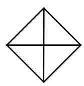


Why

We want to talk about learning associations between objects in time or space.

Definition

Let X and Y be sets. An *inductor* is a function mapping a dataset of paired records in $X \times Y$ to a function from X to Y. We commonly speak of a family of inductors indexed by $\mathbf{N}$, one for each natural number n which is the size of the dataset.

We call the elements of X the *inputs* and the elements of Y the *outputs*. A *predictor* is a function from the inputs to the outputs and the result of an input under a predictor is a *prediction*. Using this language, an inductor maps datasets to predictors. A predictor maps inputs to outputs.

Notation

Let D be a dataset of size n in $X \times Y$. Let $g : X \to Y$, a predictor, which makes prediction $g(x)$ on precept $x \in X$. Let $G : (X \times Y)^n \to (A \to B)$ be an inductor. Then $G(D)$ is the predictor which the inductor associates with dataset D.

Other terminology

Other terms for the inputs include *independent variables*, *explanatory variables*, *precepts*, *covariates*, *patterns*, *instances*, or *observations*. Other terms for the outputs include *dependent* *variables*, *explained variables*, *postcepts*, *targets*, *outcomes*, *labels* or *observational outcomes*.

Other terms for a predictor include *input-output* mapping, *prediction rule*, *hypothesis*, *concept*, or *classifier*. Some authors refer to a prediction as a *guess*.

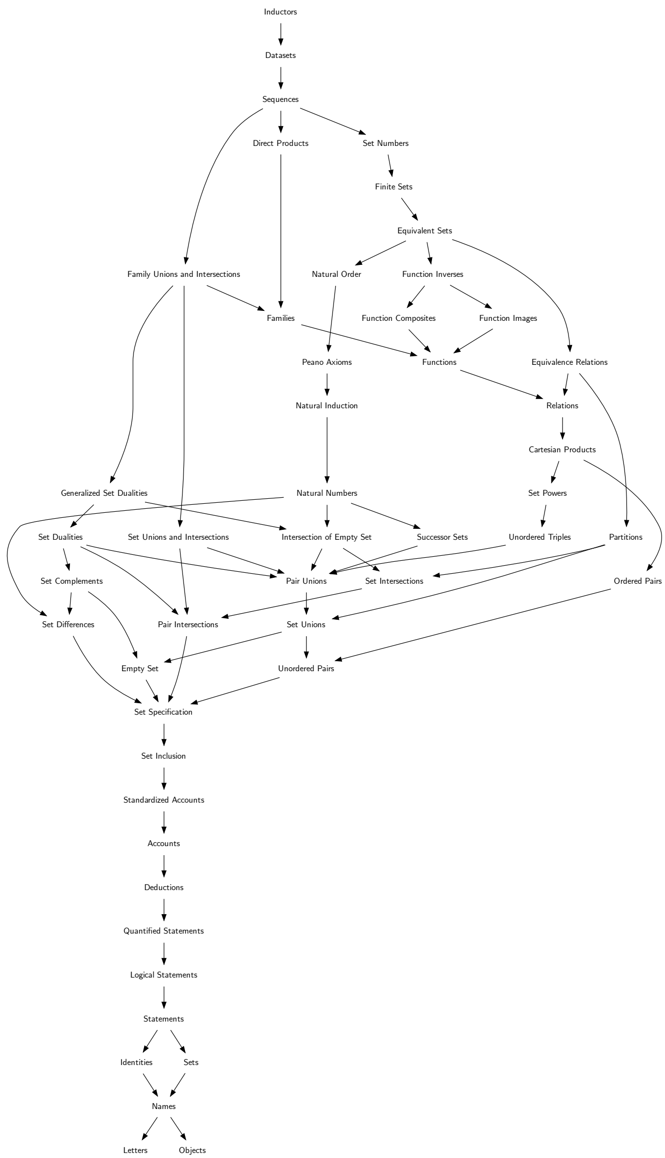
(Supervised) learning algorithms

Since we use a predictor to guess inputs which do not necessarily appear in the dataset, some authors call an inductor a *learner* or *learning algorithm*. In accordance with this usage, they refer to the argument of an inductor as the *training data* or *training dataset* or *training set*. The word "set", however, may mislead since since we are speaking of a sequence.

It is common to refer to the construction a predictor from a dataset a *learning problem*. In this case, the learning problem is said to be *supervised learning*. By supervision, we mean to indicate that we have the outputs corresponding to the inputs. In line with this usage, the outputs are often called *labels* and the labels are said "to provide supervision."

Consistent and complete datasets

Let $D = ((x_i, y_i))_{i=1}^n$ and $f : X \to Y$. D is *consistent with* f if $f(x_i) = y_i$ for all $i = 1, \dots, n$. D is *consistent* if there exists a predictor with which it is consistent. If D is consistent, then $x_i = x_j \longrightarrow y_i = y_j$. D is *complete* if $\cup_i \{x_i\} = X$. If a dataset is complete, then it includes every input.


Inductors
Datasets
Sequences
Direct Products
Set Numbers
Finite Sets
Equivalent Sets
Family Unions and Intersections
Natural Order
Function Inverses
Families
Function Composites
Function Images
Peano Axioms
Functions
Equivalence Relations
Natural Induction
Relations
Cartesian Products
Generalized Set Dualities
Natural Numbers
Set Powers
Set Dualities
Set Unions and Intersections
Intersection of Empty Set
Successor Sets
Unordered Triples
Partitions
Set Complements
Pair Unions
Set Intersections
Ordered Pairs
Set Differences
Pair Intersections
Set Unions
Empty Set
Unordered Pairs
Set Specification
Set Inclusion
Standardized Accounts
Accounts
Deductions
Quantified Statements
Logical Statements
Statements
Identities
Sets
Names
Letters
Objects