p-Adic Neural Networks and Quantum Field Theory

W. A. Zúñiga-Galindo

University of Texas Rio Grande Valley

Introduction

- The deep neural networks have been sucessfully applied to a very large tasks, including self driving cars, natural language processing, visual recognition, among many others.

Introduction

- The deep neural networks have been sucessfully applied to a very large tasks, including self driving cars, natural language processing, visual recognition, among many others.
- There is consensus about the need of developing a theoretical framework to understand how the deep learning architectures work.

Introduction

- The deep neural networks have been sucessfully applied to a very large tasks, including self driving cars, natural language processing, visual recognition, among many others.
- There is consensus about the need of developing a theoretical framework to understand how the deep learning architectures work.
- Recently, pysicists have proposed the existence of a correspondence between neural networks (NNs) and quantum field theories (QFTs), more precisely with Euclidean QFTs

Introduction

- We are studying of the mentioned correspondence in the framework of the non-Archimedean statistical field theory (SFT). The aim of this talk is to report the results obtained.

Introduction

- We are studying of the mentioned correspondence in the framework of the non-Archimedean statistical field theory (SFT). The aim of this talk is to report the results obtained.
- The p-adics are organized in a tree-like structure, this feature facilitates the description of hierarchical architectures.

Introduction

- We are studying of the mentioned correspondence in the framework of the non-Archimedean statistical field theory (SFT). The aim of this talk is to report the results obtained.
- The p-adics are organized in a tree-like structure, this feature facilitates the description of hierarchical architectures.
- We have shown that the p-adic versions of the convolutional restricted Boltzmann machines (RBMs) are universal approximators.

Introduction

- We are studying of the mentioned correspondence in the framework of the non-Archimedean statistical field theory (SFT). The aim of this talk is to report the results obtained.
- The p-adics are organized in a tree-like structure, this feature facilitates the description of hierarchical architectures.
- We have shown that the p-adic versions of the convolutional restricted Boltzmann machines (RBMs) are universal approximators.
- We also have implemented p-adic RBMs to recognize handwritten digits.

The p-adic numbers

- From now on, p denotes a fixed prime number. Any non-zero $p-$adic number x has a unique expansion of the form

$$x = x_{-k}p^{-k} + x_{-k+1}p^{-k+1} + \ldots + x_0 + x_1 p + \ldots, \tag{1}$$

with $x_{-k} \neq 0$, where k is an integer, and the x_js are numbers from the set $\{0, 1, \ldots, p-1\}$.

The p-adic numbers

- From now on, p denotes a fixed prime number. Any non-zero $p-$adic number x has a unique expansion of the form

$$x = x_{-k}p^{-k} + x_{-k+1}p^{-k+1} + \ldots + x_0 + x_1 p + \ldots, \tag{1}$$

  with $x_{-k} \neq 0$, where k is an integer, and the x_js are numbers from the set $\{0, 1, \ldots, p-1\}$.

- The set of all possible sequences of form (1) constitutes the field of p-adic numbers $\mathbb{Q}_p$. There are natural field operations, sum and multiplication, on series of form (1).

The p-adic numbers

- From now on, p denotes a fixed prime number. Any non-zero $p-$adic number x has a unique expansion of the form

$$x = x_{-k}p^{-k} + x_{-k+1}p^{-k+1} + \ldots + x_0 + x_1 p + \ldots, \tag{1}$$

with $x_{-k} \neq 0$, where k is an integer, and the x_js are numbers from the set $\{0, 1, \ldots, p-1\}$.

- The set of all possible sequences of form (1) constitutes the field of p-adic numbers $\mathbb{Q}_p$. There are natural field operations, sum and multiplication, on series of form (1).

- There is also a natural norm in $\mathbb{Q}_p$ defined as $|x|_p = p^k$, for a nonzero p-adic number x of the form (1).

The p-adic numbers

- The field of p-adic numbers with the distance induced by $|\cdot|_p$ is a complete ultrametric space.

The p-adic numbers

- The field of p-adic numbers with the distance induced by $|\cdot|_p$ is a complete ultrametric space.
- The ultrametric property refers to the fact that $|x-y|_p \leq \max\left\{|x-z|_p, |z-y|_p\right\}$ for any x, y, z in $\mathbb{Q}_p$.

The p-adic numbers

- The field of p-adic numbers with the distance induced by $|\cdot|_p$ is a complete ultrametric space.
- The ultrametric property refers to the fact that $|x-y|_p \leq \max\left\{|x-z|_p, |z-y|_p\right\}$ for any x, y, z in $\mathbb{Q}_p$.
- A p-adic integer is a sequence of the form $x_k p^k + x_{k+1} p^{k+1} + \ldots + x_0 + x_1 p + \ldots$,with $k \geq 0$. All these sequences constitute the unit ball $\mathbb{Z}_p$.

The p-adic numbers

- The field of p-adic numbers with the distance induced by $|\cdot|_p$ is a complete ultrametric space.
- The ultrametric property refers to the fact that $|x-y|_p \leq \max\left\{|x-z|_p, |z-y|_p\right\}$ for any x, y, z in $\mathbb{Q}_p$.
- A p-adic integer is a sequence of the form $x_k p^k + x_{k+1} p^{k+1} + \ldots + x_0 + x_1 p + \ldots$,with $k \geq 0$. All these sequences constitute the unit ball $\mathbb{Z}_p$.

- A function $\varphi : \mathbb{Z}_p \to \mathbb{R}$, supported in the unit ball, is called a locally constant function if $\varphi(a+x) = \varphi(a)$ for $|x|_p \leq p^{-m}$, where the integer $m \geq 0$ is independent of a.

The p-adic numbers

- The field of p-adic numbers with the distance induced by $|\cdot|_p$ is a complete ultrametric space.
- The ultrametric property refers to the fact that $|x-y|_p \leq \max\left\{|x-z|_p, |z-y|_p\right\}$ for any x, y, z in $\mathbb{Q}_p$.
- A p-adic integer is a sequence of the form $x_k p^k + x_{k+1} p^{k+1} + \ldots + x_0 + x_1 p + \ldots$,with $k \geq 0$. All these sequences constitute the unit ball $\mathbb{Z}_p$.
- A function $\varphi : \mathbb{Z}_p \to \mathbb{R}$, supported in the unit ball, is called a locally constant function if $\varphi(a+x) = \varphi(a)$ for $|x|_p \leq p^{-m}$, where the integer $m \geq 0$ is independent of a.
- We denote by $\mathcal{D}(\mathbb{Z}_p)$ the real vector space of test functions supported in the unit ball.

The p-adic numbers

- There is a natural integration theory so that $\int_{\mathbb{Z}_p} \varphi(x)\, dx$ gives a well-defined real number. The measure dx is the Haar measure of $\mathbb{Q}_p$.

The p-adic numbers

- There is a natural integration theory so that $\int_{\mathbb{Z}_p} \varphi(x)\, dx$ gives a well-defined real number. The measure dx is the Haar measure of $\mathbb{Q}_p$.
- The computation with p-adic integers requires a truncation process: $x = x_0 + x_1 p + \ldots + x_{l-1} p^{l-1}$, $l \geq 1$.

The p-adic numbers

- There is a natural integration theory so that $\int_{\mathbb{Z}_p} \varphi(x)\, dx$ gives a well-defined real number. The measure dx is the Haar measure of $\mathbb{Q}_p$.
- The computation with p-adic integers requires a truncation process: $x = x_0 + x_1 p + \ldots + x_{l-1} p^{l-1}$, $l \geq 1$.
- The set all truncated integers mod p^l is denote as $G_l = \mathbb{Z}_p / p^l \mathbb{Z}_p$. This set is a rooted tree with l levels.

The p-adic numbers

- There is a natural integration theory so that $\int_{\mathbb{Z}_p} \varphi(x)\, dx$ gives a well-defined real number. The measure dx is the Haar measure of $\mathbb{Q}_p$.
- The computation with p-adic integers requires a truncation process: $x = x_0 + x_1 p + \ldots + x_{l-1} p^{l-1}$, $l \geq 1$.
- The set all truncated integers mod p^l is denote as $G_l = \mathbb{Z}_p / p^l \mathbb{Z}_p$. This set is a rooted tree with l levels.
- The unit ball $\mathbb{Z}_p$ is an infinite rooted tree with fractal structure.

The p-adic numbers

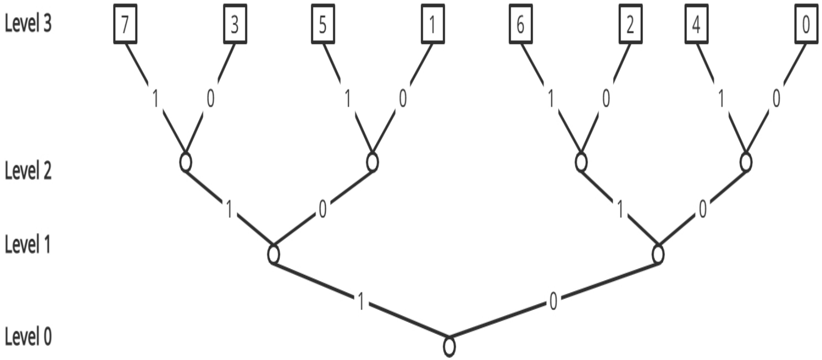


Figure: The rooted tree associated with the group $\mathbb{Z}_2/2^3\mathbb{Z}_2$. The elements of $\mathbb{Z}_2/2^3\mathbb{Z}_2$ have the form $i = i_0 + i_1 2 + i_2 2^2$, i_0, i_1, $i_2 \in \{0, 1\}$.

The Bruhat-Schwartz space

We denote by $\Omega\left(p^{l}\left|x-i\right|_{p}\right)$, the characteristic function of the ball $i+p^{l}\mathbb{Z}_{p}$.

A function φ in $\mathcal{D}(\mathbb{Z}_p)$ can be written as $\varphi\left(x\right)=\sum\limits_{i\in G_{l}}\varphi\left(i\right)\Omega\left(p^{l}\left|x-i\right|_{p}\right)$, for some $l\geq 1$, with $i\in\mathbb{Z}_{p}$, $\varphi\left(i\right)\in\mathbb{R}$. Then

$$\begin{aligned}\int\limits_{\mathbb{Z}_{p}}\varphi\left(x\right)dx &= \sum\limits_{i\in G_{l}}\varphi\left(i\right)\int\limits_{\mathbb{Z}_{p}}\Omega\left(p^{l}\left|x-i\right|_{p}\right)dx\\ &= \sum\limits_{i\in G_{l}}\varphi\left(i\right)\int\limits_{i+p^{-l}\mathbb{Z}_{p}}dx=p^{-l}\sum\limits_{i\in G_{l}}\varphi\left(i\right).\end{aligned}$$

Non-Archimedean statistical field theories

- We fix $a(x)$, $b(x)$, $c(x)$, $d(x) \in \mathcal{D}(\mathbb{Z}_p)$, $e \in \mathbb{R}$, and an integrable function $w(x) : \mathbb{Z}_p \rightarrow \mathbb{R}$.

Non-Archimedean statistical field theories

- We fix $a(x)$, $b(x)$, $c(x)$, $d(x) \in \mathcal{D}(\mathbb{Z}_p)$, $e \in \mathbb{R}$, and an integrable function $w(x) : \mathbb{Z}_p \to \mathbb{R}$.
- A *p-adic continuous Boltzmann machine (or a p-adic continuous RBM)* is a $\{\mathbf{v}, \mathbf{h}\}^4$-statistical field theory in $\mathcal{D}(\mathbb{Z}_p) \times \mathcal{D}(\mathbb{Z}_p)$.

Non-Archimedean statistical field theories

- We fix $a(x)$, $b(x)$, $c(x)$, $d(x) \in \mathcal{D}(\mathbb{Z}_p)$, $e \in \mathbb{R}$, and an integrable function $w(x) : \mathbb{Z}_p \to \mathbb{R}$.
- A *p-adic continuous Boltzmann machine (or a p-adic continuous RBM)* is a $\{\mathbf{v}, \mathbf{h}\}^4$-statistical field theory in $\mathcal{D}(\mathbb{Z}_p) \times \mathcal{D}(\mathbb{Z}_p)$.
- The function $\mathbf{v}(x) \in \mathcal{D}(\mathbb{Z}_p)$ is called the *visible field* and the function $\mathbf{h}(x) \in \mathcal{D}(\mathbb{Z}_p)$ is called the *hidden field*.

Non-Archimedean statistical field theories

- We fix $a(x)$, $b(x)$, $c(x)$, $d(x) \in \mathcal{D}(\mathbb{Z}_p)$, $e \in \mathbb{R}$, and an integrable function $w(x) : \mathbb{Z}_p \to \mathbb{R}$.
- A *p-adic continuous Boltzmann machine (or a p-adic continuous RBM)* is a $\{\mathbf{v}, \mathbf{h}\}^4$-statistical field theory in $\mathcal{D}(\mathbb{Z}_p) \times \mathcal{D}(\mathbb{Z}_p)$.
- The function $\mathbf{v}(x) \in \mathcal{D}(\mathbb{Z}_p)$ is called the *visible field* and the function $\mathbf{h}(x) \in \mathcal{D}(\mathbb{Z}_p)$ is called the *hidden field*.
- The field $\{\mathbf{v}, \mathbf{h}\}$ performs thermal fluctuations, assuming that the expectation value of the field is zero, the fluctuations take place around zero.

Non-Archimedean statistical field theories

The size of the fluctuations is controlled by an energy functional of the form $E(\mathbf{v},\mathbf{h};\boldsymbol{\theta}):=E(\mathbf{v},\mathbf{h})=E_0(\mathbf{v},\mathbf{h})+E_{\text{int}}(\mathbf{v},\mathbf{h})$, where $\boldsymbol{\theta}=(w,a,b,c,d,e)$. The first term

$$E_0(\mathbf{v},\mathbf{h})=-\int\limits_{\mathbb{Z}_p}a(x)\mathbf{v}(x)\,dx-\int\limits_{\mathbb{Z}_p}b(x)\mathbf{h}(x)\,dx+$$

$$\frac{e}{2}\int\limits_{\mathbb{Z}_p}\mathbf{v}^2(x)\,dx+\frac{e}{2}\int\limits_{\mathbb{Z}_p}\mathbf{h}^2(x)\,dx,$$

is an analogue of the *free-field energy*. The second term

$$E_{\text{int}}(\mathbf{v},\mathbf{h})=-\iint\limits_{\mathbb{Z}_p\times\mathbb{Z}_p}\mathbf{h}(y)\,w(x-y)\,\mathbf{v}(x)\,dxdy$$

$$+\int\limits_{\mathbb{Z}_p}c(x)\mathbf{v}^4(x)\,dx+\int\limits_{\mathbb{Z}_p}d(x)\mathbf{h}^4(x)\,dx$$

is an analogue of the *interaction energy*.

Non-Archimedean statistical field theories

- All thermodynamic properties of the system are described by the partition function of the fluctuating fields, which is defined as

$$Z^{\text{phys}} = \int d\mathbf{v}d\mathbf{h}\ e^{-\frac{E(\mathbf{v},\mathbf{h})}{K_B T}},$$

where K_B is the Boltzmann constant and T is the temperature.

Non-Archimedean statistical field theories

- All thermodynamic properties of the system are described by the partition function of the fluctuating fields, which is defined as

$$Z^{\mathrm{phys}} = \int d\mathbf{v}d\mathbf{h}\ e^{-\frac{E(\mathbf{v},\mathbf{h})}{K_B T}},$$

  where K_B is the Boltzmann constant and T is the temperature.
- The measure $d\mathbf{v}d\mathbf{h}$ is ill-defined. It is expected that such measure can be defined rigorously by a limit process.

Non-Archimedean statistical field theories

- All thermodynamic properties of the system are described by the partition function of the fluctuating fields, which is defined as

$$Z^{\text{phys}} = \int d\mathbf{v}d\mathbf{h}\ e^{-\frac{E(\mathbf{v},\mathbf{h})}{K_B T}},$$

  where K_B is the Boltzmann constant and T is the temperature.
- The measure $d\mathbf{v}d\mathbf{h}$ is ill-defined. It is expected that such measure can be defined rigorously by a limit process.
- The statististical field theory corresponding to the energy functional $E(\mathbf{v}, \mathbf{h}; \boldsymbol{\theta})$ is the ill-defined probability measure

$$\mathbf{P}^{\text{phys}}(\mathbf{v}, \mathbf{h}; \boldsymbol{\theta}) = d\mathbf{v}d\mathbf{h}\frac{\exp\left(-E(\mathbf{v}, \mathbf{h})\right)}{Z^{\text{phys}}}$$

  on space of functions $\mathcal{D}(\mathbb{Z}_p) \times \mathcal{D}(\mathbb{Z}_p)$, where $\boldsymbol{\theta} = (\mathbf{w}, \mathbf{a}, \mathbf{b})$.

Discrete SFTs and p-adic discrete Boltzmann machines

- In the p-adic world, *the discrete functions are a particular case of the p-adic continuous functions*, more precisely, $\mathcal{D}(\mathbb{Z}_p) = \cup_l \mathcal{D}^l(\mathbb{Z}_p)$ and $\mathcal{D}^l(\mathbb{Z}_p) \subset \mathcal{D}^{l+1}(\mathbb{Z}_p)$.

Discrete SFTs and p-adic discrete Boltzmann machines

- In the p-adic world, *the discrete functions are a particular case of the p-adic continuous functions*, more precisely, $\mathcal{D}(\mathbb{Z}_p) = \cup_l \mathcal{D}^l(\mathbb{Z}_p)$ and $\mathcal{D}^l(\mathbb{Z}_p) \subset \mathcal{D}^{l+1}(\mathbb{Z}_p)$.
- The test functions in $\mathcal{D}^l(\mathbb{Z}_p)$ have the form $\varphi(x) = \sum_{i \in G_l} \varphi(i)\, \Omega\left(p^l |x-i|_p\right)$, $\varphi(i) \in \mathbb{R}$, where $i = i_0 + i_1 p + \ldots + i_{l-1} p^{l-1} \in G_l = \mathbb{Z}_p / p^l \mathbb{Z}_p$, $l \geq 1$, and $\Omega\left(p^l |x-i|_p\right)$ is the characteristic function of the ball $i + p^l \mathbb{Z}_p$. Here, it is important to notice that G_l is a finite, Abelian, additive group.

Discrete SFTs and p-adic discrete Boltzmann machines

- In the p-adic world, *the discrete functions are a particular case of the p-adic continuous functions*, more precisely, $\mathcal{D}(\mathbb{Z}_p) = \cup_l \mathcal{D}^l(\mathbb{Z}_p)$ and $\mathcal{D}^l(\mathbb{Z}_p) \subset \mathcal{D}^{l+1}(\mathbb{Z}_p)$.
- The test functions in $\mathcal{D}^l(\mathbb{Z}_p)$ have the form
$\varphi(x) = \sum_{i \in G_l} \varphi(i)\, \Omega\left(p^l \left|x - i\right|_p\right)$, $\varphi(i) \in \mathbb{R}$, where
$i = i_0 + i_1 p + \ldots + i_{l-1} p^{l-1} \in G_l = \mathbb{Z}_p / p^l \mathbb{Z}_p$, $l \geq 1$, and
$\Omega\left(p^l \left|x - i\right|_p\right)$ is the characteristic function of the ball $i + p^l \mathbb{Z}_p$.
Here, it is important to notice that G_l is a finite, Abelian, additive group.
- The discretization of a p-adic SFT is constructed by restricting the energy functional $E(\mathbf{v}, \mathbf{h}; \boldsymbol{\theta})$ to a finite dimensional vector subspace $\mathcal{D}^l(\mathbb{Z}_p)$ of the space of test functions $\mathcal{D}(\mathbb{Z}_p)$.

Discrete SFTs and p-adic discrete Boltzmann machines

By taking $\mathbf{v}, \mathbf{h} \in \mathcal{D}^{l}(\mathbb{Z}_p)$ and l succiently large, the restriction of the energy functional $E(\mathbf{v}, \mathbf{h}; \boldsymbol{\theta})$ to $\mathcal{D}^{l}(\mathbb{Z}_p)$ has the form

$$
\begin{aligned}
E_l(\mathbf{v}, \mathbf{h}; \boldsymbol{\theta}) = & -\sum_{j \in G_l} \sum_{k \in G_l} w_k v_{j+k} h_j - \sum_{j \in G_l} a_j v_j \\
& - \sum_{j \in G_l} b_j h_j + \frac{e}{2} \sum_{j \in G_l^N} v_j^2 + \frac{e}{2} \sum_{j \in G_l} h_j^2 \\
& + \sum_{j \in G_l} c_j v_j^4 + \sum_{j \in G_l} d_j h_j^4 .
\end{aligned} \tag{2}
$$

This is the energy functional for a discrete $\{\mathbf{v}, \mathbf{h}\}^4$-STF.

Discrete SFTs and p-adic discrete Boltzmann machines

- By identifying $\mathbf{v} = [v_i]_{i \in G_l}$ and $\mathbf{h} = [h_i]_{i \in G_l}$, the Boltzmann probability distribution attached to $E_l(\mathbf{v}, \mathbf{h}; \boldsymbol{\theta})$ is given by

$$\mathbf{P}_l(\mathbf{v}, \mathbf{h}; \theta) = \frac{\exp\left(-E_l\left((\mathbf{v}, \mathbf{h}; \boldsymbol{\theta})\right)\right)}{\sum\limits_{\mathbf{v},\mathbf{h}} \exp\left(-E_l\left(\mathbf{v}, \mathbf{h}; \boldsymbol{\theta}\right)\right)}.$$

A p-adic discrete RBM is the RBM attached to $\mathbf{P}_l(\mathbf{v}, \mathbf{h}; \boldsymbol{\theta})$.

Restricted Boltzmann machines

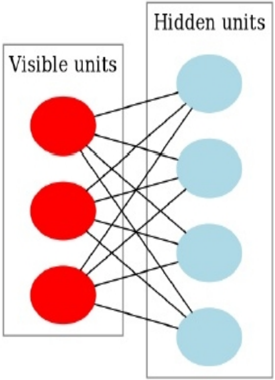
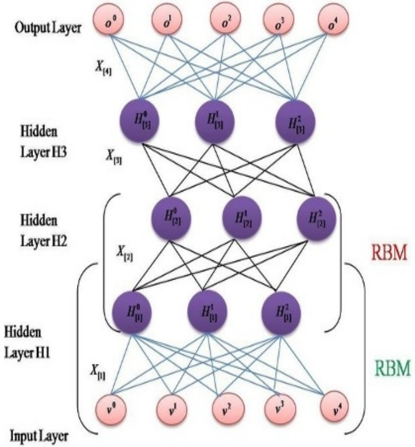


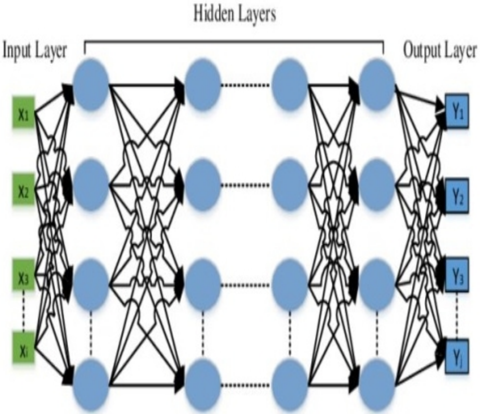
Diagram of a restricted Boltzmann machine with three visible units and four hidden units (no bias units).

$$E(v,h) = -\sum_i a_i v_i - \sum_j b_j h_j - \sum_i \sum_j v_i w_{i,j} h_j$$

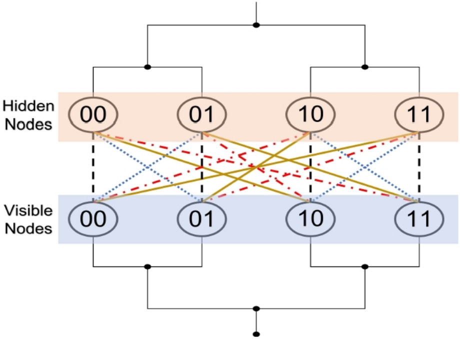
$$P(v,h) = \frac{1}{Z} e^{-E(v,h)}$$

Deep Boltzmann machines (Deep belief networks)





p-Adic deep belief networks



In a p-adic discrete RBM the visible and hidden states are functions on a finitite tree. Only the vertices at the top level, the orange and blue balls, are allowed to have states. The rest of the vertices in the trees codify the hierarchical relations between the states.

p-Adic deep belief networks

- The standard RBMs and the ϕ^4- neural networks introduced by Batchits et al. are particular classes of p-adic discrete RBMs.

p-Adic deep belief networks

- The standard RBMs and the ϕ^4- neural networks introduced by Batchits et al. are particular classes of p-adic discrete RBMs.
- The properties of the p-adic discrete RBMs depend strongly on the kernel w used in the definition of the interaction energy.

p-Adic deep belief networks

- The standard RBMs and the ϕ^4- neural networks introduced by Batchits et al. are particular classes of p-adic discrete RBMs.
- The properties of the p-adic discrete RBMs depend strongly on the kernel w used in the definition of the interaction energy.
- If $w(x, y)$ is a test function and the interaction between the visible and hidden field has the form $-\iint_{\mathbb{Z}_p \times \mathbb{Z}_p} \mathbf{h}(y)\, w(x, y)\, \mathbf{v}(x)\, dxdy$, then in the corresponding discrete energy functional, the interaction of the visible and hidden states takes the form $-\sum_{j \in G_l} \sum_{k \in G_l} w_{k,j} v_k h_j$, after a suitable rescaling of the weights $w_{k,j}$.

p-Adic deep belief networks

- The standard RBMs and the ϕ^4- neural networks introduced by Batchits et al. are particular classes of p-adic discrete RBMs.
- The properties of the p-adic discrete RBMs depend strongly on the kernel w used in the definition of the interaction energy.
- If $w(x,y)$ is a test function and the interaction between the visible and hidden field has the form $-\iint_{\mathbb{Z}_p \times \mathbb{Z}_p} \mathbf{h}\left(y\right) w\left(x,y\right) \mathbf{v}\left(x\right) dxdy$, then in the corresponding discrete energy functional, the interaction of the visible and hidden states takes the form $-\sum_{j \in G_l} \sum_{k \in G_l} w_{k,j} v_k h_j$, after a suitable rescaling of the weights $w_{k,j}$.
- Here $\left[w_{k,j}\right]$ is an ordinary matrix, which means that its entries $w_{k,j}$ do not depend on the algebraic structure of G_l neither on its topology.

p-Adic deep belief networks

- In the case in which the interaction between the visible and hidden field has the form $-\iint_{\mathbb{Z}_p\times\mathbb{Z}_p}\mathbf{h}\left(y\right)w\left(x-y\right)\mathbf{v}\left(x\right)dxdy$, then corresponding discrete energy functional depends on the group structure of G_l, and the corresponding neural network is a particular case of a DBN.

p-Adic deep belief networks

- In the case in which the interaction between the visible and hidden field has the form $-\iint_{\mathbb{Z}_p\times\mathbb{Z}_p} \mathbf{h}(y)\, w(x-y)\, \mathbf{v}(x)\, dxdy$, then corresponding discrete energy functional depends on the group structure of G_l, and the corresponding neural network is a particular case of a DBN.
- The parameter $l \geq l_0$, for some constant l_0, this means that a p-adic discrete RBM admits arbitratly large copies; this is a consequence of the fact of p-adic numbers has a tree-like structure.

p-Adic deep belief networks

- In the case in which the interaction between the visible and hidden field has the form $-\iint_{\mathbb{Z}_p \times \mathbb{Z}_p} \mathbf{h}(y)\, w(x-y)\, \mathbf{v}(x)\, dxdy$, then corresponding discrete energy functional depends on the group structure of G_l, and the corresponding neural network is a particular case of a DBN.
- The parameter $l \geq l_0$, for some constant l_0, this means that a p-adic discrete RBM admits arbitratly large copies; this is a consequence of the fact of p-adic numbers has a tree-like structure.
- In the p-adic framework the discretization of SFTs can be done in simple and rigorous way.

p-Adic deep belief networks

- We implement a p-adic discrete Boltzmann machine for processing binary images, then $\mathbf{v},\mathbf{h}:\mathbb{Z}_p\rightarrow\{0,1\}\subset\mathbb{R}$, in this case, we use an energy functional of the form

$$E_l(\mathbf{v},\mathbf{h};\theta)=-\sum_{j\in G_l}\sum_{\substack{k\in G_l\\|k|_p\leq p^{-N}}}w_k v_{j+k}h_j-\sum_{j\in G_l}a_j v_j-\sum_{j\in G_l}b_j h_j,$$

where $N\geq l$, which means that we take $e=0$, $c_j=d_j=0$ for $j\in G_l$ because the quadratic and biquadratic terms do not play any role in the case in which $\mathbf{v},\mathbf{h}$ are binary variables.

p-Adic deep belief networks

- We implement a p-adic discrete Boltzmann machine for processing binary images, then $\mathbf{v}, \mathbf{h} : \mathbb{Z}_p \to \{0,1\} \subset \mathbb{R}$, in this case, we use an energy functional of the form

$$E_l(\mathbf{v}, \mathbf{h}; \theta) = -\sum_{j \in G_l} \sum_{\substack{k \in G_l \\ |k|_p \leq p^{-N}}} w_k v_{j+k} h_j - \sum_{j \in G_l} a_j v_j - \sum_{j \in G_l} b_j h_j,$$

  where $N \geq l$, which means that we take $e = 0$, $c_j = d_j = 0$ for $j \in G_l$ because the quadratic and biquadratic terms do not play any role in the case in which $\mathbf{v}, \mathbf{h}$ are binary variables.

- The condition $N > l$ implies that the convolution operation involves only a small neighborhood of a given pixel, while the condition $N = l$ means that the convolution involves all the pixels in the image.

p-Adic deep belief networks

- We use $p = 3$ and $l = 6$, which means that we implement a six layer feature detector. **The prime number p is determined by sized of the images to be processed.**

p-Adic deep belief networks

- We use $p = 3$ and $l = 6$, which means that we implement a six layer feature detector. **The prime number p is determined by sized of the images to be processed.**
- We implement two different types of networks. In the first type, the function w_k is supported in the entire tree G_6; in the second type, the function w_k is supported in a proper subset of the tree G_6.

p-Adic deep belief networks

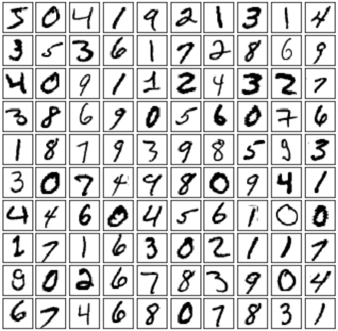
- We use $p = 3$ and $l = 6$, which means that we implement a six layer feature detector. **The prime number p is determined by sized of the images to be processed.**
- We implement two different types of networks. In the first type, the function w_k is supported in the entire tree G_6; in the second type, the function w_k is supported in a proper subset of the tree G_6.
- We use the full MNIST hadwritten digits, without considering labels, to train a six layer 3-adic feature detector.

p-Adic deep belief networks

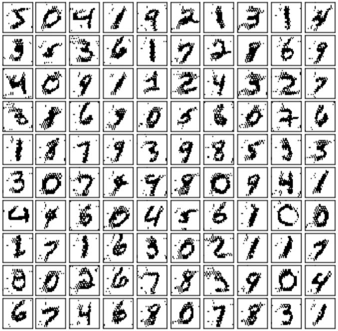
- We use $p = 3$ and $l = 6$, which means that we implement a six layer feature detector. **The prime number p is determined by sized of the images to be processed.**
- We implement two different types of networks. In the first type, the function w_k is supported in the entire tree G_6; in the second type, the function w_k is supported in a proper subset of the tree G_6.
- We use the full MNIST hadwritten digits, without considering labels, to train a six layer 3-adic feature detector.
- The results are show in Figure 4. The choice of $p = 3$ and $l = 6$ is due to the fact that we use $3^3 \times 3^3$ images. In general p and l depend on the size of the images to be processed. Thus, typically p is small, 2 or 3.

p-Adic deep belief networks

Original Images

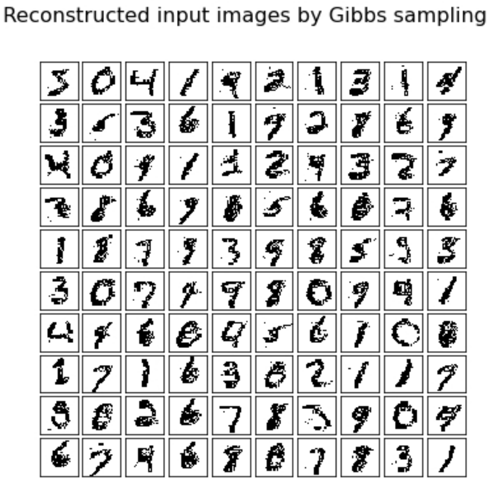


Reconstructed input images by Gibbs sampling



Left side is the original input image. Right side is the image reconstructed using Gibbs sampling with w having $\mathbb{Z}_3$ as support

p-Adic deep belief networks



Recosntructed image with w supported in $3^4\ \mathbb{Z}_3$.

The p-adic DBNs are universal approximators

We consider the problem of approximating $\mathbf{Q}(\mathbf{v})$ by the marginal distribution $\mathbf{P}_l(\mathbf{v})$ of an $DBN(p, l, \boldsymbol{\theta})$. To measure the "distance" between $\mathbf{Q}(\mathbf{v})$ and $\mathbf{P}_l(\mathbf{v})$ we use the Kullback-Leibler (KL) divergence:

$$KL(\mathbf{Q} \mid \mathbf{P}_l) = \sum_{\mathbf{v}} \mathbf{Q}(\mathbf{v}) \ln \frac{\mathbf{Q}(\mathbf{v})}{\mathbf{P}_l(\mathbf{v})} = -H(\mathbf{Q}) - \frac{1}{\# G_l} \sum_{j \in G_l} \ln \mathbf{P}_l(\mathbf{v}_j),$$

where $H(\mathbf{Q})$ is the entropy of $\mathbf{Q}$. We recall that $KL(\mathbf{Q} \mid \mathbf{P}_l) = 0$ if and only if $\mathbf{Q} = \mathbf{P}_l$. We construct an improved version of $DBN(p, l, \boldsymbol{\theta}_l)$ by increasing the the number of levels (or layers) l, and consequently, the number of hidden variables (units), but keeping the number of visible variables fixed.

Theorem

Assume that $p \geq 3$. *Let* $\mathbf{Q}(\mathbf{v})$ *be an arbitrary probability distribution on* $\{0,1\}^m$. *As discussed above, we assume without loss of generality that* $m = p^{l_0}$. *We identify* $\mathbf{v}$ *with* $\mathbf{v}_l = \left(v_j^l\right)_{j \in G_l}$, *and* $\mathbf{Q}(\mathbf{v}_l)$ *with a probability distribution on the* $\mathbf{v}_l$*s. Let* $DBN(p, l, \boldsymbol{\theta}_l)$ *be a p-adic discrete DBN, with* $l \geq l_0$, *such that* $KL(\mathbf{Q}(\mathbf{v}_l) \| \mathbf{P}_l(\mathbf{v}_l; \boldsymbol{\theta}_l)) > 0$. *Then the two following assertions hold true.*

(i) There exists an $DBN(p, l+1, \boldsymbol{\theta}_l, \mathbf{w}_{l+1}, b_{j0}^{l+1})$ *constructed from* $DBN(p, l, \boldsymbol{\theta}_l)$ *by adding one layer with marginal probability distribution* $\mathbf{P}_{l+1}\left(\mathbf{v}_l; \boldsymbol{\theta}_l, \mathbf{w}_{l+1}, b_{j0}^{l+1}\right)$ *satisfying*

$$KL(\mathbf{Q}(\mathbf{v}_l) \| \mathbf{P}_{l+1}\left(\mathbf{v}_l; \boldsymbol{\theta}_l, \mathbf{w}_{l+1}, b_{j0}^{l+1}\right)) < KL(\mathbf{Q}(\mathbf{v}_l) \| \mathbf{P}_l(\mathbf{v}_l; \boldsymbol{\theta}_l)), \tag{3}$$

for some $\boldsymbol{\theta}_l, \mathbf{w}_{l+1}, b_{j0}^{l+1}$.

Theorem

(ii) Given $\epsilon > 0$ *arbitrarily small, there exists an*

$$DBN(p, l+k, \boldsymbol{\theta}_l, \mathbf{w}_{l+1}, \ldots, \mathbf{w}_{l+k}, b_{j_0}^{l+1}, \ldots, b_{j_{k-1}}^{l+k})$$

with marginal probability distribution
$\mathbf{P}_{l+k}\left(\mathbf{v}_l; \boldsymbol{\theta}_l, \mathbf{w}_{l+1}, \ldots, \mathbf{w}_{l+k}, b_{j_0}^{l+1}, \ldots, b_{j_{k-1}}^{l+k}\right)$ *satisfying*

$$KL(\mathbf{Q}(\mathbf{v}_l) \,\|\, \mathbf{P}_{l+k}\left(\mathbf{v}_l; \boldsymbol{\theta}_l, \mathbf{w}_{l+1}, \ldots, \mathbf{w}_{l+k}, b_{j_0}^{l+1}, \ldots, b_{j_{k-1}}^{l+k}\right)) < \epsilon, \tag{4}$$

where k *is a positive integer depending on* ϵ*, for some*

$$\boldsymbol{\theta}_l, \mathbf{w}_{l+1}, \ldots, \mathbf{w}_{l+k}, b_{j_0}^{l+1}, \ldots, b_{j_{k-1}}^{l+k}.$$

Theorem

Assume that $p \geq 3$. *Let* $\mathbf{Q}(\mathbf{v})$ *be an arbitrary probability distribution on* $\{0,1\}^m$. *As discussed above, we assume without loss of generality that* $m = p^{l_0}$. *We identify* $\mathbf{v}$ *with* $\mathbf{v}_l = \left(v_j^l\right)_{j \in G_l}$, *and* $\mathbf{Q}(\mathbf{v}_l)$ *with a probability distribution on the* $\mathbf{v}_l$*s. Then* $\mathbf{Q}\left(\mathbf{v}_l\right)$ *can be approximated arbitrarily well, in the sense of the KL divergence, by an* $DBN(p, l_0 + k, \boldsymbol{\theta}_{l_0+k})$, *where* k *is the number of input vectors whose probability in not zero.*

Zúñiga-Galindo W. A., He Cuiyu, Zambrano-Luna B. A., p-Adic statistical field theory and restricted Boltzmann machines. In progress.

Zúñiga-Galindo W. A., p-Adic statistical field theory and deep belief networks. https://doi.org/10.48550/arXiv.2207.13877.

Zúñiga-Galindo W. A., Non-Archimedean statistical field theory, *Rev. Math. Phys.* **34** (8), Paper No. 2250022, 41 pp. (2022).


MERSI
obrigado
takk
mesi
ありがとう
MOLTE GRAZIE
ARIGATO
danke
TEŞEKKÜR EDERIM
THANK YOU
PALDIES
grazas
спасибо
muchas gracias
どうも多謝
merci
danke schön
DANK U
THANKS
GRACIAS
ARIGATO
MOLTE GRAZIE
THANK YOU
takk
gracias
TAK
hvala
danke schön
GO RAIBH MAITH AGAT
muchas gracias
감사합니다
どうも
grazie
obrigado
OBRIGADO
muchas gracias
TACK
Gràcies
www.bigstock.com · 404149934