a2 written part

peter

2022 年 8 月 24 日

(a) notice that

$$\hat{y}_x = \frac{\exp(u_x^T v_c)}{\sum\limits_{w \in \text{vocab}} \exp(u_w^T v_c)} \tag{1}$$

then

$$\begin{aligned} J_{\text{naive-softmax}} &= -\log P(O = o, C = c) &(2)\\ &= -\log \frac{\exp(u_o^T v_c)}{\sum\limits_{w \in \text{vocab}} \exp(u_w^T v_c)} &(3)\\ &= -\log \hat{y}_o &(4)\\ &= -\sum_{w \in \text{vocab}} y_w \log \hat{y}_w &(5) \end{aligned}$$

(b)

$$\frac{\partial J}{\partial v_c} = U(\hat{y} - y) \tag{6}$$

(c)(d)

$$\frac{\partial J}{\partial U} = (\hat{y} - y) v_c \tag{7}$$

(e)

$$f'(x) = \mathbf{1}(x > 0) \tag{8}$$

(f)

$$\sigma'(x) = \sigma(x)(1 - \sigma(x)) \tag{9}$$

(g)

$$J = -\log(\sigma(u_o^T v_c)) - \sum_{s=1}^{K} \log(\sigma(-u_{w_s}^T v_c)) \tag{10}$$

$$\frac{\partial J}{\partial v_c} = -(1-\sigma(u_o^T v_c))u_o + \sum_{s=1}^{K}(1-\sigma(-u_s^T v_c))u_s \tag{11}$$

$$\frac{\partial J}{\partial u_o} = (\sigma(u_o^T v_c) - 1)v_c \tag{12}$$

$$\frac{\partial J}{\partial u_{w_s}} = -(\sigma(u_{w_s}^T v_c) - 1)v_c \tag{13}$$

(h)

$$\frac{\partial J}{\partial u_{w_s}} = \sum_{x=1}^{k} \mathbf{1}(x = w_s)(1-\sigma(-u_{w_s}^T v_c))v_c \tag{14}$$

(i)

$$J = \sum_{-m \leq j \leq m, j \neq 0} J(v_c, w_{t+j}, U) \tag{15}$$

$$\frac{\partial J}{\partial U} = \sum_{-m \leq j \leq m, j \neq 0} \frac{\partial J(v_c, w_{t+j}, U)}{\partial U} \tag{16}$$

$$\frac{\partial J}{\partial v_c} = \sum_{-m \leq j \leq m, j \neq 0} \frac{\partial J(v_c, w_{t+j}, U)}{\partial v_c} \tag{17}$$

$$\frac{\partial J}{\partial v_w} = 0 \tag{18}$$