Centro de Investigación en Matemáticas (CIMAT). Unidad Monterrey

Tarea 5: Problema 1

Victor Manuel Gómez Espinosa

30 de noviembre de 2020

1 a)

Recordando de la Tarea 1, haciendo backpropagation hacia la capa de salida $a_2 = \hat{y}$, con función softmax $\hat{y}_j = soft\max\left(z_{2,j}\right)$, donde $z_2 = U'a_1$ (a_1 es la capa oculta), tenemos lo siguiente: $dz = \frac{\partial L}{\partial z_2} = \frac{\partial L}{\partial \hat{y}}\frac{\partial \hat{y}}{\partial z_2} = \left(\hat{y} - y\right)$.

Ahora si para este caso sabemos que $z_{2,j} = u'_{w_j} v_{w_i}$, y su derivada parcial es $\frac{\partial z_{2,j}}{\partial v_{w_i}} = u'_{w_j}$ o $\frac{\partial z_2}{\partial v_{w_i}} = U'$, entonces $\frac{\partial L}{\partial v_{w_i}} = \frac{\partial L}{\partial \hat{y}}\frac{\partial \hat{y}}{\partial z_2}\frac{\partial z_2}{\partial v_{w_i}} = U\left(\hat{y} - y\right)$.

1 b)

De forma similar, su derivada parcial es $\frac{\partial z_{2,j}}{\partial u_{w_j}} = v_{w_i}$ o $\frac{\partial z_2}{\partial U} = v_{w_i}$, entonces $\frac{\partial L}{\partial U} = \frac{\partial L}{\partial \hat{y}}\frac{\partial \hat{y}}{\partial z_2}\frac{\partial z_2}{\partial U} = v_{w_i}\left(\hat{y} - y\right)'$.

1 c)

Si ahora se tiene la siguiente función de costo:

$L\left(v_{w_i}, u_{w_j}\right) = -\ln\left(\sigma\left(u'_{w_j} v_{w_i}\right)\right) - \sum_{k=1}^{K} \ln\left(\sigma\left(-u'_{w_k} v_{w_i}\right)\right)$, o con un cambio de variable:

$L(g) = -\ln(g) - \sum_{k=1}^{K} \ln(g)$, $g(f) = \sigma(f)$, $f\left(v_{w_i}, u_{w_j}\right) = u'_{w_j} v_{w_i}$ para las palabras en el contexto, o $f\left(v_{w_i}, u_{w_r}\right) = -u'_{w_r} v_{w_i}$ si no lo están, y sus respectivas derivadas son:

$L' = -\frac{g'}{g} - \sum \frac{g'}{g}$, $g' = g(1-g) f'$ donde $f' = \frac{\partial f}{\partial v_{w_i}} = u'_{w_j}$ o $f' = \frac{\partial f}{\partial u_{w_j}}$, en esta ultima hay dos posibles casos, donde la palabra esta en el contexto o no lo está, por lo tanto $f' = \frac{\partial f}{\partial u_{w_{j=j}}} = v_{w_i}$ o $f' = \frac{\partial f}{\partial u_{w_{j\neq j}}} = 0$

Entonces: $L' = -\left(1-\sigma(f)\right) f' - \sum\left(1-\sigma(f)\right) f'$,

Para la parte derecha hay dos opciones, si las palabras están en el contexto o no lo estan y sabiendo que $1-\sigma\left(-u'_{w_k} v_{w_i}\right) = \sigma\left(u'_{w_k} v_{w_i}\right)$:

$$L' = \begin{cases} \left(\sigma\left(u'_{w_j} v_{w_i}\right) - 1\right) f' + \sum_{k=1}^{k} \left(\sigma\left(u'_{w_k} v_{w_i}\right) - 1\right) f' \\ \left(\sigma\left(u'_{w_k} v_{w_i}\right)\right) f' + \sum_{k=1}^{k} \left(\sigma\left(u'_{w_k} v_{w_i}\right)\right) f' \end{cases} , si, \begin{matrix} k = j \\ k \neq j \end{matrix}$$

Y, por lo tanto:

$$\frac{\partial L}{\partial v_{w_i}} = \begin{cases} \sum_{k=1}^{k-1} \left(\sigma\left(u'_{w_k} v_{w_i}\right) - 1\right) u'_{w_j} \\ \sum_{k=1}^{k-1} \left(\sigma\left(u'_{w_k} v_{w_i}\right)\right) u'_{w_j} \end{cases} , si, \begin{matrix} k = j \\ k \neq j \end{matrix} ,$$

y

$$\frac{\partial L}{\partial u'_{w_j}} = \begin{cases} \left(\sigma\left(u'_{w_j} v_{w_i}\right) - 1\right) v_{w_i} \\ \left(\sigma\left(u'_{w_j} v_{w_i}\right)\right) v_{w_i} \end{cases} , si \begin{matrix} k = j \\ k \neq j \end{matrix}$$

1 d)

Esta ultima función de costo es mucho mas eficiente ya que reduce el problema multinomial (softmax) a uno binario. Por ejemplo, si se tiene un vocabulario de 10,000 palabras, cada que se llame a la función softmax se tendría que hacer la sumatoria sobre ese tamaño enorme del vocabulario, al contrario, en este ultimo caso, se reduce a un problema binario, es decir, de regresión logística, con una ventana de palabras fuera de contexto de mucho menor tamaño.