

Novel Neural Networks

Notes from a Heilbronn Summer Workshop

Richard Pinch

Heilbronn Institute for Mathematical Research

rgep@chalconedon.co.uk

September 2026



Abstract

We report on successful practical experiments carried out on 2-adic neural networks at the HIMR Summer Workshop 2024; work on a suitable analogue of the sign function, a universal approximation theorem and backpropagation.

Introduction

One of the topics of the Heilbronn Summer Workshop held in Bristol in 2024 was to explore novel neural networks based on 2-adic arithmetic. Although radically different from floating point, this can be implemented efficiently using conventional integer arithmetic and so, without the need for floating-point circuits, might represent a practical competitor to current neural network technologies.

The proposal developed a prototype able to compare the behaviour and performance of such a network with a conventional neural net.

Background

The p -adics

We recall the p -adic valuation on the integers to be $\nu_p(n)$ to be the power of p that exactly divides n : that is, $\nu_p(n) = r$ if $p^r | n$ and $p^{r+1} \nmid n$. This extends to the rational numbers by $\nu_p(a/b) = \nu_p(a) - \nu_p(b)$.

We define the p -adic integers $\mathbb{Z}_p$ to be compatible sequences of integers modulo increasing powers of p ; equivalently, as infinite sequences of base p digits extending to the left. The p -adic rationals $\mathbb{Q}_p$ form the field of fractions of $\mathbb{Z}_p$: base p expansions with finitely many digits to the right of the p -ary point.

We define the p -adic *sign* to be the character $\mathbb{Q}_p^*/\mathbb{Q}_p^{*2}$: note that this has four values for odd p and eight for $p = 2$.

Neural networks

We consider a neural network as a function from some input set I to some output set O , obtained by composition in layers of basic many-to-one functions: each such function is a multilinear function of its inputs with coefficients (“weights”) followed by a fixed unary non-linear “activation” function. The network is trained on a set of inputs with prescribed outputs by assigning weights that minimise a suitable loss function.

Incidence networks and VC dimension

We define an incidence system $\mathcal{I}$ between a set R of rows and a set C of columns as a function from $R \times C \rightarrow \{0, 1\}$. We think of the rows as being hypotheses about the columns. The VC dimension of $\mathcal{I}$ is the largest d such that there exists a set of d columns for which all 2^d possible tuples occur on some row (or ∞ if unbounded).

As an example, let R be the linear functionals $P(\mathbb{R}^{n+1})$ and C be the points of R^n . Consider the incidence system $\mathcal{I}$ where

$\mathcal{I}((w_0, w_1, \dots, w_n), (x_1, \dots, x_n)) = 1$ iff $w_0 + w_1x_1 + \dots + w_nx_n \geq 0$. The VC dimension of this system is $n + 1$: that is, $n + 1$ points in R^n can be shattered by hyperplanes, but not $n + 2$.

The VC dimension enters into explicit formulae for the rate of convergence of statistical learning for neural networks implemented in this way.

We observe that this definition represents a simple formula in the first-order theory of the real numbers as an ordered field.

Models with NIP

Consider a language of terms and a model with a base set B : let $\phi(\mathbf{w}, \mathbf{x})$ be a fixed formula in the language. Consider the incidence system $\mathcal{I}_\phi$ defined by taking all possible interpretations of ϕ over B with 0, 1 denoting false, true respectively. If all such incidence systems have finite VC dimension, then the model has the *Non-Independence Property*.

Such models have been extensively studied, and sufficient conditions for NIP include o -minimality, including the theory of the real as an ordered field (Tarski) and that theory extended by exponential (Wilkie) or subanalytic functions. As a corollary we obtain the finite dimensionality of almost all currently implemented neural networks.

We consider the theory of C -minimal systems, based on closeness or C -relations (Adeleke and Neumann): the ultrametric property of p -adic distance is an example. We can embed the p -adic numbers in such a system and obtain a corresponding guarantee of statistical learnability.

Results

The workshop developed the details of a 2-adic neural network and experimented with a prototype code base intended to provide a direct comparison between real and 2-adic implementations on standard datasets such as MNIST. We concluded:

- 2-adic neural networks are sound and possess classification properties similar to real arithmetic implementations;
- there should be Universal Approximation Theorems;
- we were unable to devise an efficient backpropagation comparable to gradient descent;
- open publication of these results was appropriate to stimulate further research and development.

Acknowledgements

The work on this project was led by Richard Pinch with participation from Dan Fretwell, Kevin Hughes, Wilfrid Kendall and others at the HIMR SWAMP in Summer 2024. Code was generously shared by Andy R and Sam F.

I am solely responsible for the opinions expressed in this poster.