Why

[1]

Definition

We want to estimate a random variable $x : \Omega \to \mathbf{R}^n$ from a random variable $y : \Omega \to \mathbf{R}^n$ using an estimator $\phi : \mathbf{R}^m \to \mathbf{R}^n$ which is affine.[2] In other words, $\phi(\xi) = A\xi + b$ for some $A \in \mathbf{R}^{n\times m}$ and $b \in \mathbf{R}^n$. We will use the mean squared error cost.

We want to find A and b to minimize

$$\mathbf{E}\|Ax + b - y\|^2.$$

Proof. Express $\mathbf{E}(\|Ax + b - y\|^2)$ as $\mathbf{E}((Ax + b - y)^\top(Ax + b - y))$

$$\begin{array}{ccccccc} + & \operatorname{tr}(A\mathbf{E}(xx^\top)A^\top) & + & \mathbf{E}(x)^\top A^\top b & - & \operatorname{tr}(A^\top\mathbf{E}(yx^\top)) \\ + & b^\top A\mathbf{E}(x) & + & b^\top b & - & b^\top\mathbf{E}(y) \\ - & \operatorname{tr}(A\mathbf{E}(xy^\top)) & - & \mathbf{E}(y)^\top b & + & \mathbf{E}(yy^\top) \end{array}$$

The gradients with respect to b are

$$\begin{array}{ccccccc} + & 0 & + & A\mathbf{E}(x) & - & 0 \\ + & A\mathbf{E}(x) & + & 2b & - & \mathbf{E}(y) \\ - & 0 & - & \mathbf{E}(y) & + & 0 \end{array}$$

so $2A\mathbf{E}(x) + 2b - 2\mathbf{E}(y)$. The gradients with respect to A are

$$\begin{array}{ccccccc} + & \mathbf{E}(xx^\top)A^\top + \mathbf{E}(xx^\top)^\top A^\top & + & \mathbf{E}(x)b^\top & - & \mathbf{E}(yx^\top)^\top \\ + & \mathbf{E}(x)b^\top & + & 0 & - & 0 \\ - & \mathbf{E}(xy^\top) & - & 0 & + & 0 \end{array}$$

so $2\mathbf{E}(xx^\top)A^\top + 2\mathbf{E}(x)b^\top - 2\mathbf{E}(xy^\top)$. We want A and b solutions to

$$\begin{aligned} A\mathbf{E}(x) + b - \mathbf{E}(y) &= 0 \\ \mathbf{E}(xx^\top)A^\top + \mathbf{E}(x)b^\top - \mathbf{E}(xy^\top) &= 0 \end{aligned}$$

[1] Future editions will include an account.

[2] Actually, the development flips this. Future editions will correct.

so first get $b = \mathbf{E}(y) - A\mathbf{E}(x)$. Then express

$$\begin{aligned}
\mathbf{E}(xx^\top)A^\top + \mathbf{E}(x)(\mathbf{E}(y) - A\mathbf{E}(x))^\top - \mathbf{E}(xy^\top) &= 0. \\
\mathbf{E}(xx^\top)A^\top + \mathbf{E}(x)\mathbf{E}(y)^\top - \mathbf{E}(x)\mathbf{E}(x)^\top A^\top - \mathbf{E}(xy^\top) &= 0. \\
(\mathbf{E}(xx^\top) - \mathbf{E}(x)\mathbf{E}(x)^\top)A^\top &= \mathbf{E}(xy^\top) - \mathbf{E}(x)\mathbf{E}(y)^\top. \\
\mathbf{cov}(x,x)A^\top &= \mathbf{cov}(x,y).
\end{aligned}$$

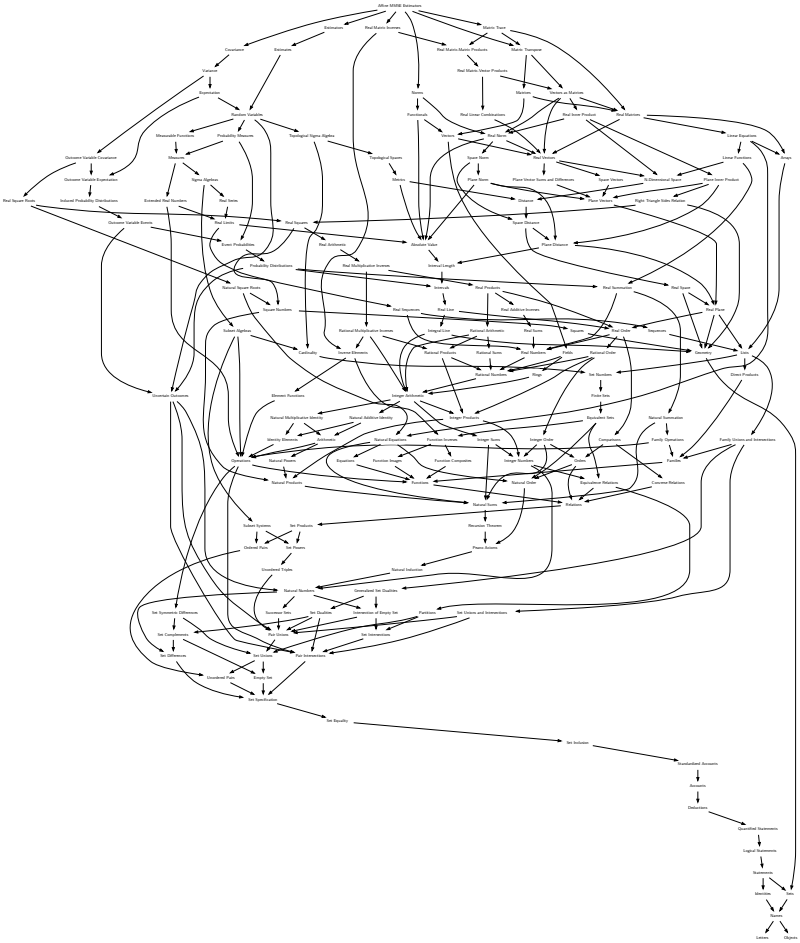
So $A^\top = \mathbf{cov}(x,x)^{-1}\mathbf{cov}(x,y)$ means $A = \mathbf{cov}(y,x)\mathbf{cov}(x,x)^{-1}$ is a solution. Then $b = \mathbf{E}(y) - \mathbf{cov}(y,x)\mathbf{cov}(x,x)^{-1}\mathbf{E}(x)$. So to summarize, the estimator $\phi(x) = Ax + b$ is

$$\mathbf{cov}(y,x)\left(\mathbf{cov}x,x\right)^{-1}x + \mathbf{E}(y) - \mathbf{cov}(y,x)\mathbf{cov}(x,x)^{-1}\mathbf{E}(x)$$

or

$$\mathbf{E}(y) + \mathbf{cov}(y,x)\left(\mathbf{cov}x,x\right)^{-1}(x - \mathbf{E}(x))$$

$\square$


Affine MMSE Estimators
Estimators
Real Matrix Inverses
Matrix Trace
Covariance
Estimands
Real Matrix Matrix Products
Matrix Transpose
Variance
Real Matrix Vector Products
Expectation
Matrix Vector
Matrix as Matrices
Random Variables
Sums
Functionals
Real Linear Combinations
Real Inner Products
Real Matrices
Measurable Functions
Probability Measures
Topological Spaces, Algebra
Norms
Real Vectors
Linear Equations
Outcome Variable Covariance
Measures
Topological Spaces
Norm Spaces
Real Norms
Linear Functions
Images
Outcome Variable Expectation
Signed Algebras
Metrics
Plane Norms
Plane Vectors Sums and Differences
Space Vectors
3-Dimensional Spaces
Plane Inner Product
Real Square Roots
Indicator Probability Distributions
Extended Real Numbers
Real Series
Distances
Norms
Right Triangle Sides Relations
Outcome Variable Events
Real Limits
Real Squares
Inner Distances
Event Probabilities
Real Subtractions
Absolute Value
Plane Distance
Probability Distributions
Real Multiplication Inverses
Interval Length
Real Powers
Real Summations
Real Spaces
Natural Square Roots
Intervals
Real Sums
Square Numbers
Real Integrations
Real Line
Real Addition Inverses
Real Plane
Indent Algebras
Rational Multiplication Inverses
Integral Line
Rational Additions
Real Series
Inequality
Real Order
Sequences
Cardinals
Integer Exponents
Rational Products
Rational Sums
Real Numbers
Fields
Rational Order
Integers
Rational Numbers
Rings
Set Operations
Direct Products
Quantifier Operators
Element Functions
Integer Multiples
Finite Sets
Natural Multiplication Identity
Natural Addition Identity
Integer Products
Subsidiary Sets
Natural Summation
Natural Multiplication Inverses
Natural Subtractions
Integer Sums
Integer Order
Compositions
Finite Summations
Finite Unions and Intersections
Operators
Natural Powers
Equations
Function Sums
Function Composites
Integer Numbers
Orders
Functions
Natural Products
Functions
Natural Order
Equivalence Relations
Concrete Relations
Natural Sums
Partitions
Subset Inclusion
Set Products
Recursion Theorems
Ordered Pairs
Set Pairs
Proof Assumes
Unordered Triples
Natural Induction
Natural Numbers
Generalized Set Qualifiers
Set Symmetric Differences
Successor Sets
Set Qualifiers
Intersection of Empty Set
Partitions
Set Unions and Intersections
Set Unions
Set Intersections
Set Complements
Set Unions
Set Intersections
Set Differences
Unordered Pairs
Empty Set
Set Specification
Set Equality
Set Inclusion
Translatable Accounts
Accounts
Deductions
Quantified Statements
Logical Statements
Statements
Identities
Rules
Theorems
Lemmas
Objects