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

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

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

Independent Sigma Algebra

Random Variables

Dependent Events

Measurable Functions

Probability Measure

Topological Sigma Algebra

Approximators

Measure

Conditional Event Probabilities

Similarity Functions

Sigma Algebras

Event Probabilities

Topological Spaces

Metrics

Real Series

Extended Real Numbers

Distance

Real Limits

Space Distance

Absolute Value

Plane Distance

Probability Distributions

Interval Length

Intervals

Real Space

Real Sequences

Real Line

Real Plane

Sequences

Real Summation

Real Optimizers

Real Order

Integral Line

Lists

Subset Algebras

Real Numbers

Geometry

Direct Products

Set Numbers

Cardinality

Rational Order

Finite Sets

Rational Numbers

Equivalent Sets

Optimizers

Integer Arithmetic

Family Unions and Intersections

Function Inverses

Family Operations

Comparisons

Natural Multiplicative Identity

Natural Additive Identity

Integer Products

Families

Function Composites

Function Images

Operations

Orders

Identity Elements

Arithmetic

Integer Order

Integer Sums

Function

Converse Relations

Equivalence Relations

Natural Order

Natural Powers

Integer Numbers

Relations

Natural Sums

Natural Products

Subset Systems

Set Products

Recursion Theorem

Ordered Pairs

Set Powers

Peano Axioms

Uncertain Outcomes

Natural Induction

Generalized Set Dualities

Unordered Triples

Natural Numbers

Set Unions and Intersections

Intersection of Empty Set

Set Dualities

Partitions

Successor Sets

Set Symmetric Differences

Set Intersections

Pair Unions

Set Complements

Pair Intersections

Set Unions

Set Differences

Unordered Pairs

Empty Set

Set Specification

Set Inclusion

Standardized Accounts

Accounts

Deductions

Quantified Statements

Logical Statements

Statements

Identities

Sets

Names

Letters

Objects