B51503|
김인제

(a) $-\sum_{w \in Vocab} y_w \log \hat{y}_w = -y_o \log \hat{y}_o = -\log \hat{y}_o$

(y는 one-hot vector)

(b) $\frac{\partial J_{\text{naive-softmax}}}{\partial v_c}$ 를 구하라. $J_{naive-softmax}(v_c, o, U)$

sol) $J = -\log \hat{y}_o = -\log \frac{\exp(u_o^T v_c)}{\sum_{w \in V} \exp(u_w^T v_c)} = -\left[u_o^T v_c - \log \sum_{w \in V} e^{u_w^T v_c}\right]$

$\frac{\partial J}{\partial v_c} = -\left[u_o - \frac{\sum_{t \in V} e^{u_t^T v_c} \cdot u_t}{\sum_{w \in V} e^{u_w^T v_c}}\right] = \sum_{t \in V} \hat{y}_t \cdot u_t - u_o$, $\left(u_o = U \cdot y,\ \sum_{t \in V} u_t \cdot \hat{y}_t = U^T \cdot \hat{y}\right)$

$= U^T(\hat{y} - y)$

$\left(\frac{\partial J}{\partial v_c} = \frac{\partial J}{\partial z} \cdot \frac{\partial z}{\partial v_c}\right.$, $(z = U^T v_c)$

$-\left[1 - \frac{\sum_{t \in V} e^{u_t^T v_c}}{\sum_{w \in V} e^{u_w^T v_c}}\right] \circ U = -\left[u_o - \frac{\sum_{t \in V} e^{u_t^T v_c} \cdot u_t}{\sum_{w \in V} e^{u_w^T v_c}}\right]$

scalar 가 아니라 one-hot vector로 보면 된다.

chain rule, $\frac{\partial J}{\partial z} = \delta$

(c) $\frac{\partial J_{\text{naive-softmax}}}{\partial u_w}$ 를 구하라.

sol) by chain rule, we can use δ.

$\frac{\partial J}{\partial u_w} = \delta \cdot \frac{\partial z}{\partial u_w}$, $\frac{\partial z}{\partial u_w} = v_c$

if $w = o$, $-\left[v_c - \frac{e^{u_w^T v_c} \cdot v_c}{\sum_{t \in V} e^{u_t^T v_c}}\right] = -v_c\left[1 - \hat{y}_w\right]$

$= -v_c\left[y - \hat{y}_w\right] = v_c \cdot (\hat{y}_w - y)$

if $w \neq o$, $-\left[0 - \frac{e^{u_w^T v_c} \cdot v_c}{\sum_{t \in V} e^{u_t^T v_c}}\right] = v_c \cdot \hat{y}_w$

(d) $\sigma(x) = \frac{1}{1+e^{-x}} = \frac{e^x}{e^x+1}$ x가 scalar로 1변수 함수.

$$\frac{d\sigma(x)}{dx} = \frac{e^x(e^x+1) - e^x \cdot e^x}{(e^x+1)^2} = \frac{e^x}{(e^x+1)^2} = \underbrace{\frac{e^x}{e^x+1}}_{\sigma(x)} \cdot \underbrace{\frac{1}{e^x+1}}_{1-\sigma(x)}$$

$$= \sigma(x)(1-\sigma(x))$$

(f) $\boldsymbol{J}(\boldsymbol{v}_c, w_{t+j}, \boldsymbol{U})$는 $\boldsymbol{J}_{\text{naive-softmax}}(\boldsymbol{v}_c, w_{t+j}, \boldsymbol{U})$ 으로 가정하자 (참고: Negative Sampling 의 loss 도 같은 방법으로 구할 수 있다) . 이 때 $\boldsymbol{J}_{\text{skip-gram}}$의 $\boldsymbol{U}$, $\boldsymbol{v}_c$, $\boldsymbol{v}_w$ 에 대한 편미분을 각각 구하라.

(i) $\partial \boldsymbol{J}_{\text{skip-gram}}(\boldsymbol{v}_c, w_{t-m}, \ldots, w_{t+m}, \boldsymbol{U}) / \partial \boldsymbol{U}$

(ii) $\partial \boldsymbol{J}_{\text{skip-gram}}(\boldsymbol{v}_c, w_{t-m}, \ldots, w_{t+m}, \boldsymbol{U}) / \partial \boldsymbol{v}_c$ $(w_t = c)$

(iii) $\partial \boldsymbol{J}_{\text{skip-gram}}(\boldsymbol{v}_c, w_{t-m}, \ldots, w_{t+m}, \boldsymbol{U}) / \partial \boldsymbol{v}_w$ when $w \neq c$

답은 $\partial \boldsymbol{J}(\boldsymbol{v}_c, w_{t+j}, \boldsymbol{U}) / \partial \boldsymbol{U}$ 와 $\partial \boldsymbol{J}(\boldsymbol{v}_c, w_{t+j}, \boldsymbol{U}) / \partial \boldsymbol{v}_c$ 를 이용하여 기술하라.

$$J(v_c, w_{t-m}, \cdots, w_{t+m}, U)$$

$$= -\log P(w_{t-m}, \cdots, w_{t+m} \mid w_t; U)$$

$$= -\log \prod_{\substack{-m \le j \le m \\ (j \neq 0)}} P(w_{t+j} \mid w_t; U) \qquad (\because c = w_t)$$

$$= \sum_{\substack{-m \le j \le m \\ (j \neq 0)}} -\log \frac{w_{t+j}^T v_c}{\sum_{i \in U} w_w^T v_c} = \sum_{\substack{-m \le j \le m \\ (j \neq 0)}} J(v_c, w_{t+j}, U)$$

using this, if $J_{\text{skip-gram}}(v_c, w_{t-m}, \cdots, w_{t+m}, U) = F$,

(i) $\partial F / \partial U = \sum_{\substack{-m \le j \le m \\ (j \neq 0)}} \partial J(v_c, w_{t+j}, U) / \partial U$

(ii) $\partial F / \partial v_c = \sum_{\substack{-m \le j \le m \\ (j \neq 0)}} \partial J(v_c, w_{t+j}, U) / \partial v_c$

(iii) $\partial F / \partial v_w = \sum_{\substack{-m \le j \le m \\ (j \neq 0)}} \partial J(v_w, w_{t+j}, U) / \partial v_w$ $(w \neq c)$

there's no v_w. $\Rightarrow$ $\therefore 0$