Homework 2

Michael Pollack

April 13, 2024

Q1

$2|V| * d_e$
Each word in the vocabulary is featured twice, in vectors of dimension d_e

Q2

—What is the range of $\sigma(x)$?—
The range is (0, 1)

—How is it used in the SGNS model?—
It is used to "squash" the values of the model so that they approach 1 for two similar vectors and approach 0 for two dissimilar vectors.

—Compute $\frac{d\sigma}{dx}$; show your work.—
$\sigma(x) = \frac{1}{1+e^{-x}}$

let $u(x) = 1 + e^{-x}$

$\frac{du}{dx} = e^{-x}$

$\frac{d\sigma}{dx} = \frac{d}{dx}(\frac{1}{u(x)})$

$= -\frac{1}{u(x)^2} * \frac{du}{dx}$

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

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

$\therefore$

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

Simplifying further, back to terms of $\sigma(x)$:

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

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

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

$\therefore$

$\frac{du}{dx} = \sigma(x)(1 - \sigma(x))$

Q3

—Rewrite this loss in terms of the parameter matrices E and C—
$L_{CE} = -\log \sigma(E_w \cdot C_{c+}) - \sum_{i=1}^{k} \log(1 - \sigma(E_w \cdot C_{c-i}))$

—Using the chain rule, compute $\frac{d}{dx}(-\log \sigma(x))$.—
First, we prove what is shown in the hint:

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

$\frac{1}{1+e^{-x}} * \frac{e^x}{e^x} = \frac{e^x}{e^x+e^0}$

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

$\therefore$

$\sigma(x) = \frac{e^x}{e^x+1}$

Now, we apply this to the chain rule to find the derivative:

$\frac{d}{dx}(-\log(\sigma(x))) = -\frac{1}{\sigma(x)} * \frac{d}{dx}(\sigma(x))$

$= -\frac{e^x+1}{e^x} * \frac{d}{dx}(\frac{e^x}{e^x+1})$

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

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

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

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

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

Now, we put this in terms of $\sigma(x)$:

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

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

$\therefore$

$\frac{d}{dx}(-\log(\sigma(x))) = \sigma(x) - 1$

—Show that $\nabla_x x \cdot y = y$ (where $x \cdot y$ is the dot product of two vectors)—

$\nabla_x x \cdot y = \langle \frac{\partial(x \cdot y)}{\partial x_1}, \frac{\partial(x \cdot y)}{\partial x_2}, ..., \frac{\partial(x \cdot y)}{\partial x_n} \rangle$

$\frac{\partial(x \cdot y)}{\partial x_i} = \frac{\partial(x_1 y_1 + x_2 y_2 + ... + x_n y_n)}{\partial x_i}$

In calculating this partial derivative, all x_j where $j \neq i$ are not dependent on x_i and therefore are treated as constants, giving them a derivative of 0. This means:

$\frac{\partial(x \cdot y)}{\partial x_i} = \frac{\partial(x_i y_i)}{\partial x_i}$

$= y_i$

$\therefore$

$\nabla_x x \cdot y = \langle y_1, y_2, ..., y_n \rangle$

$= y$

—Compute $\nabla_{C_{e_+}} L_{CE}$—

$\nabla_{C_{e_+}} L_{CE} = \frac{\partial}{\partial C_{e_+1}}(-\log \sigma(E_w \cdot C_{e+}) - \sum_{i=1}^{k} \log(1 - \sigma(E_w \cdot C_{c-i}))), ..., \frac{\partial}{\partial C_{e_+n}}(-\log \sigma(E_w \cdot C_{c+}) - \sum_{i=1}^{k} \log(1 - \sigma(E_w \cdot C_{c-i})))$

Using the chain rule shown above:

$\frac{\partial}{\partial C_{e_+i}}(-\log \sigma(E_w \cdot C_{e_+i}) - \sum_{i=1}^{k} \log(1 - \sigma(E_w \cdot C_{c-i}))) = (\sigma(E_w \cdot C_+) - 1)(E_w) - 0$

$= (\sigma(E_w \cdot C_+) - 1)(E_w)$

—Compute $\nabla_{C_{e_{-i}}} L_{CE}$—

$\nabla_{C_{e_{-i}}} L_{CE} = \frac{\partial}{\partial C_{e_{-i}1}}(-\log \sigma(E_w \cdot C_{c+}) - \sum_{i=1}^{k} \log(1 - \sigma(E_w \cdot C_{c-i}))), ...$

$\frac{\partial}{\partial C_{e_{-i}}}(-\log \sigma(E_w \cdot C_+) - \sum_{i=1}^{k} \log(1 - \sigma(E_w \cdot C_{c-i}))) = 0 - \sum_{i=1}^{k} \frac{1}{1-\sigma(E_w \cdot C_{c-i})} * (-(1 - \sigma(E_w \cdot C_{c-i}))) * \sigma(E_w \cdot C_{c-i}) * E_w$

$= -\sum_{i=1}^{k} -\sigma(E_w \cdot C_{c_{-i}})(E_w)$

$= \sum_{i=1}^{k} \sigma(E_w \cdot C_{c_{-i}})(E_w)$

—Compute $\nabla_{E_w} L_{CE}$—

$\nabla_{E_w} L_{CE} = \frac{\partial}{\partial E_w}(-\log \sigma(E_w \cdot C_{c+}) - \sum_{i=1}^{k} \log(1 - \sigma(E_w \cdot C_{c-i}))), ...$

$\frac{\partial}{\partial E_w}(...) = (\sigma(E \cdot C_{c+}) - 1)C_{c+}) - \sum_{i=1}^{k} \frac{1}{1-\sigma(E_w \cdot C_{c-i})} * (-(1 - \sigma(E_w \cdot C_{c-i})) * \sigma(E_w \cdot C_{c-i})) * C_{c-i}$

$= (\sigma(E \cdot C_{c+}) - 1)C_{c+}) - \sum_{i=1}^{k} -\sigma(E \cdot C_{c-i})C_{c-i}$

$= (\sigma(E \cdot C_{c+}) - 1)C_{c+}) + \sum_{i=1}^{k} \sigma(E \cdot C_{c-i})C_{c-i}$

Q4

Runtime: 8036.447621107101

Findings:

Looking at the vectors produced by this plot, one can see that most of the words appear to be grouped towards the lower left, with the main outliers being “amazing”, “director”, and arguably “woman”. It is interesting to see as well that there are a few cases where seeming opposites received very similar placements, such as “boring” next to ”cool”, “enjoyable” next to “dumb” (I guess not very mutually exclusive there), and “sweet” next to “annoying”.

Plot: (See next page)

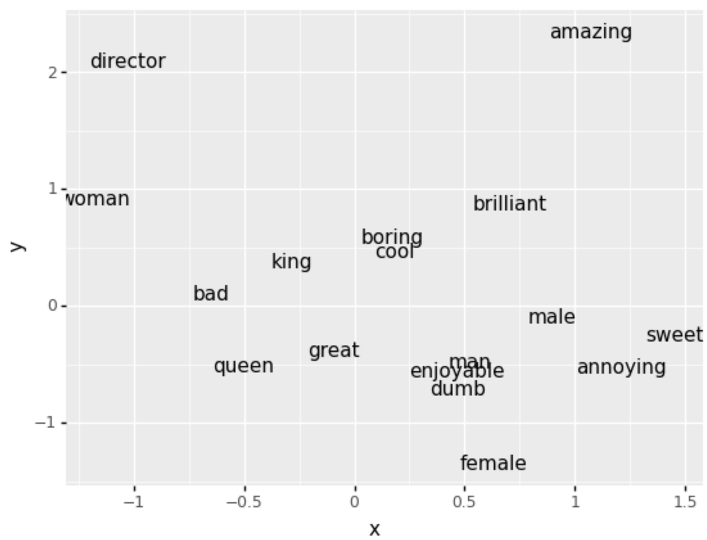

Figure 1: Generated plot