_{ij} \log_2 e, 1)$$

LNS gradients

LNS Soft-Max

Gradient calculation

$$p_{ij} = \frac{e^{a_{ij}}}{\sum_{j=1}^{N} e^{a_{ij}}}$$

$$\delta_{ij} = p_{ij} - y_{ij}$$

Log-probabilities

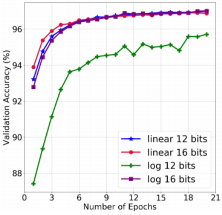
$$\log_2 p_{ij} = (a_{ij} \log_2 e) - \underset{j=1}{\overset{N}{\boxplus}} (a_{ij} \log_2 e, 1)$$

LNS gradients

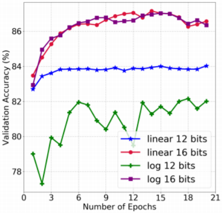
$$\left(\log_2 |\delta_{ij}|, s_{\delta_{ij}}\right) = \underline{P}_{ij} \boxminus \left(\log_2 |y_{ij}|, s_{y_{ij}}\right)$$

Numerical Results

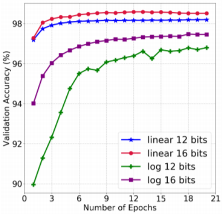
Neural Network Training with Approximate Logarithmic Computations



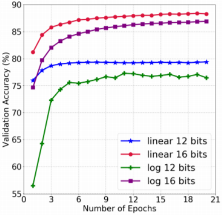
(a) MNIST



(b) Fashion-MNIST



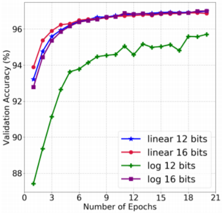
(c) EMNIST-digits



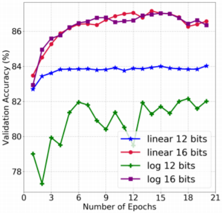
(d) EMNIST-letters

Numerical Results

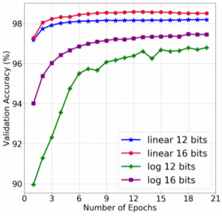
Neural Network Training with Approximate Logarithmic Computations



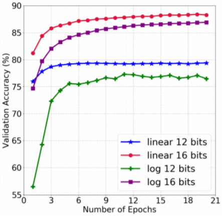
(a) MNIST



(b) Fashion-MNIST



(c) EMNIST-digits



(d) EMNIST-letters

Datasets	Float	**Linear-domain** fixed-point		**Log-domain** fixed-point look-up tables		**Log-domain** fixed-point bit-shifts	
		12b	16b	12b	16b	12b	16b
MNIST	97.4	97.3	96.9	96.0	97.2	95.5	96.5
FMNIST	87.1	82.8	88.0	80.5	87.1	79.3	85.7
EMNISTD	98.6	98.3	98.7	96.9	97.5	96.2	97.4
EMNISTL	88.1	79.7	88.7	76.4	86.7	73.7	82.5

Number of epochs trained $= 20$
Size of tables $= 20(r = \frac{1}{2})$; soft-max uses 640 element tables$(r = \frac{1}{64})$
Table shows test-set accuracy

Future Works

- Extend Future work to CNNs on harder datasets

- Extend Future work to CNNs on harder datasets
- Better Approximation Design

Future Works

- Extend Future work to CNNs on harder datasets
- Better Approximation Design
    - Functional Approximations using constrained optimization

Future Works

- Extend Future work to CNNs on harder datasets
- Better Approximation Design
  - Functional Approximations using constrained optimization
  - Replacing Soft-max Layer with multi-class Sigmoid

Future Works

Neural Network Training with Approximate Logarithmic Computations

- Extend Future work to CNNs on harder datasets
- Better Approximation Design
  - Functional Approximations using constrained optimization
  - Replacing Soft-max Layer with multi-class Sigmoid
- Reliability and Robustness Analysis

Future Works

- Extend Future work to CNNs on harder datasets
- Better Approximation Design
  - Functional Approximations using constrained optimization
  - Replacing Soft-max Layer with multi-class Sigmoid
- Reliability and Robustness Analysis
  - Cost-Accuracy trade-off across different approximations

Future Works

- Extend Future work to CNNs on harder datasets
- Better Approximation Design
  - Functional Approximations using constrained optimization
  - Replacing Soft-max Layer with multi-class Sigmoid
- Reliability and Robustness Analysis
  - Cost-Accuracy trade-off across different approximations
  - Weight-Activation Relation mapping for LNS-neurons using Supervised Learning (LDA, QDA)

Acknowledgements

This work is supported in part by the National Science Foundation (CCF-1763747)