Tree Distribution Approximation

1 Why

We approximate with tree distributions. These distributions require tabulating fewer numbers in order to express the probability of an outcome, which may save us tabulating many more numbers.

2 Problem

We approximate a distribution by a tree distribution using the relative entropy as a criterion.

2.1 Notation

Let $A_1, \dots, A_n$ be finite non-empty sets and define $A = \prod_{i=1}^{n} A_i$. Let $q : A \to [0, 1]$ a distribution. Let d denote the relative entropy.

We want to find a distribution p on A and tree T on $\{1, \dots, n\}$ to

$$\begin{array}{rl} \text{minimize} & d(q, p) \\ \text{subject to} & p \text{ factors according to the tree } T \end{array}$$

3 Solution

Proposition 1. *Let q be a distribution on A. Let T be a tree on $\{1, \dots, d\}$. Let p_j be the parent of vertex j for the T rooted at vertex i, $j = 1, \dots, n$ and*

$j \neq i$. Then the distribution p on A defined by

$$p = q_i \prod_{j \neq i} q_{j|p_j}$$

achieves minimum entropy relative to q among all distributions which factor according to T.

Proposition 2. *Let q be a distribution on A. Let T be a tree on $\{1, \ldots, d\}$. Let p_j be the parent of vertex j for the T rooted at vertex i, $j = 1, \ldots, n$ and $j \neq i$. Then the distribution p on A defined by*

$$p = q_i \prod_{j \neq i} q_{j|p_j}$$

achieves minimum entropy relative to q among all distributions which factor according to T.

Proposition 3. *Let q be a distribution on A. A tree T is a solution to the problem above if and only if it is a minimum spanning tree of the mutual information graph of q.*