1) Simplest possible classificator always choose the class with highest probability: label 5

$$\text{accuracy} = \frac{213}{2 \cdot 136 + 6 \cdot 180 + 213} = 13{,}6\ \%$$

This dataset is not well balanced.
The class with the larges count (class 5 with 213 instances) and the class with the smallest count (class 0 with 136 instances) has a relationship of instances of 213 / 136 = 1.57, which is far away from 1.0 (for perfectly balanced data).

2) $$y_j = \frac{e^{x_j}}{\sum_{i=0}^{I-1} e^{x_i}}$$

$j=i$: $$\frac{dy_j}{dx_i} = \frac{e^{x_j} \cdot \sum_{i=0}^{I-1} e^{x_i} - e^{x_j} \cdot e^{x_i}}{\left(\sum_{i=0}^{I-1} e^{x_i}\right)^2}$$

$$= \frac{e^{x_j} \sum_{i=0}^{I-1} e^{x_i}}{\left(\sum_{i=0}^{I-1} e^{x_i}\right)^2} - \frac{e^{x_j}}{\sum_{i=0}^{I-1} e^{x_i}} \cdot \frac{e^{x_i}}{\sum_{i=0}^{I-1} e^{x_i}}$$

$$= y_j - y_i^2 = y_j (1 - y_i)$$

$j \neq i$: $$\frac{dy_j}{dx_i} = \frac{0 \cdot \sum_{i=0}^{I-1} e^{x_i} - e^{x_j} \cdot e^{x_i}}{\left(\sum_{i=0}^{I-1} e^{x_i}\right)^2}$$

$$= - \frac{e^{x_j}}{\sum_{i=0}^{I-1} e^{x_i}} \frac{e^{x_i}}{\sum_{i=0}^{I-1} e^{x_i}}$$

$$= y_j \cdot y_i$$

with Kronecker Delta: $\delta_{ij} = \begin{cases} 1, & \text{for } i=j \\ 0, & \text{for } i \neq j \end{cases}$

$$\frac{dy_j}{dx_i} = y_j \cdot (\delta_{ij} - y_i)$$

3) Cross entropy: $L = -\sum_{j=0}^{I-1} \sigma_j \ln y_j$

$$\frac{dL}{dy_j} = -\frac{\sigma_j}{y_j}$$

$$\frac{dy_j}{dx_i} = y_j\left(\delta_{ij} - y_i\right)$$

$\Rightarrow$ Chain rule: $\frac{dL}{dx_i} = \frac{dL}{dy_j} \cdot \frac{dy_j}{dx_i}$

$$= -\frac{\sigma_j}{y_j} \cdot y_j\left(\delta_{ij} - y_i\right)$$

$$= -\sigma_j \cdot \left(\delta_{ij} - y_i\right)$$