Neural Networks, NIP and Definable Valuations

Lothar Sebastian Krapp

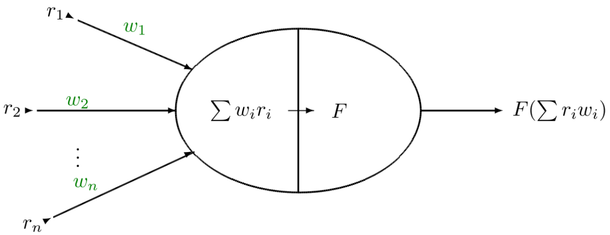
Universität Konstanz, Fachbereich Mathematik und Statistik

02 March 2021

Purdue University Model Theory and Applications Seminar

1. Neural Networks
2. NIP
3. Definable Valuations

Artificial Neurons



r_i: real numbers, **input**

w_i: real numbers, **weights**

$\sum w_i r_i$: weighted sum

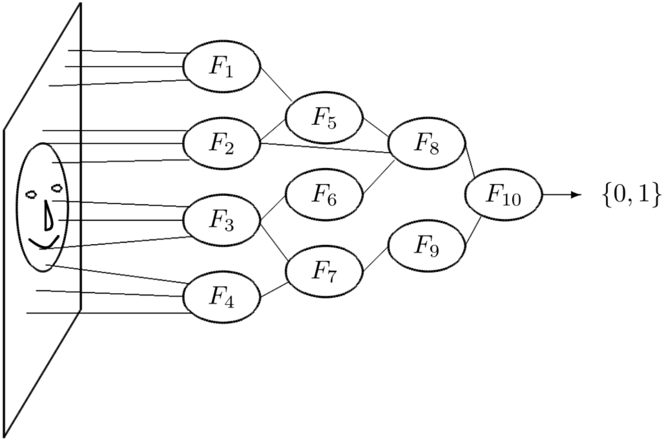
F: real valued function,
activation function

Activation Functions

Typical activation functions F:

- characteristic functions on intervals (a, ∞)
- piecewise linear functions
- sigmoid function $F(t) = \frac{1}{1+\exp(-t)}$

Artificial Neural Network



(10 neurons, 4 layers)

Artificial Neural Network

X: input space, e.g. $(\mathbb{R}^2 \times \{0, \ldots, 255\})^{12}$
Y: output space, e.g. $\{0, 1\}$
$F_i(\underline{x}, \underline{w})$: activation functions

The network computes a class of functions $X \to Y$.

Learning Cycle

1. network is in an initial state h coded by the weights
2. training sample $(x, y) \in X \times Y$ is chosen
3. $h(x)$ is computed
4. the weights are adjusted depending on $h(x) = y$ or $h(x) \neq y$ (also considering previous training samples)

Goal: After finitely many training samples the network is in a state h which gives a good approximation to recognising the pattern.

Formal Learning

Neural network H: set of all possible functions depending on the weights

Sample space $Z = X \times Y$

Learning algorithm L:

$$L : \bigcup_{m=1}^{\infty} Z^m \to H.$$

Learning Algorithm

p – probability measure on Z measuring the probability that a sample is chosen as training sample

$\mathrm{er}_p(h) = p\{(x, y) \in Z \mid h(x) \neq y\}$ – error of $h \in H$

$\mathrm{opt}_p(H) = \inf_{h \in H} \mathrm{er}_p(h)$ – best approximation in H for given p

Learning Algorithm

Definition

Let H be a collection of functions $X \to Y$ for a given sample space $Z = X \times Y$. A **learning algorithm** L is a map

$$L : \bigcup_{m=1}^{\infty} Z^m \to H$$

with the following property:
$\forall \varepsilon, \delta \in (0,1)\ \exists m_0 \in \mathbb{N}\ \forall m \geq m_0 :$
for any probability measure p on Z we have

$$p^m\{z \in Z^m \mid \mathrm{er}_p(L(z)) < \mathrm{opt}_p(H) + \varepsilon\} \geq 1 - \delta,$$

where p^m is the product measure on Z^m.
H is called **learnable** if there exists a learning algorithm for H.

Model Theoretic Setting

- first-order languages $\mathcal{L}$:
  $\mathcal{L}_{\mathrm{r}} = (+, -, \cdot, 0, 1)$, $\mathcal{L}_{\mathrm{or}} = (+, -, \cdot, 0, 1, <)$, $\mathcal{L}_{\exp} = (+, -, \cdot, 0, 1, <, \exp)$
- $\mathcal{L}$-structures:
  $(\mathbb{Z}, +, -, \cdot, 0, 1)$, $(\mathbb{Q}, +, -, \cdot, 0, 1, <)$, $(\mathbb{R}, +, -, \cdot, 0, 1, <, \exp)$
- $\mathcal{L}$-formulas and $\mathcal{L}$-sentences:
  $\forall x \exists y \ x + y = 0$
  $\exists y \ y < x$
  $\forall x \exists y \ \exp(y) < x$
- complete $\mathcal{L}$-theories:
  $\mathrm{Th}(\mathbb{R}, +, -, \cdot, 0, 1, <)$ (the theory of **real closed fields**) – set of all $\mathcal{L}_{\mathrm{or}}$-sentences which are true in $(\mathbb{R}, +, -, \cdot, 0, 1, <)$

Model Theoretic Setting

- first-order languages $\mathcal{L}$:
  $\mathcal{L}_{\mathrm{r}} = (+, -, \cdot, 0, 1)$, $\mathcal{L}_{\mathrm{or}} = (+, -, \cdot, 0, 1, <)$, $\mathcal{L}_{\mathrm{exp}} = (+, -, \cdot, 0, 1, <, \exp)$
- $\mathcal{L}$-structures:
  $(\mathbb{Z}, +, -, \cdot, 0, 1)$, $(\mathbb{Q}, +, -, \cdot, 0, 1, <)$, $(\mathbb{R}, +, -, \cdot, 0, 1, <, \exp)$
- $\mathcal{L}$-formulas and $\mathcal{L}$-sentences:
  $\forall x \exists y \; x + y = 0$ – true in $(\mathbb{R}, +, -, \cdot, 0, 1, <)$
  $\exists y \; y < x$
  $\forall x \exists y \; \exp(y) < x$ – false in $(\mathbb{R}, +, -, \cdot, 0, 1, <, \exp)$
- complete $\mathcal{L}$-theories:
  $\mathrm{Th}(\mathbb{R}, +, -, \cdot, 0, 1, <)$ (the theory of **real closed fields**) – set of all $\mathcal{L}_{\mathrm{or}}$-sentences which are true in $(\mathbb{R}, +, -, \cdot, 0, 1, <)$

VC Dimension

Let $\mathcal{L}$ be a first-order language, let $\mathcal{M} = (M, \ldots)$ be an $\mathcal{L}$-structure and let $\varphi(\underline{x}, \underline{y})$ be an $\mathcal{L}$-formula.

Definition

The **Vapnik–Chervonenkis dimension** (**VC dimension**) of φ (in $\mathcal{M}$) is defined as

$$\mathrm{vc}(\varphi(\underline{x}; \underline{y})) := \max(S_\varphi)$$

where $S_\varphi \subseteq \mathbb{N}$ consists of all $n \in \mathbb{N}$ with the following property:

There exist tuples $(\underline{a}_i)_{i<n}$ and $(\underline{b}_J)_{J \subseteq \{0,\ldots,n-1\}}$ in M such that for any $i < n$ and any $J \subseteq \{0, \ldots, n-1\}$

$$\varphi(\underline{a}_i; \underline{b}_J) \text{ holds in } \mathcal{M} \text{ if and only if } i \in J.$$

If S_φ has no maximum, then $\mathrm{vc}(\varphi(\underline{x}; \underline{y})) = \infty$.

VC Dimension

Example

Consider the $\mathcal{L}_{\mathrm{r}}$-structure $(\mathbb{Z}, +, -, \cdot, 0, 1)$. The $\mathcal{L}_{\mathrm{r}}$-formula $\varphi(x; y)$ given by

$$\exists d \; d \cdot x = y$$

expresses "x divides y".

For any $i \in \mathbb{N}$, let $a_i \in \mathbb{Z}$ be the i-th prime number. For any $n \in \mathbb{N}$ and any $J \subseteq \{0, \ldots, n-1\}$, let

$$b_J = \prod_{i \in J} p_i.$$

Then $\varphi(a_i; b_J)$ holds in $(\mathbb{Z}, +, -, \cdot, 0, 1)$ if and only if $i \in J$. Hence, $\mathrm{vc}(\varphi(x; y)) = \infty$.

NIP

Definition

A formula $\varphi(\underline{x};\underline{y})$ has the **independence property** (or is **IP**) if $\mathrm{vc}(\varphi)=\infty$. If φ does not have the independence property, it is called **NIP**.

An $\mathcal{L}$-structure $\mathcal{M}$ is NIP: For **every** $\mathcal{L}$-structure $\mathcal{N}$ satisfying $\mathrm{Th}(\mathcal{M})$, **every** $\mathcal{L}$-formula $\varphi(\underline{x};\underline{y})$ is NIP in $\mathcal{N}$.

Examples of NIP structures:

- o-minimal structures
- weakly o-minimal structures
- C-minimal structures
- ...

O-minimality

Definition

An ordered structure $(M, <, \ldots)$ is called **o-minimal** if every parametrically definable subset of M is a finite union of points and open intervals in M.

Theorem (Wilkie 1996)

The real exponential field $\mathbb{R}_{\text{exp}} = (\mathbb{R}, +, -, \cdot, 0, 1, <, \exp)$ is o-minimal.

Example: The formula $\exists y\ x^2 > \exp(y) + \pi$ parametrically defines the set $\{x \in \mathbb{R} \mid \exists y\ x^2 > \exp(y) + \pi\} = (-\infty, -\sqrt{\pi}) \cup (\sqrt{\pi}, \infty)$ over $\mathbb{R}_{\text{exp}}$.

NIP Implies Learnability

Theorem

Let $\mathcal{R} = (\mathbb{R}, +, \cdot, <, \dots)$ be an NIP expansion of $(\mathbb{R}, +, \cdot, <)$, let $X \subseteq \mathbb{R}^d$ be a (parametrically) definable set over $\mathcal{R}$ and let H be a collection of activation functions of a neural network $X \to \{0, 1\}$ (parametrically) definable over $\mathcal{R}$. Then H is learnable.

Since $\mathbb{R}_{\text{exp}}$ is o-minimal and thus NIP, any set H of $\mathbb{R}_{\text{exp}}$-definable activation functions of a neural network is learnable.

Brief Historical Overview

- **Vapnik, Chervonenkis, 1971:** paper on statistical learning theory introducing the notion of VC dimensions (for sets rather than formulas)
- **Shelah, 1971:** paper on model theory introducing the independence property
- **Pillay, Steinhorn, 1986:** proof that o-minimality implies NIP
- **Laskowski, 1992:** connecting the notion of VC dimension to the independence property
- **Wilkie, 1996:** proof that $\mathbb{R}_{\mathrm{exp}}$ is o-minimal

Classification Question

What are necessary and sufficient conditions on a general ordered field $(K, +, -, \cdot, 0, 1, <)$ to be NIP?

With view to neural network learning: What ordered subfields

$$(K, +, -, \cdot, 0, 1, <) \subseteq (\mathbb{R}, +, -, \cdot, 0, 1, <)$$

are NIP?

(Note that $(\mathbb{Q}, +, -, \cdot, 0, 1, <)$ is **not** NIP!)

Shelah–Hasson Conjecture

Shelah–Hasson Conjecture (specialised to ordered fields)

Any (strongly) NIP ordered field $(K, +, -, \cdot, 0, 1, <)$ is either real closed or contains a non-trivial henselian valuation ring that is definable in the language $\mathcal{L}_{\mathrm{or}}$.

This conjecture would imply that the NIP ordered subfields of $(\mathbb{R}, +, -, \cdot, 0, 1, <)$ are exactly its real closed subfields!

1. Neural Networks
2. NIP
3. Definable Valuations

Convex Valuation Rings

Definition

Let $(K, +, -, \cdot, 0, 1, <)$ be an ordered field. A **convex valuation ring** in K is a convex subring $R \subseteq K$ such that for any $a \in K^\times$

$$a \in R \text{ or } a^{-1} \in R.$$

Maximal ideal (of infinitesimals): $I = \{a \in K^\times \mid a^{-1} \notin R\} \cup \{0\}$.

Residue field: $\overline{K} = R/I$.

Residue elements: For $a \in R$ we write $\overline{a} = a + I$.

Convex Valuation Rings

Examples

- For any ordered subfield $(K, +, -, \cdot, 0, 1, <)$ of $(\mathbb{R}, +, -, \cdot, 0, 1, <)$, the only convex valuation ring in K is $R = K$ (the trivial valuation ring).
- The field of Laurent series $\mathbb{R}(\!(x)\!)$ in one variable over $\mathbb{R}$ can be ordered by setting $\frac{1}{x} > r$ for any $r \in \mathbb{R}^{>0}$. Its only non-trivial convex valuation ring is the power series ring $\mathbb{R}[\![x]\!]$.
- The only $\mathcal{L}_{\mathrm{or}}$-definable convex valuation ring in an o-minimal (or, equivalently, real closed) ordered field is the trivial valuation ring.

Henselian Valuation Rings

Definition

Let $(K, +, -, \cdot, 0, 1, <)$ be an ordered field. A convex valuation ring R in K is **henselian** if for any polynomial

$$p(x) = a_n x^n + \ldots + a_0 \in R[x]$$

and any simple zero $a \in \overline{K}$ of

$$\overline{p}(x) = \overline{a_n} x^n + \ldots + \overline{a_0} \in \overline{K}[x],$$

there exists a zero $b \in R$ of $p(x)$ with $\overline{b} = a$.

Motto: *Simple zeros can be lifted from the residue field.*

Henselian Valuation Rings

Examples

- Any trivial valuation ring is henselian, as the residue field coincides with the field.
- $\mathbb{R}[\![x]\!]$ is a henselian valuation ring of $\mathbb{R}(\!(x)\!)$.

Approach towards Shelah–Hasson Conjecture

Refinements of the property 'NIP':

o-minimal $\rightarrow$ **weakly o-minimal**
$\rightarrow$ **dp-minimal** $\rightarrow$ **dp-finite**
$\rightarrow$ **strongly NIP** $\rightarrow$ **NIP**

Currently the Shelah–Hasson Conjecture specialised to ordered fields has been verified for the dp-minimal case. Hence, an ordered subfield of $(\mathbb{R}, +, -, \cdot, 0, 1, <)$ is dp-minimal if and only if it is real closed.

Strongly NIP Ordered Fields

Theorem (K., Kuhlmann, Lehéricy)

The following are equivalent:

- The Shelah–Hasson Conjecture specialised to ordered fields: Any strongly NIP ordered field is either real closed or admits a non-trivial $\mathcal{L}_{\mathrm{or}}$-definable henselian valuation ring.
- Any strongly NIP ordered field admits a henselian valuation ring with real closed residue field.

References

- M. ANTHONY and P. L. BARTLETT, *Neural Network Learning: Theoretical Foundations* (Cambridge University Press, Cambridge, 1999).
- K. DUPONT, A. HASSON and S. KUHLMANN, 'Definable Valuations induced by multiplicative subgroups and NIP Fields', *Arch. Math. Logic* 58 (2019) 819–839.
- F. JAHNKE, P. SIMON and E. WALSBERG, 'Dp-minimal valued fields', *J. Symb. Log.* 82 (2017) 151–165.
- L. S. KRAPP, 'Algebraic and Model Theoretic Properties of O-Minimal Exponential Fields', doctoral thesis, Universität Konstanz, 2019.
- L. S. KRAPP, S. KUHLMANN and G. LEHÉRICY, 'Strongly NIP Almost Real Closed Fields', to appear in *Math. Log. Q.*, arXiv:2010.14770.
- L. S. KRAPP, S. KUHLMANN and G. LEHÉRICY, 'Ordered Fields Dense in Their Real Closure and Definable Convex Valuations', accepted, arXiv:2010.11832.
- M. C. LASKOWSKI, 'Vapnik–Chervonenkis classes of definable sets', *J. London Math. Soc. (2)* 45 (1992) 377–384.

References

- A. Pillay and C. Steinhorn, 'Definable sets in ordered structures', I, *Trans. Amer. Math. Soc.* 295 (1986) 565–592.
- S. Shelah, 'Stability, the f.c.p., and superstability; model theoretic properties of formulas in first order theory', *Ann. Math. Logic* 3 (1971) 271–362.
- V. N. Vapnik and A. Ya. Chervonenkis, 'The uniform convergence of frequencies of the appearance of events to their probabilities', *Teor. Verojatnost. i Primenen.* 16 (1971) 264–279 (Russian), *Theor. Probability Appl.* 16 (1971) 264–280 (English).
- A. J. Wilkie, 'Model completeness results for expansions of the ordered field of real numbers by restricted Pfaffian functions and the exponential function', *J. Amer. Math. Soc.* 9 (1996) 1051–1094.
- *Graphics from:* M. Tressl, 'Introduction to o-minimal structures and an application to neural network learning', Manuscript, 2010.