Why

A natural and (conceptually) simple inductor is to select a predictor which achieves zero error on the training dataset.

Definition

Let $(X, \mathcal{X}, \mu)$ and $f : X \to Y$ be a probabilistic supervised data model. For a dataset $(x_1, y_1), \ldots, (x_n, y_n)$ in $X \times Y$, the *empirical error* of a predictor $h : X \to Y$ is

$$(1/n) \left|\{i \in \{1, 2, \ldots, n\} \mid h(x_i) \neq y_i\}\right|,$$

An *empirical error minimizer* is a predictor whose empirical error is minimal. Since we assume that the dataset is consistent, the correct labeling function is always an empirical error minimizer, and achieves zero. If the dataset is incomplete, however, there may be other empirical error minimizers.

A family of inductors $(A^n)_n$ is an *empirical error minimizer* if, for each $n \in N$, and each dataset $D = ((x_i, y_i))_{i=1}^n$, $A^n(D)$ is an empirical risk minimizer of D.

Overfitting and incutive bias

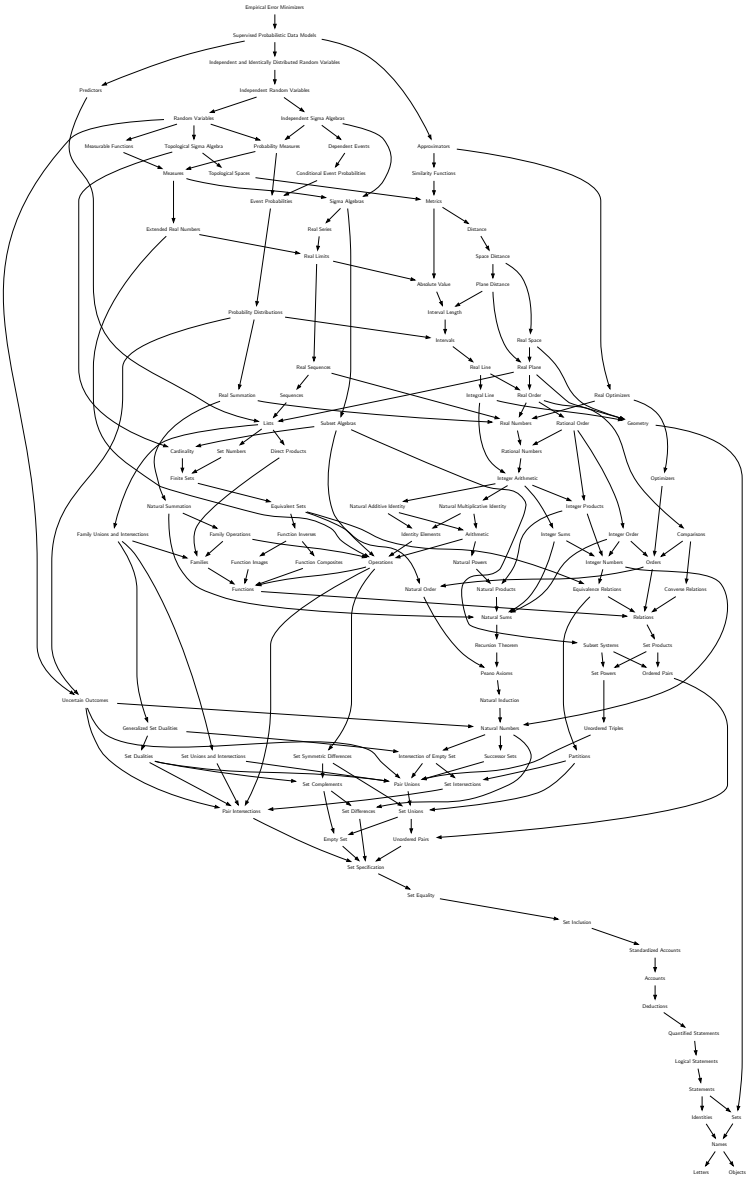
Since a dataset may be incomplete, an empirical risk minimizer may be a "poor" predictor. For example, let $A \subset X \subset \mathbf{R}^2$ and $Y = \{0, 1\}$. Define $f : X \to Y$ by $f(x) = 1$ if $x \in A$ and $f(x) = 0$ otherwise. Suppose $A \in \mathcal{X}$ and $\mu(A) = 1/2$.

For any training set $(x_1, y_1), \ldots, (x_n, y_n)$ in $X \times Y$, the hypothesis $h : X \to Y$ defined by $h(x) = y_i$ if $x = x_i$ and $h(x) = 0$ otherwise achieves zero empirical error but has "actual" error $1/2$. The predictor is said to be *overfit* or to exhibit *overfitting*. It is said to fit the training dataset "too well."

By adjusting the hypothesis class one may "guard against" overfitting. In this context, the constrained hypothesis class is called an *inductive bias*.

Other terminology

Other terminology for the empirical error includes *empirical risk*. For these reasons, the learning paradigm of selecting a predictor h to minimizer the empirical risk is called *empirical risk minimization* or *ERM*.


Empirical Error Minimizers
Supervised Probabilistic Data Models
Independent and Identically Distributed Random Variables
Predictors
Independent Random Variables
Random Variables
Independent Sigma Algebras
Approximators
Measurable Functions
Topological Sigma Algebra
Probability Measure
Dependent Events
Similarity Functions
Measures
Topological Spaces
Conditional Event Probabilities
Event Probabilities
Sigma Algebras
Metrics
Extended Real Numbers
Real Series
Distance
Real Limits
Space Distance
Absolute Value
Plane Distance
Interval Length
Probability Distributions
Intervals
Real Space
Real Sequence
Real Line
Real Plane
Real Summation
Sequences
Integral Line
Real Order
Real Optimizers
Real Numbers
Rational Order
Geometry
Lists
Subset Algebras
Cardinality
Set Numbers
Direct Products
Rational Numbers
Finite Sets
Integer Arithmetic
Optimizers
Natural Summation
Equivalent Sets
Natural Additive Identity
Natural Multiplicative Identity
Integer Products
Family Unions and Intersections
Family Operations
Function Inverses
Identity Elements
Arithmetic
Integer Sums
Integer Order
Comparisons
Families
Function Images
Function Composites
Operations
Natural Powers
Integer Numbers
Orders
Functions
Natural Order
Natural Products
Equivalence Relations
Converse Relations
Natural Sums
Relations
Recursion Theorem
Subset Systems
Set Products
Peano Axioms
Set Powers
Ordered Pairs
Natural Induction
Uncertain Outcomes
Natural Numbers
Unordered Triples
Generalized Set Dualities
Intersection of Empty Set
Successor Sets
Partitions
Set Dualities
Set Unions and Intersections
Set Symmetric Differences
Pair Unions
Set Intersections
Set Complements
Pair Intersections
Set Differences
Set Unions
Empty Set
Unordered Pairs
Set Specification
Set Equality
Set Inclusion
Standardized Accounts
Accounts
Deductions
Quantified Statements
Logical Statements
Statements
Identities
Sets
Names
Letters
Objects