

CS 6347

Lecture 4

Markov Random Fields (a.k.a Undirected Graphical Models)

Bayesian Network

- A **Bayesian network** is a directed graphical model that captures independence relationships of a given probability distribution
 - Encodes **local Markov** independence assumptions that each node is independent of its non-descendants given its parents
 - Corresponds to a **factorization** of the joint distribution

$$p(x_1, \dots, x_n) = \prod_i p(x_i | x_{parents(i)})$$



Limits of Bayesian Networks

- Not all sets of independence relations can be captured by a Bayesian network
 - $-A \perp C \mid B, D$ $-B \perp D \mid A, C$
- Possible DAGs that represent these independence relationships?







Markov Random Fields (MRFs)

- A Markov random field is an undirected graphical model
 - Undirected graph G = (V, E)
 - One node for each random variable
 - One potential function or "factor" associated with cliques, C, of the graph
 - Nonnegative potential functions represent interactions and need not correspond to conditional probabilities (may not even sum to one)



Markov Random Fields (MRFs)

- A Markov random field is an undirected graphical model
 - Corresponds to a **factorization** of the joint distribution

$$p(x_1, \dots, x_n) = \frac{1}{Z} \prod_{c \in C} \psi_c(x_c)$$

$$Z = \sum_{x'_1, \dots, x'_n} \prod_{c \in C} \psi_c(x'_c)$$



Markov Random Fields (MRFs)

- A Markov random field is an undirected graphical model
 - Corresponds to a **factorization** of the joint distribution

$$p(x_1, \dots, x_n) = \frac{1}{Z} \prod_{c \in C} \psi_c(x_c)$$

$$Z = \sum_{x'_1, \dots, x'_n} \prod_{c \in C} \psi_c(x'_c)$$

Normalizing constant, *Z*, often called the partition function



An Example



- $p(x_A, x_B, x_C) = \frac{1}{Z} \psi_{AB}(x_A, x_B) \psi_{BC}(x_B, x_C) \psi_{AC}(x_A, x_C)$
- Each potential function can be specified as a table as before $x_A = 0$ $x_A = 1$

$$\psi_{AB}(x_A, x_B) = \begin{array}{c|c} x_B = 0 & 1 \\ x_B = 1 & 1 \end{array}$$



1

0

The Ising Model

- Mathematical model of ferromagnets
- Each atom has an associated spin that is biased by both its neighbors in the material and an external magnetic field
 - Spins can be either +1 or -1
 - Edge potentials capture the local interactions
 - Singleton potentials capture the external field

$$p(x_V) = \frac{1}{Z} \exp(\sum_{i \in V} h_i x_i + \sum_{(i,j) \in E} J_{ij} x_i x_j)$$





Independence Assertions

- Instead of d-separation, we need only consider separation:
 - If $X \subseteq V$ is graph separated from $Y \subseteq V$ by $Z \subseteq V$, (i.e., all paths from X to Y go through Z) then $X \perp Y \mid Z$
 - What independence assertions follow from this MRF?





Independence Assertions



$$p(x_A, x_B, x_C) = \frac{1}{Z} \psi_{AB}(x_A, x_B) \psi_{BC}(x_B, x_C)$$

- How does separation imply independence?
- Show that $A \perp C \mid B$



Independence Assertions

- In particular, each variable is independent of all of its non-neighbors given its neighbors
 - All paths leaving a single variable must pass through some neighbor
- If the joint probability distribution, *p*, factorizes with respect to the graph *G*, then *G* is an I-map for *p*
- If *G* is an I-map of a positive distribution *p*, then *p* factorizes with respect to the graph *G*
 - Hamersley-Clifford Theorem



BNs vs. MRFs

Property	Bayesian Networks	Markov Random Fields
Factorization	Conditional Distributions	Potential Functions
Distribution	Product of Conditional Distributions	Normalized Product of Potentials
Cycles	Not Allowed	Allowed
Partition Function	1	Potentially NP-hard to Compute
Independence Test	d-Separation	Graph Separation



Moralization

- Every Bayesian network can be converted into an MRF with some possible loss of independence information
 - Remove the direction of all arrows in the network
 - If A and B are parents of C in the Bayesian network, we add an edge between A and B in the MRF
- This procedure is called "moralization" because it "marries" the parents of every node





Moralization

- Every Bayesian network can be converted into an MRF with some possible loss of independence information
 - Remove the direction of all arrows in the network
 - If A and B are parents of C in the Bayesian network, we add an edge between A and B in the MRF
- This procedure is called "moralization" because it "marries" the parents of every node





Moralization



• What independence information is lost?



Factorizations

- Many factorizations over the same graph may represent the same joint distribution
 - Some are better than others (e.g., they more compactly represent the distribution)
 - Simply looking at the graph is not enough to understand which specific factorization is being assumed





Factor Graphs

- Factor graphs are used to explicitly represent a given factorization over a given graph
 - Not a different model, but rather different way to visualize an MRF
 - Undirected bipartite graph with two types of nodes: variable nodes (circles) and factor nodes (squares)
 - Factor nodes are connected to the variable nodes on which they depend



Factor Graphs

$$p(x_A, x_B, x_C) = \frac{1}{Z} \psi_{AB}(x_A, x_B) \psi_{BC}(x_B, x_C) \ \psi_{AC}(x_A, x_C)$$





Conditional Random Fields (CRFs)

- Undirected graphical models that represent conditional probability distributions p(Y | X)
 - Potentials can depend on both X and Y, typically only the observed variables are considered in the model

$$p(Y \mid X) = \frac{1}{Z(x)} \prod_{c \in C} \psi_c(x_c, y_c)$$
$$Z(x) = \sum_{y'} \prod_{c \in C} \psi_c(x_c, y'_c)$$



Log-Linear Models

 CRFs often assume that the potentials are log-linear functions

$$\psi_c(x_c, y_c) = \exp(w \cdot f_c(x_c, y_c))$$

where f_c is referred to as a feature vector and w is some vector of feature weights

- The feature weights are typically learned from data
- CRFs don't require us to model the full joint distribution (which may not be possible anyhow)



Conditional Random Fields (CRFs)

- Binary image segmentation
 - Label the pixels of an image as belonging to the foreground or background
 - +/- correspond to foreground/background
 - Interaction between neighboring pixels in the image depends on how similar the pixels are
 - Similar pixels should preference having the same spin (i.e., being in the same part of the image)



Conditional Random Fields (CRFs)

- Binary image segmentation
 - This can be modeled as a CRF where the image information (e.g., pixel colors) is observed, but the segmentation is unobserved
 - Because the model is conditional, we don't need to describe the joint probability distribution of (natural) images and their foreground/background segmentations
 - CRFs will be particularly important when we want to learn graphical models



Low Density Parity Check Codes

 Want to send a message across a noisy channel in which bits can be flipped with some probability – use error correcting codes



- ψ_A, ψ_B, ψ_C are all parity check constraints: they equal one if their input contains an even number of ones and zero otherwise
- $\phi_i(x_i, y_i) = p(y_i|x_i)$, the probability that the *i*th bit was flipped during transmission

Low Density Parity Check Codes



- The parity check constraints enforce that the y's can only be one of a few possible codewords: 000000, 001011, 010101, 011110, 100110, 101101, 110011, 111000
- Decoding the message that was sent is equivalent to computing the most likely codeword under the joint probability distribution



Low Density Parity Check Codes



- Most likely codeword is given by MAP inference $\arg \max_{y} p(y|x)$
- Do we need to compute the partition function for MAP inference?

