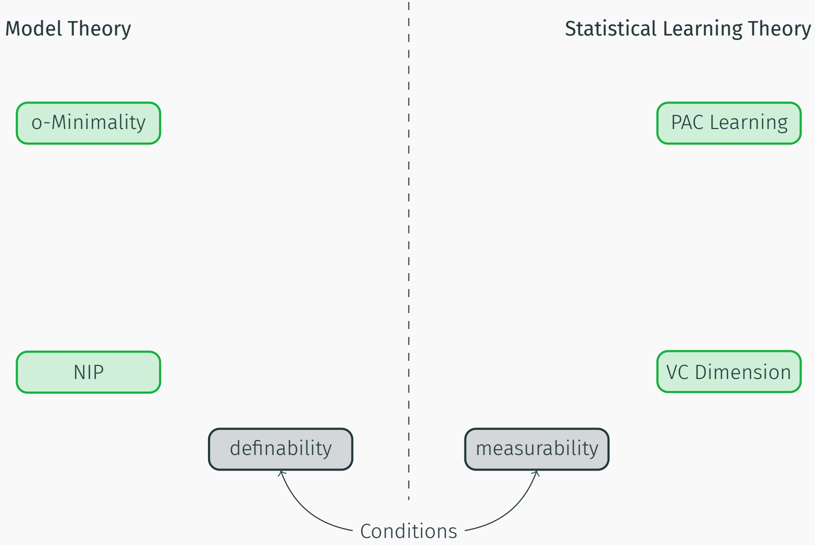
Learnability: A Model-Theoretic Perspective

(joint work with L. S. Krapp: arXiv:2410.10243)

Laura Wirth
University of Konstanz

Logic Seminar
Hebrew University of Jerusalem

Central Notions & Conditions



Measurability

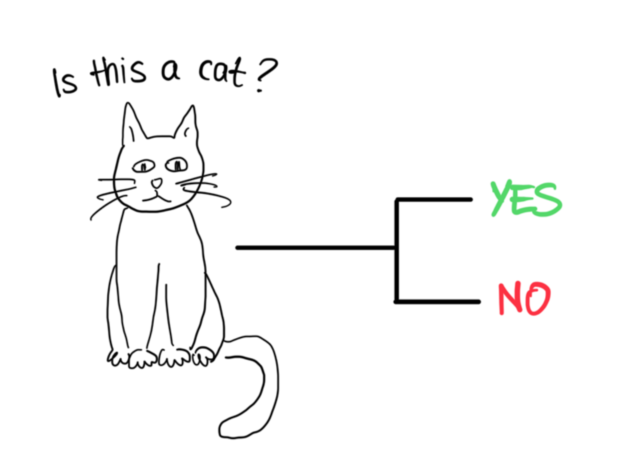
Recall.

A probability space $(\Omega, \Sigma, \mathbb{P})$ consists of

- a domain Ω,
- a σ–algebra $\Sigma \subseteq \mathcal{P}(\Omega)$,
- and a probability measure $\mathbb{P}\colon \Sigma \to [0, 1]$ (also called distribution).

Given a probability space $(\Omega, \Sigma, \mathbb{P})$, a map $g\colon \Omega \to \mathbb{R}$ is called Σ–measurable if $g^{-1}(B) \in \Sigma$ for any Borel set $B \subseteq \mathbb{R}$.

Binary Classification



Learning Framework

Ingredients of a Learning Problem.

- $\emptyset \neq \mathcal{X}$ – instance space
- $\mathcal{Z} = \mathcal{X} \times \{0,1\}$ – sample space
- $\emptyset \neq \mathcal{H} \subseteq \{0,1\}^{\mathcal{X}}$ – hypothesis space
- $\Sigma_{\mathcal{Z}}$ – σ–algebra on $\mathcal{Z}$ with $\mathcal{P}_{\text{fin}}(\mathcal{Z}) \subseteq \Sigma_{\mathcal{Z}}$
- $\mathcal{D}$ – set of distributions on $(\mathcal{Z}, \Sigma_{\mathcal{Z}})$

Assumption.

For any hypothesis $h \in \mathcal{H}$ we have

$$\Gamma(h) := \{(x, y) \in \mathcal{Z} \mid h(x) = y\} \in \Sigma_{\mathcal{Z}}.$$

Learning from Examples: Mathematically

A learning function is a map of the form $\mathcal{A}\colon \bigcup_{m\in\mathbb{N}} \mathcal{Z}^m \to \mathcal{H}$.

The input for $\mathcal{A}$ is generated according to an arbitrary distribution $\mathbb{D} \in \mathcal{D}$:

$$\boldsymbol{z} = \underbrace{((x_1, y_1), \ldots, (x_m, y_m))}_{\text{iid samples } \sim \mathbb{D}^m} \in \mathcal{Z}^m.$$

$\mathcal{A}$ then predicts a generalization hypothesis $h = \mathcal{A}(\boldsymbol{z}) \in \mathcal{H}$ based on the multi-sample $\boldsymbol{z}$.

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

Learning from Examples – Goal

The goal is to minimize the (true) error of h given by

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

More precisely, we want to achieve an error that is close to

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

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

Definition.
A learning function

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

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

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

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

Definition.

A learning function

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

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

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

The hypothesis space $\mathcal{H}$ is said to be PAC learnable if there exists a learning function for $\mathcal{H}$ that is PAC.

Definition.

A learning function

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

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

$$\forall \varepsilon, \delta \in (0,1)\ \exists m_0 \in \mathbb{N}\ \forall m \geq m_0\ \forall \mathbb{D} \in \mathcal{D}\ \exists C \in \mathbf{\Sigma}_{\mathcal{Z}}^m :$$

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

and $\mathbb{D}^m(C) \geq 1 - \delta$.

The hypothesis space $\mathcal{H}$ is said to be PAC learnable if there exists a learning function for $\mathcal{H}$ that is PAC.

Sample Error

The sample error of h on a multi-sample $\mathbf{z} = (z_1, \ldots, z_m) \in \mathcal{Z}^m$, given by

$$\hat{\mathrm{er}}_{\mathbf{z}}(h) := \frac{1}{m} \sum_{i=1}^{m} \mathbb{1}_{\mathcal{Z} \setminus \Gamma(h)}(z_i),$$

provides a useful estimate for the true error.

Remark.

The map

$$\mathcal{Z}^m \to \left\{ \tfrac{k}{m} \;\middle|\; k \in \{0, 1, \ldots, m\} \right\}, \; \mathbf{z} \mapsto \hat{\mathrm{er}}_{\mathbf{z}}(h)$$

is $\Sigma_{\mathcal{Z}}^m$–measurable.

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

A Simple Learning Principle

The sample error of h on a multi-sample $\mathbf{z} = (z_1, \ldots, z_m) \in \mathcal{Z}^m$, given by

$$\hat{\text{er}}_{\mathbf{z}}(h) := \frac{1}{m} \sum_{i=1}^{m} \mathbb{1}_{\mathcal{Z} \setminus \Gamma(h)}(z_i),$$

provides a useful estimate for the true error.

Sample Error Minimization (SEM).

Choose a learning function $\mathcal{A}$ such that

$$\hat{\text{er}}_{\mathbf{z}}(\mathcal{A}(\mathbf{z})) = \min_{h \in \mathcal{H}} \hat{\text{er}}_{\mathbf{z}}(h)$$

for any multi-sample $\mathbf{z}$.

Fundamental Theorem of Statistical Learning

In modern textbooks one finds formulations such as:

Theorem.
$\mathcal{H}$ is PAC learnable if and only if $\mathcal{H}$ has finite VC dimension.

Originally, this equivalence result is due to Blumer, Ehrenfeucht, Haussler and Warmuth 1989.

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

Theorem.

$\mathcal{H}$ is PAC learnable if and only if $\mathcal{H}$ has finite VC dimension.

Proof Remarks.

- Technical Heart: $\boxed{\Leftarrow}$

V. N. Vapnik and A. Ya. Chervonenkis, 'On the Uniform Convergence of Relative Frequencies of Events to Their Probabilities', *Teor. Veroyatn. Primen.* **16** (1971) 264–279 (Russian), *Theory Probab. Appl.* **16** (1971) 264–280 (English).

Fundamental Theorem of Statistical Learning

Theorem.

$\mathcal{H}$ is PAC learnable if and only if $\mathcal{H}$ has finite VC dimension.

Proof Remarks.

- Technical Heart: $\boxed{\Leftarrow}$
- Measurability requirements must be taken into account.

  $\longrightarrow$ Available literature lacks measure-theoretic analysis for agnostic PAC learning.

Takeaway.

The theorem only applies to well-behaved hypothesis spaces!

Definition.

A hypothesis space $\mathcal{H}$ is called well-behaved (with respect to $\mathcal{D}$) if it satisfies the following conditions:

- $\Gamma(h) \in \Sigma_{\mathcal{Z}}$ for any $h \in \mathcal{H}$.
- There exists $m_{\mathcal{H}} \in \mathbb{N}$ such that:
  The map

$$U \colon \mathcal{Z}^m \to [0,1],\; \mathbf{z} \mapsto \sup_{h \in \mathcal{H}} \left| \mathrm{er}_{\mathbb{D}}(h) - \hat{\mathrm{er}}_{\mathbf{z}}(h) \right|$$

  is $\Sigma_{\mathcal{Z}}^m$–measurable for any $m \geq m_{\mathcal{H}}$ and any $\mathbb{D} \in \mathcal{D}$,
  and the map

$$V \colon \mathcal{Z}^{2m} \to [0,1],\; (\mathbf{z}, \mathbf{z}') \mapsto \sup_{h \in \mathcal{H}} \left| \hat{\mathrm{er}}_{\mathbf{z}'}(h) - \hat{\mathrm{er}}_{\mathbf{z}}(h) \right|$$

  is $\Sigma_{\mathcal{Z}}^{2m}$–measurable for any $m \geq m_{\mathcal{H}}$.

Sufficient Conditions for Well-Behavedness

Remark.

Sufficient conditions for the measurability of the maps U and V:

- $\mathcal{X}$ resp. $\mathcal{Z}$ is countable.
- $\mathcal{H}$ is countable.
- $\mathcal{H}$ is universally separable.

Definition.

The hypothesis space $\mathcal{H}$ is called universally separable if there exists a countable subset $\mathcal{H}_0 \subseteq \mathcal{H}$ such that for any $h \in \mathcal{H}$ there exists a sequence $\{h_n\}_{n \in \mathbb{N}} \subseteq \mathcal{H}_0$ converging pointwise to h.

Fundamental Theorem of Statistical Learning: *Agnostic* Version

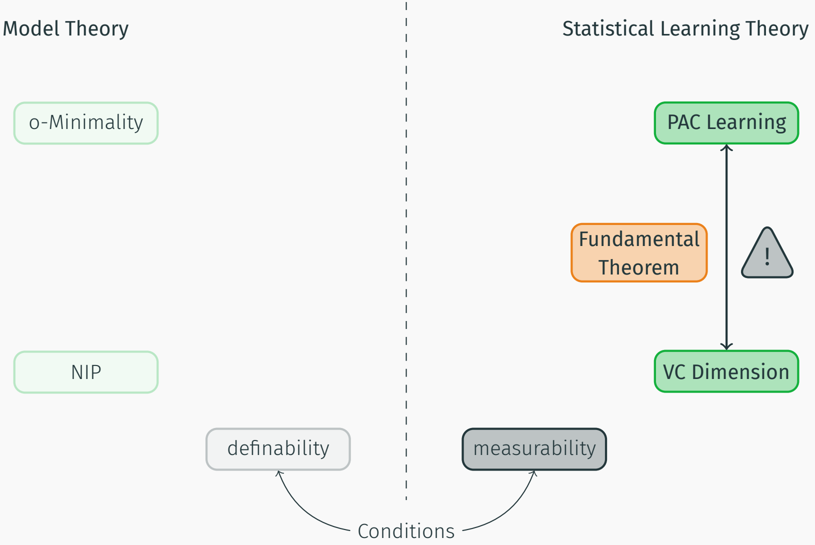
Based on Blumer, Ehrenfeucht, Haussler and Warmuth 1989, we could prove the following:

Theorem.

Let $\mathcal{D}$ contain all discrete uniform distributions and let $\mathcal{H}$ be well-behaved with respect to $\mathcal{D}$. Then $\mathcal{H}$ is PAC learnable with respect to $\mathcal{D}$ if and only if $\mathcal{H}$ has finite VC dimension.

L. S. Krapp and L. Wirth, 'Measurability in the Fundamental Theorem of Statistical Learning', Preprint, 2024, arXiv:2410.10243.

Statistical Learning Theory: Summary



Definable Hypothesis Spaces

Definiton.

Let $\mathcal{L}$ be a language, let $\mathcal{M}$ be an $\mathcal{L}$–structure and let $\varphi(x_1, \ldots, x_n; p_1, \ldots, p_\ell)$ be an $\mathcal{L}$–formula. For any $\boldsymbol{w} \in M^\ell$, set

$$\varphi(\mathcal{M}, \boldsymbol{w}) = \{\boldsymbol{a} \in M^n \mid \mathcal{M} \models \varphi(\boldsymbol{a}; \boldsymbol{w})\}.$$

Then the hypothesis space $\mathcal{H}^\varphi \subseteq \{0,1\}^{M^n}$ is given by

$$\mathcal{H}^\varphi := \left\{ \mathbb{1}_{\varphi(\mathcal{M}; \boldsymbol{w})} \,\middle|\, \boldsymbol{w} \in M^\ell \right\}.$$

Further, given a non-empty set $\mathcal{X} \subseteq M^n$ that is definable over $\mathcal{M}$, the hypothesis space $\mathcal{H}^\varphi_\mathcal{X} \subseteq \{0,1\}^\mathcal{X}$ is given by

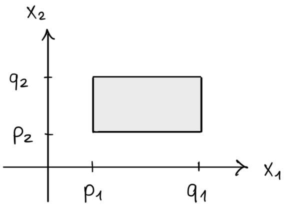
$$\mathcal{H}^\varphi_\mathcal{X} := \{h{\restriction}_\mathcal{X} \mid h \in \mathcal{H}^\varphi\}.$$

Example

Set $\mathcal{L}_{\text{or}} := \{+, \cdot, -, 0, 1, <\}$, $\mathbb{R}_{\text{or}} := (\mathbb{R}, +, \cdot, -, 0, 1, <)$ and consider the $\mathcal{L}$–formula $\varphi(x_1, x_2; p_1, q_1, p_2, q_2)$ given by

$$p_1 \leq x_1 \leq q_1 \wedge p_2 \leq x_2 \leq q_2.$$

For $w = (p_1, q_1, p_2, q_2) \in \mathbb{R}^4$, the set $\varphi(\mathbb{R}_{\text{or}}; w)$ is an axis-aligned rectangle in $\mathbb{R}^2$ of the form:

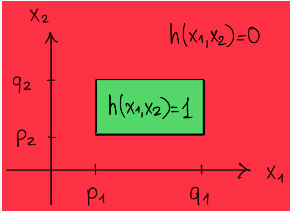


Example

Set $\mathcal{L}_{\mathrm{or}} := \{+, \cdot, -, 0, 1, <\}$, $\mathbb{R}_{\mathrm{or}} := (\mathbb{R}, +, \cdot, -, 0, 1, <)$ and consider the $\mathcal{L}$–formula $\varphi(x_1, x_2; p_1, q_1, p_2, q_2)$ given by

$$p_1 \leq x_1 \leq q_1 \wedge p_2 \leq x_2 \leq q_2.$$

For $\boldsymbol{w} = (p_1, q_1, p_2, q_2) \in \mathbb{R}^4$, the set $\varphi(\mathbb{R}_{\mathrm{or}}; \boldsymbol{w})$ is an axis-aligned rectangle in $\mathbb{R}^2$ of the form:



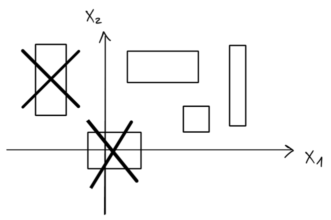
The hypothesis $h = \mathbb{1}_{\varphi(\mathbb{R}_{\mathrm{or}}; \boldsymbol{w})} \in \mathcal{H}^{\varphi}$ sends all points inside this rectangle to 1 and all points outside to 0.

Example

Set $\mathcal{L}_{\text{or}} := \{+, \cdot, -, 0, 1, <\}$, $\mathbb{R}_{\text{or}} := (\mathbb{R}, +, \cdot, -, 0, 1, <)$ and consider the $\mathcal{L}$–formula $\varphi(x_1, x_2; p_1, q_1, p_2, q_2)$ given by

$$p_1 \leq x_1 \leq q_1 \wedge p_2 \leq x_2 \leq q_2.$$

For $w = (p_1, q_1, p_2, q_2) \in \mathbb{R}^4$, the set $\varphi(\mathbb{R}_{\text{or}}; w)$ is an axis-aligned rectangle in $\mathbb{R}^2$.



Restricting to the $\mathcal{L}_{\text{or}}$–definable set $\mathcal{X} = \mathbb{R}_{\geq 0} \times \mathbb{R}_{\geq 0} \subseteq \mathbb{R}^2$ means considering only rectangles in the first quadrant.

The following notion is due to Pillay and Steinhorn 1986.

Recall.

Given a language $\mathcal{L} = \{<, \dots\}$ and an $\mathcal{L}$–structure $\mathcal{M} = (M, <, \dots)$ for which $(M, <)$ is a linear order, $\mathcal{M}$ is called o-minimal if any set $A \subseteq M$ that is definable over $\mathcal{M}$ can be expressed as a finite union of points and open intervals.

They further related it to the notion *NIP*, which was introduced by Shelah 1971.

Proposition.

If $\mathcal{M} = (M, <, \dots)$ is o-minimal, then $\mathcal{M}$ has NIP.

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.

The following result is due to Laskowski 1992.

Proposition.

Let $\mathcal{L}$ be a language and let $\mathcal{M}$ be an $\mathcal{L}$–structure. Then the following conditions are equivalent:

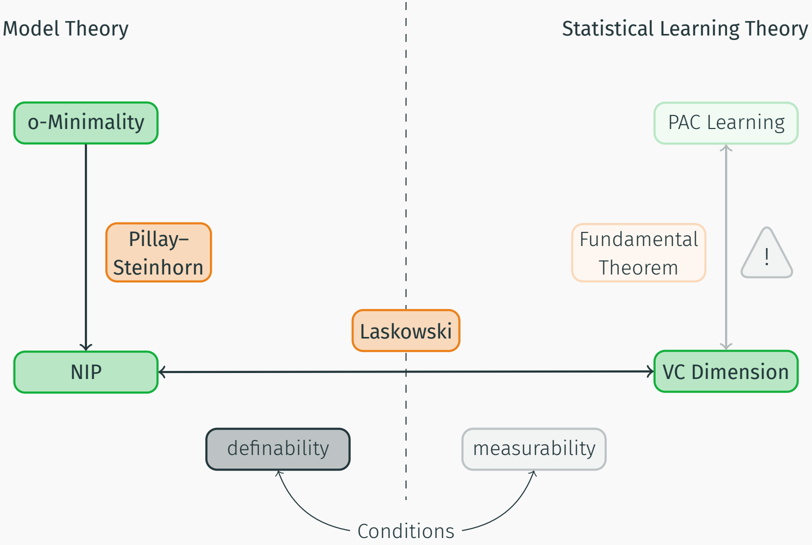
(1) $\mathcal{M}$ has NIP.

(2) The hypothesis space $\mathcal{H}^{\varphi}$ has finite VC dimension for any $\mathcal{L}$–formula $\varphi(x; p)$.

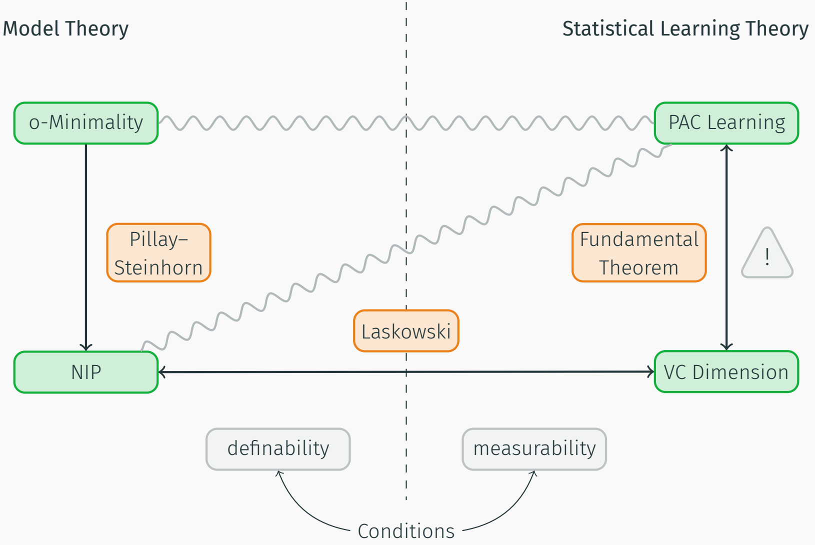
(3) The hypothesis space $\mathcal{H}^{\varphi}_{\mathcal{X}}$ has finite VC dimension for any $\mathcal{L}$–formula $\varphi(x; p)$ and any non-empty set $\mathcal{X}$ definable over $\mathcal{M}$.

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

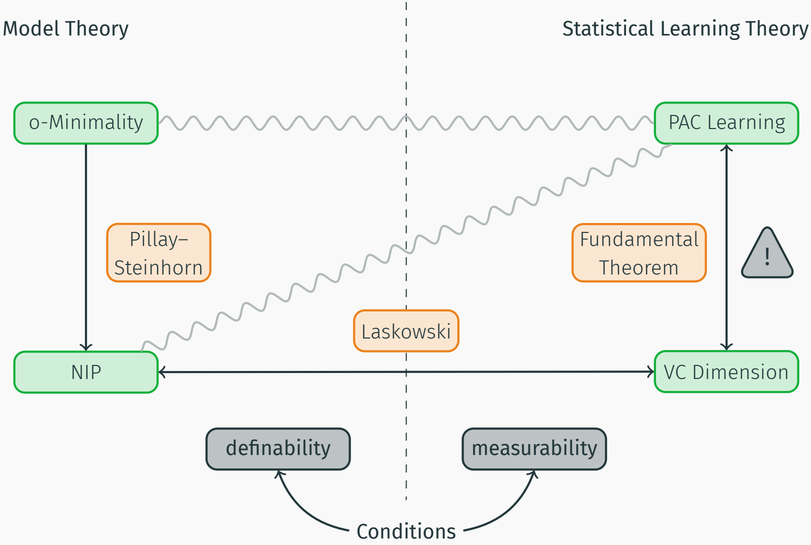
o-Minimality, NIP and VC Dimension: Summary



Jumping to Conclusions



Jumping to Conclusions – Be aware of the measurability!



Recall.

Given $k \in \mathbb{N}$, the Borel σ–algebra $\mathcal{B}(\mathbb{R}^k)$ of $\mathbb{R}^k$ is the smallest σ–algebra containing all open sets in $\mathbb{R}^k$.
For $\mathcal{Y} \subseteq \mathbb{R}^k$, we consider the trace σ–algebra given by

$$\mathcal{B}(\mathcal{Y}) := \{B \cap \mathcal{Y} \mid B \in \mathcal{B}(\mathbb{R}^k)\}.$$

o-Minimal Expansions of the Reals: Measurability

Set $\mathcal{L}_{\mathrm{or}} := \{+, \cdot, -, 0, 1, <\}$ and $\mathbb{R}_{\mathrm{or}} := (\mathbb{R}, +, \cdot, -, 0, 1, <)$.

Lemma.

Let $\mathcal{L}$ be a language expanding $\mathcal{L}_{\mathrm{or}}$, let $\mathcal{R}$ be an o-minimal $\mathcal{L}$–expansion of $\mathbb{R}_{\mathrm{or}}$, let $\varphi(x_1, \ldots, x_n; p_1, \ldots, p_\ell)$ be an $\mathcal{L}$–formula and let $w \in \mathbb{R}^\ell$. Then $\varphi(\mathcal{R}; w) \in \mathcal{B}(\mathbb{R}^n)$.

M. Karpinski and A. Macintyre, 'Approximating Volumes and Integrals in o-Minimal and p-Minimal Theories', *Connections between model theory and algebraic and analytic geometry* (ed. Macintyre), Quad. Mat. 6 (2000) 149–177.

Learning over o-Minimal Expansions of $\mathbb{R}_{\mathrm{or}}$

Guided by the work of Karpinski and Macintyre 2000 we proved the following learnability result:

Theorem. Let

- $\mathcal{L}$ be a language expanding $\mathcal{L}_{\mathrm{or}}$,
- $\mathcal{R}$ be an o-minimal $\mathcal{L}$–expansion of $\mathbb{R}_{\mathrm{or}}$,
- $\mathcal{X} \subseteq \mathbb{R}^n$ be a non-empty set that is definable over $\mathcal{R}$,
- $\varphi(x_1, \ldots, x_n; p_1, \ldots, p_\ell)$ be an $\mathcal{L}$–formula,
- $\Sigma_{\mathcal{Z}}$ be a σ–algebra on $\mathcal{Z} = \mathcal{X} \times \{0, 1\}$ with $\mathcal{B}(\mathcal{Z}) \subseteq \Sigma_{\mathcal{Z}}$, and
- $\mathcal{D}$ be a set of distributions on $(\mathcal{Z}, \Sigma_{\mathcal{Z}})$ such that $(\mathcal{Z}^m, \Sigma_{\mathcal{Z}}^m, \mathbb{D}^m)$ is a complete probability space for any $\mathbb{D} \in \mathcal{D}$ and any $m \in \mathbb{N}$.

Then $\mathcal{H}_{\mathcal{X}}^{\varphi}$ is PAC learnable with respect to $\mathcal{D}$.

L. S. Krapp and L. Wirth, 'Measurability in the Fundamental Theorem of Statistical Learning', Preprint, 2024, arXiv:2410.10243.

Proof Sketch.

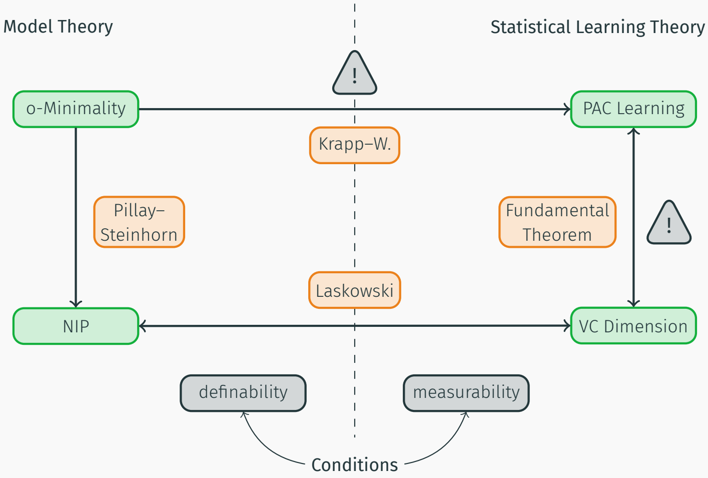
- o-Minimality implies NIP.
- Thus, $\mathcal{H}_{\mathcal{X}}^{\varphi}$ has finite VC dimension.
- Aim: Apply Fundamental Theorem.
- To this end: Verify well-behavedness.
- $\Gamma(h) \in \Sigma_{\mathcal{Z}}$ for any $h \in \mathcal{H}_{\mathcal{X}}^{\varphi}$.
- Technical analysis and application of Pollard's arguments regarding measurability of suprema establish measurability of the maps U and V.

□

Exkurs.

D. Pollard, *Convergence of Stochastic Processes*, Springer Ser. Stat. (Springer, New York, 1984).

Summary



L. S. Krapp and L. Wirth, 'Measurability in the Fundamental Theorem of Statistical Learning', Preprint, 2024, arXiv:2410.10243.

Question.

Are there a hypothesis space $\mathcal{H}$ with finite VC dimension and a set $\mathcal{D}$ of distributions such that $\mathcal{H}$ is not PAC learnable with respect to $\mathcal{D}$?

Further Work

Question.

Are there a hypothesis space $\mathcal{H}$ with finite VC dimension and a set $\mathcal{D}$ of distributions such that $\mathcal{H}$ is not PAC learnable with respect to $\mathcal{D}$?

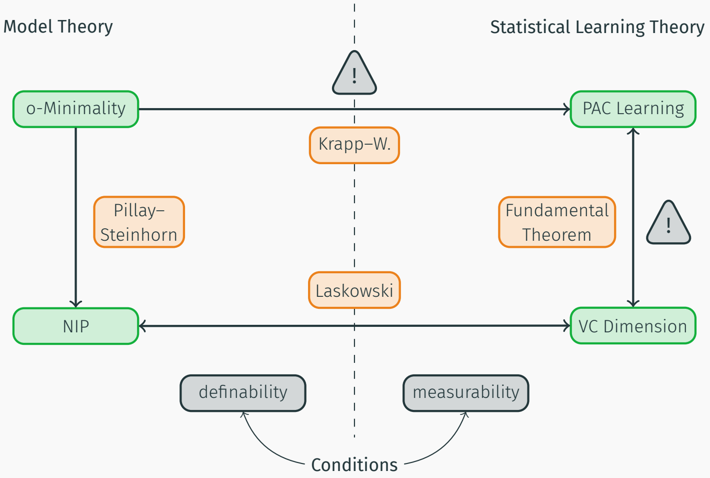
Special Case: Consider $\mathcal{Z} = \mathbb{R} \times \{0, 1\}$.

Question.

Are there an $\mathcal{L}_{\mathrm{or}}$–formula $\varphi(x; p)$ and a set $\mathcal{D}$ of distributions on $(\mathcal{Z}, \mathcal{B}(\mathcal{Z}))$ such that $\mathcal{H}^{\varphi} = \{\mathbb{1}_{\varphi(\mathbb{R}_{\mathrm{or}}; w)} \mid w \in \mathbb{R}\}$ is not PAC learnable with respect to $\mathcal{D}$?

Note: Such a hypothesis space would not be well-behaved, and the distribution set would not fulfill the completeness condition.

Thank you for your attention. Questions?



L. S. KRAPP and L. WIRTH, 'Measurability in the Fundamental Theorem of Statistical Learning', Preprint, 2024, arXiv:2410.10243.

Appendix

Definition.

A probability space $(\Omega, \Sigma, \mathbb{P})$ is called complete if

$$\forall N \in \Sigma\ (\mathbb{P}(N) = 0 \Rightarrow \forall A \subseteq N\ A \in \Sigma).$$

Learning over o-Minimal Expansions of $\mathbb{R}_{\mathrm{or}}$ – Remarks

Definition.

A probability space $(\Omega, \Sigma, \mathbb{P})$ is called complete if

$$\forall N \in \Sigma \; (\mathbb{P}(N) = 0 \Rightarrow \forall A \subseteq N \; A \in \Sigma).$$

Remarks.

- Completeness condition is crucial for deducing measurability of the maps U and V.
- Completeness condition is trivially satisfied if $\mathcal{X}$ resp. $\mathcal{Z}$ is countable, since then $\Sigma_{\mathcal{Z}} = \mathcal{B}(\mathcal{Z}) = \mathcal{P}(\mathcal{Z})$.
- Potential Solution: Replace product spaces with their respective completions at all relevant places.

  $\longrightarrow$ All impacted definitions and proofs need to be adjusted accordingly!

Jump back.

Further Remarks.

- The measurability of V can be established without imposing further conditions (like e.g. the completeness condition), since it can be shown to be definable.

  $\longrightarrow$ Unfortunately, this approach does not work for U.
- Further Work: Extend theorem to general o-minimal structures.

Jump back.
Jump to Summary.

VC Dimension

Definition.

Given $A \subseteq \mathcal{X}$, we say that $\mathcal{H}$ shatters A if

$$\{h\restriction_A \mid h \in \mathcal{H}\} = \{0,1\}^A.$$

If $\mathcal{H}$ cannot shatter sets of arbitrarily large size, then we say that $\mathcal{H}$ has finite VC dimension. In this case:

$$\mathrm{vc}(\mathcal{H}) := \max\{d \in \mathbb{N} \mid \exists A \subseteq \mathcal{X}, |A| = d \colon \mathcal{H} \text{ shatters } A\}.$$

Jump back.

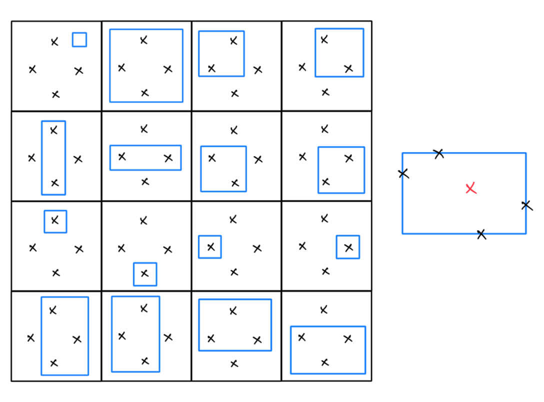
V. N. Vapnik and A. Ja. Červonenkis, 'Uniform Convergence of Frequencies of Occurrence of Events to Their Probabilities', *Dokl. Akad. Nauk SSSR* **181** (1968) 781–783 (Russian), *Sov. Math. Dokl.* **9** (1968) 915–918 (English).

Example

Consider $\mathbb{R}_{\mathrm{or}}$ and the hypothesis space $\mathcal{H}^{\varphi}$ defined by the $\mathcal{L}_{\mathrm{or}}$–formula $\varphi(x_1, x_2; p_1, q_1, p_2, q_2)$ given by

$$p_1 \leq x_1 \leq q_1 \wedge p_2 \leq x_2 \leq q_2.$$

We compute $\mathrm{vc}(\mathcal{H}^{\varphi}) = 4$:



Jump back.

Discrete Uniform Distributions

Definition

A discrete uniform distribution on a measurable space (Ω, Σ) with $\mathcal{P}_{\mathrm{fin}}(\Omega) \subseteq \Sigma$ is a probability measure $\mathbb{P}\colon \Sigma \to [0, 1]$ of the form

$$\mathbb{P} = \sum_{j=1}^{\ell} \frac{1}{\ell} \delta_{\omega_j},$$

where $\ell \in \mathbb{N}$ and $\omega_1, \ldots, \omega_\ell \in \Omega$.

Jump back.

NIP

Notation.

$[m] := \{1, \ldots, m\}$ for $m \in \mathbb{N}$.

Definition.

Let $\mathcal{L}$ be a language and let $\mathcal{M}$ be an $\mathcal{L}$–structure.
A (partitioned) $\mathcal{L}$–formula $\varphi(x_1, \ldots, x_n; p_1, \ldots, p_\ell)$ has NIP over $\mathcal{M}$ if there is $m \in \mathbb{N}$ such that for any object set $\{\boldsymbol{a}_1, \ldots, \boldsymbol{a}_m\} \subseteq M^n$ and any parameter set $\{\boldsymbol{w}_I \mid I \subseteq [m]\} \subseteq M^\ell$, there is some $J \subseteq [m]$ such that

$$\mathcal{M} \not\models \underbrace{\bigwedge_{i \in J} \varphi(\boldsymbol{a}_i; \boldsymbol{w}_J) \wedge \bigwedge_{i \in [m] \setminus J} \neg\varphi(\boldsymbol{a}_i; \boldsymbol{w}_J)}_{\varphi(\boldsymbol{a}_i; \boldsymbol{w}_J) \text{ is true iff } i \in J}.$$

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 a language and let $\mathcal{M}$ be an $\mathcal{L}$–structure.
A (partitioned) $\mathcal{L}$–formula $\varphi(x_1,\dots,x_n;p_1,\dots,p_\ell)$ has NIP over $\mathcal{M}$ if there is $m \in \mathbb{N}$ such that for any object set $\{\boldsymbol{a}_1,\dots,\boldsymbol{a}_m\} \subseteq M^n$ and any parameter set $\{\boldsymbol{w}_I \mid I \subseteq [m]\} \subseteq M^\ell$, there is some $J \subseteq [m]$ such that

$$\mathcal{M} \not\models \underbrace{\bigwedge_{i\in J} \varphi(\boldsymbol{a}_i;\boldsymbol{w}_J) \wedge \bigwedge_{i\in[m]\setminus J} \neg\varphi(\boldsymbol{a}_i;\boldsymbol{w}_J)}_{\varphi(\boldsymbol{a}_i;\boldsymbol{w}_J) \text{ is true iff } i\in J}.$$

The $\mathcal{L}$–structure $\mathcal{M}$ has NIP if every $\mathcal{L}$–formula has NIP over $\mathcal{M}$.

Jump back.

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.