IE 598 SO - Inference in Graphical Models

- Tue-Thu 2:00pm - 3:15pm, TB206
- Sewoong Oh (swoh@illinois.edu)
- http://web.engr.illinois.edu/~swoh/courses/IE598/
- Homeworaks
  - One mid-term quizz
  - Final project
Topics

- Provide a unifying framework for inference tasks in complex systems

- **Graphical models**: random variables sit on vertices

- **Probability distributions**: that can be ‘decomposed’ or ‘factorized’

- **Inference tasks**: draw a conclusion based on the distribution

- **Applications**: Images, error-correcting codes, machine learning, etc.
Example: Medical Decision

Quick Medical Reference—Decision-Theoretic network

- $d_i \in \{0, 1\}$: diseases
- $f_j \in \{0, 1\}$: symptoms of findings
- Graphical model: Bayesian network (e.g. $\mathbb{P}(f_2 = 1|d_1 = 0, d_3 = 1)$)
- Inference task: Given symptoms (e.g. $f = 01010010$), what disease is likely ($\arg \max_d \mathbb{P}(d|f)$)?
Example: Navigation

Navigating Spacecrafts (e.g. lunar landing, guiding shuttles)

- Linear system as Gaussian graphical models (e.g. $P(x_2|x_1, u_1)$)
- Inference task: Given noisy sensor readings, what is the current state? (compute $P(x_4|y_1, \ldots, y_4)$)
- Kalman filtering
Example: Image processing
Can computers generate/classify handwritten letters/numbers?
[R.Salakhutdinov, G. Hinton, 2009 AISTATS]
Example: Image processing

Can computers generate/classify handwritten letters/numbers?
[R. Salakhutdinov, G. Hinton, 2009 AISTATS]

10,000 Training data

28 × 28 Pixel images

Reconstruction by sampling
Example: Image processing

- Pairwise Markov random fields (deep Boltzmann machines)
- Gibbs sampling
Example: Communication

$2^k$ messages

<table>
<thead>
<tr>
<th>Codebook</th>
<th>Channel</th>
<th>Decoder</th>
</tr>
</thead>
</table>

$X^n_1 \rightarrow Y^n_1 \rightarrow$ recover

Error-correcting codes (e.g. Low-Density Parity Check codes)

- Factor graphs
- (loopy) Belief propagation
- Inference task: Received $y^n_1$, what $x^n_1$ is most likely?
  \[ \text{arg max}_x \mathbb{P}(x|y) \]
### General theme

Probability distribution over $X = (X_1, X_2, \ldots, X_n)$ given observations $Y = (Y_1, \ldots, Y_m)$

$$
\mu_y(x) = \mathbb{P}_{X_1, \ldots, X_n | Y_1, \ldots, Y_m}(x_1, x_2, \ldots, x_n | y_1, y_2, \ldots, y_m)
$$

from a set $x_i \in X$ and $y_j \in Y$, typically $|X| < \infty$

- Finding the most probable realization
  $$
  \hat{x} \in \arg \max_{x \in X^n} \mu_y(x)
  $$

- Calculate marginals
  $$
  \mu_y(x_1) = \sum_{x_2, \ldots, x_n} \mu_y(x)
  $$

- Sampling

---

**Key challenge: $n \gg 1$**

Computational complexity is $O(|X|^n)$ and there is no efficient method for general distributions
Structure
Suppose the variables are independent

\[ \mu_y(x) = \mu_1(x_1)\mu_2(x_2) \cdots \mu_n(x_n) \]

then, computational complexity is only \(|X| \cdot n\)

- Finding the most probable realization

\[ \hat{x}_i \in \arg \max_{x_i \in X} \mu_i(x_i) \]

- Calculate marginals

\[ \mu_y(x_1) = \mu_1(x_1) \]

- Sampling: \(X_1, X_2, \ldots, X_n\) independently

When the probability distribution factorizes, we can achieve huge computational gains
Graphical models

- Undirected pairwise graphical models
- Factor graphs
- Bayesian networks
Topics include

- Representing inference tasks using graphical models
- General and powerful framework for **efficient** inference
- Belief propagation
- Hidden Markov models, Kalman filtering
- Plenty of math: convex analysis, random processes, Markov chains, etc.