DL for NLP - Exercise 2

Itamar Trainin - 315425967

Dec 2019

Question 1:

Show that: $softmax(x) = softmax(x + c)$

Proof:

$$\begin{aligned}
softmax(x + c) &= \\
&= \frac{e^{x_i+c}}{\sum_j e^{x_j+c}} = \\
&= \frac{e^{x_i}e^c}{\sum_j e^{x_j}e^c} = \\
&= \frac{e^{x_i}e^c}{e^c\sum_j e^{x_j}} = \\
&= \frac{e^{x_i}}{\sum_j e^{x_j}} = \\
&= softmax(x)
\end{aligned} \tag{1}$$

Question 2:

Given $l(y, f(x)) = -\sum_i y_i log(f_i(x)),\ f_i(x) = softmax(xW + b)$ find $\frac{\partial l(y,f(x))}{\partial W}$ and $\frac{\partial l(y,f(x))}{\partial W}$

Derivation:

First note that if we define: $l(y, f_i) = -\sum_i y_i log(f_i(x)),\ f_i = softmax(h_i) = \frac{e^{x_i}}{\sum_j e^{x_j}}$, $h_i = xW + b$, then:

$$\begin{aligned}
\frac{\partial l(y, f_i)}{\partial W} &= \frac{\partial l(y, f_i)}{\partial f_i}\frac{\partial f_i}{\partial h_i}\frac{\partial h_i}{\partial W} \\
\frac{\partial l(y, f_i)}{\partial b} &= \frac{\partial l(y, f_i)}{\partial f_i}\frac{\partial f_i}{\partial h_i}\frac{\partial h_i}{\partial b}
\end{aligned} \tag{2}$$

Therefore, we can compute each derivative separately:

$$\frac{\partial h_i}{\partial W} = \frac{\partial}{\partial W} xW + b = x$$
$$\frac{\partial h_i}{\partial b} = \frac{\partial}{\partial b} xW + b = 1$$

$$\frac{\partial l(y, f_i)}{\partial f_i} = \frac{\partial}{\partial f_i} - \sum_i y_i log(f_i(x)) = - \sum_i y_i \frac{\partial}{\partial f_i} log(f_i(x)) = - \sum_i y_i \frac{1}{f_i(x)}$$

$$\frac{\partial f_i}{\partial h_i} = \frac{\partial}{\partial h_i} softmax(h_i) = \frac{\partial}{\partial h_i} \frac{e^{x_i}}{\sum_j e^{x_j}} = \frac{(\frac{\partial}{\partial h_i} e^{x_i})(\sum_j e^{x_j}) - (e^{x_i})(\frac{\partial}{\partial h_i} \sum_j e^{x_j})}{(\sum_j e^{x_j})^2} =$$
$$= \frac{e^{x_i}}{\sum_j e^{x_j}} (\frac{\sum_j e^{x_j}}{\sum_j e^{x_j}} - \frac{e^{x_i}}{\sum_j e^{x_j}}) = softmax(h_i)(1 - softmax(h_i))$$

we are only left with multiplying and pluging in the results to get:

$$\frac{\partial l(y, f_i)}{\partial W} = - \sum_i y_i \frac{1}{softmax(xW + b)} softmax(xW + b)(1 - softmax(xW + b))x =$$
$$= - \sum_i y_i (1 - softmax(xW + b))x$$

similarly,

$$\frac{\partial l(y, f_i)}{\partial b} = - \sum_i y_i (1 - softmax(xW + b))$$