TP 2

Jouve Vincent 3671083 Lenoir Romain 3670199

$$x \times y \times z \times \left\lceil \frac{2p+x-k}{s} + 1 \right\rceil \times \left\lceil \frac{2p+y-k}{s} + 1 \right\rceil \times 1$$

$$
\begin{aligned}
-\sum_i y_i \log \frac{e^{\tilde{y}_i}}{\sum_k e^{\tilde{y}_k}} &= -\sum_i y_i(\log(e^{\tilde{y}_i}) - \log(\sum_k e^{\tilde{y}_k})) \\
&= -\sum_i y_i\tilde{y}_i - y_i \log\,(\sum_k e^{\tilde{y}_k}) \\
&= -\sum_i y_i\tilde{y}_i + \sum_i y_i \log\,(\sum_k e^{\tilde{y}_k}) \\
&= -\sum_i y_i\tilde{y}_i + \log\,(\sum_k e^{\tilde{y}_k}) \text{ because } y_i = 0 \text{ for all classes except one}
\end{aligned}
$$

$$
\begin{aligned}
\frac{\partial l}{\partial \tilde{y}_i} &= -y_i + \frac{e^{\tilde{y}_i}}{\sum_k e^{\tilde{y}_k}} \\
&= \hat{y}_i - y_i
\end{aligned}
$$

$$\nabla_{\tilde{y}} l = \begin{bmatrix} \frac{\partial l}{\partial \tilde{y}_1} \\ \vdots \\ \frac{\partial n}{\partial \tilde{y}_n} \end{bmatrix} = \begin{bmatrix} \hat{y}_1 - y_1 \\ \vdots \\ \hat{y}_n - y_n \end{bmatrix}$$

$$\tilde{y}_k = \sum_m w_{y,km} \times h_m + b_k$$

so $\frac{\partial \tilde{y}_i}{\partial w_{y,ij}} = h_j$ and $\frac{\partial \tilde{y}_k}{\partial w_{y,ij}} = 0$ k $\neq i$

$$\text{so } \frac{\partial l}{\partial w_{y,ij}} = \sum_k \frac{\partial l}{\partial \tilde{y}_k} \frac{\partial \tilde{y}_k}{\partial w_{y,ij}}$$
$$= \frac{\partial l}{\partial \tilde{y}_i} \times h_j$$

and $\frac{\partial \tilde{y}_i}{\partial b_{y,i}} = 1$ and $\frac{\partial \tilde{y}_k}{\partial b_{y,i}} = 0$ k $\neq i$

$$\frac{\partial l}{\partial b_{y,i}} = \sum_k \frac{\partial l}{\partial \tilde{y}_k} \frac{\partial \tilde{y}_k}{\partial b_{y,i}}$$
$$= \frac{\partial l}{\partial \tilde{y}_i}$$

$$\nabla_{w_{y,ij}} l = \begin{bmatrix} \frac{\partial l}{\partial w_{y,11}} & \cdots & \frac{\partial l}{\partial w_{y,1n_h}} \\ \vdots & \ddots & \vdots \\ \frac{\partial l}{\partial w_{y,n_y 1}} & \cdots & \frac{\partial l}{\partial w_{y,n_y n_h}} \end{bmatrix} = \begin{bmatrix} (\hat{y}_1 - y_1)h_1 & \dots & (\hat{y}_1 - y_1)h_{n_h} \\ \vdots & \ddots & \vdots \\ (\hat{y}_{n_y} - y_{n_y})h_1 & \dots & (\hat{y}_{n_y} - y_{n_y})h_{n_h} \end{bmatrix}$$

$$\nabla_{b_{y,j}} l = \begin{bmatrix} \frac{\partial l}{\partial b_{y,1}} \\ \vdots \\ \frac{\partial l}{\partial b_{y,n_y}} \end{bmatrix} = \begin{bmatrix} \hat{y}_1 - y_1 \\ \vdots \\ \hat{y}_{n_y} - y_{n_y} \end{bmatrix}$$

$$\nabla_{\tilde{h}} l = \begin{bmatrix} (1 - tanh^2(\tilde{h}_1)) \sum_i \frac{\partial l}{\partial w_{y,i1}} \\ \vdots \\ (1 - tanh^2(\tilde{h}_{n_h})) \sum_i \frac{\partial l}{\partial w_{y,in_h}} \end{bmatrix}$$

$$\nabla_{w_{h,ij}} l = \begin{bmatrix} \frac{\partial l}{\partial h_1} x_1 & \cdots & \frac{\partial l}{\partial h_1} x_{n_x} \\ \vdots & \ddots & \vdots \\ \frac{\partial l}{\partial h_{n_h}} x_1 & \cdots & \frac{\partial l}{\partial h_{n_h}} x_{n_x} \end{bmatrix}$$

$$\nabla_{b_{h,i}} l = \begin{bmatrix} \frac{\partial l}{\partial h_1} \\ \vdots \\ \frac{\partial l}{\partial h_{n_h}} \end{bmatrix}$$