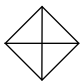


Why

We have precepts in R^d and want to predict postcepts in R. We put a probability measure on a set of linear statistical models.[1]

Setup

We work over a probability space $(\Omega, \mathcal{A}, \mathbf{P})$. We have n precepts in R^d. So let $a^1, \dots, a^n \in \mathsf{R}^d$ with data matrix $A \in \mathsf{R}^{n \times d}$.

Let $x : \Omega \to \mathsf{R}^d$ and $e : \Omega \to \mathsf{R}^n$ be random vectors with normal density (mean zero and covariances Σ_x and Σ_e respectively). For each $\omega \in \Omega$, define the map $f : \Omega \to (\mathsf{R}^d \to \mathsf{R})$ by $f(\omega)(a) = \sum_j a^i_j x_j(\omega) + e_i(\omega)$.

Define $y : \Omega \to \mathsf{R}^n$ by $y(\omega) = Ax(\omega) + e(\omega)$. In other notation,

$$y = Ax + e.$$

Let $g : \mathsf{R}^d \times \mathsf{R}^n \to \mathsf{R}$ be the density of (θ, y). Let $g_{\theta|y} : \mathsf{R}^d \times \mathsf{R}^n \to \mathsf{R}$ be the conditional density of θ given y.

Proposition 1. *(θ, y) has covariance* $\begin{pmatrix} \Sigma_\theta & \Sigma_\theta X^\top \\ X\Sigma_\theta & X\Sigma_\theta X^\top + \Sigma_e \end{pmatrix}$

Proposition 2. *There exists $c \in \mathsf{R}$, so that for all $\alpha \in \mathsf{R}^d$ and $\gamma \in \mathsf{R}^n$, $\log g(\alpha, \gamma)$ is*

$$-\frac{1}{2}(\alpha^\top \Sigma \alpha + \alpha^\top \Sigma X^\top \gamma + \gamma^\top X \Sigma \alpha + \gamma^\top X \Sigma X^\top \gamma) + c.$$

[1] The name of this sheet will change in future editions. And future editions will include accounts.

Proposition 3. *A solution to maximize $g(\alpha, \gamma)$ with respect to α is $\alpha = -\Sigma^{-1}\Sigma X^\top \gamma$.*

Proposition 4. *$g_{\theta|y}(\alpha, \gamma)$ is normal with mean*

$$\tilde{\mu}(\gamma) = \Sigma X^\top \left(X \Sigma X^\top\right)^{-1} \gamma$$

and covariance

$$\tilde{\Sigma} = \Sigma - \Sigma X^\top \left(X \Sigma X^\top\right)^{-1} X \Sigma.$$

Proposition 5. *A solution to maximize $g_{\theta|y}(\alpha, \gamma)$ w.r.t. α is*

$$\tilde{\Sigma}\tilde{\Sigma}^{-1}\tilde{\mu}(\gamma).$$

But, of course, y also has a density. Denote the density of y by $g : \mathsf{R}^n \to \mathsf{R}$. In other words, $g \geq 0$ and $\int g = 1$.

Proposition 6.

$$\log g(\gamma) = -1/2(\gamma^\top \left(X \Sigma X^\top\right)^{-1} \gamma) - \frac{d}{2} \log 2\pi - \frac{1}{2} \log \mathbf{det} \left(X \Sigma X^\top\right)$$

Test

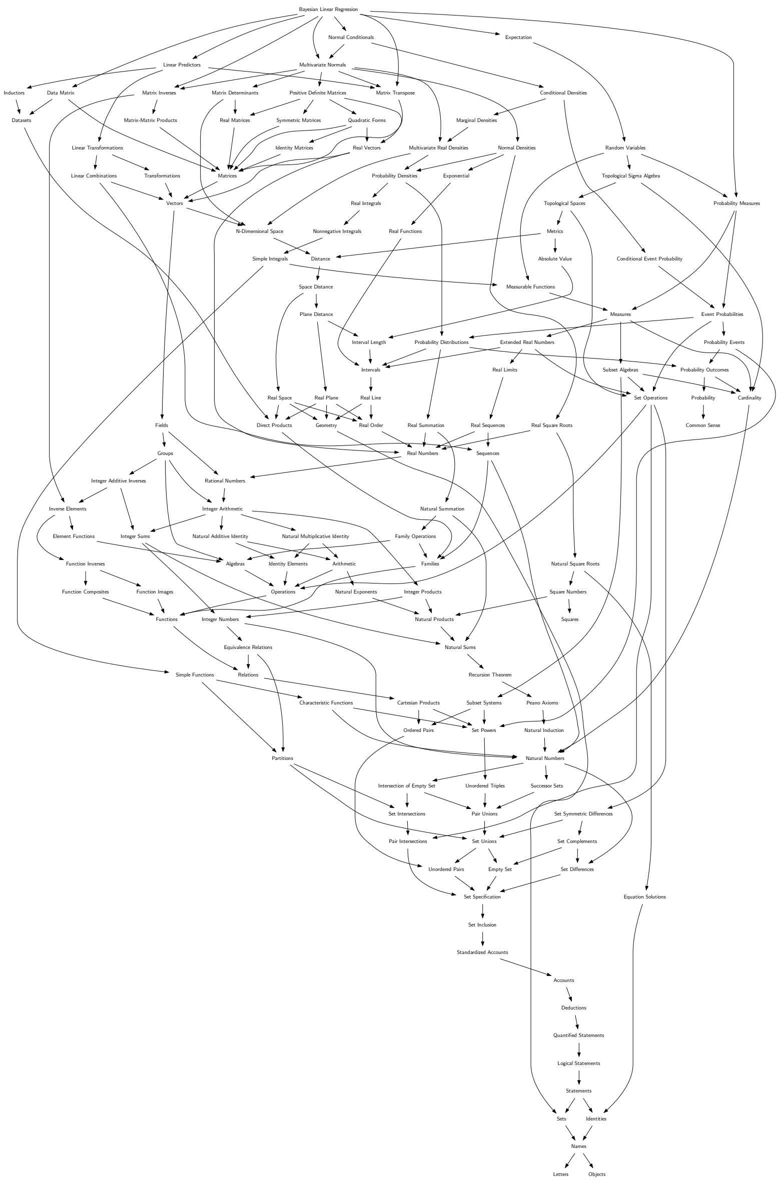
This expression makes clear that y is has a normal density with mean $X\,\mathbf{E}(x)$ and covariance $X\,\mathbf{E}(x)X^\top$.

Let $w : \Omega \to \mathsf{R}^d$ be a random vector with mean 0 and covariance ηI. Let $x^1, \dots, x^n \in \mathsf{R}^d$ Define $y^i : \Omega \to \mathsf{R}$ by $y_i(\omega) = w(\omega)^\top x^i$ for $i = 1, \dots, d$.

Noise setup

Let $e : \Omega \to \mathsf{R}^n$ be a normal random vector with mean 0 and covariance σI. Define $\tilde{y} : \Omega \to \mathsf{R}^n$ by $\tilde{y} = y(\omega) + e(\omega)$.

Proposition 7. *$\tilde{y}$ is a normal random vector with mean zero and covariance $X \Sigma X^\top + \sigma I$.*


Bayesian Linear Regression
Normal Conditionals
Expectation
Linear Predictors
Multivariate Normals
Inductors
Data Matrix
Matrix Inverses
Matrix Determinants
Positive Definite Matrices
Matrix Transpose
Conditional Densities
Datasets
Matrix-Matrix Products
Real Matrices
Symmetric Matrices
Quadratic Forms
Marginal Densities
Linear Transformations
Identity Matrices
Real Vectors
Multivariate Real Densities
Normal Densities
Random Variables
Linear Combinations
Transformations
Matrices
Probability Densities
Exponential
Topological Sigma Algebra
Vectors
Real Integrals
Topological Spaces
Probability Measures
N-Dimensional Space
Nonnegative Integrals
Real Functions
Metrics
Simple Integrals
Distance
Absolute Value
Conditional Event Probability
Space Distance
Measurable Functions
Plane Distance
Measures
Event Probabilities
Interval Length
Probability Distributions
Extended Real Numbers
Probability Events
Intervals
Real Limits
Subset Algebras
Probability Outcomes
Real Space
Real Plane
Real Line
Set Operations
Probability
Cardinality
Fields
Direct Products
Geometry
Real Order
Real Summation
Real Sequences
Real Square Roots
Common Sense
Groups
Real Numbers
Sequences
Integer Additive Inverses
Rational Numbers
Inverse Elements
Integer Arithmetic
Natural Summation
Element Functions
Integer Sums
Natural Additive Identity
Natural Multiplicative Identity
Family Operations
Function Inverses
Algebras
Identity Elements
Arithmetic
Families
Natural Square Roots
Function Composites
Function Images
Operations
Natural Exponents
Integer Products
Square Numbers
Functions
Integer Numbers
Natural Products
Squares
Equivalence Relations
Natural Sums
Simple Functions
Relations
Recursion Theorem
Characteristic Functions
Cartesian Products
Subset Systems
Peano Axioms
Ordered Pairs
Set Powers
Natural Induction
Partitions
Natural Numbers
Intersection of Empty Set
Unordered Triples
Successor Sets
Set Intersections
Pair Unions
Set Symmetric Differences
Pair Intersections
Set Unions
Set Complements
Unordered Pairs
Empty Set
Set Differences
Set Specification
Equation Solutions
Set Inclusion
Standardized Accounts
Accounts
Deductions
Quantified Statements
Logical Statements
Statements
Sets
Identities
Names
Letters
Objects