1 שאלה - NLP

$$\text{Softmax}(x+c)_i = \frac{e^{x_i+c}}{\sum_j e^{x_j+c}} = \frac{e^{x_i}e^{c}}{e^{c}\sum e^{x_j}} = \text{softmax}(x)_i \quad (1)(2)$$

$$\sum_{w\in W} y_w \log \hat{y}_w = 0+0+\dots+0+1\cdot\log \hat{y}_o+\dots+0 = \quad (2)$$

$$= \log \hat{y}_o = P(o=o \mid c=c)$$

$$\frac{J_{naive\text{-}sm}(v_c,o,U)}{\partial v_c} = -\frac{\partial}{\partial v_c}\left(-\log \frac{e^{u_o^T v_c}}{\sum_{w\in W} e^{u_w^T v_c}}\right) = \quad (3)$$

$$= \frac{\partial}{\partial v_c}\left(-u_o^T v_c + \log \sum e^{u_w^T v_c}\right) = -u_o + \frac{1}{\sum_{w\in W} e^{u_w^T v_c}} \cdot \sum_{w\in W} \frac{\partial}{\partial v_c}\left(e^{u_w^T v_c}\right) =$$

$$= -u_o + \frac{1}{\sum_{w\in W} e^{u_w^T v_c}} \sum_{k\in W} e^{u_k^T v_c}\cdot u_k = \sum_{w} \frac{e^{u_k^T v_c}}{\sum_{w} e^{u_w^T v_c}}\cdot u_k - u_o =$$

$$= \sum P(k\mid c)\, u_k - u_o = \sum_k \hat{y}_k u_k - u_o$$

$$\frac{\partial}{\partial u_w} J_{naive\text{-}sm}(v_c,o,U) = \frac{\partial}{\partial u_w}\left(-u_o^T v_c + \log \sum_{k\in w} e^{u_k^T v_c}\right) = \quad (3)$$

$$= -v_c \cdot \underbrace{1}_{w=o} \cdot + \frac{1}{\sum_x e^{u_x^T v_c}} e^{u_w^T v_c}\cdot v_c =$$

$$= v_c\left(P(w\mid c) - 1\right) = v_c(\hat{y}_w - y_w)$$

$$\sigma'(x) = \frac{\partial}{\partial x}\left((1+e^{-x})^{-1}\right) = -(1+e^{-x})^{-2}(-e^{-x}) = \quad (4)$$

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

$$\frac{\partial J_{naive\text{-}sm}(v_c,o,u)}{\partial v_c} = -\frac{1}{\sigma(u_o^T v_c)}\sigma(u_o^T v_c)\left(1-\sigma(u_o^T v_c)\right)u_o - \quad (1)$$

$$-\sum_x \frac{1}{\sigma(-u_x^T v_c)}\cdot\sigma(-u_x^T v_c)\left(1-\sigma(-u_o^T v_c)\cdot(-u_x)\right) =$$

$$= u_o\left(\sigma(u_o^T v_c)-1\right) + \sum_x u_x\left(1-\sigma(-u_x^T v_c)\right)$$

$$\frac{\partial J_{neg\text{-}sample}(v_c, o, U)}{\partial u_o} = -\frac{1}{\sigma(u_o^T v_c)} \sigma(u_o^T v_c)\left(1-\sigma(u_o^T v_c)\right) \cdot v_c$$

$$= -v_c\left(1-\sigma(u_o^T v_c)\right)$$

$$\frac{\partial J_{neg\text{-}samp}(v_c, o, U)}{\partial u_k} = -\frac{1}{\sigma(-u_k^T v_c)} \sigma(-u_k^T v_c)\left(1-\sigma(-u_k^T v_c)\right) \cdot (-v_c)$$

$$= v_c\left(1-\sigma(-u_k^T v_c)\right) = v_c\, \sigma(u_k^T v_c)$$

בגישה הנאיבית צריך לעבור על כל המילון כדי לנרמל את ההסתברויות המותנות וכו' כאן לא - לכן זה יותר יעיל.

$$\frac{\partial J_{skip\text{-}gram}(v_c, w_{t-m}, \ldots w_{t+m}, U)}{\partial U} = \sum_{\substack{-m \le j \le m \\ j \ne 0}} \frac{\partial}{\partial U} J(v_c, w_{t+j}, U)$$ (1) (5)

$$\frac{\partial J_{sg}(v_c, w_{t-m}, \ldots w_{t+m}, U)}{\partial v_c} = \sum_{\substack{-m \le j \le m \\ j \ne 0}} \frac{\partial}{\partial v_c} J(v_c, w_{t+j}, U)$$ (2)

$$\left.\frac{\partial J_{sg}(v_c, w_{t-m}, \ldots w_{t+m}, U)}{\partial v_w}\right|_{w \ne c} = 0$$ (3)

$$L(\theta) = \prod_{t=1}^{T} \prod_{j=-m}^{m} p_\theta(w_{t+j} | w_t)$$

$$\arg\max_\theta L(\theta) = \arg\max_\theta \prod_{t=1}^{T} \prod_{j=-m}^{m} p_\theta(w_{t+j} | w_t)^{\#(w_{t+j}, w_t)}$$

$$= \arg\max_\theta \sum_{t=1}^{T} \sum_{j=-m}^{m} \#(w_{t+j}, w_t) \cdot \log p_\theta(w_{t+j} | w_t)$$

$$= \arg\max_\theta \sum_{j=-m}^{m} \#(w_{t+j}, w_t) \log p_\theta(w_{t+j} | w_t)$$

$$\arg\max_\theta \left[\sum_{j=-m}^{m} \#(w_{t+j}, w_t) \cdot \log p_\theta(w_{t+j} | w_t) \right] - \lambda \left(1 - \sum_{j=-m}^{m} p_\theta(w_{t+j} | w_t) \right)$$

$$\nabla_{p(o|c)} \Rightarrow \frac{\#(o,c)}{p_\theta(o|c)} - \lambda = 0$$

$$\Rightarrow p_\theta(o|c) = \frac{\#(o,c)}{\lambda}$$

$$\sum_{i=-m}^{m} p_\theta(o_i | c) = 1 = \sum_{i=-m}^{m} \frac{\#(o_i, c)}{\lambda} \Rightarrow \lambda = \sum_{i=-m}^{m} \#(o_i, c)$$

$$p_\theta(o|c) = \frac{\#(o,c)}{\sum_{i=-m}^{m} \#(o_i, c)}$$

{house, bank}

"house house house"
"bank bank bank"

corpus :

$p(\text{house} \mid \text{house}) = 1$
$p(\text{house} \mid \text{bank}) = 0$
$p(\text{bank} \mid \text{bank}) = 1$
$p(\text{bank} \mid \text{house}) = 0$

$$p(o \mid c) = \frac{\exp(u_o^T \cdot v_c)}{\sum_w \exp(u_w^T \cdot v_c)}$$

skip gram

c = bank o = house

$\exp(x) > 0$

$$p(o \mid c) = \frac{\exp(u_o^T v_c)}{\sum \exp(u_w^T \cdot v_c)} > 0$$

$p(\text{house} \mid \text{bank}) \neq 0$

5.

$$relu(\vec{X}) = max(0, \vec{X}) \Rightarrow \geq 0$$

$$relu(\vec{x_1})^T \cdot relu(\vec{x_2}) \geq 0$$

thereshold = 0.5

לפי הנתון $sigmoid(0)=0.5$ מוגדר, הקודם ולכן

$$\forall \vec{x_1}, \vec{x_2} \quad sigmoid(relu(\vec{x_1}) \cdot relu(\vec{x_2})) \geq 0.5$$

והמודל יחזיר 1 לכל x_1, x_2 כי הם פראפרזות.

לכן יהיה הדיוק שלו ב: 1 ⇐ שלל זוג x_1, x_2 | פראפרזה
המודל יחזיר תשובה נכונה. כמו שלל זוג x_1, x_2 שאינם פראפרזה.
המודל ישגה. ולכן

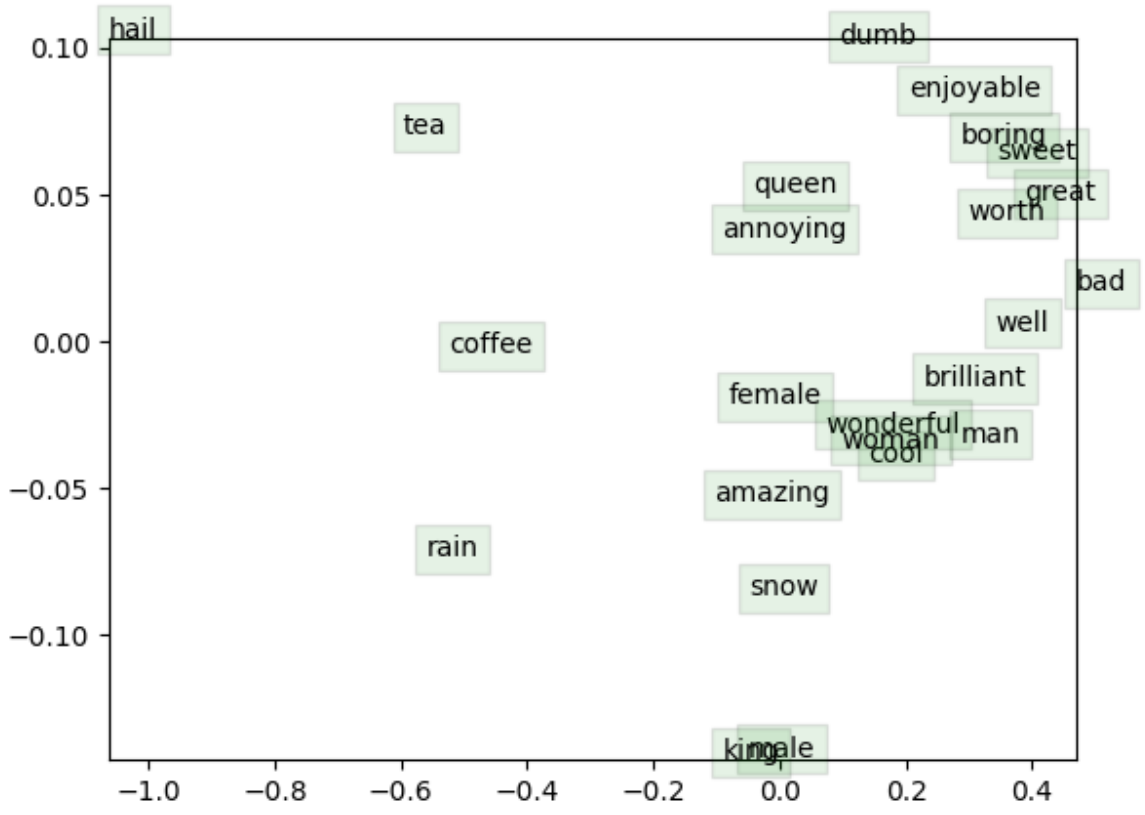
$$Acc(M) = 0.25$$

ב: נגדיר:

$$p(paraphrase | x_1, x_2) = sigmoid(\vec{X_1}^T \cdot \vec{X_2})$$

כי אם $\vec{X_1}^T \cdot \vec{X_2} > 0$ המודל יחזיר 1,
אחרת 0.

ולכן הווקטורים $\vec{x_1}$ ו- $\vec{x_2}$ צריכים להיות קרובים במרחב
וקטורי כדי שהמודל יותר לזהות פרפרזה.



There is somewhat of a cluster for all the adjectives. Also, ‘king’ and ‘male’ are clustered in a meaningful manner (though ‘queen’ and ‘female’ are not), etc.