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 $Z = \sum_{h}\int \exp(-E(x,h))\,dx$

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

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

Q1.

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

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

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

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

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

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

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 $\Pr(h_k \mid x)$

Subject Date

Attendants

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

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

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

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

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

↑ free energy

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

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

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

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

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

since

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

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

$$F(x) = \beta - \log \sum_{h} \exp(-h^T \alpha)$$

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

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

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

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

Subject Date 7

Attendants

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:

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

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

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

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

$N=3$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

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

$$= \prod_{i=1}^{N} \left(\sum_{s_i} e^{\beta H s_i} \right)$$

(= 因子分解可能

Subject

Date 9

Attendants

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

Let objective $F(x)$

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

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

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

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

fin.

As Q4 answer

Subject

Date 10

Attendants

Q5 stochastic gradient ascent for energy based model training

maximize the log-likelihood

$\mathcal{J}_\theta$, W, b, c

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

$$\frac{\partial E(x,h)}{\partial\theta} = h^T W \nabla_\theta f_\theta(x)$$

$$\mathbb{E}_{p(x,h)}\left[\frac{\partial E(x,h)}{\partial\theta}\right]$$

$$= \mathbb{E}_{p(x,h)}\left[h^T W \nabla_\theta f_\theta(x)\right]$$

Subject Date / /

Attendants

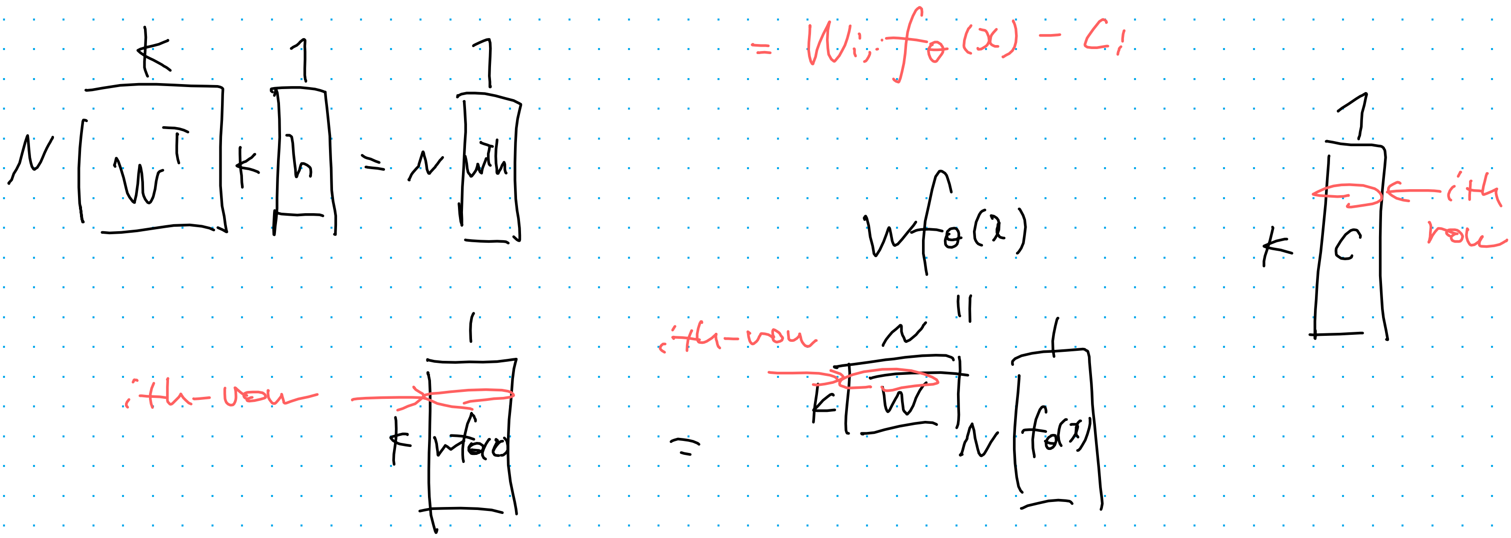
$$= \mathbb{E}_{x,h\sim p(x,h)}\left[h^T W \nabla_\theta f_\theta(x) \right]$$

$$= \mathbb{E}_{x\sim p(x)}\left[\nabla_\theta f_\theta(x)^T \left(\sum_{h\in\{0,1\}^K} p(h|x)\, W^T h \right) \right]$$

$$= \mathbb{E}_{x\sim p(x)}\left[\overbrace{\nabla_\theta f_\theta(x)}^{D\times N}{}^T \overbrace{\underbrace{\left(\sum_{h\in\{0,1\}^K} \prod_{i=1}^{K} p(h_i|x)\, W^T h \right)}}^{N\times 1} \right]$$

where $p(h_i|x) = \dfrac{\exp(-h_i\alpha_i)}{\sum_{h_i\in\{0,1\}} \exp(-h_i\alpha_i)}$

$$\alpha_i = e_i^T W f_\theta(x) - c_i$$
$$= W_{i,:} f_\theta(x) - c_i$$



Attendants

$$\boxed{\sum_{h_1\in\{0,1\}} \cdots \sum_{h_K\in\{0,1\}} \underbrace{\left(\prod_{i=1}^{K} \frac{\exp(-h_i\alpha_i)}{\sum_h \exp(-h_i\alpha_i)} \right)}_{} } w^T h$$

$$\prod_{i=1}^{K} \left(\frac{\exp(-\alpha_i)\exp(h_i)}{1+\exp(-\alpha_i)} \right)$$

$$= \prod_{i=1}^{K} \left\{ \exp(h_i) \left(\frac{e^{-\alpha_i}}{1+e^{-\alpha_i}} \right) \right\}$$

$$\sum_h \prod_{i=1}^{K} \frac{\exp(-h_i\alpha_i)}{\sum_h \exp(-h_i\alpha_i)}$$

$$= \sum_{h_1\in\{0,1\}} \cdots \sum_{h_K\in\{0,1\}} \left(\frac{e^{-h_1\alpha_1}}{\beta} \times \frac{e^{-h_2\alpha_2}}{\beta} \times \cdots \frac{e^{-h_K\alpha_K}}{\beta} \right)$$

$$= \left(\sum_{h_1\in\{0,1\}} \frac{e^{-h_1\alpha_1}}{\beta} \right) \cdots \left(\sum_{h_K\in\{0,1\}} \frac{e^{-h_K\alpha_K}}{\beta} \right)$$

$$\sum_h \left(\prod_{i=1}^{K} p(h_i|x) \right)$$

$$= \prod_{i=1}^{K} \left(\sum_{h_i \in \{0,1\}} \frac{\exp(-h_i \alpha_i)}{\exp(0) + \exp(-\alpha_i)} \right)$$

$$= \prod_{i=1}^{K} \left(\frac{1}{1+e^{-\alpha_i}} + \frac{e^{-\alpha_i}}{1+e^{-\alpha_i}} \right) = 1$$

$$\mathbb{E}_{x \sim p(x)} \left[\nabla_\theta f_\theta(x)^T \left(\underbrace{\sum_h \prod_{i=1}^{K} p(h_i|x)}_{\text{これ単体なら1}} \underbrace{W h}_{\text{これは外に まらいこ}} \right) \right]$$

$$\mathbb{E}_{x\sim p(x)}\left[\nabla_\theta f_\theta(x)^T \underbrace{\left(\sum_h \prod_{i=1}^{K} p(h_i|x)\, w^T h\right)}_{①}\right]$$

$$① = \sum_{h_1\in\{0,1\}} \cdots \sum_{h_K\in\{0,1\}} \left\{\left(\prod_{i=1}^{K} p(h_i|x)\right) w^T h\right\}$$

$$= \sum_{h_1} \cdots \sum_{h_K} \left\{\underbrace{p(h_1|x)\, p(h_2|x) \cdots p(h_K|x)}_{\text{scalar}}\, w^T h\right\}$$

$$w^T h = \begin{array}{c} K \\ \left[\; w^T \;\right] \end{array} \begin{array}{c} 1 \\ \left[\, h \,\right] \end{array} = \left[\; \right]$$ ← ith row

$$\sum_{h_1} \cdots \sum_{h_K} \left\{ ① \, w^T h \right\}$$

Subject

Date 15

Attendants EBM (Energy Based Model)

$$\frac{\partial \log p_\theta(x)}{\partial \theta} = \underbrace{\mathbb{E}_{p_\theta(x)}}_{\text{モデルからの sample}}\left[\frac{\partial E_\theta(x)}{\partial \theta}\right] - \mathbb{E}_{p_{data}(x)}\left[\frac{\partial E_\theta(x)}{\partial \theta}\right]$$

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

$$p_\theta = \frac{\exp(-E_\theta(x))}{\int \exp(-E_\theta(x))\,dx} = \frac{\exp(-E_\theta(x))}{Z}$$

$$\log p_\theta = \log_e e^{-E_\theta(x)} - \log_e \int e^{-E_\theta(x)}\,dx$$

$$= -E_\theta(x) - \log_e Z$$

$$\frac{\partial \log p_\theta}{\partial \theta} = \frac{\partial E_\theta(x)}{\partial \theta} - \frac{\partial}{\partial \theta}(\log Z)$$

Subject Date

Attendants 16

$$P_\theta = \frac{e^{-E_\theta(x)}}{\mathbb{E}_{x\sim p(x)}\left[e^{-E_\theta(x)}\right]}$$

$$\log P_\theta = -E_\theta(x) + \mathbb{E}_{x\sim p(x)}\left[-E_\theta(x)\right]$$

$$\frac{\partial \log P_\theta}{\partial \theta} = \mathbb{E}_{P_\theta(x)}\left[\frac{\partial E_\theta(x)}{\partial \theta}\right] + \mathbb{E}_{P_{data}(x)}\left[\frac{\partial E_\theta(x)}{\partial \theta}\right]$$

How to derive this

Subject

Date 17

Attendants

$$p(x) = \frac{e^{-F(x)}}{\int_{x'} \left(e^{-F(x)} \right) dx'}$$

負の対数尤度

$$L = -\log p(x) = F(x) - \mathbb{E}_{x'}\left[F(x) \right]$$

$$\frac{\partial L}{\partial \theta} = \mathbb{E}_{x \sim D}\left[\frac{\partial F_\theta(x)}{\partial \theta} \right]$$

$$- \mathbb{E}_{x' \sim p_\theta(x)}\left[\frac{\partial F_\theta(x)}{\partial \theta} \right]$$

??

Subject

Date 18

Attendants

cont. Q4

$$\mathbb{E}_{p(x,h)}\left[\frac{\partial E(x,h)}{\partial \theta}\right] = \mathbb{E}_{p(x)}\left[\nabla_\theta f_\theta(x)\,\mathbb{E}_{p(h|x)}\left[h^T W\right]\right]$$

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

$$\frac{\partial E(x,h)}{\partial \theta} = h^T W \,\nabla_\theta f_\theta(x)$$
)

$$\mathbb{E}_{p(x)}\left[\nabla_\theta f_\theta(x)\,\underbrace{\mathbb{E}_{p(h|x)}\left[h^T W \mid X = x\right]}\right]$$

$$= \mathbb{E}_{p(h|x)}\left[\sum_{l=1}^{k} W_{l:}^T h_l \mid x\right]$$

0 1

W $\quad W_{:,1}$ $\quad w_{1:}^T \leftarrow$ 0 → 1 → W $\quad (\;)^T = (\;)$

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix}\begin{pmatrix} e \\ f \end{pmatrix} = \begin{pmatrix} ae + bf \\ ce + df \end{pmatrix}$$

Subject

Date 19

Attendants

$$\mathbb{E}_{p(h|x)}\left[\sum_{i=1}^{K} W_{l,i}^{T} h_l \,\middle|\, x\right]$$

$$= \sum_{i=1}^{K} W_{l,i}^{T} \, \mathbb{E}_{p(h|x)}\left[h_l \,|\, x\right]$$

$$\sum_{h\in\{0,1\}^K} p(h|x)\, h_l$$

because $\prod_{i=1}^{K} p(h_i|x) = p(h|x)$ from Q2

$$= \sum_{h} p(h_1|x)\, p(h_2|x) \cdots p(h_K|x)\, h_l$$

$$= \sum_{h} \prod_{i=1}^{K} p(h_i|x)\, h_l$$

$$= \sum_{h} \left\{ p(h_l|x)\, h_l \prod_{\substack{i=1 \\ i\neq l}}^{K} p(h_i|x) \right\}$$

$$= \sum_{h_l\in\{0,1\}} p(h_l|x)\, h_l \sum_{\bar{h}_l\in\{0,1\}^{K-1}} \prod_{\substack{i=1 \\ i\neq l}}^{K} p(h_i|x)$$

Subject

Date 20

Attendants

$$\sum_{h_l \in \{0,1\}} p(h_l|x) h_l \underbrace{\sum_{\bar{h}_l \in \{0,1\}^{k-1}} \prod_{i=1, i\neq l}^{k} p(h_i|x)}$$

probability mass function summing up over all values equals to 1

$$= \sum_{h_l \in \{0,1\}} p(h_l|x) h_l \times 1$$

$$\mathbb{E}_p[\nabla_\theta f_\theta^T(x)$$

$$\mathbb{E}_{p(h|x)}[w^T h | x = x]]$$

$$= \mathbb{E}_{p(x)}\left[\nabla_\theta f_\theta^T(x) \sum_{i=1}^{k} w_{l,:}^T \, \mathbb{E}_{p(h|x)}[h_l|x]\right]$$

$$= \mathbb{E}_{p(x)}\left[\nabla_\theta f_\theta^T(x) \sum_{l=1}^{k} w_{l,:}^T \sum_{h_l \in \{0,1\}} p(h_l|x) h_l\right]$$

Subject

Date 21

Attendants

$$\mathbb{E}_{p(x)}\left[\nabla_\theta f_\theta^T(x)\, \mathbb{E}_{p(h|x)}\left[w^T h \mid X = x\right]\right]$$

$$= \mathbb{E}_{p(x)}\left[\nabla_\theta f_\theta^T(x) \sum_{l=1}^{K} w_{l,:}^T\, \mathbb{E}_{p(h|x)}\left[h_l \mid x\right]\right]$$

$$= \mathbb{E}_{p(x)}\left[\nabla_\theta f_\theta^T(x) \sum_{l=1}^{K} w_{l,:} \sum_{h \in \{0,1\}^K} \underbrace{p(h|x)}_{= \prod_{i=1}^{K} p(h_i|x)} h_l\right]$$

$$= \mathbb{E}_{p(x)}\left[\nabla_\theta f_\theta^T(x) \sum_{l=1}^{K} w_{l,:} \sum_{h_l \in \{0,1\}} p(h_l|x) h_l \underbrace{\boxed{\sum_{h_l \in \{0,1\}} \prod_{i=1,\, i \neq l} p(h_i|x)}}_{=1}\right]$$

$\because$ PMF

$$= \mathbb{E}_{p(x)}\left[\nabla_\theta f_\theta^T(x) \sum_{l=1}^{K} w_{l,:} \sum_{h_l \in \{0,1\}} p(h_l|x) h_l\right]$$

$$= \mathbb{E}_{p(x)}\left[\nabla_\theta f_\theta^T(x) \sum_{l=1}^{K} w_{l,:} \frac{\exp(-\alpha_l)}{1 + \exp(-\alpha_l)}\right]$$

Subject

Date 22

Attendants

$$\mathbb{E}_{p(x,h)}\left[\nabla_\theta f_\theta^T(x)\, w^T h\right]$$

$$= \mathbb{E}_{p(x)}\left[\nabla_\theta f_\theta^T(x) \sum_{l=1}^{K} w_{l,i}^T \frac{\exp(-\alpha_l)}{1+\exp(-\alpha_l)}\right]$$

$$\simeq \frac{1}{M}\sum_{i=1}^{M} \nabla_\theta f_\theta(x_i) \sum_{l=1}^{K} w_{l,i}^T \frac{\exp(-\alpha_l^i)}{1+\exp(-\alpha_l^i)}$$

where $\alpha_l^i = e_l^T w f_\theta(x_i) - c_l$

1. Draw $x_i \sim p(x)$ as Monte Carlo
2. Compute $\nabla_\theta f_\theta(x_i)$
3. Compute $\{\alpha_l^i\}_{l=1}^{K}$ as sequence
4. Take linear combination

$\uparrow$

Ans Q5