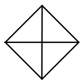


Why

What if we restrict our hypothesis class to have finite size?[1]

Definition

Let $\mathcal{D} = (\Omega, \mathcal{A}, \mathbf{Q})$ be a probability space. Let $x_i : \Omega \to \mathcal{X}$ be independent and identically distributed for $i = 1, \dots, n$. Let $f : \mathcal{X} \to \mathcal{Y}$ and define $y_i : \Omega \to \mathcal{Y}$ by $y_i = f(x_i)$. Define $S : \Omega \to (\mathcal{X} \times \mathcal{Y})^n$ by $S(\omega) = ((x_1(\omega), y_1(\omega)), \dots, (x_n(\omega), y_n(\omega)))$. Define the product space $(\Omega^n, \mathcal{A}^n, \mathbf{P})$.[2]

Let $\mathcal{H} \subset (\mathcal{X} \to \mathcal{Y})$ with $|\mathcal{H}| = m$. Then $\mathcal{H}$ is *realizable* with respect to $\mathcal{D}$ and f if there exists $h^\star \in H$ satisfying

$$\mathbf{Q}(\{\omega \in \Omega \mid h^\star(x(\omega)) \neq f(x(\omega))\}) = 0.$$[3]

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

[1] Future editions are likely to change the name of this sheet. Future editions will also discuss that this assumption is (in a sense) "mild," and may include reference to discretization or floating point arithmetic.

[2] Future editions will expand.

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

zero if and only if the sum is zero, and the set

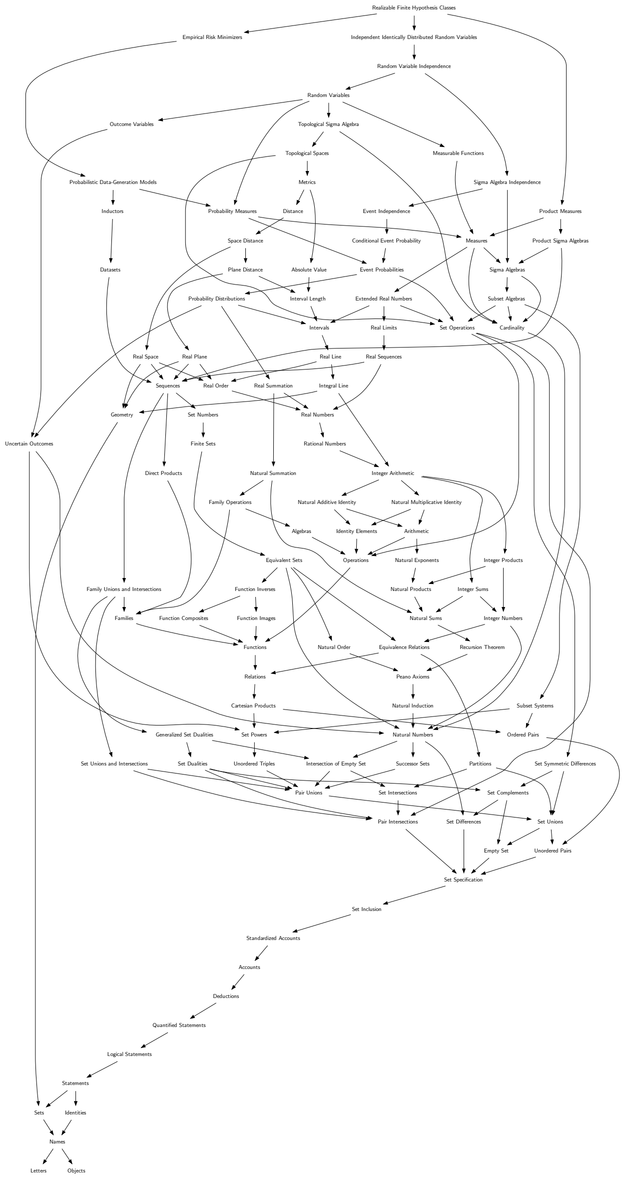
$$A = \left\{ (\omega_1, \dots, \omega_n) \in \Omega^n \;\middle|\; \sum_{i=1}^{n} e_i(\omega) = 0 \right\}$$

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 that achieves zero empirical risk with 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.[4]

There is a low probability to obtain a "nonrepresentative" dataset. This dataset may lead to selecting a hypothesis with low empirical

[4] Future editions may clarify this further.


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