Warm up ...

- Given a directed graphical model

- Prove

\[ p(X) = p(X_1, X_2, \ldots, X_N) = \prod_{i=1}^{N} p(X_i \mid X_{\pi(i)}) \]
Wake up!

- Given a HMM: \( p(Z_i | X_i) \) and \( p(X_t | X_{t-1}) \)

\[ x_1 \ x_2 \ x_3 \ \cdots \ x_t \]

\[ z_1 \ z_2 \ z_3 \ z_{t-1} \ z_t \]

- Prove:

\[
p(X_t | Z_1, \ldots, Z_t) \propto p(Z_t | X_t) \int p(X_t | X_{t-1}) p(X_{t-1} | Z_1, \ldots, Z_{t-1}) dX_{t-1}
\]

Interesting problems (i)

- Super-resolution
  - Given a low-res image, can you recover a high-res image?
  - To make it feasible, you may be given a set of training pairs of low-res and high-res images.
Interesting problems (ii)

- Capturing body motion from video
  - Given a video sequence, you need to estimate the articulated body motion from the video.

Outline

- Markov network representation
- Three issues
  - Inferring hidden variables
  - Calculating evidence
  - Learning Markov networks
- Inference
  - Exact inference
  - Approximate inference
- Applications
  - Super-resolution
  - Capturing body articulation
Representations

- Undirected graphical models
  - $G(V, E)$, $v \in V$ represents a random variable $x$, $e \in E$ indicate the relation of two random variables;
  - Each r.v. $x_i$ has a prior $\psi_i(x_i)$;
  - Each undirected link (between $x_i$ and $x_j$) is associated with a potential function $\psi_{ij}(x_i, x_j)$;
  - $v \in V, N(v)$ means the neighborhood of $v$;
- We can combine undirected and directed graphical models.

Markov Network

$$p(X) = \frac{1}{Z_c} \prod_{(i,j) \in C^2} \psi_{ij}(x_i, x_j) \prod_{i \in C^1} \psi_i(x_i)$$

$$p(Z|X) = \prod_{i=1}^{n} p_i(z_i|x_i).$$
Three issues

- Inferring hidden variables
  - Given a Markov network $\Lambda$
  - To infer the hidden variable of the evidence $p(X | Z, \Lambda)$

- Calculating evidence
  - Given a Markov network $\Lambda$
  - To calculate the likelihood of the evidence $p(Z | \Lambda)$

- Learning Markov networks
  - Given a set of training data $\{Z_1, \ldots, Z_n\}$
  - To find the best model $\Lambda^* = \arg \max_{\Lambda} \prod_{k=1}^n p(Z_k | \Lambda)$

To be covered

- The inference problem is the key among the three
  - Why?
  - Let’s discuss …

- We will focus on inference in this lecture
Given the above model, with

- \( p(z_i | x_i) \), \( \psi_i(x_i) \)
- \( \psi_{12}(x_1, x_2), \psi_{23}(x_2, x_3), \psi_{34}(x_3, x_4), \psi_{45}(x_4, x_5), \psi_{46}(x_4, x_6) \)

To solve

- \( p(x_3 | z_1, z_2, z_3, z_4, z_5, z_6) \)
What do we learn from the Ex.?

\[ p(x_i \mid Z) \propto \psi_i(x_i) \times p(z_i \mid x_i) \times \prod_{k \in N(i)} M_{ki}(x_i) \]

posterior \propto \text{local prior} \times \text{local likelihood} \times \text{messages received from neighbors}
Sum-Product Algorithm

\[ M_{ki}(x_i) = \int_{x_k} p(z_k | x_k) \psi_k(x_k) \psi_{ki}(x_k, x_i) \prod_{j \in N(k) \setminus i} M_{jk}(x_j) \, dx_k \]

Properties

- Properties of “belief propagation”
  - When the network is not loopy, the algorithm is exact;
  - When the network is loopy, the inference can be done by iterating the message passing procedures;
  - In this case, the inference is not exact,
  - and it is not clear how good is the approximation.
Application: Super-resolution

- Bill Freeman (ICCV’99)
  - X: high-frequency data
  - Z: mid-frequency data
  - “scene”: the high-frequency components
  - “evidence”: the mid-freq of the low-res image input
  - The “scene” is decomposed by a network of scene patches; and each scene patch is associated with an evidence patch
  - Learning $p(z|x)$ from training samples

Results (Freeman)

- Low-res input
- Iteration of B.P. to obtain high-frequency data
- Recovered high-res image
- Actual high-res
Variational Inference

- We want to infer $p(x_i|Z)$
- It is difficult, because of the networked structure.
- We perform probabilistic variational analysis
- The idea is to find an optimal approximation $q^*(X)$ of $p(X|Z)$, such that the Kullback-Leibler (KL) divergence of these two distribution is minimized:

$$q^*(X) = \arg\min_q KL(q(X||p(X|Z)))$$

$$= \arg\min_q \int_X q(X) \log \frac{q(X)}{p(X|Z)}$$
Mean Field Theory

- When we choose a full factorization variation:
  \[ q(\mathbf{X}) = \prod_{i}^{M} q_i(\mathbf{x}_i) \]

- We end up with a very interesting result: a set of fixed point equations:
  \[ q_i(\mathbf{x}_i) = \frac{1}{Z_i} p_i(\mathbf{z}_i | \mathbf{x}_i) \psi_i(\mathbf{x}_i) M_i(\mathbf{x}_i), \text{ where} \]
  \[ M_i(\mathbf{x}_i) = \exp \left\{ \sum_{k \in N(i)} \int_{\mathbf{x}_k} q_k(\mathbf{x}_k) \log \psi_{ik}(\mathbf{x}_i, \mathbf{x}_k) \right\}, \]

- This is very similar to the Mean Field theory in statistical physics.

Computational Paradigm

- Three factors affect the posterior of a node:
  - Local prior
  - Neighborhood prior
  - Local likelihood
Application: body tracking

- Capturing body articulation (Wu et al.)
  - X: the motion of body parts
  - Z: image features (e.g., edges)
  - The motion of body parts are constrained and correlated
- Specifying $\psi(x_i, x_j)$
- Modeling $p(z | x)$ explicitly

Dynamic Markov Network

$$q_{i,t}(x_{i,t}) = \frac{1}{Z_{i,t}} p_i(z_{i,t} | x_{i,t}) \times \int p(x_{i,t} | x_{i,t-1}) q_{i,t-1}(x_i)$$

$$\times \exp \left\{ \sum_{k \in \mathcal{N}(i)} \int_{x_{k,t}} \frac{q_{k,t}(x_k)}{q_{i,t}(x_{i,t})} \log \psi_{ik}(x_{i,t}, x_{k,t}) \right\}$$
We are still investigating the convergence property of mean field algorithms. But our empirical study shows that it converges very fast, generally, less than five iterations.

Application: body tracking
Application: body tracking

Summary

- Markov networks contain undirected edges in the graph to model the non-casual correlation
- Inference is the key of analyzing Markov networks
  - Exact inference
  - Approximate inference
- Two inference techniques were introduced here:
  - Belief propagation
  - Variational inference
- Two applications
  - Super-resolution
  - Capturing articulated body motion