Subject: 2|H-Q3 Date: March 16 2022

Attendants: Undirected Graphical Model

Energy-based model

$$p(x,h) = \frac{1}{Z}\exp(-E(x,h))$$

where $Z = \sum_h \int \exp(-E(x,h))\,dx$

$$E(x,h) = x^Tx + h^TWf_\theta(x) - c^Th - b^Tx$$

$W \in \mathbb{R}^{k\times N}$

Q1.

Show that the conditional probability distribution $\Pr(h|x)$ factorizes over the k-elements of h

$x \in \mathbb{R}^N,\ h \in \{0,1\}^k$

$$\Pr(h|x) = \frac{P(x,h)}{\sum_h P(x,h)} = \frac{\exp(-E(x,h))}{\sum_h \exp(-E(x,h))} \quad \cdots ①$$

$\exp(-E(x,h))$

$= \exp(c^Th + b^Tx - x^Tx - h^TWf_\theta(x))$

$= \underbrace{\exp(b^Tx - x^Tx)}_{\text{hが含まれない}}\ \underbrace{\exp(c^Th - h^TWf_\theta(x))}_{\text{hを含む}}$

Subject Date

Attendants

$$① = \frac{\exp(c^T h - h^T W f_\theta(x))}{\sum_h \exp(c^T h - h^T W f_\theta(x))} \quad \cdots ②$$

$$\exp(c^T h - h^T W f_\theta(x)) = \exp(-h^T \underbrace{(W f_\theta(x) - c)}_{\in \mathbb{R}^K})$$

Let α as $(W f_\theta(x) - c)$

where $\alpha = (\alpha_1 \cdots \alpha_K)$

$$= \exp(-h^T \alpha)$$

$$= \exp\left(-\sum_{k=1}^{K} h_k \alpha_k\right)$$

$$= \prod_{k=1}^{K} \exp(-h_k \alpha_k)$$

$$\boxed{\begin{aligned} \exp(x+y) &= e^{x+y} \\ &= e^x e^y \\ &= \exp(x)\exp(y) \end{aligned}}$$

$$\sum_h \exp(c^T h - h^T W f_\theta(x)) = \sum_h \prod_{k=1}^{K} \exp(-h_k \alpha_k)$$

$$= \sum_{h_1 \in \{0,1\}} \cdots \sum_{h_K \in \{0,1\}} \prod_{k=1}^{K} \exp(-h_k \alpha_k)$$

Subject

Date

Attendants

入れかわる

$$\sum_{h} \prod_{k=1}^{K} \exp(-h_k \alpha_k) = \prod_{k=1}^{K} \sum_{h} \exp(-h_k \alpha_k)$$

then

$$② = \frac{\prod_{k=1}^{K} \exp(-h_k \alpha_k)}{\prod_{k=1}^{K} \sum_{h} \exp(-h_k \alpha_k)}$$

Q1's Answer (factorized).

Q2,

$$= \prod_{k=1}^{K} \left\{ \frac{\exp(-h_k \alpha_k)}{\sum_{h} \exp(-h_k \alpha_k)} \right\}$$

$$= \prod_{k=1}^{K} \Pr(h_k \mid x)$$

Q2's Answer

Derive expression for $\Pr(h_k \mid x)$

Q3, Does the conditional probability distribution $p(x|h)$ similary factorize?

$$\exp(-E(x,h))$$
$$= \exp(c^T h + b^T x - x^T x - h^T W f_\theta(x))$$
$$= \exp(c^T h)\exp(b^T x - x^T x - h^T W f_\theta(x))$$

it includes $x^T x$

$$\frac{\exp(-E(x,h))}{\sum_x \exp(-E(x,h))} = \frac{\exp(b^T x - x^T x - h^T W f_\theta(x))}{\sum_x \exp(b^T x - x^T x - h^T W f_\theta(x))}$$

we cannot facterize as Q2.

Subject

Date 5

Attendants

Q4 $F(x) = -\log \sum_h \exp(-E(x,h))$

$\uparrow$ free energy

$= -\log \sum_h \exp(c^T h + b^T x - x^T x - h^T W f_\theta(x))$

$= -\log \sum_h \{\exp(b^T x - x^T x) \exp(c^T h - h^T W f_\theta(x))\}$

$= -\log \exp(b^T x - x^T x) \sum_h \exp(c^T h - h^T W f_\theta(x))$

$= -\log_e \{ e^{(b^T x - x^T x)} \sum_h \exp(-h^T \alpha) \}$

$\alpha = -c + W f_\theta(x)$

since

$\log(XY) = \log X + \log Y$

$= (x^T x - b^T x) - \log \sum_h \exp(-h^T \alpha)$

Subject Date 6

Attendants

$$F(x) = \beta - \log \sum_h \exp(-h^T \alpha)$$

where $\beta = x^T x - b^T x$, $\alpha = W f_\theta(x) - c$

$$= \beta - \log \sum_h \exp\left(-\prod_{k=1}^{K} h_k \alpha_k\right)$$

$$= \beta - \log \left\{ \sum_{h_1} \sum_{h_2} \cdots \exp\left(-\prod_{k=1}^{K} h_k \alpha_k\right) \right\}$$

ここで

$$\sum_{\{\vec{s}\}} \prod_{i=1}^{N} e^{\beta H s_i} = \prod_{i=1}^{N} \sum_{s_i \in \{-1, 1\}} e^{\beta H s_i}$$

sum over all possible vectors $\vec{s} = (s_1, \ldots s_N)$ with the restriction that s_i can only take values $\{-1, +1\}$

$$\sum_{s_1 \in \{-1, 1\}} \cdots \sum_{s_N \in \{-1, 1\}} = \sum_{\{\vec{s}\}}$$

Then

$$\sum_{\{\vec{s}\}} \prod_{i=1}^{N} e^{\beta H s_i} = \sum_{s_1\in\{-1,1\}} \cdots \sum_{s_N\in\{-1,1\}} \underbrace{\prod_{i=1}^{N} e^{\beta H s_i}}_{e^{\beta H s_1} \times e^{\beta H s_2} \times \cdots e^{\beta H s_N}}$$

$$= \sum_{\{\vec{s}\}} \prod_{i=1}^{N} e^{\beta H s_i} = \sum_{s_1\in\{-1,1\}} \cdots \sum_{s_N\in\{-1,1\}} e^{\beta H s_1} \cdots e^{\beta H s_N}$$

Since each s_i is independent of others then we facterize the $\sum$'s as follows:

$$= \left(\sum_{s_1\in\{-1,1\}} e^{\beta H s_1} \right) \cdots \left(\sum_{s_N\in\{-1,1\}} e^{\beta H s_N} \right)$$

Subject　　Date　8

Attendants

$$\sum_{\{\vec{s}\}} \prod_{i=1}^{N} e^{\beta H s_i} = \sum_{\{\vec{s}\}} e^{\beta H s_1} e^{\beta H s_2} \cdots e^{\beta H s_N}$$

$\{\vec{s}\} = (s_1 \cdots s_N) \quad s_i \in \{p_1, p_2\}$

$$= \sum_{s_1} \cdots \sum_{s_N \in \{p_1, p_2\}} e^{\beta H s_1} \cdots e^{\beta H s_N}$$

$N=3$のとき

$$= \sum_{s_1} \sum_{s_2} \sum_{s_3} e^{\beta H s_1} e^{\beta H s_2} e^{\beta H s_3}$$

$$= \sum_{s_1} \sum_{s_2} \underbrace{\left(e^{\beta H s_1} e^{\beta H s_2} \right)} \times \sum_{s_3} e^{\beta H s_3}$$

s_3 に独立なので外に出せる

$$= \sum_{s_1} \left(e^{\beta H s_1} \right) \sum_{s_2} \left(e^{\beta H s_2} \right) \sum_{s_3} \left(e^{\beta H s_3} \right)$$

$$= \prod_{i=1}^{N} \left(\sum_{s_i} e^{\beta H s_i} \right)$$ に因子分解可能

Subject Date

Attendants

Then using interchange technique of $\sum$ and $\prod$

Let objective $F(x)$

$$= \beta - \log \left\{ \sum_{h_1} \sum_{h_2} \cdots \exp\left(- \prod_{k=1}^{K} h_k \alpha_k\right) \right\}$$

$$= \beta - \log \left\{ \prod_{i=1}^{N} \sum_{h} \left(- h_k \alpha_k\right) \right\}$$

where $\beta = x^T x - b^T x$

$$\alpha_k = w f_\theta(x) - c$$

fin.

Subject Date 10

Attendants

Q5 stochastic gradient ascent for energy based model training

maximize the log-likelihood

θ, W, b, c

$$\frac{\partial}{\partial\theta}\frac{1}{M}\sum_{i=1}^{M}\log p(x^i) = \mathbb{E}_{p(x,h)}\left[\frac{\partial E(x,h)}{\partial\theta}\right] - \frac{1}{M}\sum_{i=1}^{M}\mathbb{E}_{p(h|x^i)}\left[\frac{\partial E(x^i,h)}{\partial\theta}\right]$$