20) $C(f) = \frac{1}{m} \sum_{i=1}^{m} \hat{C}(f_i) = \frac{1}{m} \sum_{i=1}^{m} \left[- \sum_{k=1}^{K} \mathbb{1}(y_i = k) \log \frac{\exp(f_{ki})}{\sum_j \exp f_{ji}} \right]$

a) $\nabla_W \hat{C} = \sum_{k=1}^{K} \frac{\partial \hat{C}}{\partial f_{ki}} \frac{\partial f_{ki}}{\partial W} \qquad \nabla_b \hat{C} = \sum_{k=1}^{K} \frac{\partial \hat{C}}{\partial f_{ki}} \frac{\partial f_{ki}}{\partial b}$

x_j: Spaltenvektor mit M Einträgen $\mathbb{R}^{M \times 1}$

C: Skalar $\mathbb{R}$

W: $K \times M$-Matrix $\mathbb{R}^{K \times M}$

b: Spaltenvektor mit K Einträgen $\mathbb{R}^{K \times 1}$

$\nabla_W \hat{C}$: $K \times M$-Matrix $\mathbb{R}^{K \times M}$

$\nabla_b \hat{C}$: Spaltenvektor mit K Einträgen $\mathbb{R}^{K \times 1}$

$\nabla_{f_i} \hat{C}$: " " $\mathbb{R}^{K \times 1}$

$\frac{\partial f_{ki}}{\partial W}$: $K \times M$-Matrix $\mathbb{R}^{K \times M}$

$\frac{\partial f_{ki}}{\partial f_{ki}}$: $\mathbb{R}^{K \times 1}$

b) $\nabla_{f_a} C(f) = \frac{1}{m} \sum_{i=1}^{m} \left[- \underbrace{\nabla f_a \sum_{k=1}^{K} \mathbb{1}(y_i = k) \log \frac{\exp(f_{ki})}{\sum_j \exp f_{ji}}}_{=0 \text{ für } k \neq a} \right]$

$= \frac{1}{m} \sum_{i=1}^{m} - \nabla_{f_a} \log \left(\frac{e^{f_{y_i i}}}{\sum_j e^{f_{ji}}} \right)$

$= \frac{1}{m} \sum_{i=1}^{m} \left(- \frac{\sum_j e^{f_{ji}}}{e^{f_{y_i i}}} \cdot \left[\frac{\nabla_{f_a} e^{f_{y_i i}} \cdot \sum_j e^{f_{ji}} - e^{f_{y_i i}} \nabla_{f_a} \sum_j e^{f_{ji}}}{(\sum_j e^{f_{ji}})^2} \right] \right)$

$= \frac{1}{m} \sum_{i=1}^{m} - \frac{1}{e^{f_{y_i i}}} \left[\frac{e^{f_{ai}} \mathbb{1}(y_i = a) \sum_j e^{f_{ji}} - e^{f_{y_i i}} e^{f_{ai}}}{\sum_j e^{f_{ji}}} \right]$

$= \frac{1}{m} \sum_{i=1}^{m} \left(\frac{e^{f_{ai}}}{\sum_j e^{f_{ji}}} - \frac{e^{f_{ai}}}{e^{f_{y_i i}}} \mathbb{1}(y_i = a) \right) = \frac{1}{m} \sum_{i=1}^{m} \left(\frac{e^{f_{ai}}}{\sum_j e^{f_{ji}}} - \mathbb{1}(y_i = a) \right)$

c) $\frac{\partial f_{ki}}{\partial W} = \frac{\partial (W_k x_i + b_k)}{\partial W} = \sum_{j=1}^{K} \begin{pmatrix} - 0 - \\ - \delta_{jk} x_i^T - \\ - 0 - \end{pmatrix} = \begin{pmatrix} - 0 - \\ - x_i^T - \\ - 0 - \end{pmatrix} \leftarrow$ k-te Zeile

$\frac{\partial f_{ki}}{\partial b} = \frac{\partial (W_k x_i + b_k)}{\partial b} = \begin{pmatrix} 0 \\ 1 \\ 0 \\ 0 \\ \vdots \end{pmatrix} \leftarrow$ k-te Zeile