Exercise Sheet 04 a),b)

Exercise 4-a

Let $\mathbf{x} \in \mathbb{R}^{1\times N}$ and $\sigma(\mathfrak{x}) = \mathrm{softmax}(\mathbf{x}) = [\sigma_i(\mathbf{x})]_{i=1...N}$ where $\sigma_i(\mathbf{x}) = \frac{e^{x_i}}{\Sigma_{j=1}^{N} e^{x_j}}$
Prove the lecture's statement that softmax(x) = softmax(x+c) for c$\in \mathbb{R}$. This fact can be used in the softmax implementation by subtracting $\max_{i=1...N}$ x$_i$ from the input, such that all values passed to the exp function are negative.

$\mathrm{softmax}(\mathbf{x}) = \sigma(\mathbf{x}) = \frac{e^{x}}{\Sigma_{j=1}^{N} e^{x_j}}$
$\mathrm{softmax}(\mathbf{x}+\mathbf{c}) = \sigma(\mathbf{x}+\mathbf{c}) = \frac{e^{x+c}}{\Sigma_{j=1}^{N} e^{x_j+c}}$

$$= \frac{e^{x+c}}{\Sigma_{j=1}^{N} e^{x_j+c}} = \frac{e^{x} * e^{c}}{\Sigma_{j=1}^{N} e^{x_j e^{c}}} = \frac{e^{x} * e^{c}}{e^{c}\Sigma_{j=1}^{N} e^{x_j}} = \frac{e^{x}}{\Sigma_{j=1}^{N} e^{x_j}}$$

$= \mathrm{softmax}(\mathbf{x})$

Exercise 4-b

$\mathbf{x} \in \mathbb{R}^{1\times N} = [x_1, x_2, ..., x_N]$.
$\sigma(\mathbf{x}) \in \mathbb{R}^{1\times N} =$

$$[\frac{e^{x_1}}{\Sigma_{j=1}^{N} e^{x_j}}, \frac{e^{x_2}}{\Sigma_{j=1}^{N} e^{x_j}}, ..., \frac{e^{x_N}}{\Sigma_{j=1}^{N} e^{x_j}}]$$

$\log(\sigma(\mathbf{x}) \in \mathbb{R}^{1\times N}) =$

$$[log(\frac{e^{x_1}}{\Sigma_{j=1}^{N} e^{x_j}}), log(\frac{e^{x_2}}{\Sigma_{j=1}^{N} e^{x_j}}), ..., log(\frac{e^{x_N}}{\Sigma_{j=1}^{N} e^{x_j}})]$$

$= [log(e^{x_1}) - log(\Sigma_{j=1}^{N} e^{x_j})), log(e^{x_2}) - log(\Sigma_{j=1}^{N} e^{x_j})), ..., log(e^{x_N}) - log(\Sigma_{j=1}^{N} e^{x_j}))]$

$D_{\mathbf{x}} log(\sigma(\mathbf{x})) \in \mathbb{R}^{N\times N} =$

$$[\frac{\partial(log(e^{x_1}) - log(\Sigma_{j=1}^N e^{x_j}))}{\partial x_1}, \frac{\partial(log(e^{x_2}) - log(\Sigma_{j=1}^N e^{x_j}))}{\partial x_2}, ..., \frac{\partial(log(e^{x_N}) - log(\Sigma_{j=1}^N e^{x_j})}{\partial x_N}]$$

$=$

$$[\frac{\partial(log(e^{x_1})}{\partial x_1} - \frac{\partial(log(\Sigma_{j=1}^N e^{x_j}))}{\partial x_1}, \frac{\partial(log(e^{x_2})}{\partial x_2} - \frac{\partial(log(\Sigma_{j=1}^N e^{x_j}))}{\partial x_2}, , ..., \frac{\partial(log(e^{x_N})}{\partial x_N} - \frac{\partial(log(\Sigma_{j=1}^N e^{x_j}))}{\partial x_N},]$$

$=$

$$\begin{bmatrix} \frac{\partial(log(e^{x_1})}{\partial x_1} - \frac{\partial log(\Sigma_{j=1}^N e^{x_j})}{\partial x_1}, & \frac{\partial(log(e^{x_1})}{\partial x_2} - \frac{\partial log(\Sigma_{j=1}^N e^{x_j})}{\partial x_2} & \cdots & \frac{\partial(log(e^{x_1})}{\partial x_N} - \frac{\partial log(\Sigma_{j=1}^N e^{x_j})}{\partial x_N} \\ \frac{\partial(log(e^{x_2})}{\partial x_1} - \frac{\partial log(\Sigma_{j=1}^N e^{x_j})}{\partial x_1} & \frac{\partial(log(e^{x_2})}{\partial x_2} - \frac{\partial log(\Sigma_{j=1}^N e^{x_j})}{\partial x_2} & \cdots & \frac{\partial(log(e^{x_2})}{\partial x_N} - \frac{\partial log(\Sigma_{j=1}^N e^{x_j})}{\partial x_N} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial(log(e^{x_N})}{\partial x_1} - \frac{\partial log(\Sigma_{j=1}^N e^{x_j})}{\partial x_1}, & \frac{\partial(log(e^{x_N})}{\partial x_2} - \frac{\partial log(\Sigma_{j=1}^N e^{x_j})}{\partial x_2} & \cdots & \frac{\partial(log(e^{x_N})}{\partial x_N} - \frac{\partial log(\Sigma_{j=1}^N e^{x_j})}{\partial x_N} \end{bmatrix}$$

$=$

$$\frac{\partial(log(e^{x_i})}{\partial x_k} - \frac{\partial log(\Sigma_{j=1}^N e^{x_j})}{\partial x_k} = \begin{cases} 1 - \frac{e^{x_k}}{\Sigma_{j=1}^N e^{x_j}}), & \text{if } i = k \\ -\frac{e^{x_k}}{\Sigma_{j=1}^N e^{x_j}} & \text{otherwise} \end{cases} \tag{1}$$

Thus,

$$\begin{bmatrix} 1 - \frac{e^{x_1}}{\Sigma_{j=1}^N e^{x_j}}) & -\frac{e^{x_2}}{\Sigma_{j=1}^N e^{x_j}} & \cdots & -\frac{e^{x_N}}{\Sigma_{j=1}^N e^{x_j}} \\ -\frac{e^{x_1}}{\Sigma_{j=1}^N e^{x_j}} & 1 - \frac{e^{x_2}}{\Sigma_{j=1}^N e^{x_j}} & \cdots & -\frac{e^{x_N}}{\Sigma_{j=1}^N e^{x_j}} \\ \vdots & \vdots & \ddots & \vdots \\ -\frac{e^{x_1}}{\Sigma_{j=1}^N e^{x_j}} & -\frac{e^{x_2}}{\Sigma_{j=1}^N e^{x_j}} & \cdots & 1 - \frac{e^{x_N}}{\Sigma_{j=1}^N e^{x_j}} \end{bmatrix}$$