

Applied Machine Learning

Boltzmann Machines

Boltzmann Machines

- Boltzmann Machines
- Variational Inference Approach
- Mean Field Inference Algorithm

Boltzmann Machines

- U : Set of nodes H (hidden) and X (Observed)

- Binary $\{-1, 1\}$

- Edges

- (U_i, U_j) : Coupling. $N(i)$: Neighbors of i

- θ : Weights: Intensity of coupling

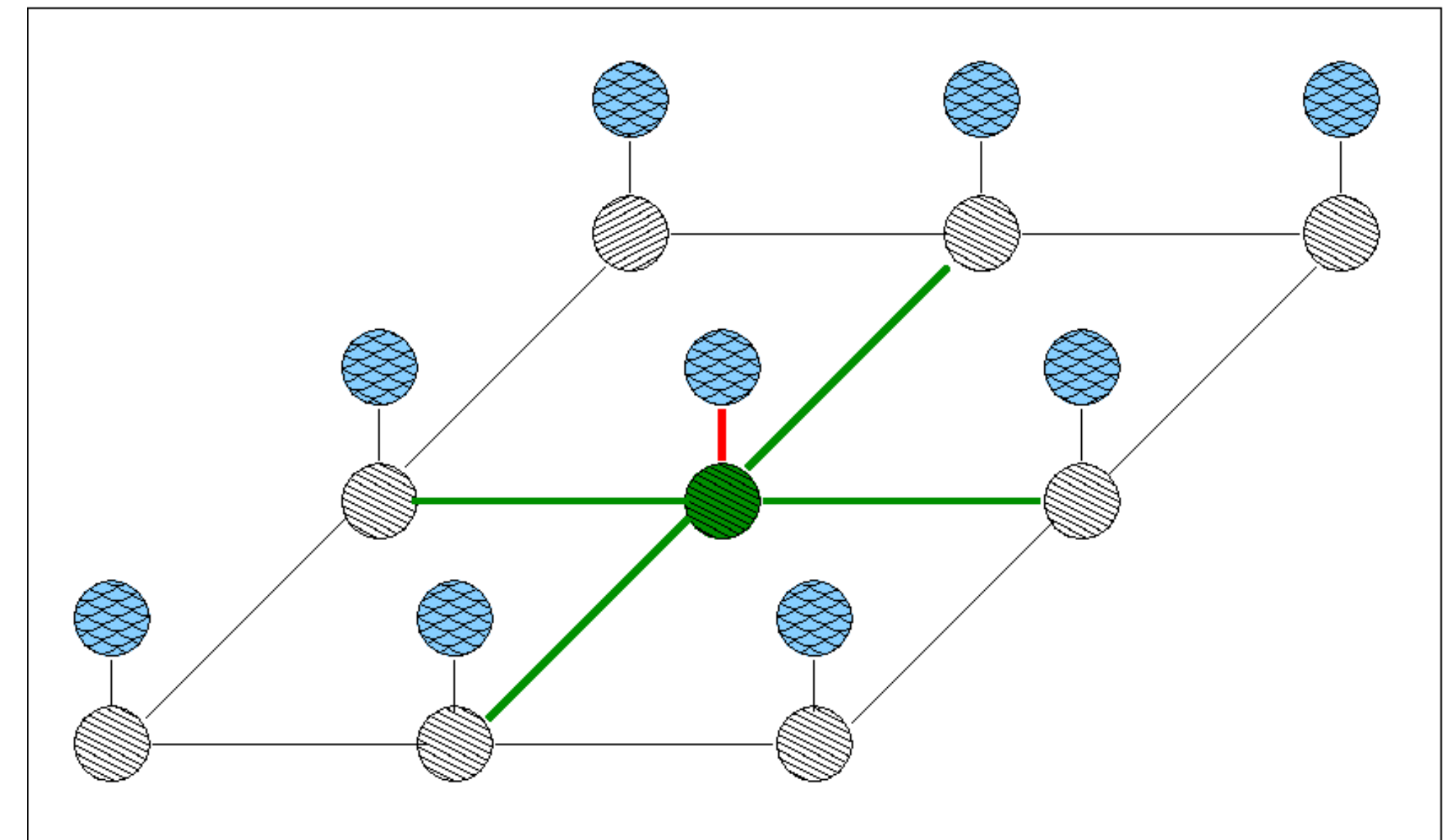
- Energy:

$$E(U|\theta) = - \sum_i \sum_{j \in N(i)} \theta_{i,j} U_i U_j$$

$$U_i U_j = \begin{cases} 1 & U_i = U_j \\ -1 & U_i \neq U_j \end{cases}$$

- $\theta > 0$: lower energy for pairs with same value

- $\theta < 0$: lower energy for pairs of different values



Boltzmann Machines

- U : Set of nodes H (hidden) and X (Observed)

- Binary $\{-1, 1\}$

- Edges

- (U_i, U_j) : Coupling

- θ : Weights: Intensity of coupling

- Energy:

- $$E(U|\theta) = - \sum_i \sum_{j \in N(i)} \theta_{i,j} U_i U_j$$

- $$U_i U_j = \begin{cases} 1 & U_i = U_j \\ -1 & U_i \neq U_j \end{cases}$$

- $\theta > 0$: lower energy for pairs with same value

- $\theta < 0$: lower energy for pairs of different values

- Log joint probability

- $\log P(U|\theta) = -E(U|\theta) - \log Z(\theta)$

- normalizing constant:
$$Z(\theta) = \sum_{U \in \{-1, 1\}} e^{-E(U|\theta)}$$

- Find H that maximizes

- $$P(H|X, \theta) = \frac{e^{-E(H, X|\theta)}}{\sum_H e^{-E(H, X|\theta)}}$$

- $$\log P(H|X, \theta) = -E(H, X|\theta) - \log \sum_H e^{-E(H, X|\theta)}$$

- $$\begin{aligned} \operatorname{argmax}_H & \left(\sum_{i,j} a_{i,j} h_i h_j \right) + \sum_j b_j h_j \\ & s.t. h_i \in \{-1, 1\} \end{aligned}$$

Variational Inference

KL Divergence

- Use $Q(H; X, \hat{\theta})$ similar to $P(H|X, \theta)$ and easier to deal with
 - find parameters $\hat{\theta}$ for Q
- Similarity measure
 - Kullback-Leibler (KL) Divergence between $P(X)$ and $Q(X)$

$$\mathbb{D}(P||Q) = \int P(X) \log \frac{P(X)}{Q(X)} dx$$

- Downsides
 - Not symmetric: $\mathbb{D}(P||Q) \neq \mathbb{D}(Q||P)$
 - Not triangle inequality:
 $\mathbb{D}(P||Q) \not\leq \mathbb{D}(P||R) + \mathbb{D}(P||Q)$
- Non-negative: $\mathbb{D}(P||Q) \geq 0$

- Entropy of $P(X)$: constant

$$H(P) = - \int P(X) \log P(X) dx = - \mathbb{E}_P[\log P]$$

$$\begin{aligned} \mathbb{D}(P||Q) &= \int P(X) \log \frac{P(X)}{Q(X)} dX \\ &= \int P(X) (\log P(X) - \log Q(X)) dx \\ &= \int P(X) \log P(X) dx - \int P(X) \log Q(X) dx \end{aligned}$$

$$= \mathbb{E}_P[\log P] - \mathbb{E}_P[\log Q]$$

$$-\mathbb{E}_P[\log Q] = \mathbb{E}_P[\log P] - \mathbb{D}(P||Q)$$

$$= -H(P) - \mathbb{D}(P||Q)$$

$$\mathcal{L}(\theta) = \sum_i \log Q(X_i | \theta)$$

$$\mathcal{L}(\theta) \approx \int P(X) \log Q(X | \theta) dX$$

$$\approx \mathbb{E}_P[\log Q(X | \theta)] \approx -H(P) - \mathbb{D}(P||Q)$$

$$-H(P) \approx \mathcal{L}(\theta) + \mathbb{D}(P||Q)$$

Variational Inference

From $\mathcal{L}(\theta)$ To $\mathbb{D}(P||Q)$

- Use $Q(H; X, \hat{\theta})$ similar to $P(H|X, \theta)$ and easier to deal with

- find parameters $\hat{\theta}$ for Q

- Similarity measure

- Kullback-Leibler (KL) Divergence between $P(X)$ and $Q(X)$

- $\mathbb{D}(P||Q) = \int P(X) \log \frac{P(X)}{Q(X)} dx$

- Downsides

- Not symmetric: $\mathbb{D}(P||Q) \neq \mathbb{D}(Q||P)$

- Not triangle inequality:
 $\mathbb{D}(P||Q) \not\leq \mathbb{D}(P||R) + \mathbb{D}(P||Q)$

- Non-negative: $\mathbb{D}(P||Q) \geq 0$

- Entropy of $P(X)$

$$H(P) = - \int P(X) \log P(X) dx = - \mathbb{E}_P[\log P]$$

- $-H(P) \approx \mathcal{L}(\theta) + \mathbb{D}(P||Q)$

- constant

- $\mathcal{L}(\theta)$ at maximum when $\mathbb{D}(P||Q)$ at minimum

- Maximizing $\mathcal{L}(\theta)$ can be achieved by minimizing $\mathbb{D}(P||Q)$

Variational Inference

Minimizing KL Divergence

- Minimizing $\mathbb{D}(Q(H) \| P(H | X))$
 - $$\begin{aligned} \mathbb{D}(Q(H) \| P(H | X)) &= \mathbb{E}_Q[\log Q] - \mathbb{E}_Q[\log P(H | X)] \\ &= \mathbb{E}_Q[\log Q] - \mathbb{E}_Q[\log P(H, X)] + \mathbb{E}_Q[\log P(X)] \\ &= \mathbb{E}_Q[\log Q] - \mathbb{E}_Q[\log P(H, X)] + \log P(X) \end{aligned}$$
- Variational Free Energy under Q:
 - $\mathbb{E}_Q = \mathbb{E}_Q[\log Q] - \mathbb{E}_Q[\log P(H, X)]$
- $\mathbb{D}(Q(H) \| P(H | X)) = \mathbb{E}_Q + \log P(X)$
- $\log P(X) = \mathbb{D}(Q(H) \| P(H | X)) - \mathbb{E}_Q$
 - constant
 - If Variational Free Energy decreases
 - KL Divergence $\mathbb{D}(Q(H) \| P(H | X))$ decreases
- Minimizing variational free energy \mathbb{E}_Q minimizes $\mathbb{D}(Q(H) \| P(H | X))$

Variational Inference

Minimizing Free Variational Energy

- Approximate distribution Q for Boltzmann Machine

- Hidden variables $H_i \in \{-1, 1\}$

- $Q(H) = q_1(H_1)q_2(H_2)\dots q_N(H_N)$

- $q_i(H_i) = \pi_i^{\frac{1+H_i}{2}}(1 - \pi_i)^{\frac{1-H_i}{2}}$

- Values of $q_i(H_i)$: $\begin{cases} \pi_i & H_i = 1 \\ 1 - \pi_i & H_i = -1 \end{cases}$

- $\pi_i = P(H_i = 1)$

- Assume that only one q_i is unknown

- $Q_{\hat{i}} = q_1(H_1)\dots q_{i-1}(H_{i-1})q_{i+1}(H_{i+1})\dots q_N(H_N)$

- find q_i that minimizes free variational energy

- Free variational energy

$$\mathbb{E}_Q = \mathbb{E}_Q[\log Q] - \mathbb{E}_Q[\log P(H, X)]$$

- $\mathbb{E}_Q[\log Q] = \mathbb{E}_{q_1(H_1)\dots q_N(H_N)}[\log q_1(H_1) + \log q_N(H_N)]$

- $= \mathbb{E}_{q_1(H_1)}[\log q_1(H_1)] + \dots \mathbb{E}_{q_N(H_N)}[\log q_N(H_N)]$

- $\mathbb{E}_Q[\log P(H, X)] = q_i(-1)p_{i,-1} + q_i(1)p_{i,1}$

- $p_{i,-1} = \mathbb{E}_{Q_{\hat{i}}}[\log P(H_1, \dots, H_i = -1, \dots, H_N, X)]$

- $p_{i,1} = \mathbb{E}_{Q_{\hat{i}}}[\log P(H_1, \dots, H_i = 1, \dots, H_N, X)]$

- Choose q_i that minimizes:

- $q_i(-1)\log q_i(-1) + q_i(1)\log q_i(1) - (q_i(-1)p_{i,-1} + q_i(1)p_{i,1})$

- s.t. $q_i(1) + q_i(-1) = 1$

Boltzmann Machines - Mean Field Inference

- Choose q_i that minimizes:

$$q_i(-1)\log q_i(-1) + q_i(1)\log q_i(1) - (q_i(-1)p_{i,-1} + q_i(1)p_{i,1})$$

- s.t. $q_i(1) + q_i(-1) = 1$

$$p_{i,-1} = \mathbb{E}_{Q_i}[\log P(H_1, \dots, H_i = -1, \dots, H_N, X)]$$

- $p_{i,1} = \mathbb{E}_{Q_i}[\log P(H_1, \dots, H_i = 1, \dots, H_N, X)]$

- Using a Lagrange Multiplier:

$$q_i(-1) = \frac{e^{p_{i,-1}}}{e^{p_{i,-1}} + e^{p_{i,1}}}$$

- $q_i(1) = \frac{e^{p_{i,1}}}{e^{p_{i,-1}} + e^{p_{i,1}}}$

- Remember:

- $\log P(U | \theta) = -E(U | \theta) - \log Z$

- $E(U | \theta) = - \sum_i \sum_{j \in N(i)} \theta_{i,j} U_i U_j$

$$p_{i,-1} = \mathbb{E}_{Q_i}[\sum_{j \in N(i) \cap H} \theta_{i,j}(-1)H_j + \sum_{j \in N(i) \cap X} \theta_{i,j}(-1)X_j + K_i]$$

- $= \sum_{j \in N(i) \cap H} \theta_{i,j}(-1)\mathbb{E}_{Q_i}[H_j] + \sum_{j \in N(i) \cap X} \theta_{i,j}(-1)X_j + K_i$

- $p_{i,-1} = \sum_{j \in N(i) \cap H} \theta_{i,j}(-1)(2\pi_j - 1) + \sum_{j \in N(i) \cap X} \theta_{i,j}(-1)X_j + K_i$

- $p_{i,1} = \sum_{j \in N(i) \cap H} \theta_{i,j}(1)(2\pi_j - 1) + \sum_{j \in N(i) \cap X} \theta_{i,j}(1)X_j + K_i$

- Parameters:

- $p_i = \sum_{j \in N(i) \cap H} \theta_{i,j}(2\pi_j - 1) + \sum_{j \in N(i) \cap X} \theta_{i,j}X_j$

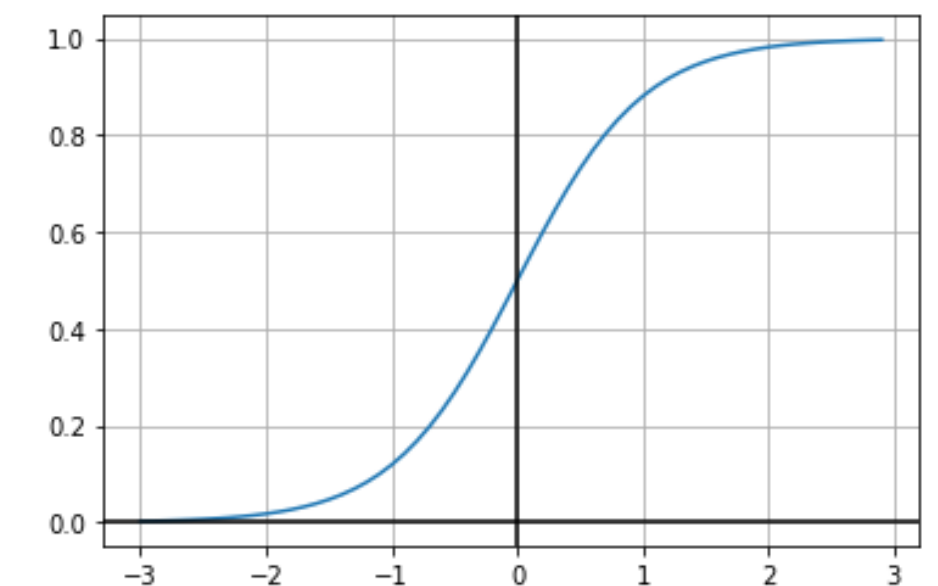
- $\pi_i = \frac{e^{p_i}}{e^{p_i} + e^{-p_i}}$

- Mean Field Inference

- While $(\pi_i$'s change more than some ϵ)

- update each π_i as defined above

- Estimated $H_i = \begin{cases} 1 & \pi_i \geq 0.5 \\ -1 & \pi_i < 0.5 \end{cases}$



Boltzmann Machines

- Boltzmann Machines
- Variational Inference Approach
- Mean Field Inference Algorithm

Applied Machine Learning

Boltzmann Machines