Why

One way to mitigate overfitting for empirical error minimization is to constrain the set of predictors considered.

Definition

Let $((\Omega, \mathcal{A}, \mathbf{P}), \{x_i : \Omega \to \mathcal{X}\}_{i=1}^n, f : \mathcal{X} \to \mathcal{Y})$ be probabilistic data-generation model with training set $S : \Omega \to (\mathcal{X} \times \mathcal{Y})^n$.

A *hypothesis class* is a subset of predictors $\mathcal{H} \subset \mathcal{X} \to \mathcal{Y}$. For a dataset $D = ((\xi_1, \gamma_1), \ldots, (\xi_n, \gamma_n)) \in (\mathcal{X} \times \mathcal{Y})^n$, a *restricted empirical error minimizer* of $\mathcal{H}$ is a hypothesis $h \in \mathcal{H}$ with minimal (among elements of $\mathcal{H}$) empirical error on D.

Inductive Bias

Many authorities call the hypothesis class a *inductive bias* and speak of "biasing" the "learning algorithm". Since one specifies the hypothesis class prior to the data it often said to "encode prior knowledge about the problem to be learned."

Properties

Let μ be the underlying distribution. A hypothesis class $\mathcal{H}$ is *realizable* if there exists $h^\star \in \mathcal{H}$ with

$$\mu(\{x \in \mathcal{X} \mid h^\star(x) \neq f(x)\}) = 0.^1$$

[1] Future editions will comment on the measurability of this set. It is no object for finite $\mathcal{X}$.

This condition is natural because of its corollaries. First, there exists $h^\star$ whose probability of achieving zero empirical risk on S is 1. Second, all empirical risk minimizers will achieve zero empirical risk on the dataset, with probability one.

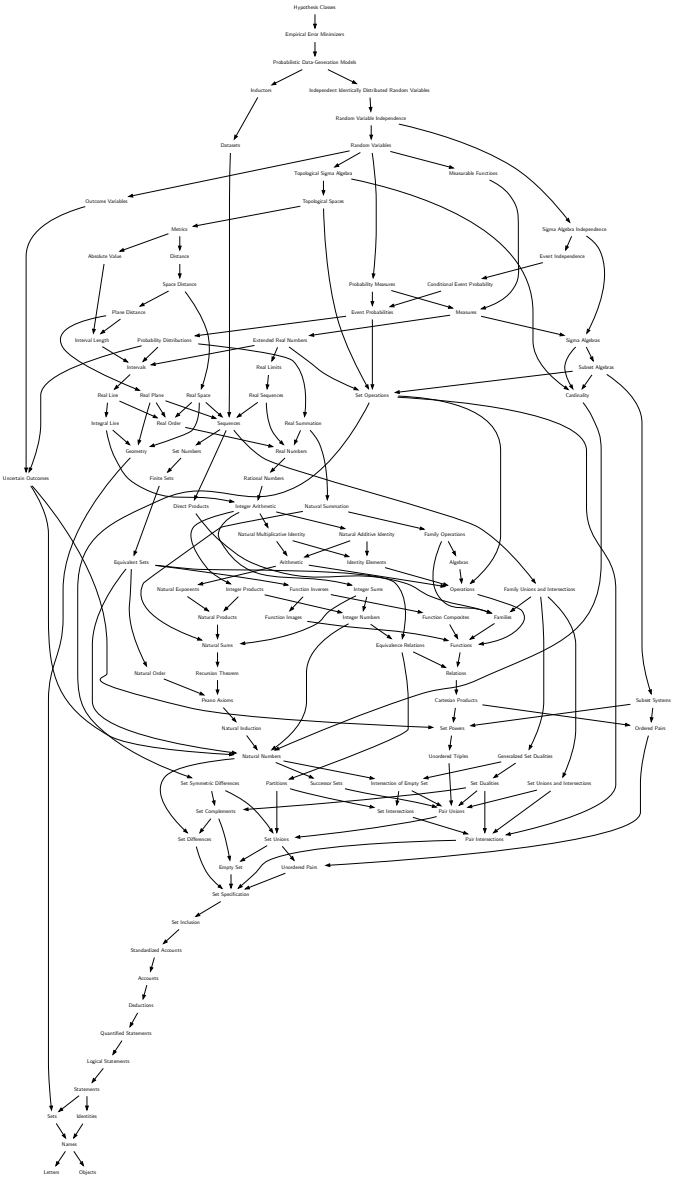
To see the first corollary, define $e_i : \Omega \to \{0, 1\}$ by $e_i(\omega_i) = 1$ if $h^\star(x_i(\omega)) \neq f(x_i(\omega))$ and $e_i(\omega) = 0$ otherwise. In other words, $(1/n) \sum_i e_i$ is the empirical risk. The empirical risk is zero if and only if the sum is zero, and the set

$$A = \left\{ (\omega_1, \ldots, \omega_n) \in \Omega^n \;\middle|\; \sum_{i=1}^{n} e_i(\omega) = 0 \right\}$$

is $\prod_i \{\omega_i \in \Omega \mid e_i(\omega_i) = 0\}$ and so $\mathbf{P}(A) = \prod_i \mathbf{Q}(\{\omega \in \Omega \mid e_i(\omega) = 0\})$ where for each i, $\mathbf{Q}[e_i = 0] = 1 - \mathbf{Q}[e_i = 1] = 1 - 0 = 0$. In other words, for a realizable hypothesis class, there exists a hypothesis obtaining zero empirical risk with probability one.

In still other words, there exists a hypothesis for which the event that it achieves zero empirical risk has probability one. In this event, every empirical risk minimizer achieves zero empirical risk. So, the set of empirical risk minimizers all achieve zero empirical risk with probability one, if the hypothesis class is realizable.[2]

[2]Future editions may clarify this further.


Hypothesis Classes
Empirical Error Minimizers
Probabilistic Data-Generation Models
Induction
Independent Identically Distributed Random Variables
Random Variable Independence
Datasets
Random Variables
Topological Sigma Algebra
Measurable Functions
Outcome Variables
Topological Spaces
Sigma Algebra Independence
Metrics
Absolute Value
Distance
Event Independence
Space Distance
Probability Measures
Conditional Event Probability
Plane Distance
Event Probabilities
Measures
Interval Length
Probability Distributions
Extended Real Numbers
Sigma Algebras
Intervals
Real Limits
Subset Algebras
Real Line
Real Plane
Real Space
Real Sequences
Set Operations
Cardinality
Integral Line
Real Order
Sequences
Real Summation
Geometry
Set Numbers
Real Numbers
Uncertain Outcomes
Finite Sets
Rational Numbers
Direct Products
Integer Arithmetic
Natural Summation
Natural Multiplicative Identity
Natural Additive Identity
Family Operations
Arithmetic
Identity Elements
Algebras
Equivalent Sets
Natural Exponents
Integer Products
Function Inverses
Integer Sums
Operations
Family Unions and Intersections
Natural Products
Function Images
Integer Numbers
Function Composites
Families
Natural Sums
Equivalence Relations
Functions
Natural Order
Recursion Theorem
Relations
Peano Axioms
Cartesian Products
Subset Systems
Natural Induction
Set Powers
Ordered Pairs
Natural Numbers
Unordered Triples
Generalized Set Dualities
Set Symmetric Differences
Partitions
Successor Sets
Intersection of Empty Set
Set Dualities
Set Unions and Intersections
Set Complements
Set Intersections
Pair Unions
Set Differences
Set Unions
Pair Intersections
Empty Set
Unordered Pairs
Set Specification
Set Inclusion
Standardized Accounts
Accounts
Deductions
Quantified Statements
Logical Statements
Statements
Sets
Identities
Names
Lemmas
Objects