The NUGGET Non-Linear Piecewise Activation

Stephen Merity [1]

Abstract

The choice of activation functions in deep neural networks has a significant impact on the training dynamics, task performance, and potential acronyms of resulting work. While numerous activation functions have been proposed, such as the Rectified Linear Unit (ReLU), most are derived from the domain of mathematics rather than by drawing inspiration from nature. We propose a non-linear piecewise activation function, the NUGGET activation function, which is a result of a complex zero-sum pricing game refined over decades of multi-agent interaction simulation. We verify the effectiveness of the activation by experimental analysis on the Modified National Institute of Standards and Technology (MNIST) digits task (Neural Numerology) and achieve state of the art results[1].

1. Introduction

The need for effective activation functions has fueled a rapid exploration of all mathematical functions. This is problematic for those of us still scared of mathematics. As such, a counter culture of human curated artisanal activation functions has emerged.

Dropout (Srivastava et al., 2014) may be the first instance of a human curated artisanal regularization technique that entered wide scale use in machine learning. Dropout, simply described, is the concept that if you can learn how to do a task repeatedly whilst drunk, you should be able to do the task even better when sober. This insight has resulted in numerous state of the art results and a nascent field dedicated to preventing dropout from being used on neural networks.

Our work seeks inspiration from the natural world in providing new and intuitive manners to frame and explore recent neural network advances. In the following sections we analyze a specific subset of these naturally occurring activation and regularization techniques, which we shall broadly refer to as NUGGET functions, to understand the impact they may have when applied to neural networks.

[1] Our state of the art results can be seen as state of the art results by ignoring the current state of the art.

2. The NUGGET n-player zero-sum game

The chicken nugget was invented in the 1950s by Robert C. Baker, a food science professor at Cornell University, and published as unpatented academic work. Since then, it has been a pivotal component in the raging fast food wars that have beseiged the nations across earth. Speculation exists that SpaceX (Musk, 2002) was started in an attempt to escape the ever looming threat of NUGGET warfare. Given the intense research, both theoretical and experimental, in determining both NUGGET pricing and strategy, the NUGGET anthologies contain rich labeled data for analysis and conversion to an ill-defined neural network component.

2.1. Data Collection

To acquire sufficient diversified samples for our task, we conducted a large scale user study. To avoid paying participants, we relied on good will (Friendship, 1901) and the unsubstantiated claim that paying participants would skew the accuracy and impartiality of the scientific results.

Our geographically diverse dataset of NUGGET pricing activations comes from multiple samples across 8 countries: 2 from Brazil, 3 from Australia, 2 from the continental United States, 1 from Germany, 1 from Malaysia, 1 from Thailand, 1 from the United Kingdom[2], and 1 from Japan. All participants in the user study found one or more instantiations of NUGGET during their search, though this might be a result of sampling bias[3].

2.2. Non-linear NUGGET pricing

Rational consumers would expect that the price of a box of NUGGET should increase linearly (or sub-linearly) as the quantity of NUGGET is increased. From both individual experiments in NUGGET acquisition and from our user study however we found this to not consistently be the case.

[2] The authors note that United Kingdom should be United Queendom whilst within a queen's reign but note this is out of the scope of this work.

[3] The authors would like to know how to handle sampling biases but carefully note that statistics is rarely used in machine learning and that the Monty Hall problem is still highly confrontational, suggesting all later forms of statistics must be equally confrontational. That's induction, right? Ugh, wait, that's math :(

We propose taking advantage of these naturally occurring non-linearities to power our activation functions and show that heavily used existing activation functions, such as the Rectified Linear Unit (ReLU), fit within this framework.

The ReLU activation, mathematically defined as

$$\mathrm{ReLU}(x) = \max(0, x)$$

represents the optimal NUGGET pricing as determined by a rational consumer. The price of a box of NUGGET should increase proportionally to the amount of NUGGET received. The $\max$ is a result of consumers being unable to return or resell any amount of NUGGET to the original producer of the NUGGET box[4].

Even this cursory analysis suggests that the ReLU function, traditionally attributed to , should be attributed to Professor Robert C. Baker, creator of the NUGGET. We feel this is a grave oversight in the current neural network literature. Our work suggests researchers have issues with maintaining and tracking long term literature depedendencies, potentially due to truncated backpropagation through time.

Motivated by this rediscovery, we investigate whether other non-linear NUGGET activations may act as a catalyst for the training and production of neutral neural networks when subjected to a generative adversarial setting[5].

In Table 1 and 2, we explore non-linear pricing for a NUGGET box in San Francisco, United States, for both McDonalds and Burger King (or Hungry Jacks in Australian). Note the price per NUGGET unit fluctuates wildly between $\$0.149$ and ∞.

3. Experiments

3.1. The Neural Numerology dataset

The Neural Numerology (MNIST) dataset contains 60,000 labeled images of digits used to specify the quantity of a given NUGGET box.

Subjects were not required to make sensible orders, resulting in orders of a zero NUGGET box and none where the NUGGET quantity exceeded nine. Future work will rectify this and allow for NUGGET boxes of ten to twenty.

[4]The authors attempted multiple times to resell uneaten NUGGET quantities to various fast food retailers. None of the initial trials resulted in success and all subsequent attempts were met with a denial of service (i.e. we were asked to leave the store).

[5]The authors do note that The Matrix (1999) can be seen as a non-continuous generative adversarial multi-agent simulation. In following work (Animatrix (2003), Reloaded (2003), Revolutions (2003)), experimentation on humans in this manner was deemed unethical. We note that the ethical treatment of neural networks when subjected to adversarial settings has not yet been thoroughly discussed in the literature but opt to ignore this insight by pretending this troubling question had never been raised in the first place.

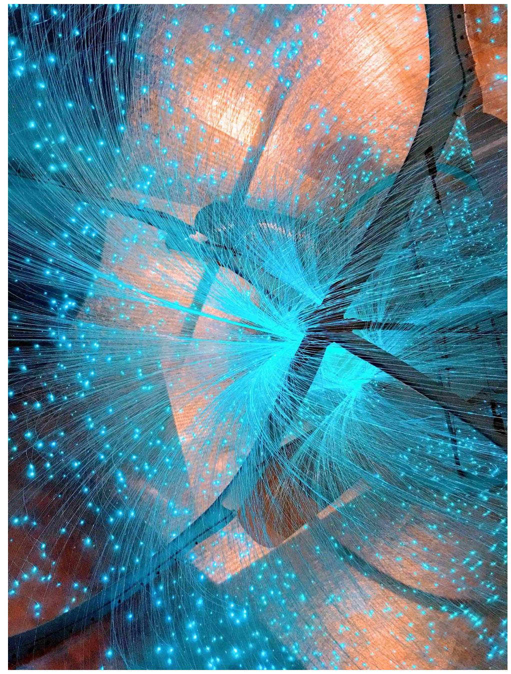

Figure 1. An architectural neuronal visualization produced when using the NUGGET activation is substantially more aesthetic than that of non-NUGGET based activation functions. Note the absence of killer robots or glowing red eyes.

Nuggets	Om nom	Dollary-doos	NUGGET unit
$\alpha = 0$	$\emptyset$	$\$0.00$	∞
$\alpha = 4$	XX	$\$1.00$	$\$0.25$
$\alpha = 6$	X	$\$4.30$	$\$0.72$
$\alpha = 10$	X	$\$4.99$	$\$0.499$
$\alpha = 20$	XXXX	$\$5.00$	$\$0.25$

Table 1. Non-linear NUGGET pricing at a McDonalds located in contintental United States. At one extreme, increasing NUGGET quantity by 2 results in $1.65 per NUGGET unit ($4 \rightarrow 6$). At the other extreme, increasing NUGGET quantity by 10 results in $0.001 per NUGGET unit ($10 \rightarrow 20$).

Nuggets	Om nom	Dollary-doos	NUGGET unit
$\alpha = 0$	$\emptyset$	$\$0.00$	∞
$\alpha = 10$	XX	$\$1.49$	$\$0.149$
$\alpha = 20$	X	$\$5.99$	$\$0.299$

Table 2. Non-linear NUGGET pricing at a Burger King located in contintental United States. Note two $n = 10$ NUGGET boxes is cheaper than an $n = 20$ NUGGET box. We are uncertain if gold or other valuable items are in the $n = 20$ NUGGET box.

Figure 2. (Left) Neural Numerology samples generated without NUGGET activations. *(Right)* Neural Numerology samples generated with NUGGET activations. Notice the zeroes (0) have similar topology to that of a traditional NUGGET blob.

3.2. Experimental setup

All experiments are implemented in PyTorch and are built upon existing codebases. The use of existing code is essential as researchers are still investigating how to make digital neurons feel warm and fuzzy [6]. We elect not to use weight or batch normalization as the authors are concerned with negatively impacting the neural network's body image. For the same reason, we avoid using L_1 or L_2 regularization.

We considered using the Hogwild lock-free approach to parallelizing stochastic gradient descent but elected against it as hogs are not operationally equivalent to chickens and thus may invalidate our results.

The neural network models were trained by a person named Adam Optimizer and used an NVIDIA Volta whilst it was mining for Ethereum. The learning rate began at 20 and was divided each time the training curator Adam desired a NUGGET box of quantity one or more. This was frequent.

All embedding weights were uniformly initialized in the interval $[-0.1, 0.1]$ and all other weights were initialized between $[-\frac{1}{\sqrt{H}}, \frac{1}{\sqrt{H}}]$, where H is the hidden size. Anyone who guessed what the hidden size was won a prize.

4. Results

Our results ... are not that bad. Like, if you hired a five year old to read the numbers in Figure 2 for you, that kid would probably do worse than our algorithm. Therefore, NUGGET based artificially intelligent models are equivalent in complexity to that of a standard human five year old.

[6]Many neural network experiments require dozens or hundreds of expensive high end GPUs, resulting in both massive expense and massive heat generation. This is necessary as it helps incubate the neural networks during their growth, with the GPUs helping heat them to their optimal temperature (i.e. acting as a catalyst) and the dollar figure spent on them ensuring the neural networks are aware of how much we love them.

That's pretty darn good. Few animals can read numbers or order nuggets, so our model is also smarter than most animals and evolution took *forever* making those things.

5. Conclusion

In this work, we revisit the ReLU activation under the framework of NUGGET based non-linear piecewise equations. The improvements that these techniques provide can likely be combined with other regularization techniques, such as the drunken dropout, and may lead to further improvements in performance as well, especially if subjected to an extensive global NUGGET hyperparameter search. We see artisanal hand crafted activation and regularization techniques the future of our field, primarily as no-one is quite certain how a neural nets anyway.

Acknowledgements

We thank Charlie Yang for funding an experimental purchase of an $n = 20$ NUGGET box that motivated this work. Additional NUGGET funders have opted to remain anonymous due to the contentious nature of artificially intelligent fast food research. Thanks to the participants in the geographical NUGGET sampling: Anton Troynikov, Joseph Stephen, Dominic Balasuriya, Georgina Wilcox, James Foster, Joshua Hall, Kenya Chan, Dominick Ng, and Vivian Li. Good research not only takes time and resources but also good friends. The authors would perform better work if they had more friends. Please be our friend.

NUGGET samples

Sydney: 3 for \$3, 6 for \$6, 10 for \$7.50, 20 for \$12.75
Sydney CBD: 3 for \$3, 6 for \$5.90, 10 for \$7.70, 20 for \$12.80
Melbourne: 3 for \$3, 6 for \$5.50, 10 for \$7.20, 20 for \$12.80, 24 for \$9.95
Japan: 5 for 200 yen, 15 for 570 yen
UK: 6 for 3.09, 9 for 3.99, 20 for 4.99
Thailand: 6 for 87B, 10 for 139B, 20 for 240B
Kuala Lumpur: 6 for 7.8RM, 9 for 10.9RM, 20 for 22RM
Germany: 6 for €3,59, 9 for €4,49, 20 for €7,59
Belo Horizonte: 4 for 6.50 reais, 10 for 16.40 reais
So Paulo: 4 for 6.50 reais, 10 for 13.90 reais
US (McDonalds): see Table 1
US (Burger King): see Table 2

References

Srivastava, Nitish, Hinton, Geoffrey E., Krizhevsky, Alex, Sutskever, Ilya, and Salakhutdinov, Ruslan. Dropout: a simple way to prevent neural networks from overfitting. *Journal of Machine Learning Research*, 15:1929–1958, 2014.