Revise Saturated Activation Function

Idriss Cabrel TSEWALO TONDJI

African Master in Machine Intelligence, AMMI-Senegal

Bootcamp-Week1

August 20, 2021

Overview

Overview

1. Introduction
   - Motivation
   - Goal and Objective
2. Approach
   - Revised activation functions
3. Experiments
4. Results
5. Conclusion

Introduction

Motivation

- Activation functions play an important role in artificial neural networks.

Introduction

Motivation

- Activation functions play an important role in artificial neural networks.
- Bad choice of an activation function can affect the performance of the neural network.

Introduction

Motivation

- Activation functions play an important role in artificial neural networks.
- Bad choice of an activation function can affect the performance of the neural network.
- Saturated activation functions are less preferred (Sigmoid, etc.).

Introduction

Motivation

- Activation functions play an important role in artificial neural networks.
- Bad choice of an activation function can affect the performance of the neural network.
- Saturated activation functions are less preferred (Sigmoid, etc.).
- Avoid the gradient vanishing/exploding problem.

Introduction

Motivation

- Activation functions play an important role in artificial neural networks.
- Bad choice of an activation function can affect the performance of the neural network.
- Saturated activation functions are less preferred (Sigmoid, etc.).
- Avoid the gradient vanishing/exploding problem.
- Two saturated functions with different performance into a network.

Overview

Introduction

Goal and Objective

- **Goal :** Provide more insights about the effect of different activation functions on the performance of the neural networks.

Introduction

Goal and Objective

- **Goal :** Provide more insights about the effect of different activation functions on the performance of the neural networks.
- **Objectives :** Revise two commonly used saturated functions, the logistic sigmoid and the hyperbolic tangent (tanh) .

AIMS African Institute for Mathematical Sciences SENEGAL

Overview

Approach

Scaled sigmoid

$$f(x) = 4 * sigmoid(x) - 2 = \frac{4}{1+\exp^{-x}} - 2$$

Approach

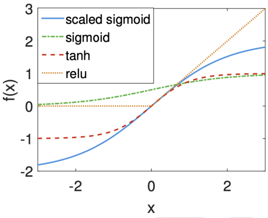
Scaled sigmoid

$$f(x) = 4 * sigmoid(x) - 2 = \frac{4}{1+\exp^{-x}} - 2$$

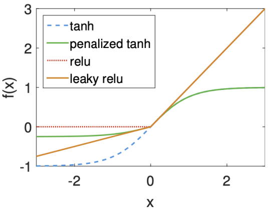
Penalized tanh

$$f(x) = \begin{cases} tanh(x) & \text{if } x < 0, \\ a * tanh(x) & \text{otherwise} \end{cases}.$$

Approach



(a) Compare the scaled logistic sigmoid with other activation function.



(b) Compare the penalized tanh with other functions.

Data set and architecture

- We use CIFAR-10 as data set.
- Implement a simple neural network architecture from scratch with two layers (forward and backward pass).

Results

Personal experiments

Activation	Train Accuracy	Test Accuracy
Sigmoid	11.68%	9.89%
Scaled sigmoid	90.19%	78.4%
tanh	97.9%	77.4%
ReLU	86.9%	72.19%
Penalized tanh ($a = 0.25$)	99.8%	84.19%
leaky ReLU ($a = 0.25$)	99.9%	83.5%

Table 1: Vary activation function for two layers nn on CIFAR-10

Conclusion

- The logistic sigmoid achieves comparable results with tanh with proper rescaling.
- "Penalized tanh" is comparable and even outperforms the state-of-the-art non-saturated functions including ReLU and leaky ReLU on deep convolution neural networks.

Conclusion

- The logistic sigmoid achieves comparable results with tanh with proper rescaling.
- "Penalized tanh" is comparable and even outperforms the state-of-the-art non-saturated functions including ReLU and leaky ReLU on deep convolution neural networks.

We can extend this work by doing more experiments in deep architectures to better understand activation functions, and also evaluate the convergence rate in each configuration.

References

Bing Xu, Ruitong Huang, Mu Li
Revise Saturated Activation Function.
2016.

Acknowledgements

Thanks for your attention!