Chapter 21: Variational inference

Thanks to machine-learning algorithms,
the robot apocalypse was short-lived.

Idea

$$p(x|\mathcal{D}) \triangleq p^*(\mathbf{x}) = \frac{1}{Z}\tilde{p}(x_1, x_2, x_3, \ldots)$$

$$q(\mathbf{x})$$

$$\text{minimize}_q \quad \text{difference}(p^*, q)$$

$$\mathbb{KL}\left(q||p^*\right) = \sum_{\mathbf{x}} q(\mathbf{x}) \log \frac{q(\mathbf{x})}{p^*(\mathbf{x})}$$

Variational inference objective function

$$J(q) \triangleq \mathbb{KL}\left(q||\tilde{p}\right)$$

$$\begin{aligned}
J(q) &= \sum_{\mathbf{x}} q(\mathbf{x}) \log \frac{q(\mathbf{x})}{\tilde{p}(\mathbf{x})} \\
&= \sum_{\mathbf{x}} q(\mathbf{x}) \log \frac{q(\mathbf{x})}{Z p^*(\mathbf{x})} \\
&= \sum_{\mathbf{x}} q(\mathbf{x}) \log \frac{q(\mathbf{x})}{p^*(\mathbf{x})} - \log Z \\
&= \mathbb{KL}\left(q||p^*\right) - \log Z
\end{aligned}$$

Exercise

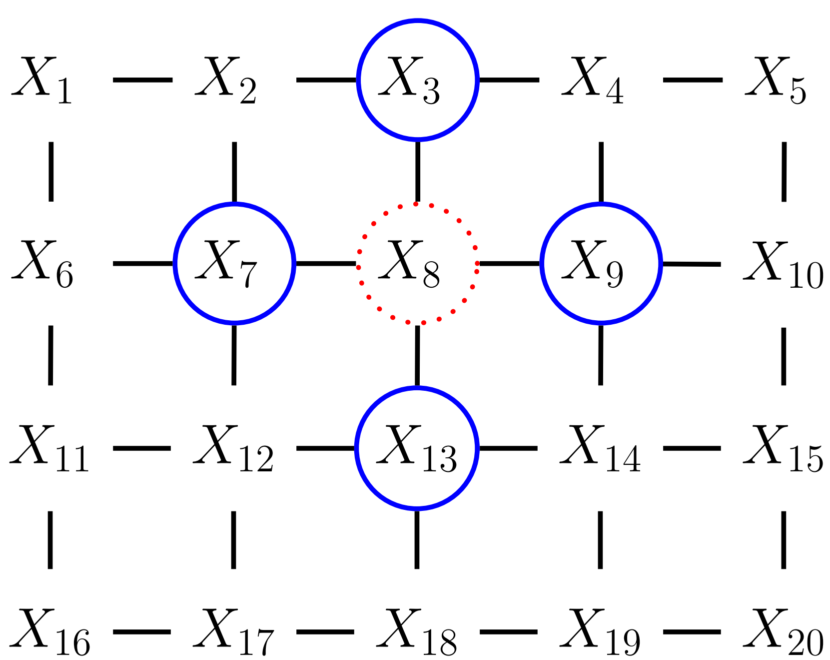
Show how to apply Gibbs sampling to a univariate mixture of Gaussians.

Exercise

Suppose $x_i \in \{-1, +1\}$

and $\phi(x_s, x_t) = \exp(J x_s x_t)$

Derive an expression for $p(x_t | x_{-t})$



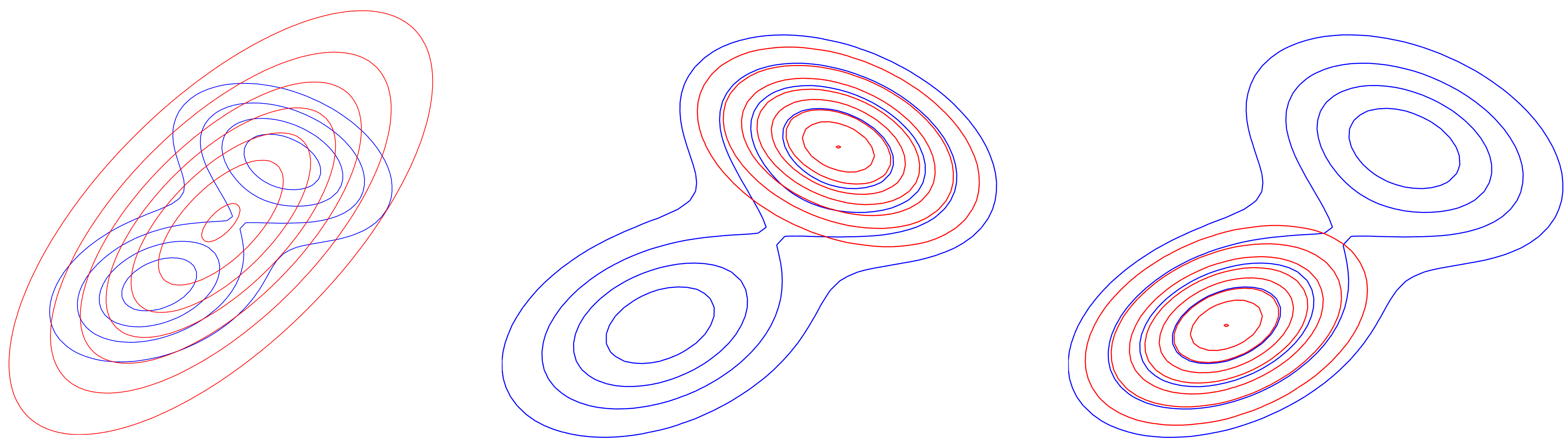
$$p(x_t | \mathbf{x}_{-t}, \boldsymbol{\theta}) \propto \prod_{s \in \mathrm{nbr}(t)} \psi_{st}(x_s, x_t)$$

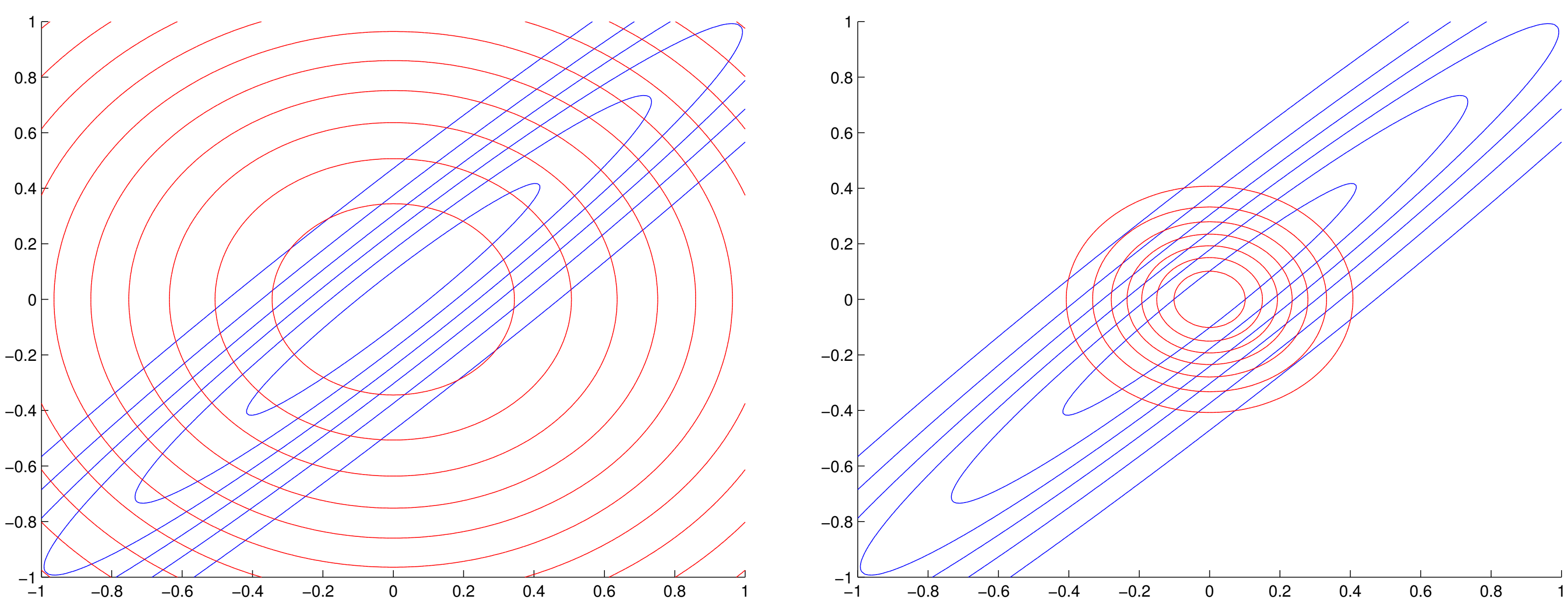
Why KL?

$$\mathbb{KL}\left(q \| p\right) = \sum_{\mathbf{x}} q(\mathbf{x}) \ln \frac{q(\mathbf{x})}{p(\mathbf{x})}$$

Forwards vs reverse KL

Forward: $\mathbb{KL}\left(p||q\right) = \sum_{\mathbf{x}} p(\mathbf{x}) \ln \frac{p(\mathbf{x})}{q(\mathbf{x})}$ Reverse: $\mathbb{KL}\left(q||p\right) = \sum_{\mathbf{x}} q(\mathbf{x}) \ln \frac{q(\mathbf{x})}{p(\mathbf{x})}$





Interpreting the variational objective

$$J(q) = \mathbb{KL}\left(q||p^*\right) - \log Z \geq -\log Z = -\log p(\mathcal{D})$$

$$\begin{aligned} J(q) &= \mathbb{E}_q\left[\log q(\mathbf{x}) - \log p(\mathbf{x})p(\mathcal{D}|\mathbf{x})\right] \\ &= \mathbb{E}_q\left[\log q(\mathbf{x}) - \log p(\mathbf{x}) - \log p(\mathcal{D}|\mathbf{x})\right] \\ &= \mathbb{E}_q\left[-\log p(\mathcal{D}|\mathbf{x})\right] + \mathbb{KL}\left(q(\mathbf{x})||p(\mathbf{x})\right) \end{aligned}$$

Mean field inference

$$p^*(\mathbf{x}) = p^*(x_1, x_2, x_3)$$

$$q(\mathbf{x}) = \prod_i q_i(x_i) = q_1(x_1)q_2(x_2)q_3(x_3)$$

$$\min_{q_1,\ldots,q_D} \mathbb{KL}(q||p)$$

$$\log q_j(\mathbf{x}_j) = \mathbb{E}_{-q_j}[\log \tilde{p}(\mathbf{x})] + \text{const}$$

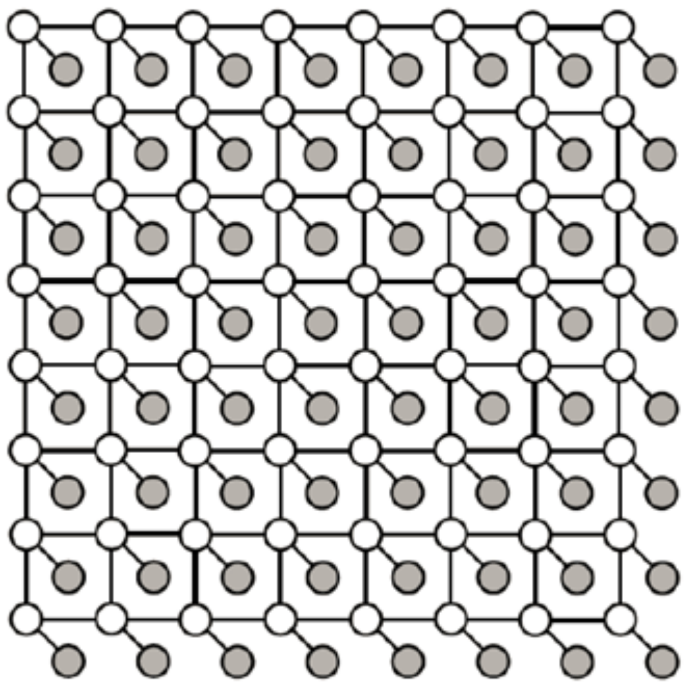
$$\mathbb{E}_{-q_2}[f(\mathbf{x})] = \sum_{x_1}\sum_{x_3} q(x_1)q_3(x_3)f(x_1, x_2, x_3)$$

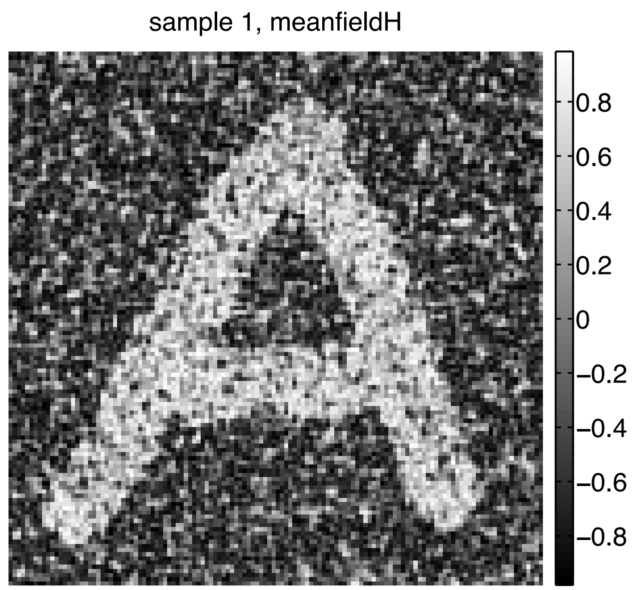
Mean field update

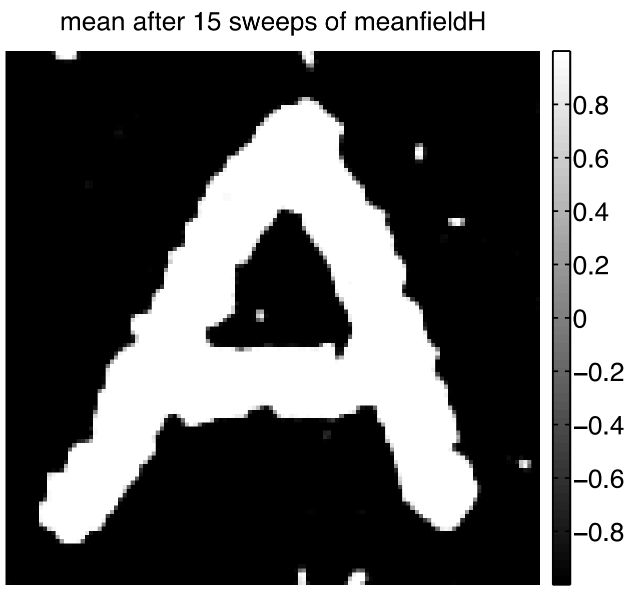
$$
\begin{aligned}
-J(q) &= \sum_{\mathbf{x}} q(\mathbf{x}) \log \frac{\tilde{p}(\mathbf{x})}{q(\mathbf{x})} \\
&= \sum_{\mathbf{x}} \prod_i q_i(\mathbf{x}_i) \left[\log \tilde{p}(\mathbf{x}) - \sum_k \log q_k(\mathbf{x}_k) \right] \\
&= \sum_{\mathbf{x}_j} \sum_{\mathbf{x}_{-j}} q_j(\mathbf{x}_j) \prod_{i \neq j} q_i(\mathbf{x}_i) \left[\log \tilde{p}(\mathbf{x}) - \sum_k \log q_k(\mathbf{x}_k) \right] \\
&= \sum_{\mathbf{x}_j} q_j(\mathbf{x}_j) \sum_{\mathbf{x}_{-j}} \prod_{i \neq j} q_i(\mathbf{x}_i) \log \tilde{p}(\mathbf{x}) \\
&\quad - \sum_{\mathbf{x}_j} q_j(\mathbf{x}_j) \sum_{\mathbf{x}_{-j}} \prod_{i \neq j} q_i(\mathbf{x}_i) \left[\sum_{k \neq j} \log q_k(\mathbf{x}_k) + \log q_j(\mathbf{x}_j) \right] \\
&= \sum_{\mathbf{x}_j} q_j(\mathbf{x}_j) \log f_j(\mathbf{x}_j) - \sum_{\mathbf{x}_j} q_j(\mathbf{x}_j) \log q_j(\mathbf{x}_j) + \text{const}
\end{aligned}
$$

$$
\log f_j(\mathbf{x}_j) \triangleq \sum_{\mathbf{x}_{-j}} \prod_{i \neq j} q_i(\mathbf{x}_i) \log \tilde{p}(\mathbf{x}) = \mathbb{E}_{-q_j} [\log \tilde{p}(\mathbf{x})]
$$

Example: Mean field inference for the Ising model







$$p(\mathbf{x}) = \frac{1}{Z} \exp\left(\sum_{i=1}^{D} \sum_{j \in \text{nbr}(i)} W_{ij} x_i x_j \right)$$

$$q(\mathbf{x}) = \prod_i q(x_i, \mu_i)$$

$$p(\mathbf{y}|\mathbf{x}) = \prod_i p(\mathbf{y}_i | x_i)$$

Example: Mean field inference for the Ising model

$$p(\mathbf{x}) = \frac{1}{Z_0}\exp(-E_0(\mathbf{x})) \qquad E_0(\mathbf{x}) = -\sum_{i=1}^{D}\sum_{j\in \mathrm{nbr}_i} W_{ij}x_i x_j$$

$$p(\mathbf{y}|\mathbf{x}) = \prod_i p(\mathbf{y}_i|x_i) = \exp(\sum_i -L_i(x_i))$$

$$p(\mathbf{x}|\mathbf{y}) = \frac{1}{Z}\exp(-E(\mathbf{x})) \qquad E(\mathbf{x}) = E_0(\mathbf{x}) - \sum_i L_i(x_i)$$

$$\log \tilde{p}(\mathbf{x}) = x_i \sum_{j\in \mathrm{nbr}_i} W_{ij}x_j + L_i(x_i) + \mathrm{const}$$

$$q_i(x_i) \propto \exp\left(x_i \sum_{j\in \mathrm{nbr}_i} W_{ij}\mu_j + L_i(x_i)\right)$$

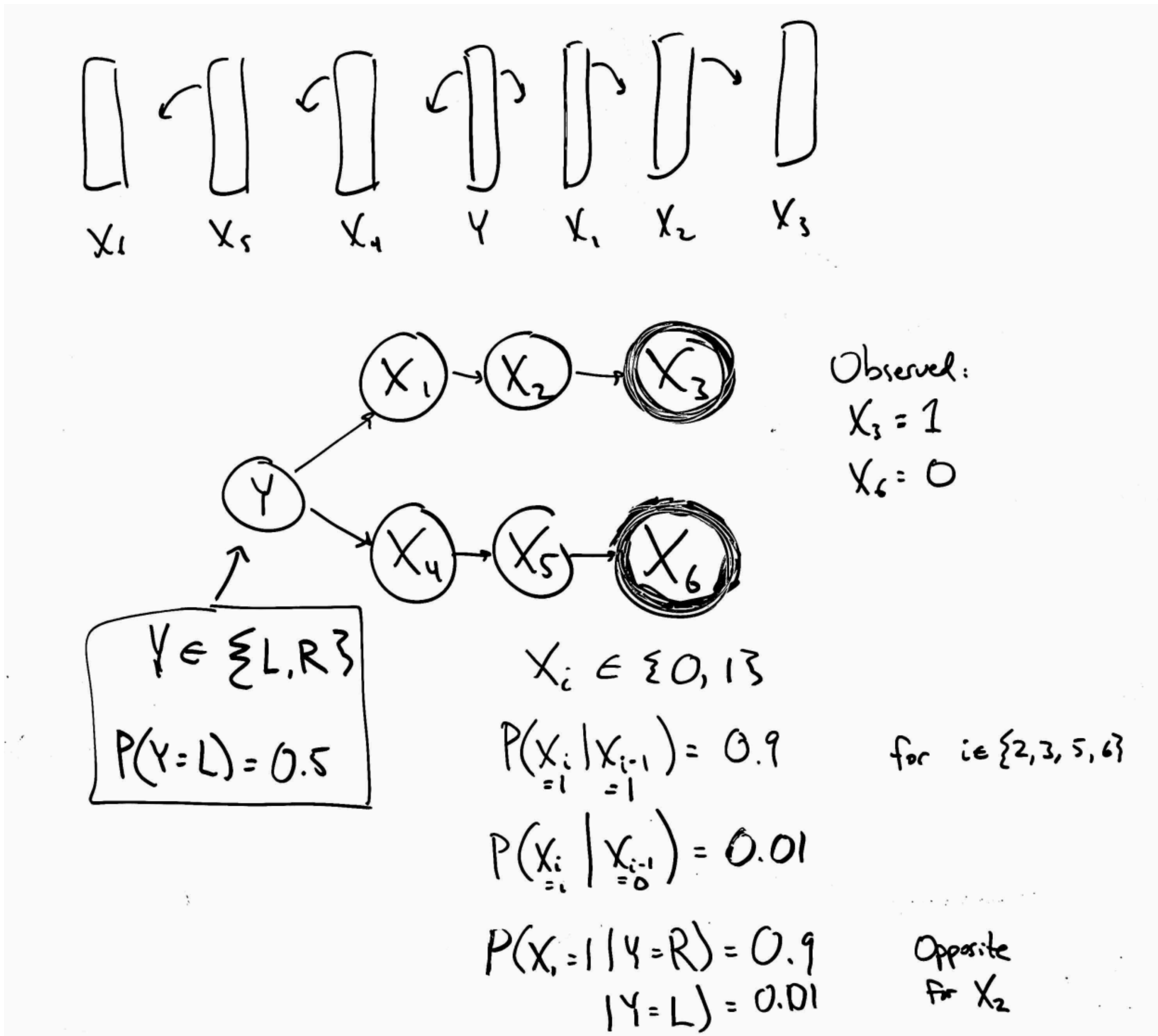
Example: Mean field inference for the Ising model

$$
\begin{aligned}
q_i(x_i = 1) &= \frac{e^{m_i + L_i^+}}{e^{m_i + L_i^+} + e^{-m_i + L_i^-}} = \frac{1}{1 + e^{-2m_i + L_i^- - L_i^+}} = \mathrm{sigm}(2a_i) \\
a_i &\triangleq m_i + 0.5(L_i^+ - L_i^-)
\end{aligned}
$$

$$
\begin{aligned}
\mu_i &= \mathbb{E}_{q_i}[x_i] = q_i(x_i = +1) \cdot (+1) + q_i(x_i = -1) \cdot (-1) \\
&= \frac{1}{1 + e^{-2a_i}} - \frac{1}{1 + e^{2a_i}} = \frac{e^{a_i}}{e^{a_i} + e^{-a_i}} - \frac{e^{-a_i}}{e^{-a_i} + e^{a_i}} = \tanh(a_i)
\end{aligned}
$$

Exercise

Using belief propagation, compute $P(Y \mid X_3=1, X_6=0)$.



$\psi_{x_6}(x_5 = 0) = P(X_6 = 0 \mid X_5 = 0) = 0.99$

$\quad\;\;(\quad = 1) = P(X_6 = 0 \mid X_5 = 1) = 0.01$

$$m_{X_5 \to X_4}(X_4 = 0) = \sum_{X_5 = X} \psi_{x_6}(x)\; \psi_{X_4, X_5}(X_4 = 0, x)$$

0.99

$0 \to \quad = \quad 0.99 \cdot P(X_4 = 0 \mid X_5 = 0) \quad = \quad 0.9811$

$1 \to \quad + \; 0.01 \cdot 0.1$

$\boxed{0}$

$(X_4 = 1) = 0.99 \cdot 0.1 + 0.01 \cdot 0.9 = 0.108$

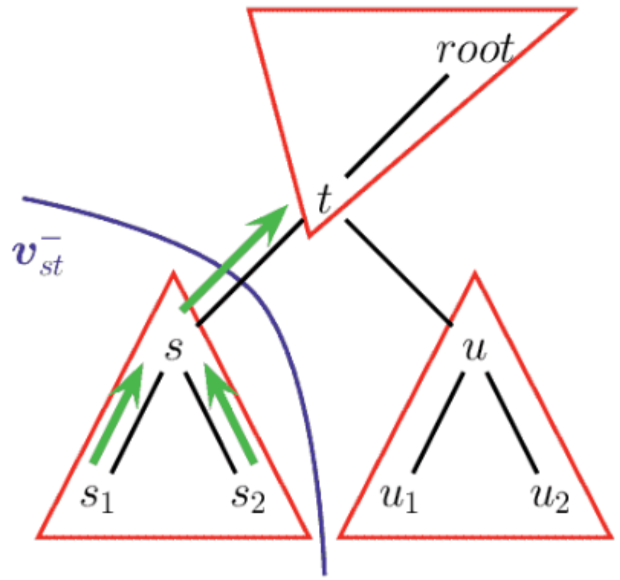
$m_{X_4 \to Y}(Y = 0) = 0.9811 \cdot 0.99 \qquad = \quad 0.972$

$\qquad\qquad\qquad 0.108 \cdot 0.01$

$(Y = 1) = 0.9811 \cdot 0.1 \qquad = \quad 0.195$

$\qquad\qquad 0.108 \cdot 0.9$

Review: belief propagation



Loopy belief propagation for pairwise MRFs

Algorithm 22.1: Loopy belief propagation for a pairwise MRF

```
Input: node potentials ψ_s(x_s), edge potentials ψ_st(x_s, x_t);
Initialize messages m_{s→t}(x_t) = 1 for all edges s − t;
Initialize beliefs bel_s(x_s) = 1 for all nodes s;
repeat
    Send message on each edge
    m_{s→t}(x_t) = Σ_{x_s} ( ψ_s(x_s) ψ_st(x_s, x_t) Π_{u∈nbr_s\t} m_{u→s}(x_s) );
    Update belief of each node bel_s(x_s) ∝ ψ_s(x_s) Π_{t∈nbr_s} m_{t→s}(x_s);
until beliefs don't change significantly;
Return marginal beliefs bel_s(x_s);
```

Input: node potentials $\psi_s(x_s)$, edge potentials $\psi_{st}(x_s, x_t)$;

Initialize messages $m_{s \to t}(x_t) = 1$ for all edges $s - t$;

Initialize beliefs $\mathrm{bel}_s(x_s) = 1$ for all nodes s;

repeat

Send message on each edge

$$m_{s\to t}(x_t) = \sum_{x_s} \left(\psi_s(x_s)\psi_{st}(x_s, x_t) \prod_{u \in \mathrm{nbr}_s \setminus t} m_{u \to s}(x_s) \right);$$

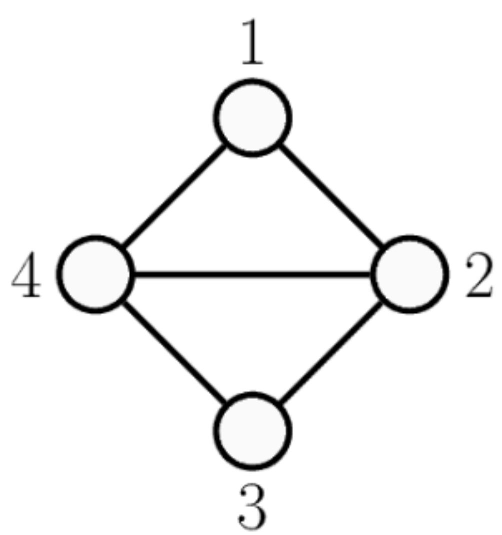
Update belief of each node $\mathrm{bel}_s(x_s) \propto \psi_s(x_s) \prod_{t \in \mathrm{nbr}_s} m_{t \to s}(x_s)$;

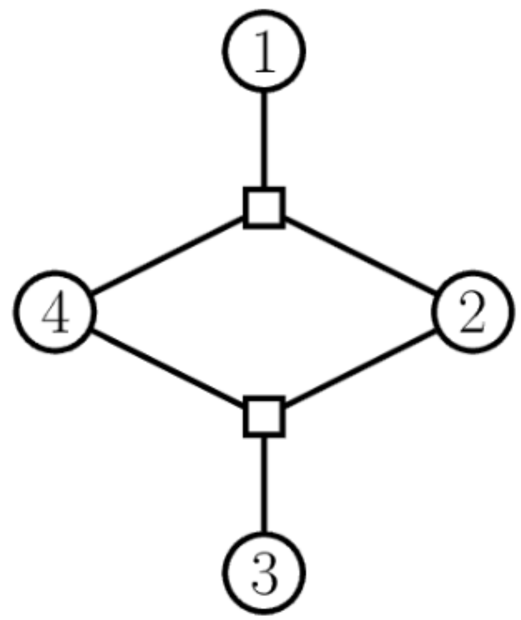
until *beliefs don't change significantly*;

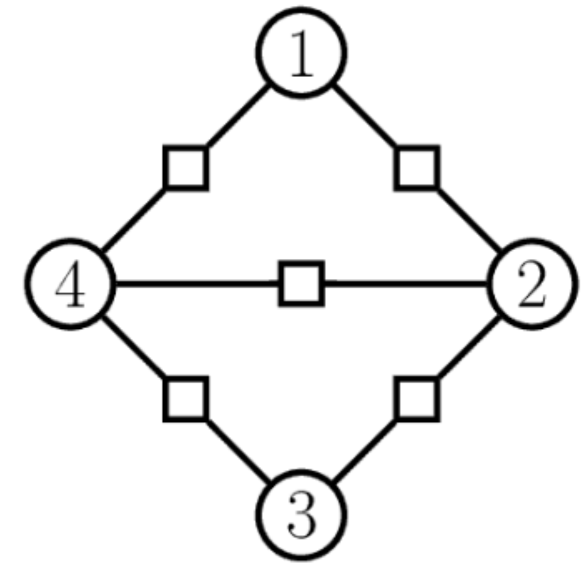
Return marginal beliefs $\mathrm{bel}_s(x_s)$;

Loopy belief propagation for pairwise MRFs

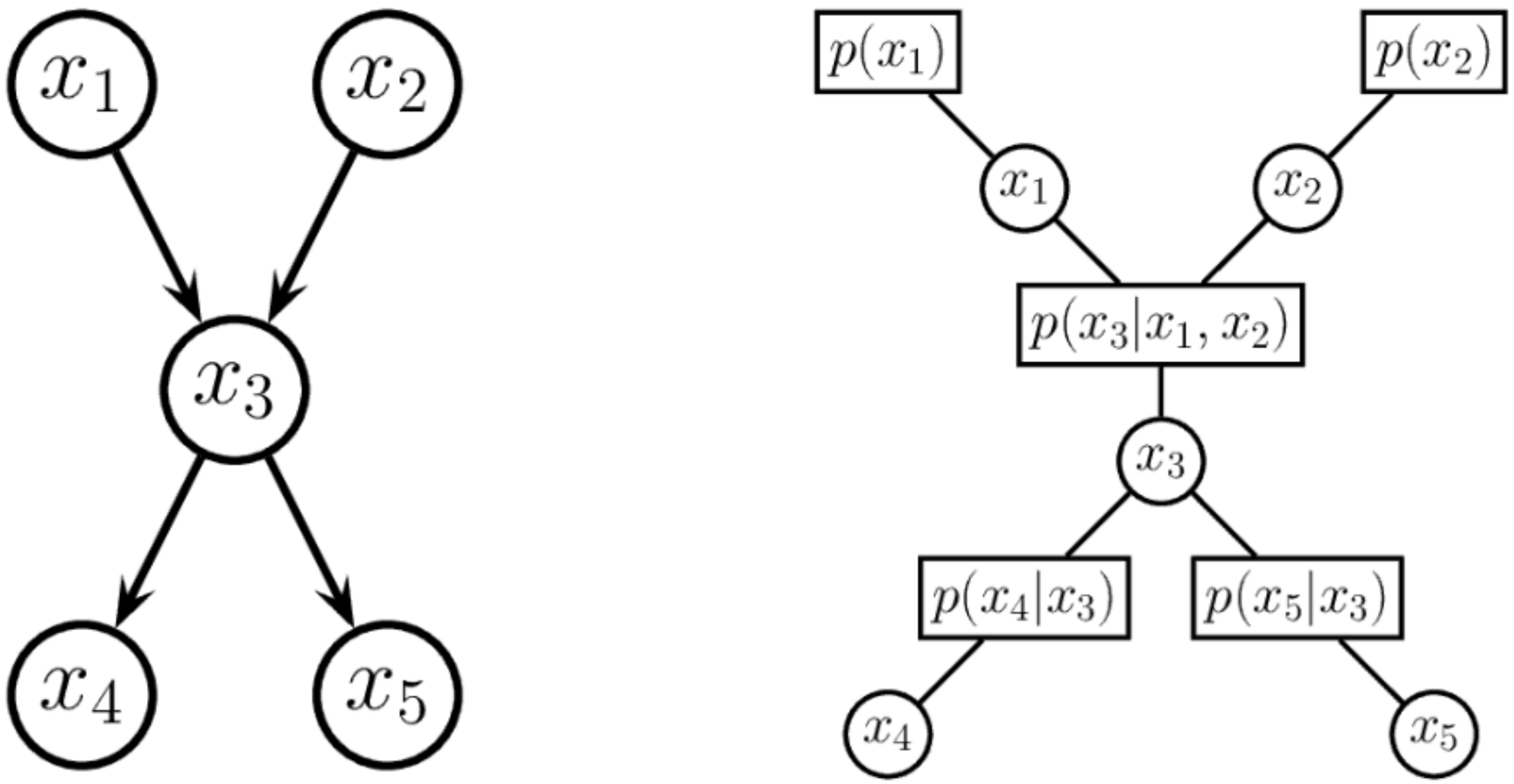
Factor graphs







Factor graphs

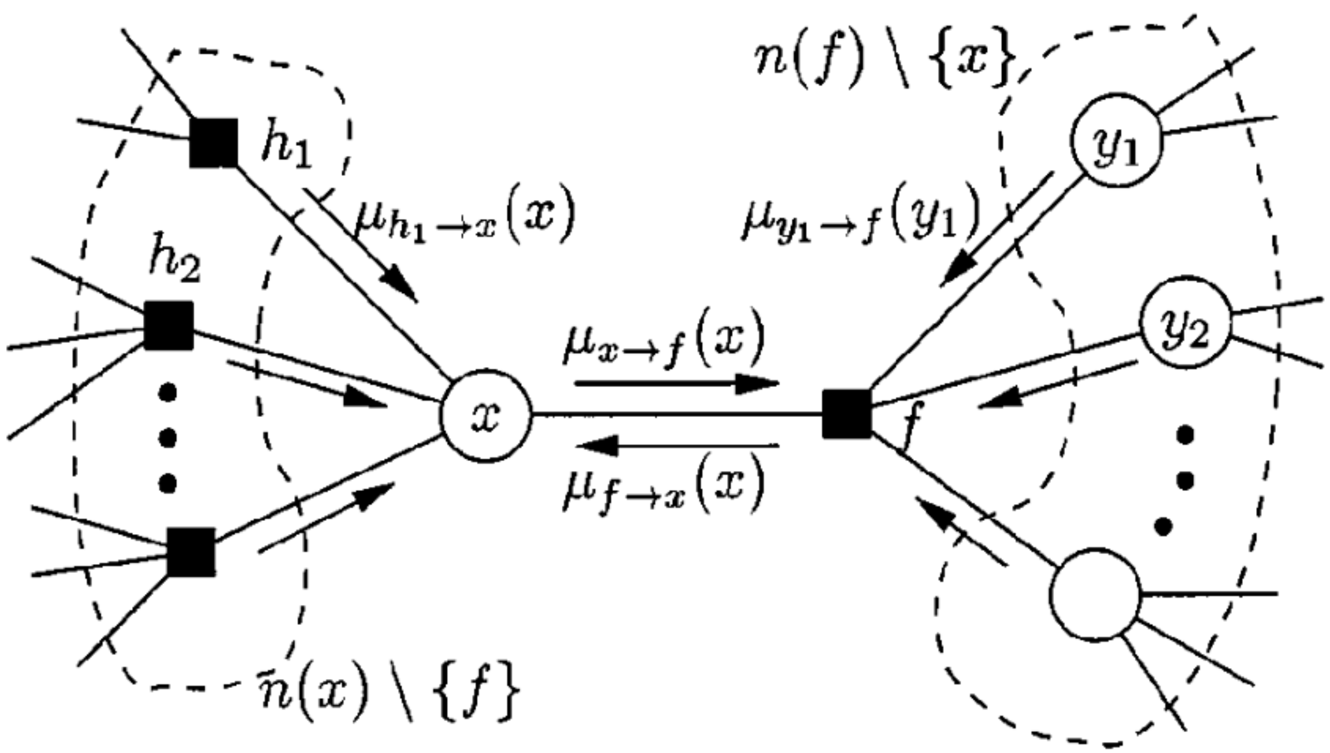


Loopy belief propagation for factor graphs

$$m_{x \to f}(x) = \prod_{h \in \mathrm{nbr}(x) \setminus \{f\}} m_{h \to x}(x)$$

$$m_{f \to x}(x) = \sum_{\mathbf{y}} f(x, \mathbf{y}) \prod_{y \in \mathrm{nbr}(f) \setminus \{x\}} m_{y \to f}(y)$$

$$\mathrm{bel}(x) \propto \prod_{f \in \mathrm{nbr}(x)} m_{f \to x}(x)$$



Exercise: Convert the following probabilistic models to factor graphs

Loopy belief propagation: practical notes