Blatt 11 Aufgabe 23

a) x_i : Spaltenvektor $(M \times 1)$
c : Kostenfunktion (1)
W : Gewichtungsmatrix $(K \times M)$
b : Biasvektor $(K \times 1)$

$\nabla_W \hat{C}$: $K \times M$ $\qquad \frac{\partial f_{ki}}{\partial W} : K \times M \qquad \frac{\partial f_i}{\partial b} : K \times 1$

$\nabla_{f_i} \hat{C}$: $K \times 1$

b) $\nabla_{f_a} C(f) = \frac{1}{m} \sum_m \nabla_{f_a} \hat{C}(f_i)$

$\nabla_{f_a} \hat{C}(f_i) = -\nabla_{f_a} \sum_k \mathbb{1}(y_i = k) \log\left(\frac{e^{f_{ki}}}{\sum_j e^{f_{ji}}}\right) = -\nabla_{f_a} \log\left(\frac{e^{f_{y_i i}}}{\sum_j e^{f_{ji}}}\right)$

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

[...] das kann man bestimmt umformen zu

$= \frac{e^{f_{a,i}}}{\sum_j e^{f_{aj}}} - \mathbb{1}(y_i = a)$

$\Rightarrow \quad \nabla_{f_a} C(f) = \frac{1}{m} \sum_m \left[\frac{e^{f_{a,i}}}{\sum_j e^{f_{aj}}} - \mathbb{1}(y_i - a)\right]$

c) $\frac{\partial f_{ki}}{\partial W} = \frac{\partial (W_k x_i + b_k)}{\partial W} = \begin{pmatrix} 0 & \cdots & 0 \\ \vdots & & \\ x_{i,1} & \cdots & x_{i,M} \\ \vdots & & \vdots \\ 0 & & 0 \end{pmatrix} \rightarrow$ k-te Zeile

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