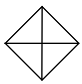


Why

What if we use the principle of maximum likelihood to select the maximum likelihood normal density which factors according to a tree?

Definition

A *maximum likelihood tree normal* of a dataset in $\mathbf{R}^d$ is a multivariate normal density that factors according to a tree and maximizes the likelihood of the dataset.

Results

Prop. 1. *Let $D = (x^1, \dots, x^n)$ be a dataset in $\mathbf{R}^d$. A normal density is a maximum likelihood tree normal of D if and only if it is an optimal tree approximator of the empirical normal density of D.*

Proof. First, let $f : \mathbf{R}^d \to \mathbf{R}$ be a normal density.

First, express the log likelihood of f on a record x^k by

$$\log f(x^k) = -\frac{1}{2}(x^k - \mu)^\top \Sigma^{-1}(x^k - \mu) - \frac{1}{2} \log \mathbf{det}\, \Sigma - \frac{d}{2} \log 2\pi.$$

Second, use the trace to rewrite the quadratic form

$$-\frac{1}{2} \mathbf{tr} \left(\Sigma^{-1}(x^k - \mu)(x^k - \mu)^\top \right).$$

Third, use these two, and the linearity of trace to express the average negative log likelihood by

$$-\frac{1}{n}\sum_{k=1}^{n}\log f(x^k) = \frac{1}{2}\,\mathbf{tr}\left(\Sigma^{-1}\left(\frac{1}{n}\sum_{i=1}^{n}(x^k-\mu)(x^k-\mu)^\top\right)\right) + \frac{1}{2}\log\mathbf{det}\,\Sigma$$

Fourth, use matrix calculus (or the derivation in Proposition 1 of Multivariate Normal Maximum Likelihood) to see that, for a minimizer of the negative average log likelihood, the mean must be $\frac{1}{n}\sum_{i=1}^{n} x^k$.

Fifth, recognize the empirical covariance matrix $\frac{1}{n}\sum_{k=1}^{n}(x^k - \mu)(x^k - \mu)^\top$; denote it by S.

Sixth, change variables with $P = \Sigma^{-1}$ and express

$$\log\left(\mathbf{det}\left(\Sigma\right)\right) = \log\left(\mathbf{det}\left(P^{-1}\right)\right) = \log\left(\left(\mathbf{det}\,P\right)^{-1}\right) = -\log\left(\mathbf{det}\left(P\right)\right)$$

Seventh, write the likelihood in simplified form (using circulant property of trace)

$$\frac{1}{2}\,\mathbf{tr}\left(SP\right) - \frac{1}{2}\log\mathbf{det}\,P - \frac{d}{2}\log 2\pi$$

Eighth, drop the constant and prefactors:

$$\mathbf{tr}\left(SP\right) - \log\mathbf{det}\,P$$

Ninth, if g is a normal with then the tree density approximation objective is the same equivalent to

$$d(g,f) = h(g,f) - h(f) \sim h(g,f) = -\int_{\mathbf{R}^d} g\log f.$$

TODO: Extra, the let g be normal and f be normal. The tree normal approximation problem

$$d(g, f) \sim h(g, f) = -\int_{R^d} g \log f.$$

The log of f is

$$-\frac{1}{2} \mathbf{tr} \left(\Sigma_f \int_{\mathbf{R}^d} (x - \mu_f)(x - \mu_f)^\top dx \right) - \frac{1}{2} \log \mathbf{det}\, \Sigma_f^{-1} - \frac{d}{2} \log 2\pi$$

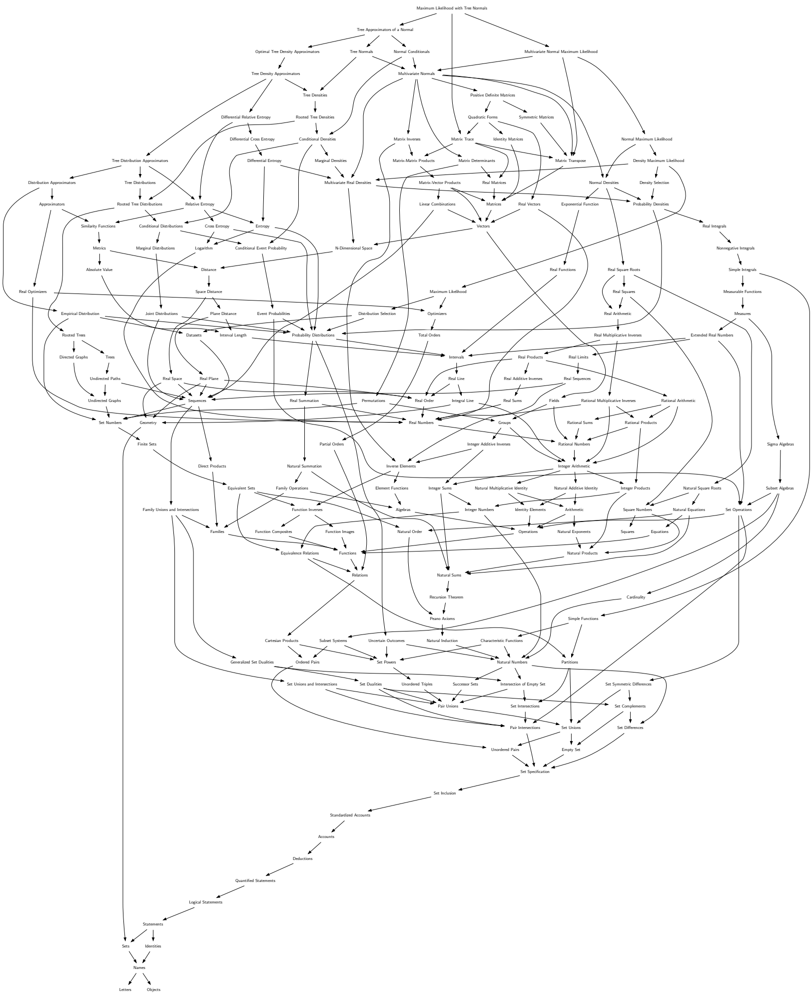
Since the set of optimal solutions for both optimization is contained in the set of normal densities which match on the mean we can assume that $\mu_f = \mu_g$. So the approximation objective is equivalent to

$$\mathbf{tr}\, P_f \Sigma_g - \log \mathbf{det}\, P_f,$$

which is exactly the maximum likelihood objective.

Thus, a solution is a maximum likelihood tree normal of a dataset if and only if it is an optimal tree approximator of the empirical normal density of the dataset.

$\square$



Maximum Likelihood with Tree Normals
Tree Approximators of a Normal
Optimal Tree Density Approximators
Tree Normals
Normal Conditionals
Multivariate Normal Maximum Likelihood
Tree Density Approximators
Multivariate Normals
Tree Densities
Positive Definite Matrices
Differential Relative Entropy
Rooted Tree Densities
Quadratic Forms
Symmetric Matrices
Differential Cross Entropy
Conditional Densities
Matrix Inverses
Matrix Trace
Identity Matrices
Normal Maximum Likelihood
Tree Distribution Approximators
Differential Entropy
Marginal Densities
Matrix-Matrix Products
Matrix Determinants
Matrix Transpose
Density Maximum Likelihood
Distribution Approximators
Tree Distributions
Multivariate Real Densities
Matrix-Vector Products
Real Matrices
Normal Densities
Density Selection
Approximators
Rooted Tree Distributions
Relative Entropy
Linear Combinations
Matrices
Real Vectors
Exponential Function
Probability Densities
Similarity Functions
Conditional Distributions
Cross Entropy
Entropy
Vectors
Real Integrals
Metrics
Marginal Distributions
Logarithm
Conditional Event Probability
N-Dimensional Space
Nonnegative Integrals
Absolute Value
Distance
Real Functions
Real Square Roots
Simple Integrals
Real Optimizers
Space Distance
Maximum Likelihood
Real Squares
Measurable Functions
Empirical Distribution
Joint Distributions
Plane Distance
Event Probabilities
Distribution Selection
Optimizers
Real Arithmetic
Measures
Rooted Trees
Datasets
Interval Length
Probability Distributions
Total Orders
Real Multiplicative Inverses
Extended Real Numbers
Directed Graphs
Trees
Intervals
Real Products
Real Limits
Undirected Paths
Real Space
Real Plane
Real Line
Real Additive Inverses
Real Sequences
Undirected Graphs
Sequences
Real Summation
Permutations
Real Order
Integral Line
Real Sums
Fields
Rational Multiplicative Inverses
Rational Arithmetic
Set Numbers
Geometry
Real Numbers
Groups
Rational Sums
Rational Products
Finite Sets
Partial Orders
Integer Additive Inverses
Rational Numbers
Sigma Algebras
Direct Products
Natural Summation
Inverse Elements
Integer Arithmetic
Equivalent Sets
Family Operations
Element Functions
Integer Sums
Natural Multiplicative Identity
Natural Additive Identity
Integer Products
Natural Square Roots
Subset Algebras
Family Unions and Intersections
Function Inverses
Algebras
Integer Numbers
Identity Elements
Arithmetic
Square Numbers
Natural Equations
Set Operations
Families
Function Composites
Function Images
Natural Order
Operations
Natural Exponents
Squares
Equations
Equivalence Relations
Functions
Natural Products
Relations
Natural Sums
Recursion Theorem
Cardinality
Peano Axioms
Simple Functions
Cartesian Products
Subset Systems
Uncertain Outcomes
Natural Induction
Characteristic Functions
Generalized Set Dualities
Ordered Pairs
Set Powers
Natural Numbers
Partitions
Set Unions and Intersections
Set Dualities
Unordered Triples
Successor Sets
Intersection of Empty Set
Set Symmetric Differences
Pair Unions
Set Intersections
Set Complements
Pair Intersections
Set Unions
Set Differences
Unordered Pairs
Empty Set
Set Specification
Set Inclusion
Standardized Accounts
Accounts
Deductions
Quantified Statements
Logical Statements
Statements
Sets
Identities
Names
Letters
Objects