ECE 6504: Advanced Topics in Machine Learning
Probabilistic Graphical Models and Large-Scale Learning

Topics
- Markov Random Fields: Representation
  - Conditional Random Fields
  - Log-Linear Models

Readings: KF 4.1-3; Barber 4.1-2

Dhruv Batra
Virginia Tech
Administrativia

• No class
  – Next week (Tue, Thu)

• Project Proposal
  – Due: Mar 12, Mar 5, 11:59pm
  – <=2pages, NIPS format

• HW2
  – Out later today
  – Due: Mar 12, 11:59pm
  – Implementation: Variable Elimination in BNs
Recap of Last Time
Markov Nets

• Set of random variables

• Undirected graph
  – Encodes independence assumptions

• Unnormalized Factor Tables

• Joint distribution:
  – Product of Factors
Pairwise MRFs

• Pairwise Factors
  – A function of 2 variables
    • Often unary terms are also allowed (although strictly speaking unnecessary)

  – On board
Pairwise MRF: Example
Computing probabilities in Markov networks vs BNs

- In a BN, can compute prob. of an instantiation by multiplying CPTs.

- In a Markov network, can only compute ratio of probabilities directly.
Normalization for computing probabilities

- To compute actual probabilities, must compute normalization constant (also called partition function)

- Computing partition function is hard! Must sum over all possible assignments
Nearest-Neighbor Grids

Low Level Vision
- Image denoising
- Stereo
- Optical flow
- Shape from shading
- Superresolution
- Segmentation

\[ y_s \rightarrow \text{unobserved or hidden variable} \]
\[ x_s \rightarrow \text{local observation} \]
General Gibbs Distribution

• Arbitrary Factors

• “Induced” MRF Graph
Factorization in Markov networks

- Given an undirected graph $H$ over variables $X = \{X_1, \ldots, X_n\}$

- A distribution $P$ factorizes over $H$ if there exist
  - subsets of variables $D_1 \subseteq X, \ldots, D_m \subseteq X$, such that $D_i$ are fully connected in $H$
  - non-negative potentials (or factors) $\phi_1(D_1), \ldots, \phi_m(D_m)$
    - also known as clique potentials
  - such that
    \[
    P(X_1, \ldots, X_n) = \frac{1}{Z} \prod_{i=1}^{m} \phi_i(D_i)
    \]

- Also called Markov random field $H$, or Gibbs distribution over $H$
MRFs

- Given a graph H, are factors unique?
Active Trails and Separation

• A path $X_1$ – … – $X_k$ is **active** when set of variables $\mathbf{Z}$ are observed
  – if none of $X_i \in \{X_1, \ldots, X_k\}$ are observed (are part of $\mathbf{Z}$)

• Variables $\mathbf{X}$ are **separated** from $\mathbf{Y}$ given $\mathbf{Z}$ in graph
  – If no active path between any $X \in \mathbf{X}$ and any $Y \in \mathbf{Y}$ given $\mathbf{Z}$
Markov networks representation Theorem 1

If joint probability distribution \( P \):
\[
P(X_1, \ldots, X_n) = \frac{1}{Z} \prod_{i=1}^{m} \phi_i(D_i)
\]

Then
\( H \) is an I-map for \( P \)

- If
  - you can write distribution as a normalized product of factors
- Then
  - Can read independencies from graph
What about the other direction for Markov networks?

If $H$ is an I-map for $P$, then

The joint probability distribution $P$:

$$P(X_1, \ldots, X_n) = \frac{1}{Z} \prod_{i=1}^{m} \phi_i(D_i)$$

- Counter-example: $X_1, \ldots, X_4$ are binary, and only eight assignments have positive probability:

  - \begin{align*}
  (0,0,0,0) & \quad (1,0,0,0) & \quad (1,1,0,0) & \quad (1,1,1,0) \\
  (0,0,0,1) & \quad (0,0,1,1) & \quad (0,1,1,1) & \quad (1,1,1,1)
  \end{align*}

- For example, $X_1 \perp X_3 | X_2, X_4$:
  - E.g., $P(X_1=0 | X_2=0, X_4=0)$

- But distribution doesn’t factorize!!
If joint probability distribution \( P \):

\[
P(X_1, \ldots, X_n) = \frac{1}{Z} \prod_{i=1}^{m} \phi_i(D_i)
\]

Then \( H \) is an I-map for \( P \)

If \( H \) is an I-map for \( P \) and \( P \) is a positive distribution

Then joint probability distribution \( P \):

\[
P(X_1, \ldots, X_n) = \frac{1}{Z} \prod_{i=1}^{m} \phi_i(D_i)
\]
= Markov Blanket of variable $x_8$ – Parents, children and parents of children
Independence Assumptions in MNs

- **Separation** defines global independencies

- **Pairwise Markov Independence:**
  - Pairs of non-adjacent variables A,B are independent given all others

- **Markov Blanket:**
  - Variable A independent of rest given its neighbors
P-map

• Perfect map

• \( G \) is a **P-map** for \( P \) if
  – \( I(P) = I(G) \)

• Question: Does every distribution \( P \) have P-map?
Structure in cliques

- Possible potentials for this graph:
Factor graphs

• Bipartite graph:
  – variable nodes (ovals) for $X_1, \ldots, X_n$
  – factor nodes (squares) for $\phi_1, \ldots, \phi_m$
  – edge $X_i - \phi_j$ if $X_i \in \text{Scope}[\phi_j]$

• Very useful for approximate inference
  – Make factor dependency explicit
Types of Graphical Models

Directed

Undirected

Factor

Undirected
Factor Graphs show Fine-grained Factorization

\[ p(x) = \frac{1}{\mathcal{Z}} \prod_{f \in \mathcal{F}} \psi_f(x_f) \]
Plan for today

- Undirected Graphical Models: Representation
  - Conditional Random Fields
  - Log-Linear Models

- Undirected Graphical Models: Inference
  - Variable Elimination
Conditional Random Fields

• What’s the difference between Naïve Bayes & Logistic Regression?
Nearest-Neighbor Grids

Low Level Vision
- Image denoising
- Stereo
- Optical flow
- Shape from shading
- Superresolution
- Segmentation

\[ y_s \rightarrow \text{unobserved or hidden variable} \]
\[ x_s \rightarrow \text{local observation} \]
Lazy Snapping

Siggraph 2004

Yin Li
Chi-Keung Tang
Hong Kong University of Science and Technology

Jian Sun
Heung-Yeung Shum
Microsoft Research Asia
Logarithmic representation

- Standard model:
  \[ P(X_1, \ldots, X_n) = \frac{1}{Z} \prod_{i=1}^{m} \phi_i(D_i) \]

- Log representation of potential (assuming positive potential):
  - also called the energy function

- Log representation of Markov net:
Log-linear Markov network (most common representation)

- **Feature (or Sufficient Statistic)** is some function $f[D]$ for some subset of variables $D$
  - e.g., indicator function

- **Log-linear model** over a Markov network $H$:
  - a set of features $f_1[D_1],..., f_k[D_k]$
    - each $D_i$ is a subset of a clique in $H$
    - two $f$'s can be over the same variables
  - a set of weights $w_1,...,w_k$
    - usually learned from data

  $P(X_1,...,X_n) = \frac{1}{Z} \exp \left[ \sum_{i=1}^{k} w_i f_i(D_i) \right]$
CRFs

Felzenszwalb, Huttenlocher, IJCV '04
CRFs

Node Features

Edge Features

\[ f_i(Y_i) \]

\[ f_j(Y_j) \]

\[ f_{ij}(Y_i, Y_j) \]
Node Feature -- Color

Superpixel

Feature Extraction
Step 1

- Rmean
- Gmean
- Bmean
- Hmean
- Smean
- Vmean
- 5-dim hist on H
- 3-dim hist on S

Holm, Efros, Hebert, IJCV 2007
Node Feature – Color Clustering

Feature Space

K-means/X-means, Pelleg, Moore
Auton Lab implementation

\{ (\mu_1, \Sigma_1), (\mu_2, \Sigma_2), \ldots, (\mu_N, \Sigma_N) \}
Node Feature -- Color

Superpixel

Feature Extraction
Step 1

\[ \begin{align*}
R_{\text{mean}} & \\
G_{\text{mean}} & \\
B_{\text{mean}} & \\
H_{\text{mean}} & \\
S_{\text{mean}} & \\
V_{\text{mean}} & \\
5\text{-dim} & \text{hist on } H \\
3\text{-dim} & \text{hist on } S
\end{align*} \]

Feature Extraction
Step 2

\[ \begin{align*}
\Pr(\text{Cluster } i \mid \phi_i) & \\
\vdots & \\
\Pr(\text{Cluster } N \mid \phi_i) &
\end{align*} \]

Hoiem, Efros, Hebert, IJCV 2007
Conditional Random Fields
Summary of types of Markov nets

• Pairwise Markov networks
  – very common
  – potentials over nodes and edges

• General MRFs

• Factor graphs
  – explicit representation of factors
    • you know exactly what factors you have
  – very useful for approximate inference

• Log-linear models
  – log representation of potentials
  – linear coefficients learned from data
  – most common for learning MNs