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 component-wise on the fields of a. These are sometimes called *basis functions*, by analogy with real function approximators (see Real Function Approximators). 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 $a = (a^1, \ldots, a^n)$ in A and a feature map $\phi : A \to \mathbf{R}^d$, the *embedded dataset* of a with respect to ϕ is the dataset $(\phi(a^1), \ldots, \phi(a^n)$ in $\mathbf{R}^d$.

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

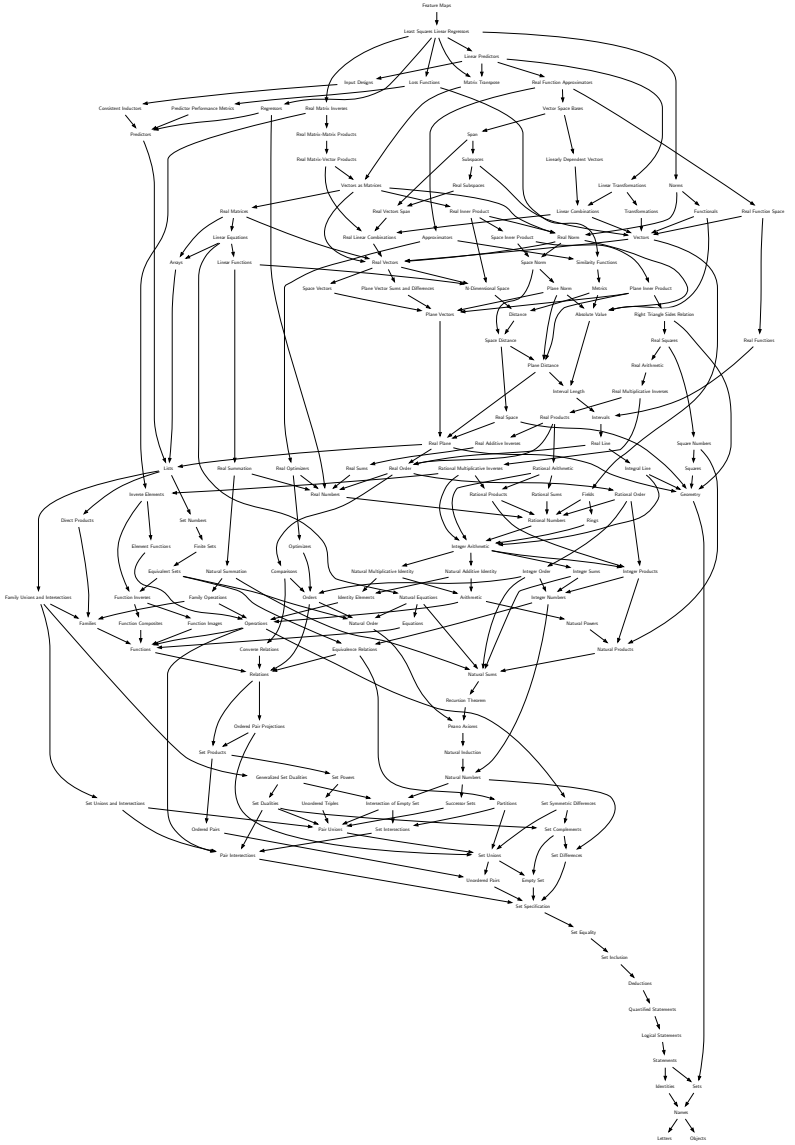
Recall that a dataset is parametrically consistent with the family $\{h_\theta : X \to Y\}_\theta$ if there exists $\theta^\star$ so that the dataset is consistent with $\theta^\star$. We saw how to pick θ if we use a linear model with a squared loss (see Least Squares Linear Regressors).

Let $\mathcal{G} = \{g_\theta : \mathbf{R}^d \to \mathbf{R}\}_\theta$. A dataset is *featurized parametrically consistent* with respect to the family $\mathcal{G}$ and the feature map $\phi : X \to \mathbf{R}^d$ 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 $\mathbf{R}^d$. 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 $h : X \to \mathbf{R}$ to the *regression function* from $\mathbf{R}^d \to \mathbf{R}$. 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 $\theta \in \mathbf{R}^d$ instead of the inputs (or experimental conditions) $x \in X$.

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


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