



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.

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*.

Result

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

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

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

Proof. Let $p : A \rightarrow [0, 1]$ be a distribution which factors ac-

cording to T . First, express

$$p = p_1 \prod_{i \neq 1} p_{i|\mathbf{pa} \, i}$$

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

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

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

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

Fourth, recall $H(\cdot, \cdot) \geq 0$ 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.

