

Tree Distribution Approximators

1 Why

We approximate a distribution with a distribution that factors according to a given tree. Such a distribution requires tabulating fewer numbers to express the probability of outcome.

2 Definition

We will use the relative entropy as a criterion of approximation. Given a distribution over a product of finite sets and a tree, we want to find the optimal approximator among distributions which factor according to the tree. We call such a distribution a tree distribution approximator or tree approximator of the given distribution for the tree. We call the tree the approximator tree.

3 Result

Proposition 1. Let A_1, \ldots, A_n be finite non-empty sets. Define $A = \prod_{i=1}^n A_i$. Let $q: A \to [0,1]$ a distribution and T a tree on $\{1,\ldots,n\}$. The distribution $p_T^*: A \to [0,1]$ defined by

$$p_T^* = q_1 \prod_{i
eq 1} q_{i|\mathsf{pa}_i}$$

minimizes the relative entropy with q among all distributions on A which factor according to T.

Proof. Let $p:A\to [0,1]$ be a distribution which factors according to T. First, express

$$p = p_1 \prod_{i \neq i} p_{i|\mathsf{pa}_i}$$

where \mathbf{pa}_i is the parent of vertex i in T rooted at vertex 1 (i = 2, ..., n).

Second, recall that the relative entropy of q with p is H(q, p) - H(q). Since H(q) does not depend on p, p is a minimizer of the relative of q with p if and only if p is a minimizer of H(q, p).

Third, express

$$\begin{split} H(q,p) &= -\sum_{a \in A} q(a) \log p(a) \\ &= -\sum_{a \in A} q(a) \left(\log p_1(a_1) + \sum_{i \neq 1} \log p_{i|\mathbf{pa}_i} \big(a_i, a_{\mathbf{pa}_i}\big) \right) \\ &= H(q_1, p_1) + \sum_{i \neq 1} \sum_{a_{\mathbf{pa}_i} \in A_{\mathbf{pa}_i}} q_{\mathbf{pa}_i}(a_{\mathbf{pa}_i}) H\big(q_{i|\mathbf{pa}_i}(\cdot, a_{\mathbf{pa}_i}), p_{i|\mathbf{pa}_i}(\cdot, a_{\mathbf{pa}_i})\big) \end{split}$$

which separates across p_1 an $p_{i|\mathbf{pa}_i}(\cdot, a_{\mathbf{pa}_i})$ for $i = 2, \ldots, n$ and $a_{pa_i} \in A_{\mathbf{pa}_i}$.

Fourth, recall $H(\cdot, \cdot) \ge 0$ and is zero on repeated pairs. By this, we mean, for example, $H(p_1, p_1) = 0$. So $p_1 = q_1$ and $p_{i|\mathbf{pa}_i} = q_{i|\mathbf{pa}_i}$ are solutions.

Proposition 1 states the form of an optimal approximator given a tree. A natural next question is to select the tree.