Why

Which tree is optimal for tree distribution approximation?

Definition

We want to choose a tree whose corresponding approximator for the given distribution minimizes the relative entropy with the given distribution among all tree distribution approximators. We call such a distribution an *optimal* tree approximator of the given distribution. We call a tree according to which an optimal tree approximator factors and *optimal* approximator tree.

Result

Proposition 1. *Let $A_1, \ldots, A_n$ be finite nonempty sets. Define $A = \prod_{i=1}^{n} A_i$. Let $q : A \to [0,1]$ a distribution. A tree T on $\{1, \ldots, n\}$ is an optimal approximator tree if and only if it is a maximal spanning tree of the mutual information graph of q.*

Proof. First, denote the optimal tree distribution approximator of q for tree T by p_T^*. Express

$$p_T^* = q_1 \prod_{i \neq 1} q_{i|\mathrm{pa}\, i}$$

Second, express $d(q, p) = H(q, p) - H(q)$. Since $H(q)$ does not depend on T, p_T^* is a minimizer (w.r.t. T) of $d(q, p_T^*)$ if and only if it is a minimizer of $H(q, p_T^*)$.

Third, express the cross entropy of p_T^* relative to q as

$$H(q, p_T^*) = H(q_1) - \sum_{i \neq 1} \sum_{a \in A} q(a) \log q_{i|pai}(a_i, a_{\text{pa}\,i})$$

$$= H(q_1) - \sum_{i \neq 1} \sum_{a \in A} q(a)(\log q_{i,\text{pa}\,i}(a_i, a_{\text{pa}\,i}) - \log q_{\text{pa}\,i}(a_{\text{pa}\,i}))$$

$$= H(q_1) - \sum_{i \neq 1} \sum_{a \in A} q(a)(\log q_{i,\text{pa}\,i}(a_i, a_{\text{pa}\,i}) - \log q_{\text{pa}\,i}(a_{\text{pa}\,i}) - \log q_i(a_i) + \log$$

$$= \sum_{i=1}^{n} H(q_i) - \sum_{i \neq 1} I(q_i, q_{\text{pa}\,i})$$

$$= \sum_{i=1}^{n} H(q_i) - \sum_{\{i,j\} \in T} I(q_i, q_j)$$

where $\text{pa}\,i$ denotes the parent of vertex i in T rooted at vertex 1 ($i = 2, \ldots, n$). For $i = 1, \ldots, n$, $H(q_i)$ does not depend on the choice of tree. Therefore selecting a tree to minimize the second term in the final expression above is equivalent to choosing a maximal spanning tree from the weighted graph with mutual information edge weights; namely, the mutual information graph of q.

$\square$

Proposition 1 says that to we should first select a maximum spanning tree of the mutual information graph of the distribution we are approximating. Then, we should pick the best approximator to q which factors according to that three.

Optimal Tree Distribution Approximators

Tree Distribution Approximators — Mutual Information Graph

Distribution Approximators — Tree Distributions — Mutual Information

Approximators — Relative Entropy

Similarity Functions

Rooted Tree Distributions

Metrics

Conditional Distributions

Distance — Absolute Value

Conditional Event Probabilities — Marginal Distributions

Space Distance

Real Optimizers — Plane Distance — Empirical Distribution — Discrete Entropy — Cross Entropy — Event Probabilities — Joint Distributions — Optimal Spanning Trees

Interval Length — Logarithm — Optimizers — Probability Distributions — Weighted Graphs — Rooted Trees — Spanning Tree

Real Space — Intervals — Directed Graphs — Trees

Real Plane — Real Line

Integral Line — Real Order — Real Summation — Lists — Undirected Paths

Geometry — Rational Order — Real Numbers — Undirected Graphs — Set Numbers

Rational Numbers — Finite Sets

Integer Arithmetic — Direct Products

Integer Products — Natural Multiplicative Identity — Natural Additive Identity

Arithmetic — Identity Elements

Natural Powers — Natural Summation — Equivalent Sets

Comparisons — Integer Order — Integer Sums — Natural Products — Family Operations — Function Inverses — Family Unions and Intersections

Orders — Integer Numbers — Natural Sums — Operations — Families — Function Composites — Function Images

Converse Relations — Recursion Theorem — Natural Order — Equivalence Relations — Functions

Relations — Peano Axioms

Set Products — Natural Induction — Uncertain Outcomes

Set Powers — Natural Numbers — Generalized Set Dualities

Set Symmetric Difference — Unordered Triples — Successor Sets — Set Dualities — Intersection of Empty Set — Partitions — Set Unions and Intersections

Ordered Pairs — Set Complements — Pair Unions — Set Intersections

Set Differences — Set Unions — Pair Intersections

Unordered Pairs — Empty Set

Set Specification

Set Equality

Set Inclusion

Standardized Accounts

Accounts

Deductions

Quantified Statements

Logical Statements

Statements

Sets — Identities

Names

Letters — Objects

4