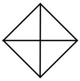


Why

Linear predictors are simple and we know how to select the parameters. The main downside is that there may not be a linear relationship between inputs and outputs.

Definition

A *feature map* (or *regression function*) for outputs A is a mapping $\phi : A \to \mathbf{R}^d$. In this setting, we call $a \in A$ the *raw input record* and we call $\phi(a)$ an *embedding*, *feature embedding* or *feature vector*. We call the components of a feature vector the *features*. We call $\phi(A)$ the *regression range*.

A feature map is *faithful* if, whenever records a_i and a_j are in some sense "similar" in the set A, the embeddings $\phi(a_i)$ and $\phi(a_j)$ are close in the vector space $\mathbf{R}^d$.

Since it is common for raw input records $a \in A$ to consist of many fields, it is regular to have several feature maps ϕ_i which operate componentwise on the fields of a. These are sometimes called *basis functions*, by analogy with real function approximators (see Real Function Apppoximators). We concatenate these field feature maps and commonly add a constant feature 1. Since $\mathbf{R}^d$ is a vector space, it is common to refer to it in this case as the *feature space*.

Given a dataset $a = (a^1, \dots, a^n)$ in A and a feature map $\phi : A \to \mathbf{R}^d$, the *embedded dataset* of a with respect to ϕ is the dataset $(\phi(a^1), \dots, \phi(a^n)$ in $\mathbf{R}^d$.

Featurized consistency: a route around $X \neq \mathbf{R}^d$

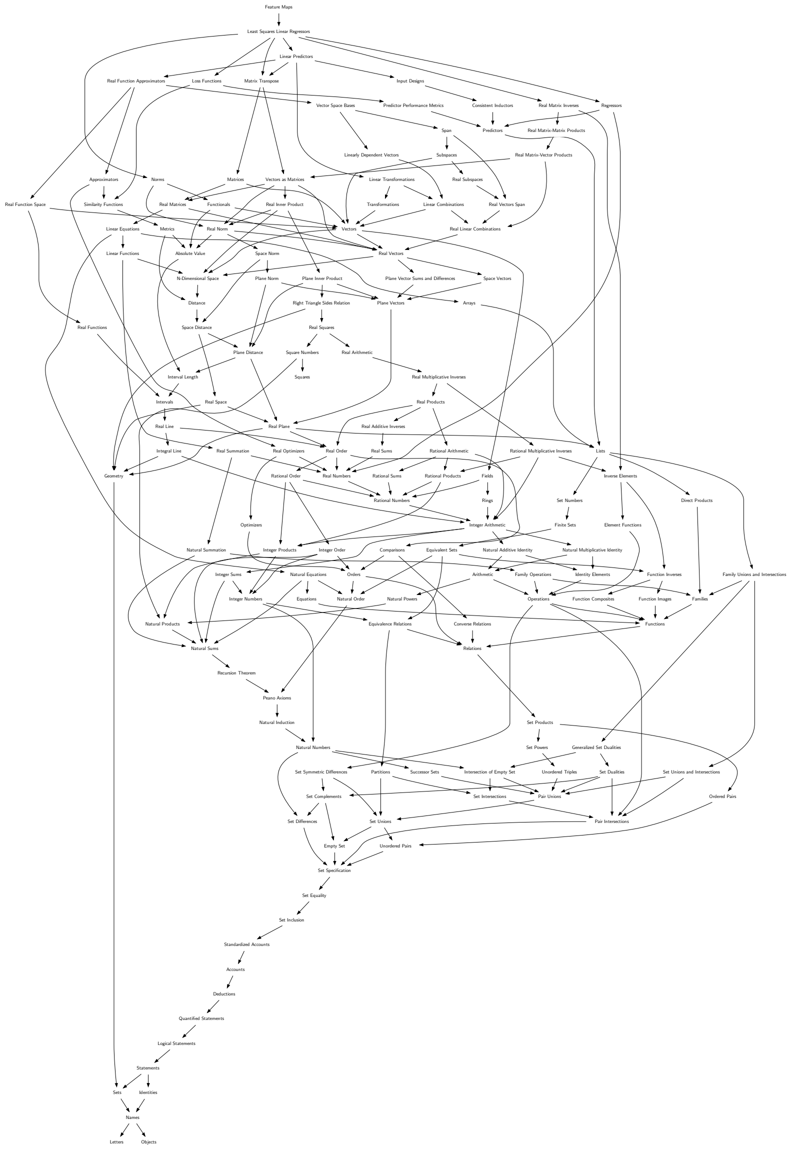
Recall that a dataset is parametrically consistent with the family $\{h_\theta : X \to Y\}_\theta$ if there exists $\theta^\star$ so that the dataset is consistent with $\theta^\star$. We saw how to pick θ if we use a linear model with a squared loss (see Least Squares Linear Regressors).

Let $\mathcal{G} = \{g_\theta : \mathsf{R}^d \to \mathsf{R}\}_\theta$. A dataset is *featurized parametrically consistent* with respect to the family $\mathcal{G}$ and the feature map $\phi : X \to \mathsf{R}^d$ if it is parametrically consistent with respect to $\mathcal{G} \circ \phi = \{g \circ \phi \mid g \in \mathcal{G}\}$.

The interpretation is that we have transformed the problem of selecting a predictor on an arbitrary space X to the problem of selecting a predictor on the space R^d. In so doing, we can continue to use simple predictors, such as those that are linear and minimize the squared error on the dataset.[1]

In other words, we have "shifted emphasis" from the model function $h : X \to \mathsf{R}$ to the *regression function* from $\mathsf{R}^d \to \mathsf{R}$. If we know the features and the input x, then we know the *regression vector* $\phi(x)$. The *regression range* is the set $\{\phi(x) \mid x \in X\}$. In this case linearity pertains to the parameters $\theta \in \mathsf{R}^d$ instead of the inputs (or experimental conditions) $x \in X$.

[1] Future editions are likely to modify this section.


Feature Maps
Least Squares Linear Regressors
Linear Predictors
Real Function Approximators
Loss Functions
Matrix Transpose
Input Designs
Vector Space Bases
Predictor Performance Metrics
Consistent Inductors
Real Matrix Inverses
Regressors
Span
Predictors
Real Matrix-Matrix Products
Linearly Dependent Vectors
Subspaces
Real Matrix-Vector Products
Approximators
Norms
Matrices
Vectors as Matrices
Linear Transformations
Real Subspaces
Real Function Space
Similarity Functions
Real Matrices
Functionals
Real Inner Product
Transformations
Linear Combinations
Real Vectors Span
Linear Equations
Metrics
Real Norm
Vectors
Real Linear Combinations
Linear Functions
Absolute Value
Space Norm
Real Vectors
N-Dimensional Space
Plane Norm
Plane Inner Product
Plane Vector Sums and Differences
Space Vectors
Distance
Right Triangle Sides Relation
Plane Vectors
Arrays
Real Functions
Space Distance
Real Squares
Plane Distance
Square Numbers
Real Arithmetic
Interval Length
Squares
Real Multiplicative Inverses
Intervals
Real Space
Real Products
Real Line
Real Plane
Real Additive Inverses
Integral Line
Real Summation
Real Optimizers
Real Order
Real Sums
Rational Arithmetic
Rational Multiplicative Inverses
Lists
Geometry
Rational Order
Real Numbers
Rational Sums
Rational Products
Fields
Inverse Elements
Rational Numbers
Rings
Set Numbers
Direct Products
Optimizers
Integer Arithmetic
Finite Sets
Element Functions
Natural Summation
Integer Products
Integer Order
Comparisons
Equivalent Sets
Natural Additive Identity
Natural Multiplicative Identity
Integer Sums
Natural Equations
Orders
Arithmetic
Family Operations
Identity Elements
Function Inverses
Family Unions and Intersections
Integer Numbers
Equations
Natural Order
Natural Powers
Operations
Function Composites
Function Images
Families
Natural Products
Equivalence Relations
Converse Relations
Functions
Natural Sums
Relations
Recursion Theorem
Peano Axioms
Natural Induction
Set Products
Natural Numbers
Set Powers
Generalized Set Dualities
Set Symmetric Differences
Partitions
Successor Sets
Intersection of Empty Set
Unordered Triples
Set Dualities
Set Unions and Intersections
Set Complements
Set Intersections
Pair Unions
Ordered Pairs
Set Differences
Set Unions
Pair Intersections
Empty Set
Unordered Pairs
Set Specification
Set Equality
Set Inclusion
Standardized Accounts
Accounts
Deductions
Quantified Statements
Logical Statements
Statements
Sets
Identities
Names
Letters
Objects