Why

We can approximate a density with a tree density similar to how we can approximate a distribution with a tree distribution.

Definition

We use the differential relative entropy as a criterion of approximation. An *optimal tree approximator* of density for a tree is a density which factors according to a tree and minimizes its differential relative entropy with the given density.

Notation

Let $g : \mathbf{R}^n \to \mathbf{R}$ be a density and T be a tree on $\{1, \ldots, n\}$. An optimal tree approximator of g for T is a density f that factors according to T and minimizes $d(g, f)$. In other words, given g and T we want to find f to

$$\text{minimize} \quad d(g, f)$$
$$\text{subject to} \quad f \text{ factors according to } T.$$

Result

Proposition 1. *Let* $g : \mathbf{R}^n \to \mathbf{R}$ *be a density and* T *be a tree on* $\{1, \ldots, n\}$. *The density* $f_T^* : \mathbf{R}^d \to \mathbf{R}$ *defined by*

$$f^*\_T = g\_1 \prod \_i \neq 1 g\_i \mid \text{pa}\, i$$

minimizes the differential relative entropy with g *among all densities on* $\mathbf{R}^n$ *which factor according to* T *(*$\text{pa}\, i$ *is the parent of* i *in* T *rooted at vertex* 1, $i = 2, \ldots, n$*).*

Proof. Let $f : \mathbf{R}^d \to \mathbf{R}$ be a density factoring according to T. First, express

$$f = f_1 \prod_{i=1} f_{i \mid \text{pa}\, i}.$$

Second, recall that $d(g, f) = h(g, f) - h(g)$. Since $h(g)$ does not depend on f, f is a minimizer of $d(g, f)$ if and only if f is a minimizer of $h(g, f)$.

Third, express

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

$$\& = -\int_{\mathbf{R}^d} g(x) \left(\log f_i(x_i) + \sum_{i \neq 1} \log f_{i|\mathrm{pa}\,i}(x_i, x_{\mathrm{pa}\,i}) \right) dx$$

$$\& = h(g_1, f_1) + \sum_{i \neq 1} \left(\int_{\mathbf{R}} g_{\mathrm{pa}\,i}(\xi) h\big(g_{i|\mathrm{pa}\,i}(\cdot, \xi), f_{i|\mathrm{pa}\,i}(\cdot, \xi) \big) d\xi \right)$$

which separates across f_1 and $f_{i|\mathrm{pa}\,i}(\cdot, \xi)$ for $i = 1, \ldots, n$ and $\xi \in \mathbf{R}$. In particular, since $g_{pai} \geq 0$, we can minimize the integrand pointwise.

Fourth, recall $h(\phi, \psi) \geq 0$ for densities ϕ, ψ of any dimension, and zero if $\phi = \psi$. So $f_1 = g_1$ and $f_{i|\mathrm{pa}\,i} = g_{i|\mathrm{pa}\,i}$ are solutions. $\qquad \square$

Tree Density Approximators

Tree Densities

Differential Relative Entropy

Rooted Tree Densities

Differential Cross Entropy

Conditional Densities

Differential Entropy

Marginal Densities

Multivariate Real Densities

N-Dimensional Space

Probability Densities

Real Integrals

Nonnegative Integrals

Tree Distribution Approximators

Simple Integrals

Tree Distributions

Distribution Approximators

Measurable Functions

Relative Entropy

Approximation

Rooted Tree Distribution

Measure

Discrete Entropy

Similarity Functions

Conditional Distributions

Sigma Algebras

Metrics

Conditional Event Probabilities

Marginal Distributions

Extended Real Numbers

Real Series

Distance

Cross Entropy

Event Probabilities

Joint Distributions

Empirical Distribution

Real Limits

Space Distance

Logarithm

Probability Distribution

Absolute Value

Simple Functions

Plane Distance

Exponential Function

Interval Length

Real Functions

Rooted Tree

Real Space

Intervals

Real Sequence

Trees

Directed Graphs

Subset Algebra

Real Line

Real Plane

Sequences

Undirected Paths

Integral Line

Real Order

Real Summation

Real Optimizer

Cardinality

Lists

Undirected Graphs

Geometry

Real Numbers

Rational Order

Set Numbers

Rational Numbers

Optimizers

Finite Sets

Direct Products

Integer Arithmetic

Natural Multiplication Identity

Natural Additive Identity

Natural Summation

Integer Products

Equivalent Sets

Comparisons

Identity Elements

Arithmetic

Family Operations

Function Inverses

Integer Order

Integer Sums

Family Unions and Intersections

Operations

Families

Function Images

Natural Powers

Function Composition

Orders

Integer Numbers

Natural Order

Converse Relation

Function

Natural Products

Equivalence Relation

Natural Sums

Relations

Subset Systems

Set Products

Characteristic Function

Recursion Theorem

Ordered Pairs

Set Powers

Peano Axioms

Unordered Triples

Natural Numbers

Generalized Set Dualities

Natural Induction

Uncertain Outcomes

Set Symmetric Differences

Set Dualities

Successor Sets

Intersection of Empty Set

Partitions

Set Unions and Intersections

Set Complements

Pair Unions

Set Intersections

Set Differences

Set Unions

Pair Intersections

Unordered Pairs

Empty Set

Set Specification

Set Equality

Set Inclusion

Deduction

Quantified Statements

Logical Statements

Statements

Sets

Identities

Names

Letters

Objects