7. Bonus: Derivative of the Softmax Function

Show that the derivative of the softmax function for an input $x \in \mathbb{R}^n$ can be written as

$$\frac{d\,\mathrm{softmax}(x)_i}{d x_j} = \begin{cases} -\mathrm{softmax}(x)_i\, \mathrm{softmax}(x)_j , & i \neq j \\ \mathrm{softmax}(x)_i\,(1-\mathrm{softmax}(x)_i) , & i = j \end{cases}$$

if $i=j$:

$$\frac{\partial \frac{e^{x_i}}{\sum_{k=1}^N e^{x_k}}}{\partial x_j} = \frac{e^{x_i}\sum_{k=1}^N e^{x_k} - e^{x_j}e^{x_i}}{\left(\sum_{k=1}^N e^{x_k}\right)^2} =$$

$$= \frac{e^{x_i}\left(\sum_{k=1}^N e^{x_k} - e^{x_j}\right)}{\left(\sum_{k=1}^N e^{x_k}\right)^2} = \frac{e^{\cancel{x_j}}}{\sum_{k=1}^N e^{x_k}} \cdot \frac{\sum_{k=1}^N e^{x_k} - e^{x_j}}{\sum_{k=1}^N e^{x_k}} = \mathrm{softmax}(x_i)(1 - \mathrm{softmax}(x_j))$$

if $i \neq j$

$$\frac{\partial \frac{e^{x_i}}{\sum_{k=1}^N e^{x_k}}}{\partial x_j} = \frac{0 - e^{x_j}e^{x_i}}{\left(\sum_{k=1}^N e^{x_k}\right)^2} = \frac{-e^{x_j}}{\sum_{k=1}^N e^{x_k}} \cdot \frac{e^{x_i}}{\sum_{k=1}^N e^{x_k}} =$$

$$= -\mathrm{softmax}(x_i)\,\mathrm{softmax}(x_j)$$