Why

Let $X = \{a, b\}$ and $Y = \{0, 1\}$. Define $f : X \to Y$ by $f \equiv 0$.

The dataset $(a, 0)$ is consistent with f. So are the datasets $((a, 0), (a, 0))$ and $((a, 0), (a, 0), (a, 0))$. Unfortunately, these datasets are "bad" in the sense that we do not see the value associated with b. Each dataset is consistent with the (functional) relations $\{(a, 0), (b, 0)\}$ and $\{(a, 0), (b, 1)\}$.

In other words, a dataset can be incomplete. In spite of this limitation, we want to discuss an inductor's performance on consistent (but possibly incomplete) datasets.

We take two steps. First, we put a measure on the set of training sets and only consider high-measure subsets. Second, we consider inductors which perform well in some tolerance.

Definition

Let $(X, \mathcal{X})$ and $(Y, \mathcal{Y})$ be measurable spaces and R be a relation on $X \times Y$. Let $\mu : \mathcal{X} \times \mathcal{Y} \to [0, 1]$ be a probability measure so that $\mathcal{D} = (X \times Y, \mathcal{X} \times \mathcal{Y}, \mu)$ is a probability space.

If μ satisfies $\mu(B) = 0$ for all $B \subset C_{X \times Y}(R)$, then we call $\mathcal{D}$ a *probabilistic dataset model* for R. In other words, μ gives zero measure to any set of points not in the relation.

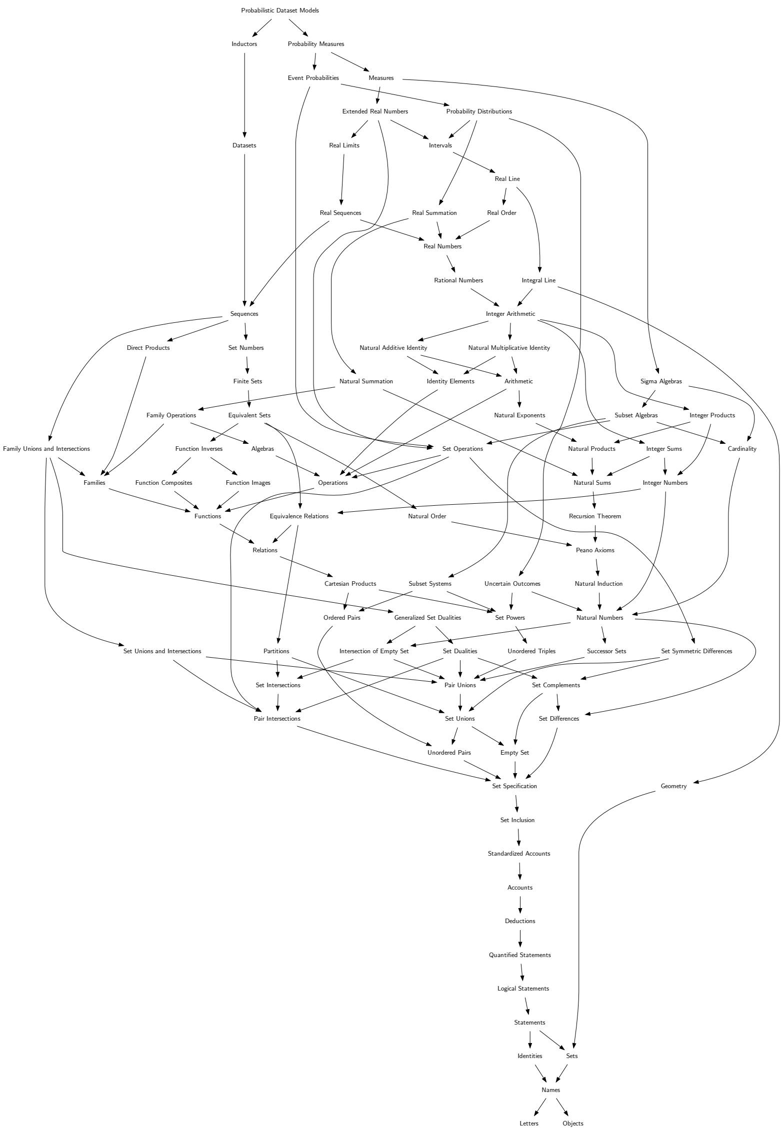
If R is functional, then we call $\mathcal{D}$ a *supervised probabilistic dataset model*. In this case, since there is a functional relation between X and Y, we call the marginal measure $\mu_X : \mathcal{X} \to$

$[0, 1]$ the *data-generating distribution* or *underlying distribution* since $\mu(A) = \mu_X(\{x \in X \mid (x, y) \in A)\})$. In this case we call the (functional) relation R the *correct labeling function.* Many authors refer to a supervised probabilistic data model as the *statistical learning (theory) framework.*

Probable datasets

For datasets of size n, we use the product measure $((X \times Y)^n, (\mathcal{X} \times \mathcal{Y})^n, \mu^n)$. We interpret this measure as modeling independent and identically distributed elements of R.

For $\delta \in (0, 1)$, $\mathcal{S} \subset (X \times Y)^n$ is $1 - \delta$*-probable* if $\mu^n(S) \geq 1 - \delta$. If δ is small, we interpret $\mathcal{S}$ as a set of "reasonable" datasets. We call δ the *confidence parameter.*


Probabilistic Dataset Models
Inductors
Probability Measures
Event Probabilities
Measures
Extended Real Numbers
Probability Distributions
Datasets
Real Limits
Intervals
Real Line
Real Sequences
Real Summation
Real Order
Real Numbers
Rational Numbers
Integral Line
Sequences
Integer Arithmetic
Direct Products
Set Numbers
Natural Additive Identity
Natural Multiplicative Identity
Finite Sets
Natural Summation
Identity Elements
Arithmetic
Sigma Algebras
Family Operations
Equivalent Sets
Natural Exponents
Subset Algebras
Integer Products
Family Unions and Intersections
Function Inverses
Algebras
Set Operations
Natural Products
Integer Sums
Cardinality
Families
Function Composites
Function Images
Operations
Natural Sums
Integer Numbers
Functions
Equivalence Relations
Natural Order
Recursion Theorem
Relations
Peano Axioms
Cartesian Products
Subset Systems
Uncertain Outcomes
Natural Induction
Ordered Pairs
Generalized Set Dualities
Set Powers
Natural Numbers
Set Unions and Intersections
Partitions
Intersection of Empty Set
Set Dualities
Unordered Triples
Successor Sets
Set Symmetric Differences
Set Intersections
Pair Unions
Set Complements
Pair Intersections
Set Unions
Set Differences
Unordered Pairs
Empty Set
Set Specification
Geometry
Set Inclusion
Standardized Accounts
Accounts
Deductions
Quantified Statements
Logical Statements
Statements
Identities
Sets
Names
Letters
Objects