Homework2

Xiong Jiao

October 30, 2019

(a)

As above known,

$$y_w = \begin{cases} 1, & k=o \\ 0, & k \neq o \end{cases}$$

So,

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

(b)

$$\frac{\partial J\left(v_c, o, U\right)}{\partial v_c} = \frac{\partial(-\log P(O=o \mid C=c)}{\partial v_c}$$

$$= \frac{\partial\left(-\log \frac{\exp\left(u_o^\top v_c\right)}{\sum_{\mathrm{w} \in \mathrm{Vocab}} \exp\left(u_w^\top v_c\right)}\right)}{\partial v_c}$$

$$= -\frac{\partial\left(\log \exp\left(u_o^\top v_c\right)\right)}{\partial v_c} + \frac{\partial\left(\log \sum_{W \in Vocab} \exp\left(u_w^\top v_c\right)\right)}{\partial v_c}$$

$$= -\frac{\exp(u_o^\top v_c) u_o^\top}{\exp(u_o^\top v_c)} + \frac{\partial\left(\log \sum_{W \in Vocab} \exp\left(u_w^\top v_c\right)\right)}{\partial v_c}$$

$$= -u_o^\top + \frac{\sum_{x \in Vocab} \exp\left(u_x^\top v_c\right) u_x^T}{\sum_{W \in Vocab} \exp\left(u_w^\top v_c\right)}$$

$$= -u_o^\top + \sum_{x \in Vocab} P(O = x \mid C = c) u_x^T$$

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

(c)

$$\frac{\partial J\left(v_c, o, U\right)}{\partial u_w} = \frac{\partial(-\log P(O = o \mid C = c)}{\partial u_w}$$

$$= \frac{\partial\left(-\log \frac{\exp\left(u_o^\top v_c\right)}{\sum_{w \in Vocab} \exp\left(u_w^\top v_c\right)}\right)}{\partial u_w}$$

$$= -\frac{\partial\left(\log \exp\left(u_o^\top v_c\right)\right)}{\partial u_w} + \frac{\partial\left(\log \sum_{W \in Vocab} \exp\left(u_w^\top v_c\right)\right)}{\partial u_w}$$

$$when\ \ w = o,$$

$$\frac{\partial J\left(v_c, o, U\right)}{\partial u_w} = -v_c + \frac{1}{\sum_{w \in Vocab} \exp\left(u_w^\top v_c\right)} * \frac{\partial\left(\sum_{x \in Vocab} \exp\left(u_x^\top v_c\right)\right)}{\partial u_o}$$

$$= -v_c + \frac{1}{\sum_{w \in Vocab} \exp\left(u_w v_c\right)} * \exp\left(u_o^\top v_c\right) * \frac{\partial\left(u_o^\top v_c\right)}{\partial u_o}$$

$$= -v_c + \frac{\exp\left(u_o^\top v_c\right)}{\sum_{w \in Vocab} \exp\left(u_w^\top v_c\right)} * v_c$$

$$= -v_c + P(O = o \mid C = c)^* v_c$$

$$= (P(O = o \mid C = c) - 1)^* v_c = (\hat{y} - y)^* v_c (y = 1)$$

$$when\ \ w \neq o,$$

$$\frac{\partial J\left(v_c, o, U\right)}{\partial u_w} = \frac{1}{\sum_{w \in Vocab} \exp\left(u_w^\top v_c\right)} * \frac{\partial\left(\sum_{x \in Vocab} \exp\left(u_x^\top v_c\right)\right)}{\partial u_w}$$

$$= \frac{1}{\sum_{w \in Vocab} \exp\left(u_w^\top v_c\right)} * \exp\left(u_w^\top v_c\right) * \frac{\partial\left(u_w^\top v_c\right)}{\partial u_w}$$

$$= \frac{1}{\sum_{w \in Vocab} \exp\left(u_w^\top v_c\right)} * \exp\left(u_w^\top v_c\right) * v_c$$

$$= \frac{\exp\left(u_w^\top v_c\right)}{\sum_{w \in Vocab} \exp\left(u_w^\top v_c\right)} * v_c$$

$$= \hat{y} v_c = (\hat{y} - y) v_c (y = 0)$$

(d)

$$\frac{\partial \sigma(x)}{\partial x} = \frac{\partial\left(\frac{1}{1+e^{-x}}\right)}{\partial x}$$

$$= -\frac{1}{\left(1+e^{-x}\right)^2} * \frac{\partial\left(1+e^{-x}\right)}{\partial x}$$

$$= \frac{e^{-x}}{\left(1+e^{-x}\right)^2}$$

$$= \frac{1}{1+e^{-x}} * \frac{e^{-x}}{1+e^{-x}}$$

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