CS281A/Stat241A: Statistical Learning Theory

Examples and Basic Concepts

Lecturer: Michael I. Jordan *Scribes: Luo*

Examples

Phylogeny, part-of-speech tagging, Kalman filtering and smoothing, disease diagnosis, Independent Component Analysis (or blind source separation), error-control coding.

Directed graphical models

Given a directed acyclic graph $\mathcal{G}(\mathcal{V}, \mathcal{E})$, associate random variables X_i indexed by the vertices $i \in \mathcal{V}$. Denote x_i as a realization of X_i, so that, if X_i is discrete, the joint distribution is

$$p(x_1, \ldots, x_n) \triangleq p(X_1 = x_1, \ldots, X_n = x_n).$$

If X_i is continuous,

$$p(x_1, \ldots, x_n) \triangleq \frac{\partial^n}{\partial x_1 \cdots \partial x_n} p(X_1 \leq x_1, \ldots, X_n \leq x_n).$$

Let π_i denote $\{j : (j, i) \in \mathcal{E}\}$, indices of the parents of node i. Restricting to the discrete case, use nonnegative functions f_i with $\sum_{x_i} f_i(x_i, x_{\pi_i}) = 1$ to define

$$p(x_1, \ldots, x_n) \triangleq \prod_{i=1}^{n} f_i(x_i, x_{\pi_i}),$$

where we allow subsets of indices to appear wherever single indices appear.

For example, consider $X_i \in \{0, 1\}$ for the graph in Figure 1. Each f_i is a table on node i. Next, consider the graph in Figure 2. What is the family of distributions associated with this graph? (Since it has no trivial conditional independencies, it is associated with the family of all probability distributions, as we later show.) Calculating conditionals from the joint is easy. For the graph of Figure 2, marginalize to get $p(x_1, x_2) = \sum_{x_3} p(x_1, x_2, x_3)$ and $p(x_1) = \sum_{x_2} p(x_1, x_2)$. Then the conditionals are $p(x_2|x_1) = \frac{p(x_1, x_2)}{p(x_1)}$ and $p(x_1|x_2, x_3) = \frac{p(x_1, x_2, x_3)}{p(x_2, x_3)}$. What if our graph is loopy? Can we have $p(x_1|x_2)$ inconsistent with $p(x_2|x_1)$? That is, are the two probabilities conditionals under the same joint?

Theorem 1. *$p(x_1, \ldots, x_n)$ defines a joint probability distribution and $f_i(x_i, x_{\pi_i}) = p(x_i|x_{\pi_i})$, the conditional distribution of x_i given its parents.*

Proof. $p(x_1, \ldots, x_n) = \prod_{i=1}^{n} f_i(x_i, x_{\pi_i}) \geq 0$ since f_is are nonnegative. Compute

$$\sum_{x_1, \cdots, x_n} \prod_{i=1}^{n} f_i(x_i, x_{\pi_i})$$

by marginalizing in reverse topological order, and using $\sum_{x_i} f_i(x_i, x_{\pi_i}) = 1$. □

X_A is conditionally independent of X_B given X_C if $p(X_A, X_B|X_C) = p(X_A|X_C)p(X_B|X_C)$. We'll show that we can find conditional indepence statements graph theoretically.

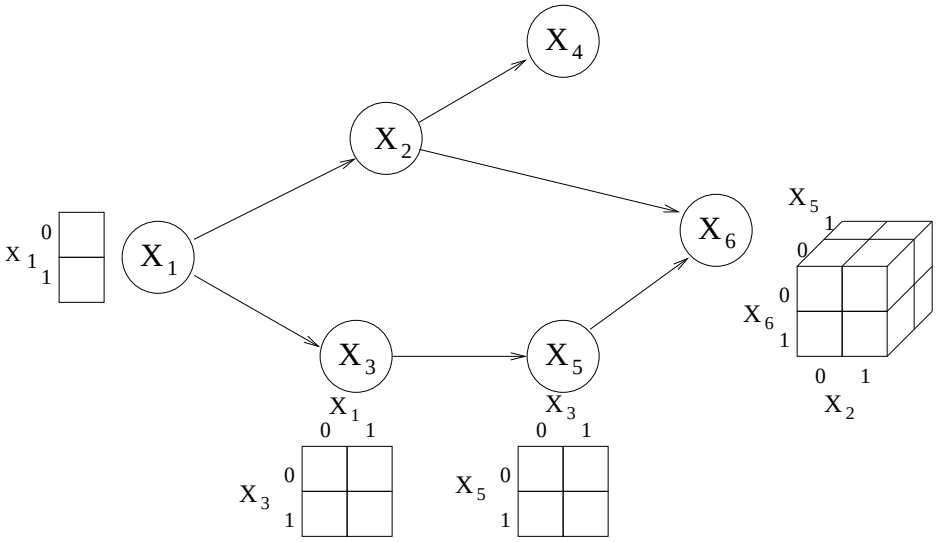

Figure 1: Example graphical model for binary X_i with tables for f_1, f_3, f_5, and f_6.

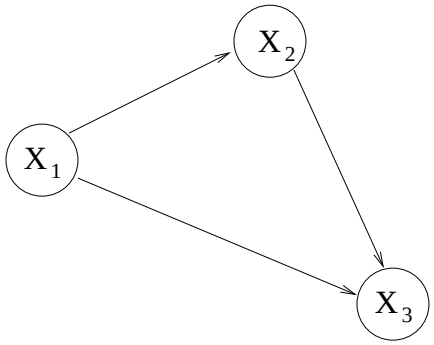

Figure 2: A graph with no trivial conditional independencies.