CS 584-A: Assignment 2 Word Vectors

Jeremiah McReynolds

March 2, 2021

1 Written

1. Softmax

$$\begin{aligned}\sigma(x_i + c) &= \frac{e^{x_i+c}}{\sum_j e^{x_j+c}} \\ &= \frac{e^{x_i}e^c}{\sum_j e^{x_j}e^c} \\ &= \frac{e^{x_i}e^c}{e^c\sum_j e^{x_j}} \\ &= \frac{e^{x_i}}{\sum_j e^{x_j}}\frac{e^c}{e^c} \\ &= \frac{e^{x_i}}{\sum_j e^{x_j}}\end{aligned}$$

2. Sigmoid

$$\begin{aligned}\sigma(x) &= \tfrac{1}{1+e^{-x}} = (1+e^{-x})^{-1} \\ \tfrac{\delta\sigma(x)}{\delta x} &= -(1+e^{-x})^{-2}(-e^{-x}) \\ &= -\tfrac{1}{(1+e^{-x})^2}(-e^{-x}) \\ &= \tfrac{e^{-x}}{(1+e^{-x})^2} \\ &= \tfrac{1}{1+e^{-x}}\tfrac{e^{-x}}{1+e^{-x}} \\ &= \sigma(x)(\tfrac{e^{-x}}{1+e^{-x}}) \\ &= \sigma(x)(\tfrac{e^{-x}+1-1}{1+e^{-x}}) \\ &= \sigma(x)(\tfrac{e^{-x}+1}{1+e^{-x}} - \tfrac{1}{1+e^{-x}}) \\ &= \sigma(x)(\tfrac{1+e^{-x}}{1+e^{-x}} - \tfrac{1}{1+e^{-x}}) \\ \tfrac{\delta\sigma(x)}{\delta x} &= \sigma(x)(1-\sigma(x))\end{aligned}$$

3. Word2vec

(a) $\frac{\delta J(o,v_c,U)}{\delta v_c} = U^T(\hat{y} - y)$

(b) $\frac{\delta J(o,v_c,U)}{\delta u_w} = v_c(\hat{y} - y)^T$

To start, we need to derive the softmax function. For simplicity, let's define $\hat{y}_i$ as the following:

$$\hat{y}_i = \frac{\exp(f_i)}{\sum_w \exp(f_w)} \tag{1}$$

where

$$f_i = u_i^T v_c \tag{2}$$

With this, its derivative with respect to its input is as follows:

$$\frac{\delta \hat{y}_i}{\delta f_i},\ c = i$$

$$\begin{gathered}
\frac{\exp(f_i)\sum_w \exp(f_w) - \exp(f_i)\exp(f_i)}{[\sum_w \exp(f_w)]^2} \\
\frac{\exp(f_i)(\sum_w \exp(f_w) - \exp(f_i))}{[\sum_w \exp(f_w)]^2} \\
\frac{\exp(f_i)}{\sum_w \exp(f_w)} \frac{\sum_j \exp(f_j) - \exp(f_i)}{\sum_w \exp(f_w)} \\
\hat{y}_i\left(\frac{\sum_w \exp(f_w)}{\sum_w \exp(f_w)} - \frac{\exp(f_i)}{\sum_w \exp(f_w)}\right) \\
\hat{y}_i(1 - \hat{y}_i)
\end{gathered}$$

$$\frac{\delta \hat{y}}{\delta f_i},\ c \neq i$$

$$\begin{gathered}
\frac{\delta \hat{y}}{\delta v_c} = \frac{0 - \exp(f_c)\exp(f_i)}{[\sum_w \exp(f_w)]^2} \\
\frac{-\exp(f_c)\exp(f_i)}{[\sum_w \exp(f_w)]^2} \\
\frac{\exp(f_c)}{\sum_w \exp(f_w)} \frac{\exp(f_i)}{\sum_w \exp(f_w)} \\
-\hat{y}_c\hat{y}_i
\end{gathered}$$

Plugging its derivative into the cross-entropy function, we can find its derivative with respect to f_i:

$$\begin{gathered}
J(o, v_c, U) = -\sum_{i=1}^{i} y_i log(\hat{y}_i) \\
= -\sum_i y_i \frac{\delta log(\hat{y}_i)}{\delta f_i} \\
= -\sum_i y_i \frac{1}{\hat{y}_i} \frac{\delta \hat{y}_i}{\delta f_i} \\
= -y(1 - \hat{y}) - \sum_{c!=i} y_i \frac{1}{\hat{y}_i}(-\hat{y}\hat{y}_i) \\
= -y + y\hat{y} + \sum_{c!=i} y_i\hat{y} \\
= y\hat{y} + \sum_{c!=i} y_i\hat{y} - y \\
= \hat{y}(y_i + \sum_{c!=i} y_i) - y \\
= \hat{y} - y
\end{gathered}$$

So, the derivative of cross-entropy with respect to f_i is as follows:

$$\frac{\delta J(o, v_c, U)}{\delta f_i} = \hat{y} - y \tag{3}$$

Armed with this equation, it is easy to see the derivative of cross-entropy loss with respect to v_c and u_w.

$$\begin{array}{c|c}
\frac{\delta J(o,v_c,U)}{\delta v_c} & \frac{\delta J(o,v_c,U)}{\delta f_i}\frac{\delta f_i}{\delta v_c} \\
 & (\hat{y} - y)U \\
 & U^T(\hat{y} - y) \\
 & \\
\frac{\delta J(o,v_c,U)}{\delta u_w} & \frac{\delta J(o,v_c,U)}{\delta f_i}\frac{\delta f_i}{\delta u_w} \\
 & (\hat{y} - y)v_c \\
 & v_c(\hat{y} - y)^T
\end{array}$$

(c) The derivatives for each term of $J_{neg-sample}$ is as follows:

i. $\frac{\delta J_{neg-sample}}{\delta v_c} = (\sigma(u_o^T v_c) - 1)u_o - \sum_{k=1}^{K}(\sigma(-u_k^T v_c) - 1)u_k$

$$\begin{gathered}
\frac{\delta J_{neg-sample}}{\delta v_c} = -\frac{\delta J}{\delta \sigma_o}\frac{\delta \sigma_o}{\delta f_o}\frac{\delta f_o}{\delta v_c} - \sum_k \frac{\delta J}{\delta \sigma_k}\frac{\delta \sigma_k}{\delta f_k}\frac{\delta f_k}{\delta v_c} \\
= -\frac{1}{\sigma_o}\sigma_o(1 - \sigma_o)u_o - \sum_k \frac{1}{\sigma_k}\sigma_k(1 - \sigma_k) - u_k \\
= -(1 - \sigma_o)u_o - \sum_k(1 - \sigma_k) - u_k \\
= (\sigma_o - 1)u_o - \sum_k(\sigma_k - 1)u_k \\
= (\sigma(u_o^T v_c) - 1)u_o - \sum_k(\sigma(-u_k^T v_c) - 1)u_k
\end{gathered}$$

ii. $\frac{\delta J_{neg-sample}}{\delta u_o} = (\sigma(u_o^T v_c) - 1)v_c$

$$\begin{gathered}
\frac{\delta J_{neg-sample}}{\delta u_o} = -\frac{\delta J}{\delta \sigma_o}\frac{\delta \sigma_o}{\delta f_o}\frac{\delta f_o}{\delta u_o} - 0 \\
= -\frac{1}{\sigma_o}\sigma_o(1 - \sigma_o)v_c \\
= -(1 - \sigma_o)v_c \\
= (\sigma_o - 1)v_c \\
= (\sigma(u_o^T v_c) - 1)v_c
\end{gathered}$$

iii. $\frac{\delta J_{neg-sample}}{\delta u_k} = -\sum_{k=1}^{K}(\sigma(-u_k^T v_c) - 1)v_c$

$$\begin{gathered}
\frac{\delta J_{neg-sample}}{\delta u_k} = 0 - \sum_k \frac{\delta J}{\delta \sigma_k}\frac{\delta \sigma_k}{\delta f_k}\frac{\delta f_k}{\delta v_c} \\
= -\sum_k \frac{1}{\sigma_k}\sigma_k(1 - \sigma_k) - v_c \\
= -\sum_k(1 - \sigma_k) - v_c \\
= -\sum_k(\sigma_k - 1)v_c \\
= -\sum_k(\sigma(u_k^T v_c) - 1)v_c
\end{gathered}$$

iv. The derivatives are as follows:

A. $\frac{\delta J(o,v_c,U)}{\delta U} = \sum\limits_{-m \leq j \leq m, j \neq 0} \frac{\delta F(w_{c+j}, v_c)}{\delta U}$

B. $\frac{\delta J(o,v_c,U)}{\delta v_c} = \sum\limits_{-m \leq j \leq m, j \neq 0} \frac{\delta F(w_{c+j}, v_c)}{\delta v_c}$

C. $\frac{\delta J(o,v_c,U)}{\delta v_j} = \sum\limits_{-m \leq j \leq m, j \neq c} \frac{\delta F(w_{c+j}, v_c)}{\delta v_j} = 0$