10-708: Probabilistic Graphical Models 10-708, Spring 2018

3 : Introduction

Lecturer: Kayhan Batmanghelich *Scribes: Arjun Sharma, Mohit Sharma*

1 Why Undirected Graphs?

The undirected graphical models (UGM) are also referred to as Markov Random Fields. In constrast to directed graphical models (DGM) which directly model causal relationship between random variables, UGM are used to mdoel pair-wise, non-causal relationship between random variables.

For instance, Figure 1 shows one particular instance where an UGM is more relevant than a DGM. Notice, that the relationship between superpixels in the above figure 1 is *not really causal* in any sense. These superpixels are related with each other but we cannot say that one superpixel "causes" another superpixel. Hence, for such relationships a UGM is a better modeling choice than a DGM.

2 Undirected Graphical Models

Similar to Bayesian Networks, undirected graphical models consists of sets of nodes and edges. Each node represents a random variable while an edge represents a direct interaction between the neighboring variables, *i.e.*, an interaction not affected by other interactions in the graph. Additionally, these edges are used to capture the affinity between interacting variables.

We now look at some other examples for undirected graphical models. Figure 2 shows a social network graph which has been modeled as a Markov Network. Notice that interactions in most social settings cannot be explained by a causal structure. Hence, undirected graphical models present a good choice to model these interactions. Also, Figure 3 shows an example of Markov Networks being used in Computer Vision. Markov Networks have been widely used for in computer vision tasks such as, image segmentation. The task of image segmentation requires partitioning a given image into multiple groups such that, each group represents a particular object entity. Markov networks are used to incorporate the spatial relationships among neighbouring labels as a markovian prior. This prior can encourage adjacent pixels to be classified into the same group. An instance of this process can be seen in Figure 3.

More formally, an undirected graphical model consists of a set of random variables $\mathcal{X} = x_1, x_2, ..., x_n$; where the joint distribution between these variables is modeled as a product of potentials on subsets of the random variables $\mathcal{X}_c \subseteq \mathcal{X}$,

$$p(x_1, x_2, ..., x_n) = \frac{1}{Z} \prod_{c=1}^{C} \phi_c(\mathcal{X}_c) \tag{1}$$

The term ϕ_c above represents a potential function which is required to be $\phi_c \geq 0$. These potential functions capture the affinity between interacting random varaibles, *i.e.*, if the The terms $\mathcal{X}_c$ represent the maximal cliques in the graph. Section 4 discusses how these cliques are defined in an undirected graphical model.

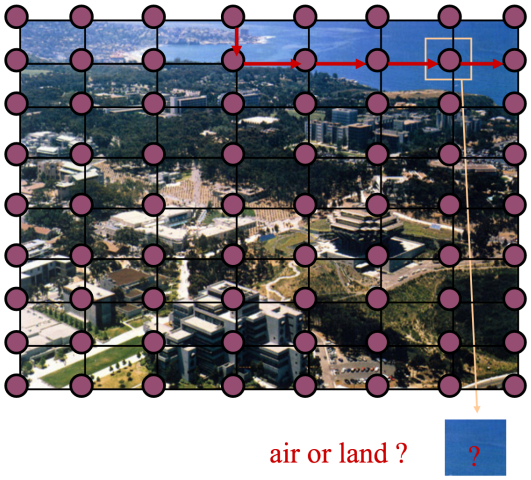


Figure 1: An example for undirected graphical model

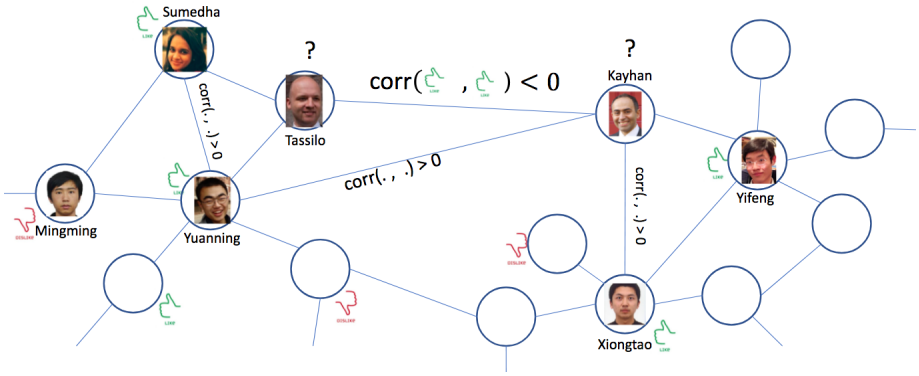


Figure 2: Social networks can be adequately represented as undirected graphical models.

3 Independence

Similar to Bayesian Networks, undirected graphical models present an approach to look at an undirected graph as a representation of independence assertions. In the following sections we look at both Global independence and local independence.

3.1 Global independence

The graph structure in a Markov Network encodes a set of independence assumptions. For this we first define the concept of active path in Markov Networks,

Active Path: Let $\mathcal{H}$ be a Markov Network structure, a path $X_1, ..., X_k$ in $\mathcal{H}$ and let $\mathbf{Z} \subseteq \mathcal{X}$ be a subset of observed variables. The path $X_1, ..., X_k$ is said to be active given $\mathbf{Z}$ if none of the $X_i, i = 1, 2, ..., k$ is in $\mathbf{Z}$.

We can now define the separation,

Separation: A set of nodes $\mathbf{Z}$ separates $\mathbf{X}$ and $\mathbf{Y}$ in $\mathcal{H}$ denoted by $sep_H(\mathbf{X}; \mathbf{Y}|\mathbf{Z})$, if there is no active path

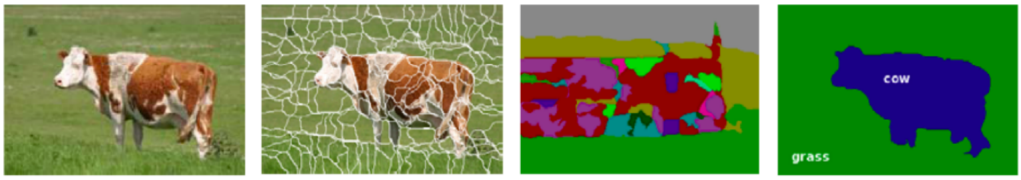


Figure 3: Many old Image segmentation algorithms are based on Markov Networks.

between any node $x \in \mathbf{X}$ and $y \in \mathbf{Y}$ given $\mathbf{Z}$. Thus, global independences associated with $\mathcal{H}$ is defined as:

$$\mathcal{I}(\mathcal{H}) = (\mathbf{X} \perp\!\!\!\perp \mathbf{Y}|\mathbf{Z} : sep_{\mathcal{H}}(\mathbf{X}; \mathbf{Y}|\mathbf{Z}) \tag{2}$$

3.2 Local independence

To define local independence for each node in a Markov Network we define the Markov Blanket for every node. Thus, for a graph with node $X_i \in \mathcal{X}$, there is a markov blanket denoted as MB_{X_i} which is the set of neighbors of X_i. Thus, local independence associated with each node can be defined as:

$$X_i \perp\!\!\!\perp \mathcal{X} - X_i - X_{MB_i}|X_{MB_i} \tag{3}$$

Thus, local independence simply states that a node is independent of all the other nodes in a Markov Network given its immediate neighbors.

4 Clique

A clique in an undirected graph is a subset of vertices such that every two distinct vertices in this subset are adjacent. A maximal clique is a clique that cannot be extended by including one more adjacent vertex, that is, a clique which does not exist exclusively within the vertex set of a larger clique. For example in Figure 4, the set $\{x_1, x_2, x_3\}$ is a clique and the set $\{x_1, x_2, x_3, x_4\}$ is a maximal clique.

5 Interpretation of clique potentials

In directed graphical models, the joint distribution over the vertices could be factorized as a product of marginal and conditional distributions. In undirected graphical models, the joint distribution can be factorized into a product of clique potentials. However, these clique potentials are not necessarily marginals or conditionals. They only represent a notion of "goodness" or "compatibility" of the variables. To illustrate this, consider the graph shown in Figure 5. While the graph implies $X \perp\!\!\!\perp Z|Y$ and hence, the joint distribution can be represented as $p(x, y, z) = p(y)p(x|y)p(z|y)$, it can also be written in other forms as shown in the figure.

6 Hammersley-Clifford theorem

For an undirected graph G with vertices $V = \{Y_1, \cdots, Y_n\}$, the following are equivalent:

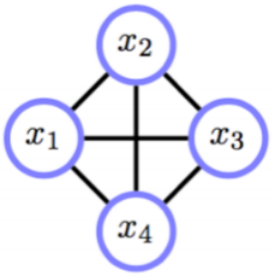


Figure 4: Figure illustrating clique and maximal clique

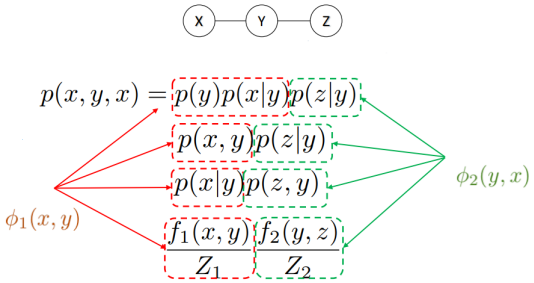


Figure 5: Interpreting clique potentials

(i) P is a distribution such that

$$P(Y_j = y_j) > 0 \quad \forall j = 1, \cdots, n \quad \Rightarrow \quad P(Y_1 = y_1, \cdots, Y_n = y_n) > 0$$

and P satisfies the Markov Property

(ii) The joint distribution can be factorized over cliques of G, i.e.

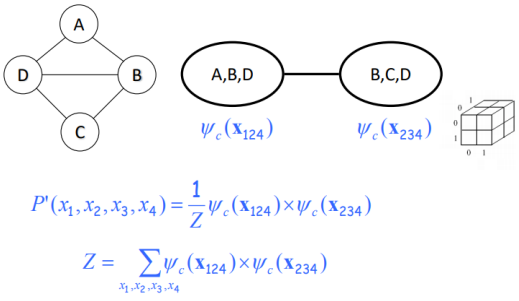
$$P(Y_1 = y_1, \cdots, Y_n = y_n) = \frac{1}{Z} \prod_{c \in \mathbb{C}} \phi_c(Y_c)$$

where $\mathbb{C}$ denotes the set of cliques of G and ϕ_c is a strictly positive function. In other words, G is a Gibbs Random Field.

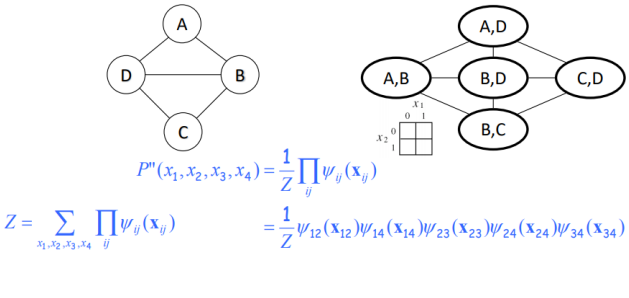
As an example, Figure 6 shows how the joint distribution for the same graph can be written in two different forms by considering maximal cliques or cliques of size 2. The representation used to factorize the joint distribution is a design choice, depending on the type of interactions that need to be captured.

7 Factor Graph

Factor graphs are a way to easily represent the factorization for both directed and undirected graphs. Consider the undirected graph shown in Figure 7a (left). Factor graphs can be easily used to graphically

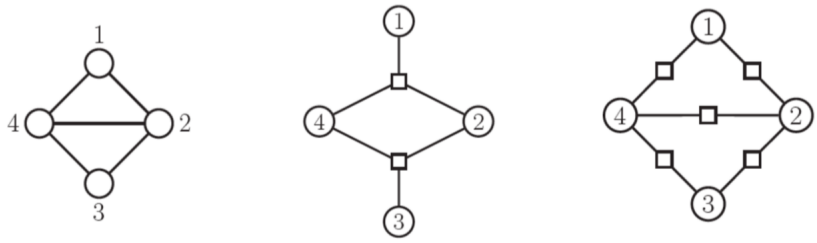


(a) Factorization into max-cliques

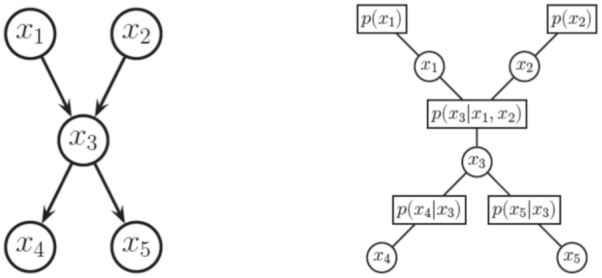


(b) Factorization into 2-cliques

Figure 6: Different factorizations based on Hammersley-Clifford Theorem



(a) Undirected Graphical Models



(b) Directed Graphical Models

Figure 7: Factor Graphs

represent the different possible factorizations of its joint distribution. Representing variables with round nodes and factors with square nodes, Figure 7a (centre) and Figure 7a (right) are factor graphs for maximal clique and pairwise factorizations respectively.

Factor graphs can also be used to represent the factorization in directed graphs as shown in Figure 7b.

8 Examples

8.1 Exponential form

Gibbs distribution is defined as

$$p(x_1, \cdots, x_n) = \frac{1}{Z} \prod_{c=1}^{C} \phi_c(X_c)$$

with $\phi_c(X_c) > 0 \ \forall c$. A canonical example of the function ϕ is the exponential function, giving the form

$$p(x_1, \cdots, x_n) = \frac{1}{Z} \exp(-\sum_c \phi_c(X_c))$$

where $H(x_1, \cdots, x_n) = \sum_c \phi_c(X_c)$ is called the free-energy of the system. When ϕ is a linear function of some features of X_c, this is called a log-linear model.

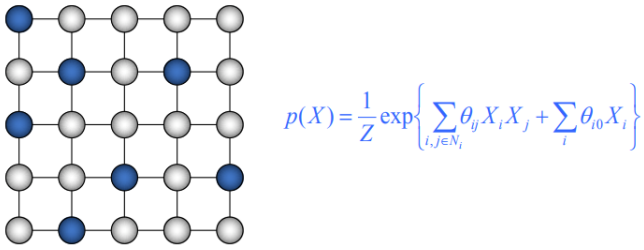


Figure 8: Ising model

8.2 Boltzmann Machines

A fully connected graph with pairwise (edge) potentials on binary-valued nodes (for $x_i \in \{-1, 1\}$ or $x_i \in \{0, 1\}$) is called a Boltzmann Machine. The joint distribution in a Boltzmann machine is given by

$$p(x_1, \cdots, x_n; \theta, \alpha) = \frac{1}{Z(\theta, \alpha)} \exp[\sum_{ij} \theta_{ij} x_i x_j + \sum_i \alpha_i x_i]$$

Hence, the overall energy has a quadratic form.

8.3 Ising models

Ising models are undirected graphs, packed in a regular topology (such as a regular packing grid), where vertices are only connected to their geometric neighbours. For example. Figure 8 shows an example of an Ising model and the form of its joint distribution. The form of the joint distribution is the same as that of a Boltzmann Machine with $\theta_{ij} \neq 0$ *iff* i, j are neighbours.

8.4 Restricted Boltzmann Machines

Restricted Boltzmann Machine (RBM) consists of a bipartite graph with hidden and visible nodes and edges between hidden and visible units only. Figure 9 shows two sets of nodes (hidden and visible) with edges

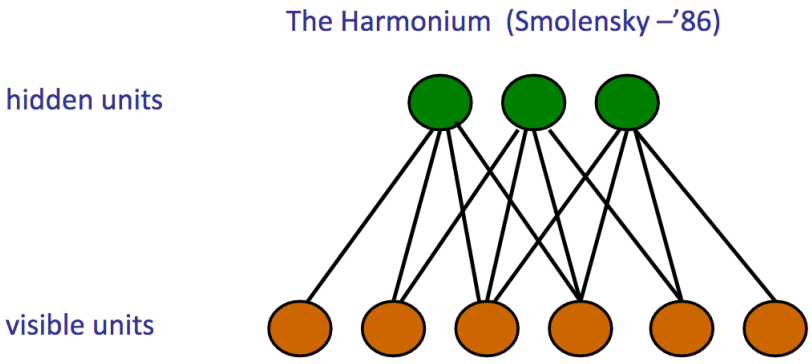


Figure 9: An Example of Restrited Boltzman Machine.

going across these set of nodes. In most general settings, these hidden nodes are used to model the latent variables in the input data distribution. The energy function for a pair of visible and hidden nodes can then be written as,

$$E(v, h|\theta) = \sum_i \theta_i \phi_i(v_i) + \sum_j \theta_j \phi_j(h_j) + \sum_{i,j} \theta_{i,j} \phi_{i,j}(v_i, h_j) \tag{4}$$

8.5 Conditional Random Fields

Rather than modelling a complex joint distribution sometimes it can be more useful to model a conditional distribution independently. Conditional Random Fields are used to model such conditional distributions $P(\mathbf{Y}|\mathbf{X})$. In the most general case, if ψ_a is the set of factors in a graph G, then the conditional distribution for a CRF is,

$$p(\mathbf{Y}|\mathbf{X}) = \frac{1}{Z(X)} \prod_{a=1}^{A} \psi_a(Y_a, X_a) \tag{5}$$

Notice, the partition function is only dependent on $\mathbf{X}$. Since, conditioning simplifies the complexity of any distribution, it is easier to estimate $Z(\mathbf{X})$ as compared to the partition function required in joint probability distribution.

9 Summary

- Undirected graphical models indicate relatedness between random variables (rather than causality).
- The graph separation criteria characterize local and global independencies in undirected graphical models.
- Clique potentials quantitatively define undirected graphical models and the Hammersley-Clifford theorem provides a way to calculate the joint distribution of such models in terms of the clique potentials.
- Factor graphs provide a convenient representation for graphical models.
- Boltzmann Machines, Ising Models, Restricted Bolzmann Machines and Conditional Random Fields are some commonly used undirected graphical models.