Course Overview

1. Probabilistic graphical models
   We study joint probability distribution over n variables:
   $$\mu(x_1, \ldots, x_n)$$
   (a) Factorization according to graph
      i. Markov random field (MRF)
         A. Undirected graph
         $$\mu(x) = \frac{1}{Z}\prod_{c\in\mathcal{C}}\psi_c(x_c)$$
      ii. Factor graph (FG)
         A. Directed acyclic graph
         $$\mu(x) = \prod_{i\in V}\mu(x_i|x_{\pi(i)})$$
      iii. Bayesian network (BN)
         A. Bipartite undirected graph
         $$\mu(x) = \frac{1}{Z}\prod_{a\in F}\psi_a(x_{\partial a})$$
   (b) Conditional independencies implied from a graphical model
      i. Markov random field (MRF)
         A. pairwise, local, global Markov property
         B. graph separation
         C. Markov blanket: neighbors of x_i
         D. Hammersley-Clifford: if μ is positive and satisfy conditional independencies according to G then μ factorizes over G
      ii. Factor graph (FG)
         A. graph separation
      iii. Bayesian network (BN)
         A. local, ordered, global Markov property
         B. d-separation, Bayes' ball algorithm
   (c) Representing $\mu(x)$ using graphical models
      i. I-map, P-map
      ii. conversion between graphical models
         A. from BN to MRF: moralization
         B. from MRF to BN, from FG to pairwise MRF
      iii. relationship between graphical models

2. Inference
   (a) Inference task
      i. Marginals

$$\mu(x_A) = \sum_{x_{V \setminus A}} \mu(x)$$

      ii. MAP estimation

$$\arg\max_x \quad \mu(x)$$

      iii. Partition function

$$Z = \sum_x \prod_{c \in \mathcal{C}} \psi_c(x_c)$$

      iv. Sampling
   (b) Belief Propagation (BP)
      i. computing marginals
         A. elimination algorithm
         B. sum-product algorithm
         C. relationship between computing marginals, computing partition function, and sampling
      ii. computing MAP estimate
         A. elimination algorithm
         B. max-product algorithm
      iii. Gaussian graphical models
         A. Gaussian belief propagation
         B. Gaussian HMM and Kalman filtering
      iv. analysis of belief propagation
         A. computation tree
         B. correctness of Gaussian BP
         C. (correctness of BP with a single loop)
         D. (correctness of BP with log-concave $\mu(x)$)
         E. (planar graph with binary variables)
         F. (associative potentials)
         G. (convergence of max-product)
         H. density evolution: asymptotic analysis of BP on random graphs
   (c) Variational methods
      i. Gibbs free energy

$$\log Z \quad = \quad \max_b \ \mathbb{E}_b[\psi(x)] + H(b)$$

      ii. naive mean field approximation

$$\log Z \quad \geq \quad \max_{b : b(x) = \prod_i b_i(x_i)} \ \mathbb{E}_b[\psi(x)] + H(b)$$

      iii. Bethe approximation

$$\log Z \quad \approx \quad \max_{b \in \mathrm{LOC(G)}} \ \mathbb{E}_b[\psi(x)] + H_{\mathrm{Bethe}}(b)$$

iv. Region-based approximation
v. Tree-based approximation
vi. (variational MAP using Linear Programming)

(d) Monte Carlo Markov Chain (MCMC)
i. Metropolis-Hastings algorithm
A. Gibbs sampling
ii. bounding mixing time via spectral analysis
iii. bounding mixing time via coupling

3. Learning

(a) Parameter estimation
i. maximum likelihood estimation

$$\max_{\theta} \sum_{(i,j)\in E} \psi_{ij}(x_i, x_j) - \log Z(\theta)$$

ii. convex in θ
iii. gradient requires inference (efficient on trees)
iv. moment matching, iterative proportional fitting
v. sample complexity of Hammersely-Clifford construction

(b) Structure learning
i. Bayesian approach: MAP estimate = ML + regularization
ii. local independence tests

(c) Restricted Boltzmann machines

4. Computational challenge

(a) computing marginal is #P-complete
(b) MAP estimation is NP-complete
(c) approximate inference with bounded error is also NP-hard