Probabilistic Graphical Model

1 Markov Network

1.1 Definition

A Markov Network (alias Markov Random Field) is defined by:

- A set of **random variables** $X = (X_1, \ldots, X_n)$
- **Undirected graph** G, with vertices corresponding to random variables.
- Non-negative **potential functions** $\{\phi_k\}$.
- Each (maximum) **clique** $c \in C$ of G is assigned with a corresponding potential function.

The joint distribution represented by a Markov network is:

$$P(X = x) = \frac{1}{Z} \prod_{c \in C} \phi_c(x_c) \tag{1}$$

- where Z normalizes the probability and is often referred to as the **partition function**.

1.2 Properties

Markov networks are often expressed as log-linear models, in which the potential functions is replaced by an exponentiated weighted sum of features:

$$P(X = x) = \frac{1}{Z} \exp\left(\sum_j w_j f_j(x)\right) \tag{2}$$

2 Markov Logical Network

(See Link)

2.1 First-order Knowledge Base

A first-order knowledge base is a set of formulas in first-order logic, constructed using four types of symbols: constants, variables, functions, and predicates. A first-order KB can also be viewed as a set of hard constraints: If a *possible world* violates even one formula, it has zero probability.

Markov logical network softens these hard constraints (formulas), and uses weights to represent how strong a constraint is.

2.2 Definition

A Markov logic network L consists of a set of pairs (F_i, w_i). F_i is a first-order logic formula and weight $w_i \in \mathbb{R}$ denotes its importance.

- Nodes: For each predicates in L (0/1 random variables)
- Factors: Each factor represents a formula F_i in L. It contains a weight w_F and a factor function $f_F : \bar{v}_F \to \{0, 1\}$, where $\bar{v}_F$ is the clique in Markov network.
- The joint probability is:

$$P(X = x) = \frac{1}{Z} \exp\left(\sum_{F \in \mathcal{F}} w_F f_F(x(\bar{v}_F))\right) \tag{3}$$

3 Bayesian Network

3.1 Definition

A Bayesian network is a DAG, in which each node represents a random variable and each edge represents a direct function relation. For each node x_i, the conditional independency is given as:

$$\forall x_i, (x_i \perp\!\!\!\perp NonDescendents\ of\ x_i | Pa_i) \tag{4}$$

Bayesian network is in 1-1 correspondence with casual structures, and it represents the joint distribution P. We say a distribution P over X factorizes according to G if:

$$P(x_1, \ldots, x_n) = \prod_{i=1}^{n} P(x_i | Pa_{x_i}^G). \tag{6}$$