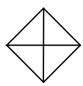


Why

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 may be incomplete. In spite of this limitation, we want to discuss an inductor's performance on consistent datasets. One route is to put a measure on the set of training sets and only consider high-measure subsets.

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.

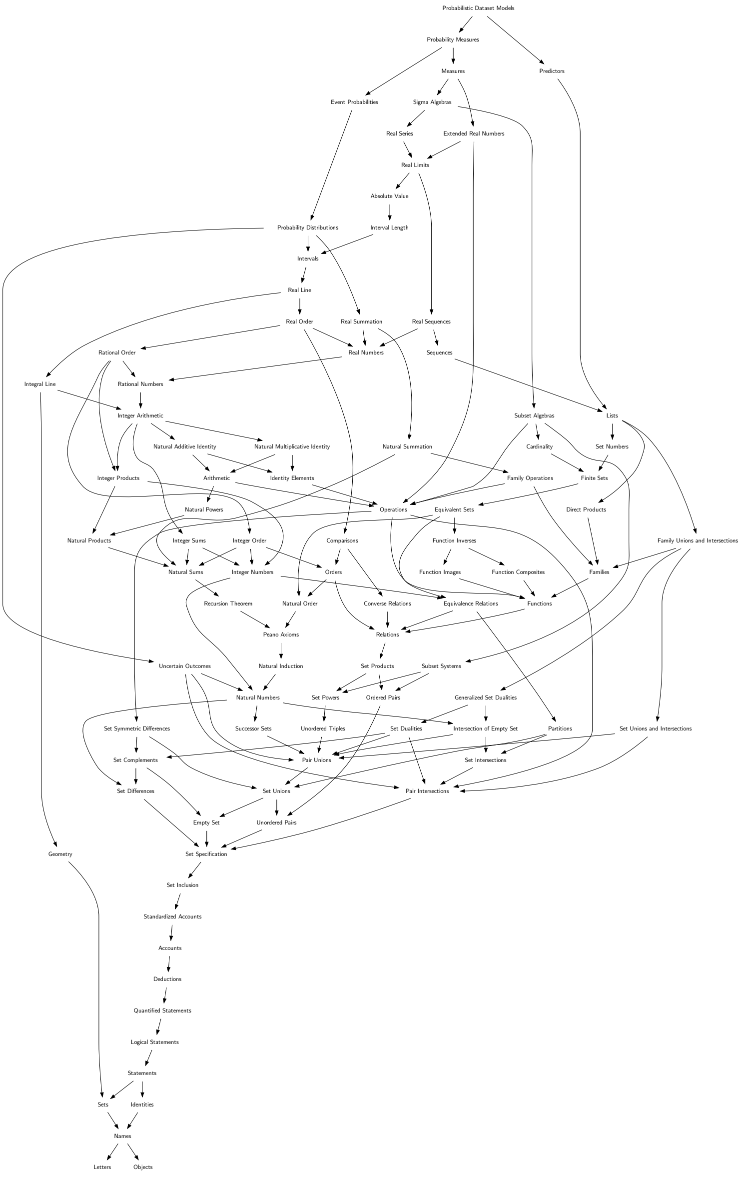
If $\mu(B) = 0$ for all $B \subset C_{X \times Y}(R)$, then we call $\mathcal{D}$ a *probabilistic dataset model* for R. In other words, μ gives zero measure to any set of points not in the relation. Equivalently, $\mu(R) = 1$; or we observe a pair in R almost surely.

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


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