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

Topics
- Markov Random Fields: Inference
  - Exact: VE
  - Exact+Approximate: BP

Readings: Barber 5

Dhruv Batra
Virginia Tech
Administrativia

- HW3
  - Out 2 days ago
  - Due: Apr 4, 11:55pm
  - Implementation: Loopy Belief Propagation in MRFs
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

\[ \phi_1[A, B] \]
\[
\begin{array}{c|c|c}
 a^0 & b^0 & 30 \\
 a^0 & b^1 & 5 \\
 a^1 & b^0 & 1 \\
 a^1 & b^1 & 10 \\
\end{array}
\] 
\[ \phi_2[B, C] \]
\[
\begin{array}{c|c|c}
 b^0 & c^0 & 100 \\
 b^0 & c^1 & 1 \\
 b^1 & c^0 & 1 \\
 b^1 & c^1 & 100 \\
\end{array}
\] 
\[ \phi_3[C, D] \]
\[
\begin{array}{c|c|c|c}
 c^0 & d^0 & 1 & d^0 a^0 100 \\
 c^0 & d^1 & 100 & d^0 a^1 1 \\
 c^1 & d^0 & 100 & d^1 a^0 1 \\
 c^1 & d^1 & 1 & d^1 a^1 100 \\
\end{array}
\] 
\[ \phi_4[D, A] \]
Normalization for computing probabilities

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

<table>
<thead>
<tr>
<th>Assignment</th>
<th>Unnormalized</th>
<th>Normalized</th>
</tr>
</thead>
<tbody>
<tr>
<td>$a^0$</td>
<td>300000</td>
<td>0.04</td>
</tr>
<tr>
<td>$b^0$</td>
<td>300000</td>
<td>0.04</td>
</tr>
<tr>
<td>$c^0$</td>
<td>300000</td>
<td>0.04</td>
</tr>
<tr>
<td>$d^0$</td>
<td>300000</td>
<td>0.04</td>
</tr>
<tr>
<td>$a^0$</td>
<td>30</td>
<td>4.1 $\cdot$ 10^{-6}</td>
</tr>
<tr>
<td>$b^0$</td>
<td>500</td>
<td>6.9 $\cdot$ 10^{-5}</td>
</tr>
<tr>
<td>$c^0$</td>
<td>500</td>
<td>6.9 $\cdot$ 10^{-5}</td>
</tr>
<tr>
<td>$d^0$</td>
<td>500</td>
<td>6.9 $\cdot$ 10^{-5}</td>
</tr>
<tr>
<td>$a^1$</td>
<td>50000000</td>
<td>0.69</td>
</tr>
<tr>
<td>$b^1$</td>
<td>500</td>
<td>6.9 $\cdot$ 10^{-5}</td>
</tr>
<tr>
<td>$c^1$</td>
<td>100</td>
<td>1.4 $\cdot$ 10^{-5}</td>
</tr>
<tr>
<td>$d^1$</td>
<td>100</td>
<td>1.4 $\cdot$ 10^{-5}</td>
</tr>
<tr>
<td>$a^0$</td>
<td>10000000</td>
<td>0.14</td>
</tr>
<tr>
<td>$b^0$</td>
<td>10</td>
<td>1.4 $\cdot$ 10^{-6}</td>
</tr>
<tr>
<td>$c^0$</td>
<td>10</td>
<td>1.4 $\cdot$ 10^{-6}</td>
</tr>
<tr>
<td>$d^0$</td>
<td>10</td>
<td>1.4 $\cdot$ 10^{-6}</td>
</tr>
<tr>
<td>$a^1$</td>
<td>10000000</td>
<td>0.014</td>
</tr>
<tr>
<td>$b^1$</td>
<td>10000000</td>
<td>0.014</td>
</tr>
<tr>
<td>$c^1$</td>
<td>10000000</td>
<td>0.014</td>
</tr>
</tbody>
</table>

• 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

\[ y_s \quad \text{unobserved or hidden variable} \]

\[ x_s \quad \text{local observation} \]
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 $H$
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
- Factor
- Undirected

(C) Dhruv Batra

Slide Credit: Erik Sudderth
Plan for today

• MRF Inference
  – Exact Inference
    • Variable Elimination
  – Exact+Approximate Inference
    • (General) Belief Propagation
    • Cluster Graphs
      – Family Preserving Property
      – Running Intersection Property
    • Message-Passing
  – Approximate Inference
    • Bethe Cluster Graph
    • Loopy BP
  – Exact Inference
    • Junction Tree
    • BP on Junction Trees
Marginal Inference Example

- Evidence: $E=e$ (e.g. $N=t$)
- Query variables of interest $Y$

- Conditional Probability: $P(Y | E=e)$
  - $P(F | N=t)$
Variable Elimination algorithm

• Given a BN and a query \( P(Y|e) \approx P(Y,e) \)
  – “Instantiate Evidence”

• Choose an ordering on variables, e.g., \( X_1, \ldots, X_n \)

• For \( i = 1 \) to \( n \), If \( X_i \notin \{Y,E\} \)
  – Collect factors \( f_1,\ldots,f_k \) that include \( X_i \)
  – Generate a new factor by eliminating \( X_i \) from these factors
    \[
    g = \sum_{X_i} \prod_{j=1}^{k} f_j
    \]
  – Variable \( X_i \) has been eliminated!

• Normalize \( P(Y,e) \) to obtain \( P(Y|e) \)
VE for MRF

- Exactly the same algorithm works!
  - Factors are no longer CPTs
  - But VE doesn’t care
Example

• Chain MRF

\[ X_5 \times X_3 \times X_4 \times X_2 \times X_1 \]

Compute:

\[ P(X_1 \mid X_5 = x_5) \]

• VE steps on board
Example

• Chain MRF

Compute:

\[ P(X_i \mid X_5 = x_5) \]
\[ \forall i \in \{1, 2, 3, 4\} \]

Variable elimination for every \( i \), what’s the complexity?

Can we do better by caching intermediate results?

Yes! via Junction-Trees
But let’s look at BP first
New Topic: Belief Propagation
What is BP?

• Technique invented by Judea Pearl in 1982
  – Initially to compute marginals in BNs

• Later generalized
  – to MRFs, Factor Graphs
  – To MAP inference; to Marginal-MAP inference

• Lots of analysis
  – Under some cases EXACT
    • Tree graphs
      – In this setting, BP equivalent to VE on Junction-Trees
    • Submodular potentials
      – In this setting, BP equivalent to Graph-Cuts! [Tarlow et al. UAI11]
  – In general Approximate
Message Passing

• Variables/Factors “talk” to each other via messages:

“I (variable $X_3$) think that you (variable $X_2$):
   - belong to state 1 with confidence 0.4
   - belong to state 2 with confidence 10
   - belong to state 3 with confidence 1.5”
Overview of BP

• Pick a graph to pass messages on
  – Cluster Graph

• Pick an ordering of edges
  – Round-robin
  – Leaves-Root-Leaves on a tree
  – Asynchronous

• Till convergence or exhaustion:
  – Pass messages on edges

• At vertices on graph compute *pseudo-marginals*
• **Cluster Graph:**
  For set of factors $F$
  - Undirected graph
  - Each node $i$ associated with a cluster $C_i$
  - Each edge $i - j$ is associated with a separator set of variables $S_{ij} \subseteq C_i \cap C_j$
Generalized BP

- **Initialization:**
  - Assign each factor $\phi$ to a cluster $\alpha(\phi)$, $\text{Scope}[\phi] \subseteq \mathbf{C}_{\alpha(\phi)}$
  - Initialize cluster: $\psi^0_i(S_i) \propto \prod_{\phi: \alpha(\phi) = i} \phi$
  - Initialize messages: $\delta_{j \rightarrow i} = 1$

- **While not converged, send messages:**
  $$\delta_{i \rightarrow j}(S_{ij}) \propto \sum_{C_i \leftarrow S_{ij}} \psi^0_i(C_i) \prod_{k \in \mathcal{N}(i) - j} \delta_{k \rightarrow i}(S_{ik})$$

- **Belief:**
  - On Board

(C) Dhruv Batra
Properties of Cluster Graphs

- **Family preserving**: For set of factors $F$
  - for each factor $f_j \in F$, $\exists$ node $i$ such that $\text{scope}[f_j] \subseteq C_i$
Properties of Cluster Graphs

- Running intersection property (RIP)
  - If \( X \in \mathbf{C}_i \) and \( X \in \mathbf{C}_j \) then
    \( \exists \text{ one and only one path from } \mathbf{C}_i \text{ to } \mathbf{C}_j \)
    where \( X \in S_{uv} \) for every edge \((u,v)\) in the path
Two cluster graph satisfying RIP with different edge sets
Overview of BP

• Pick a graph to pass messages on
  – Cluster Graph

• Pick an ordering of edges
  – Round-robin
  – Leaves-Root-Leaves on a tree
  – Asynchronous

• Till convergence or exhaustion:
  – Pass messages on edges

• At vertices on graph compute *pseudo-marginals*
Cluster Graph for Loopy BP

• Bethe Cluster Graph
  – Set of Clusters = Factors $F \cup \{X_i\}$
  – Sometimes also called “Running BP on Factor Graphs”
  – Example on board

• Does the Bethe Cluster Graph satisfy properties?
Loopy BP in Factor graphs

- From node $i$ to factor $j$:
  - $F(i)$ factors whose scope includes $X_i$
  \[
  \delta_{i \rightarrow j}(X_i) \propto \prod_{k \in F(i) - j} \delta_{k \rightarrow i}(X_i)
  \]

- From factor $j$ to node $i$:
  - $\text{Scope}[\phi_j] = Y \cup \{X_i\}$
  \[
  \delta_{j \rightarrow i}(X_i) \propto \sum_y \phi_j(X_i, y) \prod_{X_k \in \text{Scope}[\phi_j] - X_i} \delta_{k \rightarrow j}(x_k)
  \]

- Belief:
  - Node:
  - Factor:
Loopy BP on Pairwise Markov Nets

\[
\delta_{i\rightarrow j}(y_j) = \sum_{y_i} \phi_i(y_i) \phi_{ij}(y_i, y_j) \prod_{k \in \mathcal{N}(i) - j} \delta_{k\rightarrow i}(y_i)
\]