Why

We want to use the notation and language of matrices to discuss datasets.

Definition

The *data matrix* corresponding to a data set of n records which are length d tuples of real numbers is the n by d matrix whose ith row is the i element of the data set. Some authors refer to the data matrix as the *design matrix*.

Notation

Let $(a^1, \ldots, a^n)$ where $a^i \in \mathbf{R}^n$ for $i = 1, \ldots, n$. Let A be the $n \times d$ matrix whose rows are the a^i. Then A is the data matrix of $(a^1, \ldots, a^n)$. In other words,

$$A = \begin{pmatrix} (a^1)^\top \\ \cdots \\ (a^n)^\top \end{pmatrix}.$$

Data Matrix
Matrices
Vectors
Datasets
N-Dimensional Space
Distance
Space Distance
Plane Distance
Interval Length
Intervals
Real Line
Real Plane
Real Space
Geometry
Real Order
Direct Products
Real Numbers
Fields
Rational Numbers
Groups
Families
Integer Arithmetic
Integer Additive Inverses
Integer Sums
Natural Additive Identity
Natural Multiplicative Identity
Inverse Elements
Integer Products
Arithmetic
Identity Elements
Element Functions
Natural Exponents
Algebras
Function Inverses
Natural Products
Operations
Function Images
Function Composites
Integer Numbers
Natural Sums
Functions
Equivalence Relations
Recursion Theorem
Relations
Peano Axioms
Natural Induction
Cartesian Products
Natural Numbers
Set Powers
Partitions
Intersection of Empty Set
Successor Sets
Unordered Triples
Set Intersections
Pair Unions
Ordered Pairs
Set Unions
Set Differences
Pair Intersections
Empty Set
Unordered Pairs
Set Specification
Set Inclusion
Standardized Accounts
Accounts
Deductions
Quantified Statements
Logical Statements
Statements
Sets
Identities
Names
Letters
Objects