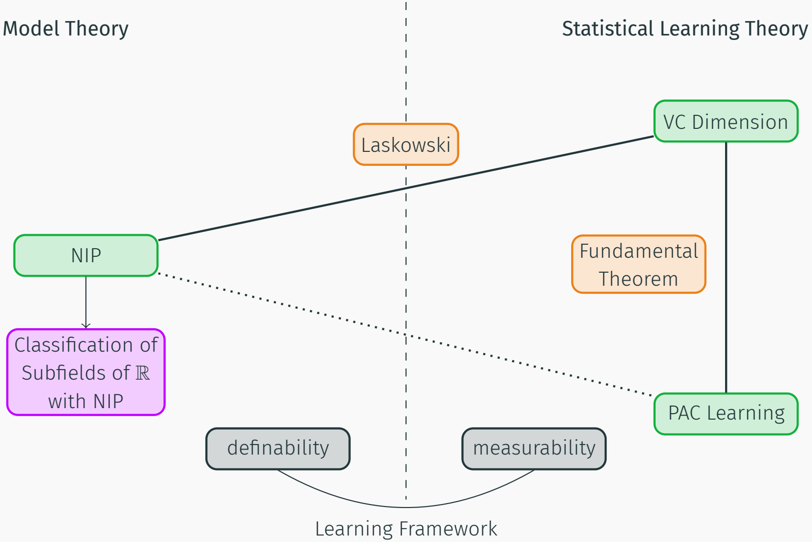
Subfields of $\mathbb{R}$: NIP, VC Dimension and PAC Learning

Laura Wirth
University of Konstanz

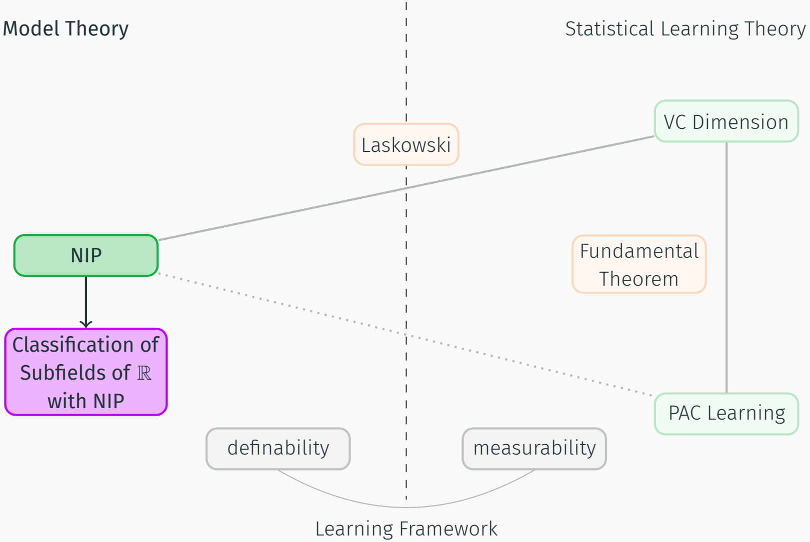
Combinatorial Problems in Model Theory and Computer Science
Leeds, November 2023

Leeds, Crown Point Bridge

Overview


Model Theory
Statistical Learning Theory
VC Dimension
Laskowski
NIP
Fundamental Theorem
Classification of Subfields of ℝ with NIP
PAC Learning
definability
measurability
Learning Framework

Model Theory of Subfields of $\mathbb{R}$



Definition.

Let $\mathcal{L}$ be any language and let T be a complete $\mathcal{L}$–theory.
A partitioned $\mathcal{L}$–formula $\varphi(x;y)$ has the independence property (IP) in T, if for any $n \in \mathbb{N}$, writing $[n] = \{1,\dots,n\}$, we have

$$T \models \exists a_1,\dots,a_n \, \exists b_\emptyset,\dots,b_{[n]} : \underbrace{\bigwedge_{\substack{i\in[n]\\ J\subseteq[n]\\ i\in J}} \varphi(a_i;b_J) \wedge \bigwedge_{\substack{i\in[n]\\ J\subseteq[n]\\ i\notin J}} \neg\varphi(a_i;b_J)}_{\varphi(a_i;b_J)\text{ is true iff } i\in J}.$$

An $\mathcal{L}$–structure $\mathcal{M}$ has the independence property (IP) if there exists a partitioned $\mathcal{L}$–formula $\varphi(x;y)$ that has IP in the theory $\mathrm{Th}_{\mathcal{L}}(\mathcal{M})$.

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.

Definition.

Let $\mathcal{L}$ be any language and let T be a complete $\mathcal{L}$–theory.
A partitioned $\mathcal{L}$–formula $\varphi(\boldsymbol{x};\boldsymbol{y})$ has the independence property (IP) in T, if for any $n \in \mathbb{N}$, writing $[n] = \{1,\dots,n\}$, we have

$$T \models \exists \boldsymbol{a}_1,\dots,\boldsymbol{a}_n \, \exists \boldsymbol{b}_\emptyset,\dots,\boldsymbol{b}_{[n]} : \underbrace{\bigwedge_{\substack{i\in[n]\\ J\subseteq[n]\\ i\in J}} \varphi(\boldsymbol{a}_i;\boldsymbol{b}_J) \wedge \bigwedge_{\substack{i\in[n]\\ J\subseteq[n]\\ i\notin J}} \neg\varphi(\boldsymbol{a}_i;\boldsymbol{b}_J)}_{\varphi(\boldsymbol{a}_i;\boldsymbol{b}_J)\text{ is true iff } i\in J}.$$

An $\mathcal{L}$–structure $\mathcal{M}$ has the independence property (IP) if there exists a partitioned $\mathcal{L}$–formula $\varphi(\boldsymbol{x};\boldsymbol{y})$ that has IP in the theory $\mathrm{Th}_{\mathcal{L}}(\mathcal{M})$.

NIP — not IP

Subfields of $\mathbb{R}$

Definition.

A subfield K of $\mathbb{R}$ is called real closed if it is relatively algebraically closed in $\mathbb{R}$.

Definition.

$\mathcal{L}_{\mathrm{r}} := \{+, -, \cdot, 0, 1\}$ — language of rings
$\mathcal{L}_{\mathrm{or}} := \{+, -, \cdot, 0, 1, <\}$ — language of ordered rings

Theorem.

A subfield of $\mathbb{R}$ has NIP if it is real closed.

S. Shelah, 'Strongly dependent theories', *Isr. J. Math.* **204** (2014) 1–83.

Subfields of $\mathbb{R}$

Definition.

A subfield K of $\mathbb{R}$ is called real closed if it is relatively algebraically closed in $\mathbb{R}$.

Definition.

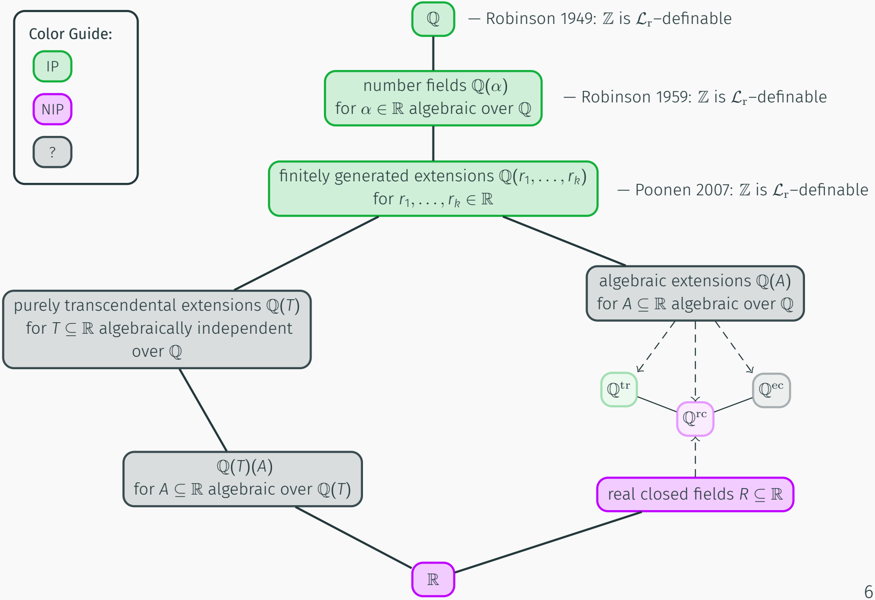
$\mathcal{L}_{\mathrm{r}} := \{+, -, \cdot, 0, 1\}$ — language of rings
$\mathcal{L}_{\mathrm{or}} := \{+, -, \cdot, 0, 1, <\}$ — language of ordered rings

Conjecture.

A subfield of $\mathbb{R}$ has NIP only if it is real closed.

S. Shelah, 'Strongly dependent theories', *Isr. J. Math.* **204** (2014) 1–83.

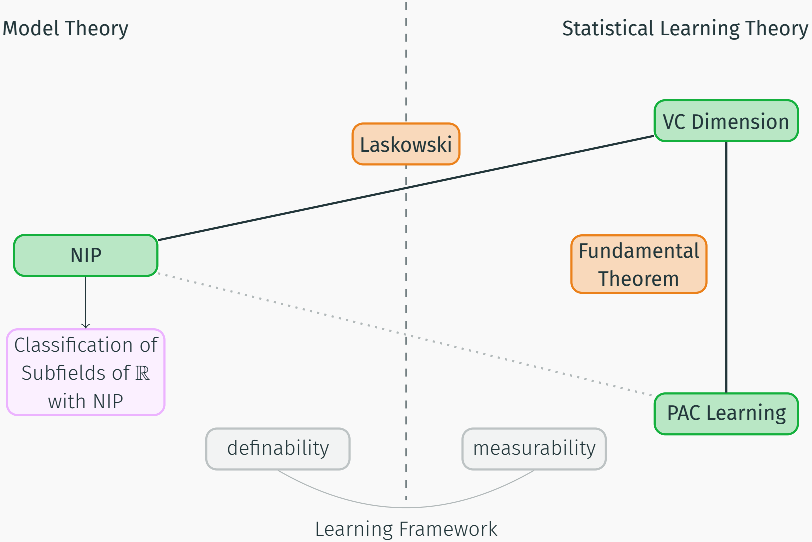
Subfields of $\mathbb{R}$ with (N)IP



References

- B. Poonen, 'Uniform first-order definitions in finitely generated fields', *Duke Math. J.* **138** (2007) 1–22.
- J. Robinson, 'Definability and decision problems in arithmetic', *J. Symb. Log.* **14** (1949) 98–114.
- J. Robinson, 'The undecidability of algebraic rings and fields', *Proc. Amer. Math. Soc.* **10** (1959) 950–957.

Relation of NIP to Statistical Learning Theory



NIP and VC Dimension

Given an $\mathcal{L}$–structure $\mathcal{M}$ and a partitioned $\mathcal{L}$–formula $\varphi(\boldsymbol{x};\boldsymbol{y})$ with $|\boldsymbol{x}| = n$ and $|\boldsymbol{y}| = \ell$, we set

$$\varphi(\mathcal{M};\boldsymbol{b}) := \{\boldsymbol{a} \in M^n \mid \mathcal{M} \models \varphi(\boldsymbol{a};\boldsymbol{b})\}$$

for any $\boldsymbol{b} \in M^\ell$, and

$$\mathcal{H}_\varphi := \{\mathbb{1}_{\varphi(\mathcal{M};\boldsymbol{b})} \mid \boldsymbol{b} \in M^\ell\}.$$

Proposition (Laskowski 1992).

The class $\mathcal{H}_\varphi$ has finite VC dimension if and only if the $\mathcal{L}$–formula $\varphi(\boldsymbol{x};\boldsymbol{y})$ has NIP in $\mathcal{M}$.

M. C. Laskowski, 'Vapnik–Chervonenkis classes of definable sets', *J. Lond. Math. Soc.* **45** (1992) 377–384.

The Setting

Ingredients of a Learning Problem.

- $\mathcal{X}$ — input space
- $\{0, 1\}$ — output space
- $\mathcal{Z} = \mathcal{X} \times \{0, 1\}$ — sample space
- $\mathcal{H} \subseteq \{0, 1\}^{\mathcal{X}}$ — hypothesis space

Learning – Basic Procedure

Using an arbitrary distribution $\mathbb{D}$ on $\mathcal{Z} = \mathcal{X} \times \{0, 1\}$, we choose a sequence of iid samples from $\mathcal{Z}$:

$$z = ((x_1, y_1), \ldots, (x_m, y_m)).$$

These samples provide the input data for a learning algorithm $\mathcal{A}$ that determines a hypothesis $h = \mathcal{A}(z)$ in $\mathcal{H}$.

S. Ben-David and S. Shalev-Shwartz, *Understanding Machine Learning: From Theory to Algorithms*, (Cambridge University Press, Cambridge, 2014).

Learning – Goal

The goal is to minimize the error of h given by

$$\mathrm{er}_{\mathbb{D}}(h) := \mathbb{D}(\{(x, y) \in \mathcal{Z} \mid h(x) \neq y\}).$$

More precisely, we want to achieve an error that is close to the approximation error of $\mathcal{H}$ given by

$$\mathrm{opt}_{\mathbb{D}}(\mathcal{H}) := \inf_{h \in \mathcal{H}} \mathrm{er}_{\mathbb{D}}(h).$$

S. Ben-David and S. Shalev-Shwartz, *Understanding Machine Learning: From Theory to Algorithms*, (Cambridge University Press, Cambridge, 2014).

PAC Learning

Definition.

A learning algorithm

$$\mathcal{A}\colon \bigcup_{m\in\mathbb{N}} \mathcal{Z}^m \to \mathcal{H}$$

for $\mathcal{H}$ is said to be probably approximately correct (PAC) if it satisfies the following condition:

$$\forall \varepsilon, \delta \in (0,1)\ \exists m_0 = m_0(\varepsilon,\delta)\ \forall m \geq m_0\ \forall \mathbb{D}:$$
$$\mathbb{D}^m(\{\mathbf{z} \in \mathcal{Z}^m \mid \mathrm{er}_{\mathbb{D}}(\mathcal{A}(\mathbf{z})) - \mathrm{opt}_{\mathbb{D}}(\mathcal{H}) \leq \varepsilon\}) > 1 - \delta.$$

L. G. Valiant, 'A Theory of the Learnable', *Comm. ACM* **27** (1984) 1134–1142.

Definition.

A learning algorithm

$$\mathcal{A}\colon \bigcup_{m \in \mathbb{N}} \mathcal{Z}^m \to \mathcal{H}$$

for $\mathcal{H}$ is said to be probably approximately correct (PAC) if it satisfies the following condition:

$$\forall \varepsilon, \delta \in (0,1)\ \exists m_0 = m_0(\varepsilon, \delta)\ \forall m \geq m_0\ \forall\, \mathbb{D}:$$
$$\mathbb{D}^m(\{\mathbf{z} \in \mathcal{Z}^m \mid \mathrm{er}_{\mathbb{D}}(\mathcal{A}(\mathbf{z})) - \mathrm{opt}_{\mathbb{D}}(\mathcal{H}) \leq \varepsilon\}) > 1 - \delta.$$

The hypothesis class $\mathcal{H}$ is said to be probably approximately correct (PAC) learnable if there exists a learning algorithm for $\mathcal{H}$ that is PAC.

The following result is due to Blumer, Ehrenfeucht, Haussler and Warmuth 1989.

Theorem.

Under certain measurability conditions, a hypothesis class $\mathcal{H}$ is PAC learnable if and only if its VC dimension is finite.

A. Blumer, A. Ehrenfeucht, D. Haussler and M. K. Warmuth, 'Learnability and the Vapnik-Chervonenkis dimension', *J. Assoc. Comput. Mach.* **36** (1989) 929–965.

Fundamental Theorem of Statistical Learning Theory

The following result is due to Blumer, Ehrenfeucht, Haussler and Warmuth 1989.

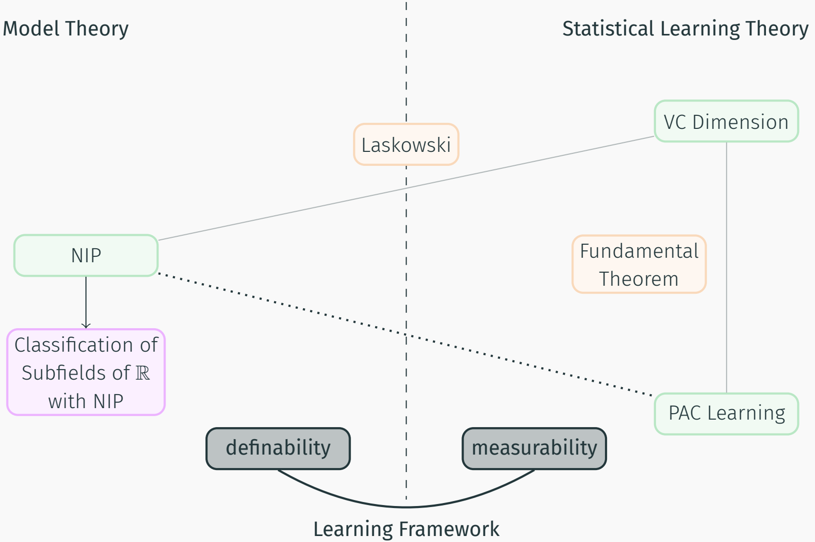
Theorem.

Under certain measurability conditions, a hypothesis class $\mathcal{H}$ is PAC learnable if and only if its VC dimension is finite.

Corollary.

Given a subfield $K \subseteq \mathbb{R}$ with NIP, any *definable* hypothesis class $\mathcal{H} \subseteq \{0,1\}^{\mathcal{X}}$ with $\mathcal{X} \subseteq K^n$ has finite VC dimension and is thus PAC learnable, provided certain measurability conditions are guaranteed.

Creating a Learning Framework



Creating a Learning Framework

Assembling the Ingredients.

- $K \subseteq \mathbb{R}$ — subfield of $\mathbb{R}$
- $\mathcal{L} \in \{\mathcal{L}_{\mathrm{r}}, \mathcal{L}_{\mathrm{or}}\}$ — language for model-theoretic examination
- $\mathcal{X} \subseteq K^n$ — *definable* subset of K^n
- $\mathcal{H} \subseteq \{0,1\}^{\mathcal{X}}$ — set of hypotheses $h = \mathbb{1}_A \colon \mathcal{X} \to \{0,1\}$
  with *definable* support $A \subseteq \mathcal{X}$

Measurability

Definition.

We denote by $\mathcal{B}_{\mathcal{X}}$ and $\mathcal{B}_{\mathcal{Z}}$ the Borel σ–algebras on $\mathcal{X}$ and $\mathcal{Z}$, respectively.

Lemma.

$\mathcal{B}_{\mathcal{Z}} = \mathcal{B}_{\mathcal{X}} \otimes \mathcal{P}(\{0, 1\})$.

The distributions $\mathbb{D}$ on $\mathcal{Z}$ that we consider are all defined on $\mathcal{B}_{\mathcal{Z}}$.

Measurability Issues

Given a hypothesis $h = \mathbb{1}_A$ with definable support $A \subseteq \mathcal{X}$ and a distribution $\mathbb{D}$ defined on $\mathcal{B}_{\mathcal{Z}}$, the error

$$\mathrm{er}_{\mathbb{D}}(h) = \mathbb{D}(\{(x, y) \in \mathcal{Z} \mid h(x) \neq y\})$$

is only well-defined if the support of h is Borel, i.e. $A \in \mathcal{B}_{\mathcal{X}}$.

Question.

Given a subfield $K \subseteq \mathbb{R}$, is any definable set $A \subseteq K^n$ Borel measurable?

Partial Answers.

Yes, if

- A is quantifier-free definable (e.g. if K is real closed),
- A is countable (e.g. if K is countable).

Summary and Future Work

Summary

Corollary.

If $K \subseteq \mathbb{R}$ has NIP, then any definable hypothesis class $\mathcal{H}$ has finite VC dimension and is thus PAC learnable, *provided certain measurability conditions are guaranteed.*

Task.

Identify and study the measurability requirements involved in the Fundamental Theorem.
For instance: Does definability guarantee Borel measurability?

Questions

Corollary.

If $K \subseteq \mathbb{R}$ has NIP, then any definable hypothesis class $\mathcal{H}$ is PAC learnable.

Question.

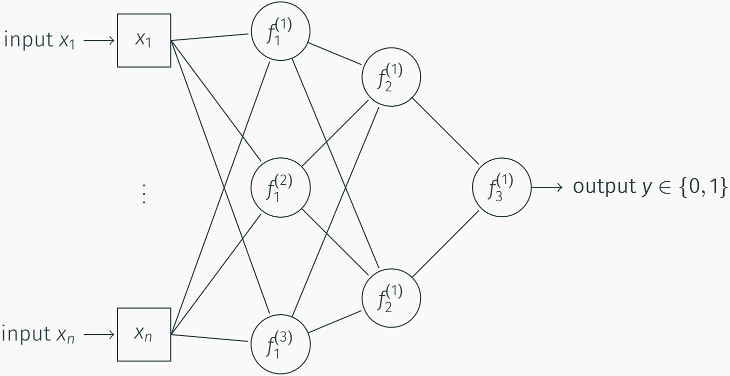
Given $K \subseteq \mathbb{R}$ with IP, does there exist a definable hypothesis class that is not PAC learnable? **Yes!**

Refined Question.

Given $K \subseteq \mathbb{R}$ with IP, does there exist a definable hypothesis class $\mathcal{H}$ generated by a neural network that is not PAC learnable?

Appendix

Neural Networks

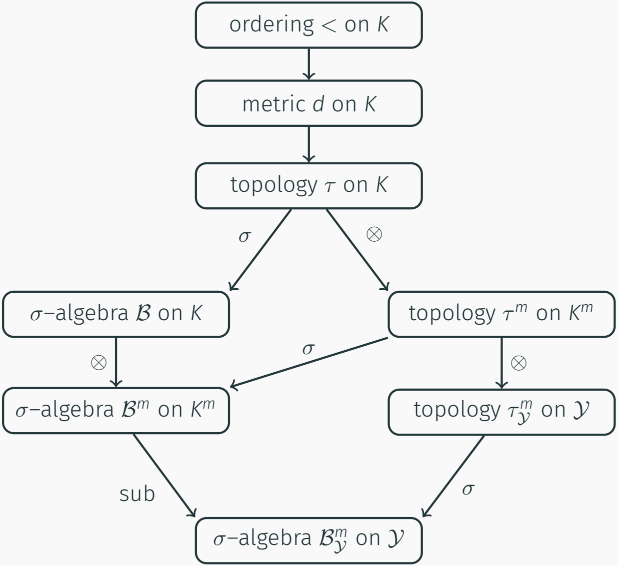


Measurability – Borel σ–Algebras

Setting:

$K \subseteq \mathbb{R}$ subfield

$\mathcal{Y} \subseteq K^m$



Measurability – Identity

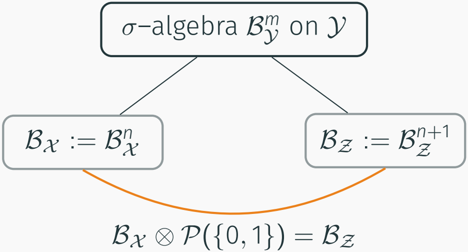
Setting:

$K \subseteq \mathbb{R}$ subfield

$\mathcal{Y} \subseteq K^m$

$\mathcal{X} \subseteq K^n$

$\mathcal{Z} = \mathcal{X} \times \{0,1\} \subseteq K^{n+1}$



Measurability – Hypotheses and Sets

Given a hypothesis $h = \mathbb{1}_A$ with support $A \subseteq \mathcal{X}$ and a distribution $\mathbb{D}$ defined on $\mathcal{B}_{\mathcal{Z}}$, the error

$$\mathrm{er}_{\mathbb{D}}(h) = \mathbb{D}(\{(x, y) \in \mathcal{Z} \mid h(x) \neq y\}) = \mathbb{D}(\mathcal{Z} \setminus \Gamma(h))$$

is only well-defined if one of the following equivalent conditions is satisfied:

- the map h is Borel measurable,
- the support A is Borel, i.e. $A \in \mathcal{B}_{\mathcal{X}}$,
- the graph $\Gamma(h)$ is Borel, i.e. $\Gamma(h) \in \mathcal{B}_{\mathcal{Z}}$.

Measurability Issues – Advanced

Problem.

Find sufficient/necessary conditions for the following measurability requirement:

$$\{z \in \mathcal{Z}^m \mid \mathrm{er}_{\mathbb{D}}(\mathcal{A}(z)) - \mathrm{opt}_{\mathbb{D}}(\mathcal{H}) \leq \varepsilon\} \overset{!}{\in} \mathcal{B}_{\mathcal{Z}}^m$$