o-Minimality and Neural Networks

Konrad Bals, Lukas Bonfert

Summer Academy Leysin
German Academic Scholarship Foundation

August 2018

1 Basic definitions

Throughout the whole paper, L denotes a first-order language (with equality). L-structures are denoted $\mathfrak{A}, \mathfrak{B}, \ldots$, with domains $A, B, \ldots$.

Definition. Let $\mathfrak{A}$ be an L-structure. A set $X \subseteq A$ is *definable* if there is an L-formula $\varphi(x)$, such that

$$X = \{a \in A \mid \mathfrak{A} \models \varphi(x)\}.$$

Definition (o-minimality)**.** Let $\mathfrak{A} = (A, \leq)$ be a totally ordered set. $\mathfrak{A}$ is *o-minimal* if every definable subset is a finite union of intervals.

Definition (VC dimension)**.** Let X be a set. A collection of subsets $\mathcal{S} = \{S_i \subseteq X \mid i \in I\}$ *shatters* a set $B \subseteq X$, if for every subset $A \subseteq B$ there is $i \in I$, such that $A = B \cap S_i$.
The *Vapnik-Chervonenkis dimension* (VC dimension) of $\mathcal{S}$ is the smallest integer d such that $\mathcal{S}$ does not shatter any subset of size d of X, or ∞ if no such d exists.

Definition (NIP)**.** Let Φ be a complete L-theory. A formula $\varphi(x_1, \ldots, x_m, y_1, \ldots, y_n)$ has the *independence property* (IP) in Φ if there is a model $\mathfrak{A}$ of Φ with an infinite subset $N \subseteq A^{\vec{y}}$ shattered by

$$\mathcal{S}_\varphi = \Big\{\{b \in A^{\vec{y}} \mid \mathfrak{A} \models \varphi(a, b)\} \;\Big|\; a \in A^{\vec{x}}\Big\},$$

where $a = (a_1, \ldots, a_m)$ and $b = (b_1, \ldots, b_n)$. If no such set X exists, φ has *not the independence property* (NIP). The theory Φ has IP if there is a formula φ having IP in Φ, and has NIP otherwise. A structure $\mathfrak{A}$ has (N)IP if its theory $\mathrm{Th}(\mathfrak{A}) = \{\varphi \mid \mathfrak{A} \models \varphi\}$ has (N)IP.

Definition. An *embedding* $i\colon \mathfrak{A} \hookrightarrow \mathfrak{B}$ is an injective map $i\colon A \to B$ such that

- $i(f^{\mathfrak{A}}(a_1, \ldots, a_n)) = f^{\mathfrak{B}}(i(a_1), \ldots, i(a_n))$ for all n-ary function symbols f and $a_1, \ldots, a_n \in A$,
- $(a_1, \ldots, a_n) \in R^{\mathfrak{A}} \iff (i(a_1), \ldots, i(a_n)) \in R^{\mathfrak{B}}$ for all n-ary relation symbols R and $a_1, \ldots, a_n \in A$.

If there is an embedding $i\colon \mathfrak{A} \hookrightarrow \mathfrak{B}$, then $\mathfrak{A}$ can be identified with a *substructure* of $\mathfrak{B}$. We also write $\mathfrak{A} \subseteq \mathfrak{B}$.
An embedding i is *elementary* if moreover for any L-formula φ with free variables $x_1, \dots, x_n$ and any $a_1, \dots, a_n \in A$, we have

$$\mathfrak{A} \models \varphi(a_1, \dots, a_n) \iff \mathfrak{B} \models \varphi(a_1, \dots, a_n).$$

In this case, we write $\mathfrak{A} \preceq \mathfrak{B}$.

Definition. Let Φ be an L-theory. A model $\mathfrak{M}$ of Φ is a *monster model* if for every model $\mathfrak{A}$ of Φ, there is an elementary embedding $\mathfrak{A} \preceq \mathfrak{M}$.

To prove the existence of a monster model, ZFC is insufficient, however, we will not worry about this.

Lemma 1.1. *Let Φ be an L-theory and $\varphi = \varphi(\vec{x}, \vec{y})$ an L-formula. Then the following are equivalent:*

1) *φ has IP in Φ.*
2) *Let $\mathfrak{M}$ be a monster model of Φ. For all $n \in \mathbb{N}$, there are elements b_i of $M^{\vec{x}}$ indexed by natural numbers $0 \leq i \leq n-1$ and elements a_J of $M^{\vec{y}}$ indexed by subsets $J \subset \{0, \dots, n-1\}$ such that*

$$\mathfrak{M} \models \varphi(a_J, b_i) \iff i \in J.$$

Proof. 1) $\implies$ 2): Choose a model $\mathfrak{A} \models \Phi$ and an infinite subset $X \subseteq A^{\vec{x}}$ as in the definition. Since $\mathfrak{M}$ is a monster model of Φ, there is an elementary embedding $\mathfrak{A} \preceq \mathfrak{M}$. Choose some arbitrary $b_1, \dots, b_n \in X$. Since X is shattered by $\mathcal{S}_\varphi$, so is the subset $\{b_1, \dots, b_n\}$, hence for any $J \subset \{0, \dots, n-1\}$ there is $a_J \in M^{\vec{x}}$ such that

$$\begin{aligned}\{b_i \mid i \in J\} &= \{b \in \{b_0, \dots, b_{n-1}\} \mid \mathfrak{A} \models \varphi(a_J, b)\}\\ &= \{b \in \{b_0, \dots, b_{n-1}\} \mid \mathfrak{M} \models \varphi(a_J, b)\},\end{aligned}$$

because $\mathfrak{A} \preceq \mathfrak{M}$ is elementary. In other words, $\mathfrak{M} \models \varphi(a_J, b_i) \iff i \in J$.
2) $\implies$ 1): It is sufficient to construct a countably infinite set $X \subseteq M^{\vec{y}}$ shattered by $\mathcal{S}_\varphi$. Expand the language L to a new language

$$L' = L \cup \{\dot{a}_J \mid J \subseteq \mathbb{N}\} \cup \{\dot{b}_i \mid i \in \mathbb{N}\},$$

where the $\dot{a}_J$ and $\dot{b}_i$ are used as shorthands for tuples of constant symbols. Let

$$\Psi = T \cup \{\varphi(\dot{a}_J, \dot{b}_i) \mid J \subseteq \mathbb{N}, i \in J\} \cup \{\neg\varphi(\dot{a}_J, \dot{b}_i) \mid J \subseteq \mathbb{N}, i \notin J\}.$$

We claim that Ψ is satisfiable. To prove this, let $\Psi_0 \subseteq \Psi$ be finite, hence there are only finitely many formulas $\varphi(\dot{a}_J, \dot{b}_i)$, $\neg\varphi(\dot{a}_J, \dot{b}_i)$ contained in Ψ_0. By assumption, $\mathfrak{M}$ with $\dot{a}_J$ and $\dot{b}_i$ interpreted by the a_J and b_i provided by 1) satisfies Ψ_0, because there are only finitely many $\dot{b}_i$ occuring in Ψ_0. By the compactness theorem, Ψ is satisfiable, so there is a model $\mathfrak{A} \models \Psi$, and by construction, $X = \{\mathfrak{A}(\dot{b}_i) \mid i \in \mathbb{N}\}$ is shattered by $\mathcal{S}_\varphi$. $\square$

Corollary 1.2. *A formula φ has IP in Φ if and only if the VC dimension of $\mathcal{S}_\varphi$ is infinite.*

Proof. If φ has IP, then $\mathcal{S}_\varphi$ shatters an infinite set, and clearly also all of its subsets. Hence the VC dimension of $\mathcal{S}_\varphi$ is inifinite.
On the other hand, $\mathcal{S}_\varphi$ shattering arbitrarily large sets is equivalent to condition 2) in Lemma 1.1. $\square$

Definition (Model completeness)**.** An L-theory Φ is *model complete* if for any two models $\mathfrak{A} \subseteq \mathfrak{B}$ of Φ, the embedding is elementary.

Lemma 1.3. *Φ is model complete if and only if for any L-formula φ there is a quantifier-free L-formula ψ such that*

$$\Phi \vdash (\varphi \leftrightarrow \exists x_1 \ldots \exists x_n \psi).$$

We say that φ is equivalent to an existential formula*.*

Theorem 1.4 (Tarski-Vaught test)**.** *Let L be a language and $\mathfrak{A} \subseteq \mathfrak{B}$ two L-structures. Then the following are equivalent:*

1) $\mathfrak{A}$ is an elementary substructure of $\mathfrak{B}$.

2) Let $\varphi(x, y_1, \ldots, y_n)$ be an L-formula and $a_1, \ldots, a_n \in A$. Then if $\mathfrak{B} \models \exists x \varphi(x, a_1, \ldots, a_n)$, there is $a \in A$ such that $\mathfrak{A} \models \varphi(a, a_1, \ldots, a_n)$.

Proof. 1) $\implies$ 2): Let $\mathfrak{A} \preceq \mathfrak{B}$. If $\mathfrak{B} \models \exists x \varphi(x, a_1, \ldots, a_n)$ with $a_1, \ldots, a_n \in A$, then by definition $\mathfrak{A} \models \exists x \varphi(x, a_1, \ldots, a_n)$, so there is $a \in A$ with $\mathfrak{A} \models \varphi(a, a_1, \ldots, a_n)$ and by definition, $\mathfrak{B} \models \varphi(a, a_1, \ldots, a_n)$.
2) $\implies$ 1): By induction on formulas.

- φ is atomic: For any term t, since $\mathfrak{A} \subseteq \mathfrak{B}$ we have $\mathfrak{B}(t) = \mathfrak{A}(t)$. Thus, for $\varphi = t_1 = t_2$ and arbitrary $a_1, \ldots, a_n \in A$ we have

  $$\mathfrak{B} \models \varphi(a_1, \ldots, a_n) \iff \mathfrak{A} \models \varphi(a_1, \ldots, a_n).$$

  Similarly, for $\varphi = Rt_1 \ldots t_n$ with R a relation symbol,

  $$\mathfrak{B} \models \varphi(a_1, \ldots, a_n) \iff \mathfrak{A} \models \varphi(a_1, \ldots, a_n)$$

  because $R^{\mathfrak{A}}(a_1, \ldots, a_n) \iff R^{\mathfrak{B}}(a_1, \ldots, a_n)$ for $a_1, \ldots, a_n \in A$.

- φ is not atomic:

  - $\varphi = \psi_1 \to \psi_2$: $\mathfrak{B} \models \varphi(a_1, \ldots, a_n)$ if and only if $\mathfrak{B} \models \psi_1(a_1, \ldots, a_n)$ implies $\mathfrak{B} \models \psi_2(a_1, \ldots, a_n)$. By induction for ψ_1 and ψ_2 this is equivalent to $\mathfrak{A} \models \psi_1(a_1, \ldots, a_n)$ implying $\mathfrak{A} \models \psi_2(a_1, \ldots, a_n)$, or equivalently $\mathfrak{A} \models \varphi(a_1, \ldots, a_n)$.
  - $\varphi = \neg\psi$: By induction, $\mathfrak{B} \models \varphi(a_1, \ldots, a_n)$ if and only if $\mathfrak{B} \not\models \psi(a_1, \ldots, a_n)$ if and only if $\mathfrak{A} \not\models \varphi(a_1, \ldots, a_n)$ if and only if $\mathfrak{A} \models \varphi(a_1, \ldots, a_n)$.
  - $\varphi = \exists x \psi$: If $\mathfrak{B} \models \exists x \psi(x, a_1, \ldots, a_n)$ for $a_1, \ldots, a_n \in A$, then by assumption there is $a \in A$ with $\mathfrak{B} \models \psi(a, a_1, \ldots, a_n)$, hence by definition $\mathfrak{A} \models \psi(a, a_1, \ldots, a_n)$ and thus $\mathfrak{A} \models \exists x \psi(x, a_1, \ldots, a_n)$.
    Conversely, if $\mathfrak{A} \models \exists x \psi(x, a_1, \ldots, a_n)$, then there is $a \in A$ with $\mathfrak{A} \models \psi(a, a_1, \ldots, a_n)$. By definition, $\mathfrak{B} \models \psi(a, a_1, \ldots, a_n)$, hence $\mathfrak{B} \models \exists x \psi(x, a_1, \ldots, a_n)$.

□

Definition (Existentially closed)**.** A model $\mathfrak{A}$ of a theory Φ is *existentially closed* if for every model $\mathfrak{B}$ with $\mathfrak{A} \subseteq \mathfrak{B}$ and $\mathfrak{B} \models \Phi$ and for any quantifier-free formula φ and $a_1, \ldots, a_m \in \mathfrak{A}$ with $\mathfrak{B} \models \exists \vec{v}\ \varphi(\vec{v}, \vec{a})$, then also $\mathfrak{A} \models \exists \vec{v}\ \varphi(\vec{v}, \vec{a})$.

Theorem 1.5 (Robinson's test)**.** *A theory Φ is model-complete if and only if every model of Φ is existentially closed.*

2 o-minimality of $\mathbb{R}_{\exp}$

Let $L = \{0, 1, +, -, \cdot, <, \exp\}$ be the language of ordered rings with the exponential function and $\mathbb{R}_{\exp} = (\mathbb{R}, 0, 1, +, -, \cdot, <, \exp)$ the L-structure, where the constants functions and predicates have the natural interpretation. We show that $\mathbb{R}_{\exp}$ is o-minimal, assuming the model completeness of $\mathbb{R}_{\exp}$.

Theorem 2.1 (Wilkie)**.** $\mathbb{R}_{\exp}$ *is model complete.*

Sketch of proof. This is a difficult application of Robinson's test. In this context, we have to prove that for any two models $\mathfrak{A} \subseteq \mathfrak{B}$ of $\mathbb{R}_{\exp}$, any exponential polynomial with coefficients in $\mathfrak{A}$ that has a zero in $\mathfrak{B}$ also has a zero in $\mathfrak{A}$.

Lemma 2.2. *Let t be an L-term with a single variable x and $R \in \{=, <\}$. Then there exists a quantifier-free L-formula φ with*

$$\mathbb{R}_{\exp} \models (t\ R\ 0 \leftrightarrow \exists y_1 \ldots \exists y_n \varphi),$$

such that all terms in φ are exponential polynomials, i.e. polynomials in the variables $x, y_1, \ldots, y_n, \exp x, \exp y_1, \ldots, \exp y_n$.

Proof. For every appearance of $\exp t'$ in t with the term t' neither a constant nor a variable we have

$$\mathbb{R}_{\exp} \models t\ R\ 0 \leftrightarrow \begin{cases} t\left[\frac{\exp t_1 \cdot \exp t_2}{\exp t'}\right]\ R\ 0 & \text{if } t' = t_1 + t_2, \\ \exists y\ t\left[\frac{y}{\exp t'}\right]\ R\ 0 \wedge y - \exp t' = 0 & \text{otherwise.} \end{cases}$$

We can repeat this finitely many times to get the desired formula φ. □

Theorem 2.3 (Wilkie)**.** $\mathbb{R}_{\exp}$ *is o-minimal.*

Proof. Let $D \subset \mathbb{R}$ be a definable set in $\mathbb{R}_{\exp}$. Because $\mathbb{R}_{\exp}$ is model complete we have an existential L-formula $\varphi(x)$ which defines D. With $\neg t_1 = t_2 \Leftrightarrow (t_1 < t_2) \vee (t_2 < t_1)$ and $\neg t_1 < t_2 \Leftrightarrow (t_1 = t_2) \vee (t_2 < t_1)$ we now write

$$\begin{array}{ll} \exists z\ t_2 - t_1 - z^2 = 0 & \text{for } t_1 < t_2, \\ t_1 - t_2 = 0 & \text{for } t_1 = t_2. \end{array}$$

for every atomic formula in φ. Applying Lemma 2.2 and moving the existential quantifiers to the front yields an existential formula ψ with atomic formulas in the form $t = 0$ with t an exponential polynomial. For $t_1 = 0$ and $t_2 = 0$ we write

$$\begin{array}{ll} t_1^2 + t_2^2 = 0 & \text{if } t_1 \wedge t_2 \text{ is in } \psi, \\ t_1 \cdot t_2 = 0 & \text{if } t_1 \vee t_2 \text{ is in } \psi. \end{array}$$

and get an existential formula with free variable x and quantified variables satisfying the only atomic formula $t = 0$ if and only if they are in the zero set of the exponential polynomial. Due to a theorem by Khovanskii every exponential variety has only finitely many connected components. Projection to the x component yields, that the set D consists of finitely many intervals. $\square$

3 o-minimality and NIP

In this section, we will show that o-minimal structures have NIP. Unless explicitly mentioned, all structures are totally ordered sets.

Definition. An *n-type* is a consistent set p of formulas with free variables $x_1, \dots, x_n$ such that for any formula $\varphi(x_1, \dots, x_n)$, either $\varphi \in p$ or $\neg\varphi \in p$.
For a 1-type p and a structure $\mathfrak{A}$, let $p^< = \{a \in A \mid x < a \in p\}$, $p^= = \{a \in A \mid x = a \in p\}$, $p^> = \{a \in A \mid a < x \in p\}$. $p^=$ contains at most one element. Hence, a 1-type induces a *cut* of $\mathfrak{A}$.

Lemma 3.1. *If $\mathfrak{A}$ is o-minimal, then any 1-type is uniquely determined by the formulas $x < a$, $b < x$ and $x = c$ contained in p.*

Proof. Let $\varphi(x)$ be a formula. Since $\mathfrak{A}$ is o-minimal, there are $a_i, b_i, c_j \in A$ with

$$\mathfrak{A} \models \varphi \leftrightarrow \left(\bigvee_{j=1}^{m} x = c_j \vee \bigvee_{i=1}^{n} (a_i < x \wedge x < b_i) \right).$$

If there is j such that $x = c_j \in p$ or i such that $a_i < x \in p$ and $x < b_i \in p$, then since p is complete, also $\varphi \in p$. Otherwise, for all j we have $\neg x = c_j \in p$ and for all i, $\neg(a_i < x \wedge x < b_i) \in p$. Hence $p \cup \{\varphi\}$ is inconsistent, so $\varphi \notin p$. $\square$

Lemma 3.2. *Let $\mathfrak{A}, \mathfrak{B}$ be two L-structures. If* $\mathrm{Th}(\mathfrak{A}) = \mathrm{Th}(\mathfrak{B})$*, then $\mathfrak{A}$ is o-minimal if and only if $\mathfrak{B}$ is o-minimal. This property is called* strong o-minimality*.*

Sketch of proof. A formula $\varphi(x)$ defines a set of elements $a \in A$ satisfying φ. By o-minimality of $\mathfrak{A}$, this set is a finite union of intervals and thus can be expressed by a different formula ψ without free variables, which is satisfied by $\mathfrak{A}$ and thus also by $\mathfrak{B}$. So the subset of B defined by φ can be written as a finite union of intervals, as described by ψ.

For an L-structure $\mathfrak{A}$, let $L_{\mathfrak{A}} = L \cup \{\dot{a} \mid a \in A\}$. The structure $\mathfrak{A}$ naturally becomes an $L_{\mathfrak{A}}$-structure via $\mathfrak{A}(\dot{a}) = a$.

Definition. Let $\mathfrak{A}$ be an L-structure. An *n-type* over $\mathfrak{A}$ is a deductively closed set p of $L_{\mathfrak{A}}$-formulas in n free variables such that any finite subset of p is satisfied in $\mathfrak{A}$. If there are elements $a_1, \dots, a_n \in A$ such that $\mathfrak{A} \models p(a_1, \dots, a_n)$, then p is *realized* in $\mathfrak{A}$.

Definition. Let $\mathfrak{A} \preceq \mathfrak{B}$ be an elementary extension. A 1-type q over $\mathfrak{B}$ is a *coheir* over $\mathfrak{A}$ if every finite subset of q is realizable in $\mathfrak{A}$.

Proposition 3.3 (Poizat)**.** *An L-theory Φ has IP if and only if there is a 1-type p over a model $\mathfrak{A}$ of Φ with $|A| \geq |L|$ and an elementary extension $\mathfrak{B} \succeq \mathfrak{A}$ such that p has $2^{2^{|A|}}$ coheirs over $\mathfrak{B}$.*

Lemma 3.4. *Let $\mathfrak{A}$ be o-minimal and p a 1-type over $\mathfrak{A}$. Then p has at most two coheirs over any $\mathfrak{B} \succeq \mathfrak{A}$.*

Proof. Assume that p has three different coheirs q_1, q_2, q_3. By Lemma 3.2, $\mathfrak{B}$ is o-minimal, so by Lemma 3.1 the q_i are determined by their induced cuts. Without loss of generality, we may assume $q_1 < q_2 < q_3$. Then $p^< \subseteq q_1^< \subseteq q_3^<$ and $p^> \subseteq q_3^> \subseteq q_1^>$. Hence for $q_1 \leq b_1 < b_2 \leq q_3$ with $b_1, b_2 \in B$, at most one of b_1, b_2 can be contained in A (because otherwise $b_1, b_2 \in p^=$ and $b_1 \neq b_2$, a contradiction). Hence, the open interval (b_1, b_2) does not intersect A, so $b_1 < x < b_2$ is not realizable in A. But $b_1 < x < b_2$ is contained in q_2, contradicting the definition of coheir. □

Theorem 3.5. *If $\mathfrak{A}$ is o-minimal, then $\mathfrak{A}$ has NIP.*

Proof. By Lemma 3.2, it is sufficient to consider only the structure $\mathfrak{A}$. If $\mathfrak{A}$ had IP, then by Prop. 3.3, $|A| \geq |L| \geq 1$ and there would be a 1-type p over $\mathfrak{A}$ having $2^{2^{|A|}} > 2$ coheirs in some structure $\mathfrak{B} \succeq \mathfrak{A}$, contradicting Lemma 3.4. □

4 Application to neural networks

In this section we link the previous results of o-minimal structures to binary neural networks.

Definition. An *artificial neuron* is a set of functions $\mathbb{R}^d \to \mathbb{R}$

$$\{x \mapsto F(\langle w, x \rangle) \mid w \in \mathbb{R}^d\},$$

where $F\colon \mathbb{R} \to \mathbb{R}$ is a fixed *activation function* and $\langle \cdot\,, \cdot \rangle$ is the canonical scalar product. w is the *weight vector*. A *neural network* consists of several artificial neurons (with different activation functions) connected to each other, and thus can be described as a set H of functions given by compositions of the functions of the neurons.

In the following, X denotes the set of allowed inputs for the neural network and Y the possible outputs.

Definition. Let p be a probability measure on $X \times Y$. The *error* of $h \in H$ with respect to p is

$$\mathrm{er}_p(h) = p(\{(x, y) \in X \times Y \mid h(x) \neq y\}).$$

The *minimal error* of H for fixed p is

$$\mathrm{opt}_p(H) = \inf_{h \in H} \mathrm{er}_p(h).$$

For a neural network to be able to adapt to a certain task we have to define a learning cycle of the network. First we have to define a measure how well the network calculates a given sample $(x, y) \in X \times Y$.

Definition. A *learning algorithm* for a neural network H is a function

$$L\colon \bigcup_{m=1}^{\infty} (X \times Y)^m \to H$$

such that for all $0 < \varepsilon, \delta < 1$ there is $m_0(\varepsilon, \delta) \in \mathbb{N}$ such that for all $m \geq m_0$, any probability measure p on $X \times Y$ and all $z \in (X \times Y)^m$, we have

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

where p^m denotes the product measure. If there is a learning algorithm for H, then H is called *learnable.* The *inherent sample complexity* $m_H(\varepsilon, \delta)$ of H is the minimum number $m_0(\varepsilon, \delta)$ over all learning algorithms L.

Theorem 4.1. *A neural network H is learnable if and only if $\mathcal{S}_H = \{h^{-1}(1) \mid h \in H\}$ has finite VC dimension. Moreover, the inherent sample complexity is*

$$m_H(\varepsilon, \delta) = \Theta\left(\frac{1}{\varepsilon^2} \log \frac{1}{\delta}\right).$$

Corollary 4.2. *Any neural network with activation functions defined in $\mathbb{R}_{\exp}$ is learnable. In particular, this includes the* sigmoid function $\sigma(x) = \frac{1}{1+\exp(-x)}$.

Proof. The set of functions H can be parameterized by some $\mathbb{R}^l$ (the weight vectors). So there is a $\mathbb{R}_{\exp}$-formula $\varphi(y_1, \ldots, y_l, x_1, \ldots, x_d)$ such that $\{h^{-1}(1) \mid h \in H\}$ are exactly the fibers $\mathcal{S}_\varphi$ of φ. By Thm. 2.3, φ has NIP in $\mathbb{R}_{\exp}$, and by Cor. 1.2, $\mathcal{S}_\varphi$ has finite VC dimension. The claim follows from Thm. 4.1. □

References

[Güt17] Linnéa Gütlein. The ordered field of real numbers with exponentiation: Model completeness, decidability and Schanuel's conjecture. Master's thesis, Universität Konstanz, 2017.

[Mar02] David Marker. *Model Theory: An Introduction*, volume 217 of *Graduate Texts in Mathematics.* Springer, 2002.

[Poi85] Bruno Poizat. *Course de théorie des modeles.* Office International de Documentation et Librairie, 1985.

[Ram08] Janak Ramakrishnan. *Types in o-minimal theories.* PhD thesis, University of California, Berkeley, 2008.

[Tre10] Marcus Tressl. Introduction to o-minimal structures and an application to neural network learning. Course outline for LMS and EPSRC Short Instructional Course: Model Theory, University of Leeds, 2010.

[TZ12] Katrin Tent and Martin Ziegler. *A Course in Model Theory.* Cambridge University Press, 2012.

[Wil96] Alex Wilkie. Model completeness results for expansions of the real field by restricted pfaffian functions and the exponential function. *J. Am. Math. Soc.*, 9, 1996.