Softmax Classifier Vectorized.

$$L_i = -\log\left(\frac{e^{f_{y_i}}}{\sum_j e^{f_j}}\right) = -f_{y_i} + \log\sum_j e^{f_j}$$

$f_j = x_i w_j$ $\quad x_i w$

$$\frac{dL_i}{dw_j} = \frac{1}{\sum_j e^{f_j}} \cdot x_i \cdot e^{x_i w_j} = \boxed{\frac{e^{f_j}}{\sum_j e^{f_j}} \cdot x_i}$$

$$\frac{dL_i}{dw_{y_i}} = -x_i + \frac{e^{f_{y_i}}}{\sum_j e^{f_j}} \cdot x_i = \left(\frac{e^{f_{y_i}}}{\sum_j e^{f_j}} - 1\right) x_i$$

$$\frac{dL_i}{dw} = \begin{bmatrix} \frac{dL_i}{dw_1} & \cdots & \frac{dL_i}{dw_j} \\ \vdots & & \vdots \\ \frac{dL_i}{dw_1} & \cdots & \frac{dL_i}{dw_j} \end{bmatrix}$$

$$= \begin{bmatrix} \frac{e^{x_1 w_1}}{\sum_j e^{x_1 w_j}} \cdot x_1 & \cdots & \frac{e^{x_1 w_j}}{\sum_j e^{x_1 w_j}} \cdot x_1 \\ \frac{e^{x_2 w_1}}{\sum_j e^{x_2 w_j}} \cdot x_2 & \cdots & \frac{e^{x_2 w_j}}{\sum_j e^{x_2 w_j}} \cdot x_2 \end{bmatrix}$$

$$= \begin{bmatrix} \frac{e^{x_1 w_1}}{\sum e^{x_1 w_j}} & \cdots & \frac{e^{x_1 w_j}}{\sum e^{x_1 w_j}} \\ \vdots & & \vdots \\ \frac{e^{x_2 w_1}}{\sum e^{x_2 w_j}} & \cdots & \frac{e^{x_2 w_j}}{\sum e^{x_2 w_j}} \end{bmatrix} \cdot \begin{bmatrix} x_1 & & \\ & \ddots & \\ & & x_2 \end{bmatrix}$$

$$= [x_1, x_2 \cdots, x_i] \cdot \begin{bmatrix} & & \\ & & \\ & & \end{bmatrix}$$

$$\frac{dL_i}{dw} = \left[\frac{e^{f_j}}{\sum e^{f_j}} \frac{dL_i}{dw_1} \cdots \frac{dL_i}{dw_j}\right]$$

$$= \left[\frac{e^{x_i w_1}}{\sum_j e^{f_j}} \cdot x_i \cdots \right]$$

$$= \left[\frac{e^{x_i w_1}}{\sum e^{f_j}} \cdots \frac{e^{x_i w_j}}{\sum e^{f_j}}\right] \cdot [x_i \cdots x_i]$$