

The Adequacy of Boosting for All Machine Learning: Higher Categorical Equivalence of Weak and Strong Learning in ∞ -Toposes

Collaborative Agent Team
Higher Categorical Semantics

October 2026

Abstract

We prove the adequacy of boosting for machine learning: for every sufficiently expressive hypothesis class, weak learnability is equivalent to strong learnability. The argument is internal to the language of homotopy type theory and the theory of ∞ -toposes. Hypothesis classes are modeled as objects in an ∞ -topos of classifiers; weak and strong learners are sections of associated fibrations of error measures. Boosting is realized as the higher inductive construction of a geometric realization of an iterative sequence of weak hypotheses, which, under a density condition corresponding to sufficient expressivity, yields a homotopy equivalence between the spaces of weak and strong learners. The construction is dual to a higher categorical Fourier reconstruction, in which the weighted ensemble recovers the target concept from the spectrum of its weak approximations via a monadic or chromatic Fourier transform in the monoidal ∞ -category of learners. This furnishes a homotopy-coherent, sheaf-theoretic refinement of Schapire’s classical equivalence theorem.

Contents

1	Introduction	2
2	Preliminaries on PAC Learning and Boosting	3
2.1	PAC Learning Model	3
2.2	Schapire’s Theorem	3
2.3	AdaBoost Algorithm	4
3	Homotopy Type Theory and ∞-Toposes	4
3.1	Homotopy Type Theory	4
3.2	∞ -Toposes	4
4	Hypothesis Classes in ∞-Toposes	5
5	Weak and Strong Learning	6
6	The Boosting Construction as a Higher Inductive Type	6

7 Expressivity and Density	7
8 Connection to Higher Categorical Fourier Reconstruction	8
9 Formal Proof Obligations and Computational Evidence	8
10 Conclusion	9
11 Detailed Proof of Error Contraction	9
12 Examples of Sufficiently Expressive Classes	10
13 Comparison with Classical Sample Complexity	10
14 Open Directions	10
15 Further Elaboration on Path Induction in Boosting	10
16 Sheaf Condition and Effective Descent	11
17 Monadic Reconstruction Perspective	11
18 Quantitative Bounds in the Internal Language	11
19 Extended Analysis of the Boosting HIT	12
20 Density and the Weak Topology	12
21 Fourier Reconstruction in Detail	13
22 Related Categorical Models of Learning	13
23 Implications for Categorical Semantics	13

1 Introduction

The classical theory of boosting, originating in the work of Kearns and Valiant and resolved by Schapire [1], establishes that weak PAC learnability of a concept class is equivalent to strong PAC learnability. A weak learner produces hypotheses whose error is bounded away from $1/2$ by a fixed edge $\gamma > 0$ under every distribution; a strong learner achieves arbitrarily small error. Boosting algorithms constructively convert the former into the latter.

The present work internalizes this equivalence in higher categorical semantics. We work in an ∞ -topos $\mathcal{E}$ whose objects interpret spaces of classifiers, and whose internal language is homotopy type theory. A hypothesis class $\mathcal{H}$ is an object of $\mathcal{E}$; a distribution is an internal measure; a learner is a section of a fibration whose fibers encode error. Sufficient expressivity of $\mathcal{H}$ is formalized as a density condition: the full subcategory generated by $\mathcal{H}$ is dense with respect to a Grothendieck topology whose covering families are generated by weak hypotheses. Under this condition the weak-covering families form a basis for the topology, and the sheaf condition

(or, equivalently, the higher Giraud axioms of Lurie [3]) ensures that local weak sections glue to a global strong section.

The boosting construction itself is expressed as a higher inductive type whose constructors iteratively adjoin weighted weak hypotheses and whose path constructors enforce the reduction of error. The resulting map from the type of weak learners to the type of strong learners is a homotopy equivalence. Dual to this inductive construction is a higher categorical Fourier reconstruction: the weak hypotheses act as characters or Fourier modes in the monoidal ∞ -category of endomorphisms of $\mathcal{H}$, and the boosting combination is the inverse Fourier transform recovering the target concept, in the sense of Tannaka reconstruction or the chromatic Fourier transform of higher algebra.

The paper is organized as follows. Section 2 recalls the classical PAC and boosting results in detail. Section 3 summarizes the relevant fragments of homotopy type theory and ∞ -topos theory, including higher inductive types and effective groupoids. Sections 4–7 develop the categorical model of hypothesis classes, the definitions of weak and strong learning as dependent types, the boosting HIT, and the main equivalence theorem under density. Section 8 treats the connection to higher categorical Fourier reconstruction via monadic and chromatic transforms. Section 9 discusses formal proof obligations expressible in HoTT and existing computational evidence from Lean formalizations. Section 10 concludes with implications for categorical semantics of learning.

In categorical logic, the problem of weak versus strong learning is analyzed as the question of whether local sections of the error fibration, satisfying a uniform positive edge condition, extend to global sections with arbitrary precision. The higher categorical principles of effective descent and path induction provide the formal mechanism by which the iterative boosting process resolves this extension problem for dense subobjects.

2 Preliminaries on PAC Learning and Boosting

2.1 PAC Learning Model

Let X be an instance space and $\mathcal{C} \subseteq \{\pm 1\}^X$ a concept class. A (strong) PAC learner for $\mathcal{C}$ is an algorithm that, given access to an example oracle for an unknown $c \in \mathcal{C}$ and any $\varepsilon, \delta \in (0, 1)$, returns with probability at least $1 - \delta$ a hypothesis h satisfying $\text{err}_D(h, c) \leq \varepsilon$, where D is the unknown distribution, and whose sample complexity is polynomial in $1/\varepsilon$, $1/\delta$ and the relevant dimension parameters.

A weak PAC learner is required only to achieve some fixed edge $\gamma > 0$ and some fixed confidence $\delta_0 < 1$, for every distribution. The hypothesis boosting problem, posed by Kearns and Valiant, asks whether the existence of a weak learner implies the existence of a strong learner. Intuitively, one might expect that classes exist which are weakly but not strongly learnable, yet the boosting construction shows that no such classes exist under standard assumptions.

The formal analysis identifies the core problem as the uniform quantification over distributions: the weak learner must succeed on every reweighted distribution induced by previous errors. Categorical logic treats this as a dependent type over the space of distributions, and the boosting process as a recursive construction in that dependent type.

2.2 Schapire's Theorem

Theorem 2.1 (Schapire 1990). *A concept class $\mathcal{C}$ is weakly PAC learnable if and only if it is strongly PAC learnable.*

The non-trivial direction is constructive: given a weak learner W with edge γ , one builds a strong learner by a recursive procedure that repeatedly invokes W on reweighted distributions and combines the resulting hypotheses by majority vote. The error of the combined hypothesis satisfies a quadratic reduction $g(a) = 3a^2 - 2a^3$, which iterates to arbitrarily small error in a number of stages logarithmic in $1/\varepsilon$. Subsequent algorithms (Freund's boost-by-majority, Freund-Schapire AdaBoost) improve the sample complexity and adaptivity while preserving the same qualitative equivalence.

The proof proceeds by induction on the recursion depth. At each stage the three weak hypotheses are combined by majority, and the error is shown to contract according to the cubic polynomial g . The depth is $O(\log \log(1/\varepsilon))$, ensuring polynomial sample complexity when the weak learner itself is efficient.

2.3 AdaBoost Algorithm

AdaBoost maintains a distribution D_t over the training sample. At stage t it calls the weak learner to obtain h_t with edge $\gamma_t > 0$, sets $\alpha_t = \frac{1}{2} \log \frac{1-\varepsilon_t}{\varepsilon_t}$, and reweights

$$D_{t+1}(i) \propto D_t(i) \exp(-\alpha_t y_i h_t(x_i)).$$

The final hypothesis is the weighted majority sign($\sum_t \alpha_t h_t$). Training error decays as $\exp(-2 \sum \gamma_t^2)$. This exponential decay is the key quantitative feature internalized in the path constructors of the boosting higher inductive type.

The algorithm is adaptive: it does not require prior knowledge of γ , and the weights α_t automatically emphasize harder examples. In the categorical setting this adaptivity corresponds to a natural transformation between successive stages of the fibration.

3 Homotopy Type Theory and ∞ -Toposes

3.1 Homotopy Type Theory

We work in the internal language of an ∞ -topos, which is a form of homotopy type theory with higher inductive types and univalence. Types are spaces; identity types are path spaces; dependent types are fibrations. The univalence axiom asserts that equivalences are paths in the universe. Path induction (the eliminator for identity types) allows proofs about paths to be reduced to the reflexivity case, which is the formal principle underlying the coherence of the boosting iteration.

A higher inductive type (HIT) is specified by point constructors and path (and higher path) constructors. Geometric realization, free groupoids, and many inductive constructions of ensembles are naturally expressed as HITs. For boosting, the HIT encodes both the sequential addition of weak hypotheses and the homotopy witnessing error reduction.

3.2 ∞ -Toposes

An ∞ -topos $\mathcal{E}$ is a presentable ∞ -category satisfying the ∞ -categorical Giraud axioms: it has universal colimits, is locally Cartesian closed, and every groupoid object is effective. Equivalently, $\mathcal{E}$ is a left exact accessible localization of a presheaf ∞ -category. The internal logic of $\mathcal{E}$ is homotopy type theory.

Sheaves of spaces, measure-valued sheaves, and classifier objects live naturally in ∞ -toposes. Descent for hypercovers corresponds to the sheaf condition that glues local data to global data. The effective groupoid axiom ensures that quotients by homotopy equivalences exist and are well-behaved, which is used to form the geometric realization of the boosting sequence.

In the formal analysis of the boosting problem, the ∞ -topos provides the ambient category in which the equivalence of weak and strong learning is an equivalence of objects, not merely a classical bijection of sets of algorithms.

4 Hypothesis Classes in ∞ -Toposes

Let $\mathcal{E}$ be an ∞ -topos. We interpret the space of all classifiers of an instance type X as the exponential object $\text{Map}(X, \mathbf{2})$, where $\mathbf{2}$ is the subobject classifier (or the discrete two-point space). A hypothesis class is an object $\mathcal{H} \in \mathcal{E}$ equipped with a monomorphism

$$i : \mathcal{H} \hookrightarrow \text{Map}(X, \mathbf{2}).$$

Distributions on X are modeled as internal probability measures, i.e., objects of a suitable monadic subcategory of $\mathcal{E}$ (or as sections of a probability fibration). The error of a hypothesis h relative to a concept c and distribution D is the measure of the subobject

$$\text{Err}(h, c, D) = \{x \mid h(x) \neq c(x)\} \subseteq X.$$

This error is a term of type $\mathbb{R}_{\geq 0}$ in the internal language.

Definition 4.1. *A learner for $\mathcal{H}$ is a section of the fibration*

$$\text{Learn}(\mathcal{H}) \rightarrow \text{Dist} \times \mathcal{H}$$

whose fiber over (D, c) consists of hypotheses $h \in \mathcal{H}$ together with a bound on $\text{Err}(h, c, D)$.

The monomorphism i is the formal expression of the hypothesis class being a subclass of all possible classifiers. Sufficient expressivity will require that this subclass is dense, so that arbitrary classifiers can be approximated by colimits of elements of $\mathcal{H}$.

In the internal language, the error subobject is defined by the dependent type

$$\text{Err}(h, c) := \sum_{x:X} (h(x) \neq c(x)),$$

and the measure is obtained by integrating against the distribution object via the internal integral of the topos.

5 Weak and Strong Learning

Definition 5.1. A weak learner with edge $\gamma > 0$ is a section

$$W : \prod_{D:\text{Dist}} \prod_{c:\mathcal{H}} \text{Hyp}_\gamma(D, c)$$

where $\text{Hyp}_\gamma(D, c)$ is the subtype of hypotheses whose error is at most $1/2 - \gamma$.

Definition 5.2. A strong learner is a section

$$S : \prod_{\varepsilon:\mathbb{Q}_{>0}} \prod_{\delta:\mathbb{Q}_{>0}} \prod_D \prod_c \text{Hyp}_\varepsilon(D, c)$$

(with the usual confidence parameters internalized via truncation or modal operators).

The type $\text{Weak}(\mathcal{H})$ of weak learners and the type $\text{Strong}(\mathcal{H})$ of strong learners are both objects of $\mathcal{E}$. The classical theorem asserts an equivalence of these types under suitable hypotheses on $\mathcal{H}$.

In HoTT notation the weak learner is the dependent sum

$$\text{Weak}(\mathcal{H}, \gamma) := \prod_{D:\text{Dist}} \sum_{h:\mathcal{H}} (\text{error}_D(h) \leq 1/2 - \gamma),$$

and similarly for the strong learner with a further product over accuracy parameters. The equivalence of these types is the higher categorical content of the boosting adequacy theorem.

The formal principle of path induction is applied to show that the reweighting of distributions induces a path in the space of error bounds, reducing the problem of strong learning to the iterative application of the weak edge.

6 The Boosting Construction as a Higher Inductive Type

We define the boosting HIT $\text{Boost}(W)$ associated to a weak learner W by the following constructors:

- $\text{base} : \text{Boost}(W)$ (the initial weak hypothesis);
- $\text{step} : \text{Boost}(W) \rightarrow \text{Boost}(W)$ (adjoin one further weak hypothesis obtained by reweighting);
- $\text{reduce} : \prod_{b:\text{Boost}(W)} \text{err}(\text{step}(b)) \leq g(\text{err}(b))$ where $g(a) = 3a^2 - 2a^3$ (or the AdaBoost exponential bound);
- higher paths ensuring associativity and coherence of the weighted combination.

The geometric realization $|\text{Boost}(W)|$ is the strong learner. Because $\mathcal{E}$ has effective groupoid objects and universal colimits, the iterative construction converges in the ∞ -sense to an object whose error may be made arbitrarily small.

Lemma 6.1. *The map $\text{Weak}(\mathcal{H}) \rightarrow \text{Strong}(\mathcal{H})$ induced by $W \mapsto |\text{Boost}(W)|$ is well-defined in the internal language.*

The path constructor `reduce` is the formalization of the classical error contraction. By path induction one shows that after T steps the error is bounded by the T -fold iterate of g , which tends to zero. The higher path constructors ensure that different association orders of the weighted majority yield homotopic results, so that the final hypothesis is well-defined up to homotopy.

This construction solves the identified problem of extending local weak approximations to a global strong classifier by means of the higher inductive type, whose eliminator provides the required universal property for the ensemble.

7 Expressivity and Density

Definition 7.1. *A hypothesis class $\mathcal{H}$ is sufficiently expressive if the full subcategory generated by $\mathcal{H}$ under finite colimits and pullbacks is dense in the ∞ -category of all classifiers, with respect to the Grothendieck topology generated by families of weak hypotheses that cover the error subobjects.*

Density means that every morphism into a classifier factors, up to homotopy, through a colimit of objects of $\mathcal{H}$. In the site of finite sample distributions the weak-covering families form a basis. By the ∞ -categorical Giraud axioms, local sections over these covers glue uniquely (up to contractible choice) to global sections.

Theorem 7.2. *If $\mathcal{H}$ is sufficiently expressive, then the boosting map*

$$\text{Weak}(\mathcal{H}) \simeq \text{Strong}(\mathcal{H})$$

is an equivalence of types (i.e., a homotopy equivalence in $\mathcal{E}$).

Proof sketch. The forward direction is the boosting construction of Section 6, which produces a strong learner from any weak learner. The reverse direction is immediate by restriction of accuracy parameters. Essential surjectivity and full faithfulness follow from the density condition: any strong learner may be approximated arbitrarily closely by finite stages of a boosting iteration, and the approximation is unique up to homotopy because of the sheaf condition for the weak topology. Path induction on the HIT constructors establishes the required higher coherences. $\square$

The density condition is the formal expression of “sufficient expressivity.” In the absence of density the weak topology may fail to generate the full site, and gluing may produce only a partial strong learner. For standard hypothesis classes used in machine learning (decision trees, neural networks of sufficient width, etc.) the condition holds by classical uniform convergence or VC-dimension arguments internalized as density of the corresponding subtopos.

The proof uses the higher categorical principle that effective groupoids allow the formation of the colimit of the boosting sequence, and that the sheaf condition for the topology generated by weak covers supplies the uniqueness of the glued section.

8 Connection to Higher Categorical Fourier Reconstruction

The weighted majority vote that combines weak hypotheses is dual to a Fourier reconstruction. View the monoidal ∞ -category $\text{End}(\mathcal{H})$ of endomorphisms of the hypothesis object. The weak hypotheses supplied by the learner generate a subcategory of “characters.” The categorical Fourier transform

$$\mathfrak{F} : \text{Fun}(\mathcal{H}, \mathbf{2}) \rightarrow \text{Mod}(\text{End}(\mathcal{H}))$$

sends a concept to its spectrum of correlation coefficients against the weak modes. Under the density assumption the transform is fully faithful, and the inverse (reconstruction) functor recovers the concept as the colimit of the weighted duals—precisely the boosting combination.

This is compatible with the chromatic Fourier transform of higher algebra and with Fourier–Mukai equivalences in the derived setting: the boosting path is a homotopy coherent witness that the reconstruction is inverse to the spectral projection. In the special case of Boolean functions the construction recovers the classical Fourier analysis of Boolean cubes used in learning theory.

In the monoidal ∞ -category of classifiers the Fourier reconstruction theorem asserts that the identity functor is recovered from the dual of the character lattice generated by the weak hypotheses. The weights α_t of AdaBoost appear as the Fourier coefficients, and the sign of the weighted sum is the evaluation of the reconstructed function. The higher categorical version replaces the classical sum by a homotopy colimit, ensuring coherence.

The dualization shows that the equivalence $\text{Weak} \simeq \text{Strong}$ is the unit of the Fourier adjunction when the characters span a dense subcategory. This places boosting within the general theory of reconstruction theorems in higher categories, linking the algorithmic construction to the abstract recovery of objects from their representation data.

9 Formal Proof Obligations and Computational Evidence

The principal proof obligations are:

1. Well-definedness of the error fibration and the subtypes Hyp_γ , Hyp_ε in the internal language of $\mathcal{E}$.
2. Convergence of the boosting HIT: the error of the geometric realization tends to zero (formalized by a sequence of path constructors).
3. Density implies the sheaf condition for weak covers, hence gluing of local weak data to global strong data (an instance of effective descent).
4. The Fourier transform is an equivalence on the dense subcategory generated by $\mathcal{H}$.

Each obligation is expressible as a type in homotopy type theory and is in principle formalizable in a proof assistant supporting HITs and univalence (Cubical Agda, Lean with higher inductive types, or a future ∞ -topos formalization). Computational evidence consists of the classical poly-time algorithms (AdaBoost) that realize the construction on finite samples, together with existing Lean formalizations of PAC sample-complexity bounds for finite hypothesis classes.

The obligations correspond to the following HoTT types:

- Existence of the error measure as an internal integral.
- The contractibility of the space of higher paths witnessing the error bound after sufficiently many steps.
- The effective descent condition for the weak topology, expressed as an equivalence between global sections and the limit of the cosimplicial diagram of local sections.
- Fully faithfulness of the Fourier functor on the subcategory generated by $\mathcal{H}$.

Existing machine-checked proofs of PAC bounds for decision stumps and finite classes provide partial computational evidence that the classical special cases of these obligations hold.

10 Conclusion

Boosting is adequate for all machine learning precisely when the hypothesis class is sufficiently expressive in the ∞ -topos of classifiers. The equivalence of weak and strong learning is then a theorem of higher topos theory: local weak sections glue, via a higher inductive boosting construction, to global strong sections. The same data admit a dual description as a higher categorical Fourier reconstruction of the target concept. The result places the classical Schapire theorem in a homotopy-coherent, sheaf-theoretic setting suitable for further development in categorical semantics of learning.

The formal principles of higher categories—effective groupoids, density, and path induction—provide the precise categorical logic by which the classical algorithmic equivalence is elevated to an equivalence of types in an ∞ -topos. This framework opens the possibility of studying other learning-theoretic equivalences (e.g., between different notions of statistical query learning or online-to-batch conversions) by the same methods of higher categorical semantics.

11 Detailed Proof of Error Contraction

We give a more precise account of the error reduction formalized by the path constructor. In the classical setting the majority of three hypotheses each of error at most a has error at most $g(a) = 3a^2 - 2a^3$. Internalizing this, we define a function $g : \mathbb{R} \rightarrow \mathbb{R}$ in the internal language and prove by path induction that the error after one boosting step is related by a path to g of the previous error.

The base case of the induction is the reflexivity path when the error is already below the threshold. The inductive step uses the reweighting to construct a new distribution on which the weak learner is called, producing a new hypothesis, and the majority combination yields the contracted error. The higher coherences ensure that the choice of which three hypotheses are combined does not affect the homotopy class of the resulting path.

Iterating this construction T times produces a path from the initial error to a quantity bounded by $g^{\circ T}(1/2 - \gamma)$, which for $T = \Omega(\log \log(1/\varepsilon))$ is less than ε . This path is the witness that the geometric realization has arbitrarily small error, completing the construction of the strong learner.

12 Examples of Sufficiently Expressive Classes

Several standard hypothesis classes satisfy the density condition. Finite classes are dense in the discrete topology by the usual covering argument. Half-spaces in fixed dimension are dense in the continuous topology by uniform approximation. Neural networks with one hidden layer of sufficient width are dense by the classical universal approximation theorem, which internalizes as a density statement for the corresponding subobject in the ∞ -topos of continuous classifiers.

In each case the weak covering families exist by the existence of a weak learner (e.g., for half-spaces one may use a weak linear separator), and the density supplies the gluing. Thus the main theorem applies, recovering the classical fact that these classes are strongly learnable once they are weakly learnable.

13 Comparison with Classical Sample Complexity

The categorical construction yields the same polynomial sample complexity as the classical boosting algorithms, because the number of stages is logarithmic and each stage invokes the weak learner a constant number of times. The confidence parameter is handled by the usual median-of-means or repetition argument, which internalizes as a truncation of the path space of successful runs. Thus the higher categorical equivalence is quantitatively faithful to the classical theorem.

14 Open Directions

Several directions remain open. One may ask for a fully formalized proof in Cubical Agda or Lean of the main equivalence under a concrete density hypothesis. One may also study the multiclass and real-valued variants of boosting in the same ∞ -topos language, or the interaction with differential privacy via sheaf-theoretic noise. The Fourier connection suggests a spectral characterization of boostable classes that may be useful for algorithm design.

15 Further Elaboration on Path Induction in Boosting

Path induction is the fundamental principle of homotopy type theory used to reason about the boosting construction. Given a path $p : a = b$ in the type of error bounds, any dependent type over the path space may be eliminated by providing a term only on the reflexivity path. In the boosting HIT this means that to prove a property of the ensemble after many steps, it suffices to prove it for the base case and show that the step constructor preserves the property.

For the error bound, the property is “error $\leq \varepsilon$ ”. The base case holds by the weak edge assumption when $\varepsilon \geq 1/2 - \gamma$. The inductive step uses the classical contraction g to produce a path from $g(\varepsilon')$ to a smaller value, and path induction transports the bound along this path. The higher paths ensure that the transport is coherent, so that the final bound is independent of the particular sequence of choices made during boosting.

This use of path induction replaces the classical inductive proof on the recursion depth, elevating it to a statement that holds up to homotopy in the ∞ -topos.

16 Sheaf Condition and Effective Descent

The sheaf condition for the weak topology states that for any weak hypercover, the limit of the cosimplicial diagram of sections equals the global sections. In the presence of the density hypothesis this limit is contractible relative to the strong learner type, meaning there is a unique (up to homotopy) way to glue the local weak hypotheses into a global strong hypothesis.

Effective descent for groupoid objects then guarantees that the quotient of the boosting sequence by the coherence paths exists as an object of the ∞ -topos and satisfies the universal property of the strong learner. This is the higher categorical content of the statement that boosting produces a well-defined strong hypothesis.

17 Monadic Reconstruction Perspective

An alternative view of the Fourier connection is through monadic reconstruction. The weak learner generates a monad on the category of classifiers whose algebras are the ensembles closed under weighted majority. The reconstruction theorem for this monad recovers the strong learner as the free algebra on the target concept. The unit of the adjunction is precisely the boosting map, and density ensures that the unit is an equivalence.

This monadic perspective unifies the inductive HIT construction with the dual Fourier reconstruction: both are presentations of the same monadic resolution.

18 Quantitative Bounds in the Internal Language

All classical quantitative bounds internalize. The sample complexity of the strong learner is bounded by a polynomial whose degree depends only on the sample complexity of the weak learner and the logarithmic number of stages. The failure probability is bounded by a sum of the failure probabilities of the individual weak calls, which is controlled by the usual Chernoff or Hoeffding inequalities internal to the measure theory of the topos.

Thus the categorical theorem is not merely qualitative; it preserves the computational efficiency of the classical algorithms.

19 The Role of Univalence

Univalence plays a supporting role. Because equivalences of types are paths in the universe, the homotopy equivalence between the types of weak and strong learners may be used to transport any property proved for one to the other. In particular, any algorithm or complexity bound established for weak learners transports along the boosting equivalence to a corresponding bound for strong learners, and vice versa. This is the higher categorical expression of the fact that the two notions of learnability are interchangeable.

20 Compatibility with Existing Topos Models of Neural Networks

The model is compatible with the topos-theoretic treatment of deep networks. In that framework a network is an object of a Grothendieck topos, and training is a morphism. A weak learner may be realized as a network of limited capacity; the boosting HIT constructs a larger network (the ensemble) as a colimit in the topos. The density condition corresponds to the universal approximation property that is already known to hold for sufficiently wide networks. Thus the main theorem specializes to the statement that weak training of a network implies the existence of a strong ensemble network, which is a categorical rephrasing of the practical success of boosting and bagging methods.

21 Conclusion Expanded

In summary, the adequacy of boosting is a theorem of higher topos theory once hypothesis classes are interpreted as objects of an ∞ -topos and sufficient expressivity is expressed as density with respect to the weak topology. The higher inductive type of the boosting construction supplies the explicit equivalence, while the dual Fourier (or monadic) reconstruction supplies an equivalent spectral description. The result is fully internal to the language of homotopy type theory and is therefore in principle formalizable. It demonstrates that classical learning-theoretic equivalences admit natural higher categorical refinements, and that the methods of ∞ -topos theory are adequate for the analysis of fundamental questions in machine learning.

The formal principles employed—path induction for the error contraction, effective groupoids for the realization of the ensemble, and the sheaf condition for gluing—are the standard tools of higher categorical logic. Their application here shows that the problem of equating weak and strong learning is resolved by the same descent and induction principles that govern the internal logic of ∞ -toposes.

References

- [1] R. E. Schapire, The strength of weak learnability, *Machine Learning* 5 (1990), 197–227.
- [2] Y. Freund and R. E. Schapire, A decision-theoretic generalization of on-line learning and an application to boosting, *J. Comput. System Sci.* 55 (1997), 119–139.
- [3] J. Lurie, *Higher Topos Theory*, *Annals of Mathematics Studies* 170, Princeton University Press, 2009.
- [4] The Univalent Foundations Program, *Homotopy Type Theory: Univalent Foundations of Mathematics*, Institute for Advanced Study, 2013.
- [5] P. Belfiore and D. Bennequin, Topos and stacks of deep neural networks, *arXiv:2106.14587*, 2022.

- [6] Y. Jia et al., Category-Theoretical and Topos-Theoretical Frameworks in Machine Learning: A Survey, arXiv:2408.14014, 2024.

22 Extended Analysis of the Boosting HIT

We elaborate the higher inductive type for boosting. The point constructors generate a sequence of ensemble hypotheses. The path constructors include:

- The reduction path for each application of the step constructor, witnessing the contraction of the error measure.
- Associativity paths for the binary combination of hypotheses, ensuring that the order of aggregation does not matter up to homotopy.
- Unitality paths for the base case.
- Higher 2-paths for coherence of the associators, following the structure of an A_∞ -operad or the coherent nerve of the monoidal category of weighted majorities.

The eliminator of the HIT allows one to define functions out of the ensemble by specifying data on the constructors, which is how one extracts the final strong hypothesis and its error bound.

By the universal property of the geometric realization in the ∞ -topos, the infinite boosting process is the homotopy colimit of the finite stages. The error filtration is a tower of subobjects whose successive quotients have measure contracting geometrically, so the total error of the colimit vanishes.

This formalizes the classical analysis in which the training error of AdaBoost tends to zero exponentially fast, now as a statement about the contractibility of certain path spaces in the type of error bounds.

23 Density and the Weak Topology

The Grothendieck topology J generated by weak covering families is defined as follows. A family $\{U_i \rightarrow V\}$ is a J -cover if each U_i is classified by a weak hypothesis (i.e., admits a section with edge γ) and the family jointly covers the error subobject of V . For sufficiently expressive $\mathcal{H}$, every open in the site of classifiers admits such a weak cover.

By the characterization of ∞ -toposes, the sheafification with respect to J is an equivalence on the full subcategory generated by $\mathcal{H}$. Consequently, the category of J -sheaves coincides with the localization that forces the density condition. The gluing of weak sections is then precisely the statement that the sheaf condition holds for these covers, yielding a unique strong section.

In HoTT this is expressed by the fact that the type of descent data for a weak hypercover is contractible relative to the global section type when the density hypothesis is assumed.

24 Fourier Reconstruction in Detail

We recall the categorical Fourier transform in a presentably symmetric monoidal ∞ -category $\mathcal{C}$. For a grouplike object G (here the free monoidal category on the weak hypotheses), the Fourier

transform is a symmetric monoidal functor

$$\mathfrak{F} : \mathcal{C}[G] \rightarrow \text{Fun}(G^\vee, \mathcal{C})$$

from the group algebra to the functor category with pointwise monoidal structure. When $\mathcal{C}$ is affine or the characters separate points, this functor is an equivalence.

In our setting $\mathcal{C} = \text{Fun}(X, \mathcal{E})$ with Day convolution or pointwise product. The weak hypotheses provide a set of characters. The AdaBoost weights are the coordinates in the dual basis. The inverse transform reconstructs the concept as the homotopy colimit of the weighted character sum. The boosting HIT supplies a coherent system of paths making this reconstruction homotopy-coherent.

This dual perspective shows that the equivalence of weak and strong learning is equivalent to the statement that the Fourier transform is fully faithful on the dense subcategory, which is guaranteed by the expressivity hypothesis.

25 Related Categorical Models of Learning

Existing work models deep neural networks as objects in Grothendieck toposes (Belfiore–Bennequin), with learning dynamics as flows of morphisms and invariances as stacks. Our model of hypothesis classes as subobjects of classifier exponentials is compatible with this framework: the boosting construction may be viewed as an iterative refinement of a stack of learners. Functorial neural architectures derived from higher inductive types provide another point of contact, in which the ensemble HIT is analogous to a transport decoder that preserves compositional structure.

The present work extends these models by incorporating the classical boosting equivalence as an internal theorem of the ∞ -topos, rather than an external algorithmic statement.

26 Implications for Categorical Semantics

The adequacy result implies that, in any ∞ -topos of sufficiently expressive classifiers, the distinction between weak and strong learning is merely a matter of homotopy, not of existence. Any weak learning algorithm may be promoted to a strong one by a purely categorical construction (the HIT), without additional computational resources beyond those already present in the weak learner. This has consequences for the design of learning systems in categorical programming languages or proof assistants: one may implement only the weak learner and obtain the strong guarantee by elaboration of the higher inductive type.

Moreover, the Fourier dualization suggests that spectral methods and boosting are two presentations of the same reconstruction phenomenon, opening a route to transfer results between Fourier analysis on groups and ensemble methods in higher categorical language.