Why

We use a loss function and a predictive density to construct a regressor on the domain of the random process.

Setup

Let $\ell : \mathbf{R} \times \mathbf{R} \to \mathbf{R}$ be a loss function. We choose a predictor to minimize the expected loss under the predictive density.

Squared Error Case

Consider $\ell(\alpha, \beta) = (\alpha - \beta)^2$. The *minimum squared error normal random function predictor* or *minimum squared error gaussian process predictor* for dataset $(a^1, \gamma_1), \ldots, (a^n, \gamma_n)$ in $A \times \mathbf{R}$ is the predictor which minimizes the squared error loss. Since the predictive density is normal, the minimizer is the conditional mean.

Absolute Error Case

Consider $\ell(\alpha, \beta) = |\alpha - \beta|$. The *minimum absolute deviation normal random function predictor* or *minimum absolute deviation gaussian process predictor* for dataset $(a^1, \gamma_1), \ldots, (a^n, \gamma_n)$ in $A \times \mathbf{R}$ is the predictor which minimizes the absolute deviation loss. For any density, the solution is the median. Since the predictive density is normal, and so symmetric, the median is the conditional mean. In other words, the minimum absolute deviation normal random function predictor coinicides with the minimium squared error normal random function predic-

tor.

Definition

For this reason, the *normal random function predictor* or *gaussian process predictor* for dataset $(a^1, \gamma_1), \ldots, (a^n, \gamma_n)$ in $A \times \mathbf{R}$ is $h : A \to \mathbf{R}$ defined by

$$h(x) = m(x) + \Big(k(x, a^1) \cdots k(x, a^n) \Big) (\Sigma_a + \Sigma_e)^{-1} (\gamma - m_a).$$

In other words, the regressor which assigns to each point its conditional mean.

Notice that h is an affine function of γ. If the mean function $m \equiv 0$ then h is linear in γ. This is sometimes called a *linear estimator*.[1] Alternatively, notice (in the zero mean setting) that h is a linear combination of n kernel function $k(x, a^i)$ for $i = 1, \ldots, n$. Specifically, h is a linear combination of

The process of using a normal random function predictor is often called *Gaussian process regression* or (especially in spatial statistics) *kriging*. The upside is that a gaussian process predicor interpolates the data, is smooth, and the so-called variance increases with the distance from the data.[2]

---

[1] We avoid the other terminology we have seen used—linear predictor—because the predictor h is not linear in its input x.

[2] This last piece is true for certain covariance kernels and will be clarified in future editions.

Normal Random Function Regression

Normal Random Function Predictive Densities

Normal Random Functions — Affine Transformations

Multivariate Normals — Transformations

Positive Definite Matrices — Matrix Determinants — Random Functions

Quadratic Forms — Symmetric Matrices — Real Matrices — Matrix Inverse — Random Vectors

Identity Matrices — Matrix Transpose — Matrix-Matrix Products — Multivariate Real Densities — Normal Densities — Random Variables

Real Vector — Matrices — Vectors — Exponential Function — Probability Densities — Topological Sigma Algebras

N-Dimensional Space — Metrics — Real Integrals — Topological Spaces

Distance — Absolute Value — Real Functions — Outcome Variables — Nonnegative Integrals

Real Square Roots — Space Distance — Simple Integrals

Real Arithmetic — Place Distance — Probability Measures — Measurable Functions

Real Multiplicative Inverses — Loss Functions — Internal Length — Probability Distributions — Intervals — Extended Real Numbers — Event Probabilities — Measures

Real Products — Inductors — Subset Algebras

Real Addition Inverses — Datasets — Real Limits — Real Space — Real Plane — Real Line — Cardinality — Set Operations

Rational Arithmetic — Rational Multiplication Inverses — Real Sums — Fields — Real Summation — Real Sequences — Direct Products — Real Order — Geometry

Rational Sums — Rational Products — Real Numbers — Groups

Rational Numbers — Integer Additive Inverse — Natural Summation — Inverse Elements — Uncertain Outcomes

Integer Arithmetic — Element Functions — Family Operations — Sequences

Natural Addition Identity — Natural Multiplicative Identity — Identity Elements — Integer Products — Integer Sums — Algebras — Function Inverses — Families

Natural Square Roots — Arithmetic — Integer Numbers — Operation — Function Composition — Function Images — Functions

Square Number — Natural Exponents — Relations

Squares — Natural Products — Equivalence Relations

Natural Sums — Simple Functions

Recursion Theorem — Cartesian Products — Subset Systems — Characteristic Functions

Peano Axioms — Ordered Pairs — Set Powers — Partitions

Natural Induction — Natural Numbers — Unordered Triples

Equation Solutions — Successor Sets — Intersection of Empty Set

Pair Union — Set Symmetric Difference — Set Intersection

Set Unions — Set Complements — Pair Intersection

Unordered Pairs — Set Differences — Empty Set

Set Specification

Set Inclusion

Standardized Accounts

Accounts

Deductions

Quantified Statements

Logical Statements

Statements

Identifies — Sets

Names

Letters — Objects