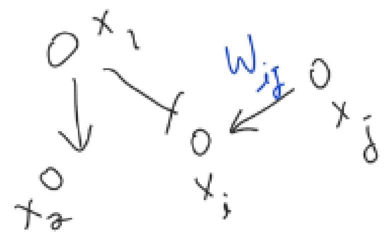
정보이론을 이용한 네트워크 추론

조정효 (jojunghyo@gmail.com)



$$\vec{x}(t) = \{x_1(t), x_2(t), \cdots, x_n(t)\}$$

시계열 데이터로부터 W_{ij} 추론

$x_i(t+1)$ 은 $H_i(t) = \sum_j W_{ij} x_j(t)$ 에 의해서 확률적으로 결정 :

$$\boxed{P(x_i(t+1) = \pm 1 \mid \vec{x}(t)) = \frac{e^{\pm H_i(t)}}{e^{H_i(t)} + e^{-H_i(t)}}}$$

kinetic Ising model

$x_i(t+1)$ 의 기대값

$$E[x_i(t+1)] = (+1) \cdot P(x_i(t+1) = +1 \mid \vec{x}(t)) + (-1) \cdot P(x_i(t+1) = -1 \mid \vec{x}(t))$$

$$= \frac{e^{H_i(t)} - e^{-H_i(t)}}{e^{-H_i(t)} + e^{H_i(t)}} = \tanh H_i(t)$$

질문: $\vec{x}(t)$, $\vec{H}(t)$ 가 주어진 경우 W_{ij}를 찾을 수 있나?

데이터

$$\begin{array}{cccccccc}
X_1(1) & X_2(1) & \cdots & X_n(1) & H_1(1) & H_2(1) & \cdots & H_n(1) \\
X_1(2) & X_2(2) & \cdots & X_n(2) & H_1(2) & H_2(2) & \cdots & H_n(2) \\
\vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\
X_1(L) & X_2(L) & \cdots & X_n(L) & H_1(L) & H_2(L) & \cdots & H_n(L) \\
X_1(L+1) & X_2(L+1) & \cdots & X_n(L+1) & & & &
\end{array}$$

$P(\vec{x}, \vec{H}(\vec{x})) = P(\vec{x})$ 분포는 w_{ij}에 의존

$P_0(\vec{x}) = \frac{1}{L} \sum_{t=1}^{L} \delta(\vec{x} - \vec{x}(t))$ 데이터의 빈도에서 추론한 $P(\vec{x})$ 의 사전확률 (Prior probability)

$x_j(t)$와 $H_i(t)$의 기대값을 생각해 보자.

$$m_j = \sum_{\vec{x}} x_j P(\vec{x}), \quad h_i = \sum_{\vec{x}} H_i(\vec{x}) P(\vec{x})$$

$P(\vec{x}) = P_0(\vec{x})$ 인 경우

$$\begin{cases} m_j^0 = \sum_{\vec{x}} x_j P_0(\vec{x}) = \frac{1}{L}\sum_{t=1}^{L} x_j(t) \\ h_i^0 = \sum_{\vec{x}} H_i(\vec{x}) \cdot P_0(\vec{x}) = \frac{1}{L}\sum_{t=1}^{L} H_i(t) \end{cases}$$

$$\left.\begin{aligned} h_i &= \sum_{\vec{x}} H_i(\vec{x}) \cdot P(\vec{x}) \\ &= \sum_{\vec{x}} \sum_{j} w_{ij} x_j \cdot P(\vec{x}) \\ &= \sum_{j} w_{ij} m_j \end{aligned}\right\} \quad w_{ij} = \frac{\partial h_i}{\partial m_j}$$

질문: $P_0(\vec{x})$에 가장 가까우면서

$m_j = \sum_{\vec{x}} x_j\, p(\vec{x})$ 와 $h_i = \sum_{\vec{x}} H_i(\vec{x}) \cdot p(\vec{x})$

두 구속조건을 만족하는 확률분포 $p(\vec{x}) = ?$

답: $$p(\vec{x}) = \frac{P_0(\vec{x}) \cdot e^{\vec{J}\cdot\vec{x} + \vec{\beta}\cdot\vec{H}}}{Z}, \quad Z = \sum_{\vec{x}} P_0(\vec{x}) \cdot e^{\vec{J}\cdot\vec{x} + \vec{\beta}\cdot\vec{H}}$$

유도

주사위의 눈 x가 나올 사전확률 $P_0(x)$가 주어진 경우, n번 시행에서 x가 나온 횟수 n_x를 관찰했다고 해보자.

이 관찰의 가능도 (Likelihood)는

$$\text{Likelihood} = \binom{n}{n_1}\binom{n-n_1}{n_2}\binom{n-n_1-n_2}{n_3}\cdots\binom{n-n_1-n_2-\cdots-n_5}{n_6}$$

$$\times P_0(1)^{n_1} \cdot P_0(2)^{n_2} \cdots P_0(6)^{n_6}$$

$$= \frac{n!}{n_1!\, n_2! \cdots n_6!} P_0(1)^{n_1} P_0(2)^{n_2} \cdots P_0(6)^{n_6}$$

Stirling의 어림 $\log n! \simeq n \cdot \log n - n$

$$\begin{aligned}\log \text{Likelihood} = {} & n \cdot \log n - n \\ & - n_1 \cdot \log n_1 + n_1 + n_1 \cdot \log P_0(1) \\ & - n_2 \cdot \log n_2 + n_2 + n_2 \cdot \log P_0(2) \\ & \vdots \\ & - n_6 \cdot \log n_6 + n_6 + n_6 \cdot \log P_0(6)\end{aligned}$$

$$= -n \cdot \sum_x P(x) \cdot \log \frac{P(x)}{P_0(x)}, \quad P(x) \equiv \frac{n_x}{n}$$

$$\equiv -n \cdot \underbrace{I(P \| P_0)}_{\text{relative information}}$$

(Kullback-Leibler divergence)

log likelihood를 최대화

$= p(x)$와 $P_0(x)$ 사이의 거리를 최소화

앞서 질문은 다음 Lagrangian을 최소화 하는 문제임.

$$\mathcal{L} = I(p\|p_0) - \sum_j J_j \cdot \left(E[x_j] - m_j \right) - \sum_i \beta_i \cdot \left(E[H_i] - h_i \right) - \lambda \left(\sum_{\vec{x}} p(\vec{x}) - 1 \right)$$

$$= \sum_{\vec{x}} p(\vec{x}) \cdot \log \frac{p(\vec{x})}{p_0(\vec{x})} - \sum_j J_j \cdot \left(\sum_{\vec{x}} x_j \cdot p(\vec{x}) - m_j \right) - \sum_i \beta_i \cdot \left(\sum_{\vec{x}} H_i(\vec{x}) \cdot p(\vec{x}) - h_i \right) - \lambda \left(\sum_{\vec{x}} p(\vec{x}) - 1 \right)$$

$$\frac{\delta L}{\delta p(\vec{x})} = \log \frac{p^*(\vec{x})}{p_0(\vec{x})} + 1 - \sum_j J_j \cdot x_j - \sum_i \beta_i \cdot H_i - \lambda = 0$$

$$\Rightarrow p^*(\vec{x}) = \frac{p_0(\vec{x}) \cdot e^{\vec{J} \cdot \vec{x} + \vec{\beta} \cdot \vec{H}}}{Z}, \quad Z = \sum_{\vec{x}} p_0(\vec{x}) \cdot e^{\vec{J} \cdot \vec{x} + \vec{\beta} \cdot \vec{H}}$$

$$I(p\|p_0) \geq \underbrace{I(p^*\|p_0)}_{\text{Minimum discrimination information}}$$

볼츠만 분포

$$\frac{\partial \log Z}{\partial J_j} = \sum_{\vec{x}} \frac{x_j \cdot p_0(\vec{x}) \cdot e^{\vec{J} \cdot \vec{x} + \vec{\beta} \cdot \vec{H}}}{Z} = \sum_{\vec{x}} x_j \, p^*(\vec{x}) = m_j$$

$$\frac{\partial \log Z}{\partial \beta_i} = \sum_{\vec{x}} \frac{H_i \, p_0(\vec{x}) \cdot e^{\vec{J} \cdot \vec{x} + \vec{\beta} \cdot \vec{H}}}{Z} = \sum_{\vec{x}} H_j(\vec{x}) \cdot p^*(\vec{x}) = h_j$$

$p^*(\vec{x})$는 m_j와 h_i와 관련된 구속조건을 만족하는 $p(\vec{x})$ 가운데 $p_0(\vec{x})$에 가장 가까운 분포!

그 때, $p^*(\vec{x})$와 $p_0(\vec{x})$ 사이의 거리는?

$$I(p^* \| p_0) = \sum_{\vec{x}} p^*(\vec{x}) \cdot \log \frac{p^*(\vec{x})}{p_0(\vec{x})}$$

$$= \sum_{\vec{x}} p^*(\vec{x}) \cdot \left(\vec{J} \cdot \vec{x} + \vec{\beta} \cdot \vec{H} - \log Z \right)$$

$$G(\vec{m}, \vec{h}) = \vec{J} \cdot \vec{m} + \vec{\beta} \cdot \vec{h} - \underbrace{\log Z(\vec{J}, \vec{\beta})}_{F(\vec{J}, \vec{\beta})}$$

르장드르 변환

$$\frac{\partial G}{\partial m_j} = \frac{\partial J_j}{\partial m_j} \cdot m_j + J_j - \underbrace{\frac{\partial \log Z}{\partial J_j}}_{m_j} \cdot \frac{\partial J_j}{\partial m_j} = J_j$$

$$\frac{\partial G}{\partial h_i} = \frac{\partial \beta_i}{\partial h_i} \cdot h_i + \beta_i - \underbrace{\frac{\partial \log Z}{\partial \beta_i}}_{h_i} \cdot \frac{\partial \beta_i}{\partial h_i} = \beta_i$$

Taylor series expansion

$$\begin{cases} \vec{m} = \vec{m}_0 \\ \vec{h} = \vec{h}_0 \end{cases} \text{ at } \vec{J} = \vec{\beta} = 0$$

$$G(\vec{m}, \vec{h}) \simeq G(\vec{m}_0, \vec{h}_0) + \cancel{\sum_j \left[\frac{\partial G}{\partial m_j}\right]_{\vec{m}_0, \vec{h}_0}}^{J_j = 0} (m_j - m_j^0)$$

$$+ \cancel{\sum_i \left[\frac{\partial G}{\partial h_i}\right]_{\vec{m}_0, \vec{h}_0}}^{\beta_i = 0} (h_i - h_i^0)$$

$$+ \frac{1}{2} \sum_{i,j} \left[\frac{\partial^2 G}{\partial m_i \partial m_j}\right]_{\vec{m}_0, \vec{h}_0} (m_i - m_i^0)(m_j - m_j^0)$$

$$+ \frac{1}{2} \sum_{i,j} \left[\frac{\partial^2 G}{\partial m_i \partial h_j}\right]_{\vec{m}_0, \vec{h}_0} (m_i - m_i^0) \cdot (h_j - h_j^0)$$

$$+ \frac{1}{2} \sum_{i,j} \left[\frac{\partial^2 G}{\partial h_i \partial h_j}\right]_{\vec{m}_0, \vec{h}_0} (h_i - h_i^0) \cdot (h_j - h_j^0)$$

$\vec{m}$과 $\vec{h}$ 사이의 관계에 대한 정보를 G 함수의 curvature에 포함

$$W_{ij} = \frac{\partial h_i}{\partial m_j} = \sum_k \left(\frac{\partial^2 G}{\partial h_i \partial m_k}\right)^{-1} \cdot \left(\frac{\partial^2 G}{\partial m_k \partial m_j}\right)$$

① $$\frac{\partial^2 G}{\partial h_i \partial m_k} = \frac{\partial}{\partial h_i}\left(\frac{\partial G}{\partial m_k}\right) = \frac{\partial J_k}{\partial h_i}$$

$$\left[\frac{\partial^2 G}{\partial h_i \partial m_k}\right]^{-1} = \frac{\partial h_i}{\partial J_k} = \frac{\partial}{\partial J_k}\left[\sum_{\vec{x}} \frac{H_i(\vec{x}) \cdot P_0(\vec{x}) \cdot e^{\vec{J}\cdot\vec{x} + \vec{\beta}\cdot\vec{H}}}{Z}\right]$$

$$= \sum_{\vec{x}} \frac{H_i(\vec{x}) \cdot x_k \cdot P_0(\vec{x}) \cdot e^{\vec{J}\cdot\vec{x} + \vec{\beta}\cdot\vec{H}}}{Z}$$

$$- \sum_{\vec{x}} \frac{H_i(\vec{x})\, P_0(\vec{x}) \cdot e^{\vec{J}\cdot\vec{x} + \vec{\beta}\cdot\vec{H}}}{Z} \cdot \sum_{\vec{x}} \frac{x_k \cdot P_0(\vec{x}) \cdot e^{\vec{J}\cdot\vec{x} + \vec{\beta}\cdot\vec{H}}}{Z}$$

$$= \langle H_i \cdot x_k \rangle - \langle H_i \rangle \cdot \langle x_k \rangle$$

$$\equiv \langle H_i\, x_k \rangle_c$$ connected correlation

② $$\frac{\partial^2 G}{\partial m_k \partial m_j} = \frac{\partial}{\partial m_k}\left(\frac{\partial G}{\partial m_j}\right) = \frac{\partial J_j}{\partial m_k}$$

$$\left[\frac{\partial^2 G}{\partial m_k \partial m_j}\right]^{-1} = \frac{\partial m_k}{\partial J_j} = \frac{\partial}{\partial J_j}\left[\sum_{\vec{x}} \frac{x_k \cdot P_0(\vec{x}) \cdot e^{\vec{J}\cdot\vec{x} + \vec{\beta}\cdot\vec{H}}}{Z}\right]$$

$$= \sum_{\vec{x}} \frac{x_k \cdot x_j \cdot P_0(\vec{x}) \cdot e^{\vec{J}\cdot\vec{x} + \vec{\beta}\cdot\vec{H}}}{Z}$$

$$- \sum_{\vec{x}} \frac{x_k \cdot P_0(\vec{x}) \cdot e^{\vec{J}\cdot\vec{x} + \vec{\beta}\cdot\vec{H}}}{Z} \cdot \sum_{\vec{x}} \frac{x_j \cdot P_0(\vec{x}) \cdot e^{\vec{J}\cdot\vec{x} + \vec{\beta}\cdot\vec{H}}}{Z}$$

$$= \langle x_k \cdot x_j \rangle - \langle x_k \rangle \langle x_j \rangle$$

$$= \langle x_k \cdot x_j \rangle_c$$

$$\Rightarrow \boxed{W_{ij} = \sum_k \langle H_i \cdot x_k \rangle_c \cdot \langle x_k \cdot x_j \rangle_c^{-1}}$$

다중 선형회귀 식의 결과랑 일치.

$$H = W \cdot X + b$$

$X(t), H(t)$ 가 주어졌을때 W 와 b 를 추론

$$\text{Error} = \sum_t \left(H(t) - W \cdot X(t) - b \right)^2$$

① $\dfrac{\partial \text{Error}}{\partial b} = 0 \qquad \sum_t 2\left(H(t) - W \cdot X(t) - b\right) \cdot (-1) = 0$

② $\dfrac{\partial \text{Error}}{\partial W} = 0 \qquad \sum_t 2\left(H(t) - W \cdot X(t) - b\right) \cdot \left(-X(t)\right) = 0$

① $\langle H \rangle \equiv \dfrac{1}{L} \sum_{t=1}^{L} H(t) \quad , \quad \langle X \rangle \equiv \dfrac{1}{L} \sum_{t=1}^{L} X(t)$

$$\langle H \rangle = W \cdot \langle X \rangle + b$$

② $\langle H \cdot X \rangle - \langle H \rangle \langle X \rangle = W \left(\langle X \cdot X \rangle - \langle X \rangle \langle X \rangle \right)$

$$\Rightarrow W = \frac{\langle H \cdot X \rangle_c}{\langle X \cdot X \rangle_c}$$

일반화

$H_i = \sum_j W_{ij} \cdot X_j + b_i$ 인 다중선형회귀 문제에서는

$$\boxed{W_{ij} = \sum_k \langle H_i \cdot X_k \rangle_c \cdot \langle X_k \cdot X_j \rangle_c^{-1}}$$

이 결과는 이용해서 kinetic Ising model의 W_{ij} 를 추론 해보자.

$$P\left(X_i(t+1) = \pm 1 \mid \vec{X}(t) \right) = \frac{e^{\pm H_i(t)}}{e^{H_i(t)} + e^{-H_i(t)}}$$

$$E\left[X_i(t+1) \right] = \tanh H_i(t)$$

recipe

① $H_i(t) = \sum_{j=1}^{n} w_{ij} \; X_j(t)$

임의의 w_{ij} 에 대해 모든 시간의 $H_i(t)$ 를 계산

② $H_i'(t) = \frac{X_i(t+1)}{E[X_i(t+1)]} \cdot H_i(t) = X_i(t+1) \cdot \frac{H_i(t)}{\tanh H_i(t)}$

를 이용해서 $H_i(t) \to H_i'(t)$ 로 업데이트

③ $w_{ij} = \sum_k \langle H_i' \cdot X_k \rangle_c \cdot \langle X_k \cdot X_j \rangle_c^{-1}$

을 이용해서 w_{ij} 를 업데이트

④ $D_i = \sum_{t=0}^{L-1} \Big[\underbrace{X_i(t+1)}_{\text{관찰}} - E[\underbrace{X_i(t+1)}_{\text{예측}}] \Big]^2$

관찰값과 예측값 사이의 거리 D_i 가 최소가
되면 ① → ② → ③ → ① ··· 의
Iteration 을 끝냄.
(index i 마다 독립적으로 병렬계산 가능!)

마지막 수정: 오전 12:55