$$\frac{\partial}{\partial x}\max(0, x) = \begin{cases} 0, & x<0 \\ 1, & x>0 \end{cases}$$

$$\frac{\partial}{\partial x}\max(\alpha x, x) = \begin{cases} \alpha, & x<0 \\ 1, & x>0 \end{cases}$$

$$\frac{\partial}{\partial x}\frac{1}{1+e^{-x}} = \left(1+e^{-x}\right)^{-1\,'} =$$

$$= -\left(1+e^{-x}\right)^{-2} \cdot e^{-x} \cdot (-1) =$$

$$= \frac{e^{-x}}{(1+e^{-x})^2} = \frac{1}{1+e^{-x}} \cdot \frac{e^{-x}}{1+e^{-x}} =$$

$$= \sigma(x) \cdot \left(\frac{1+e^{-x}-1}{1+e^{-x}} = \frac{1+e^{-x}}{1+e^{-x}} - \sigma(x)\right)$$

$$= \sigma(x)\left(1-\sigma(x)\right)$$

4) $$\frac{d}{dx}\frac{e^x - e^{-x}}{e^x+e^{-x}} =$$

$$\frac{(e^x-e^{-x})'(e^x+e^{-x}) - (e^x-e^{-x})(e^x+e^{-x})'}{(e^x+e^{-x})^2} =$$

$$= \frac{(e^x + e^{-x})^2 - (e^x - e^{-x})^2}{(e^x + e^{-x})^2}$$

$$= 1 - \left(\frac{e^x - e^{-x}}{e^x + e^{-x}}\right)^2 = \underline{1 - \tanh(x)}$$

$$f(x) = x^T \cdot A x, \quad A \in \mathbb{R}^{n \times n}, \quad x \in \mathbb{R}^{1 \times n}$$

$$x^T A x = \sum_{i=1}^{N} \sum_{j=1}^{N} a_{ij} x_i x_j =$$

$$\frac{\partial f}{\partial x_k} = \frac{\partial}{\partial x_k} \sum_{j=1}^{N} a_{kj} \cdot x_k x_j + \frac{\partial}{\partial x_k} \sum_{i=1}^{N} a_{ik} \cdot x_i x_k =$$

$$= \sum_{j=1}^{N} a_{kj} x_j + \sum_{i=1}^{N} a_{ik} \cdot x_i$$

$$\nabla_{x_0} f = A x + A^T x = (A + A^T) x$$

Write backprop. for my neural network

$$n \times 3$$

X

$\cdot$

$W_1^{3\times 10}$

$+$

$b^{1\times 10}$

$\|$

$Z_1^{n\times 10}$

$\downarrow$

$\text{ReLU}(Z_1) = A_1$

$\downarrow$

$Z_2^{N\times 6}$

$\downarrow_{N\times 6}$

$\text{ReLU}(Z_2) = A_2$ — $k = 3$ ← classes count

$\downarrow$

$Z_3^{N\times 3}$

$\downarrow$

$\text{SOFTmax}(Z_3) = \hat{y}^{N\times 3}$

$\downarrow$

$\text{Loss} - \text{CCE}(y, \hat{y}) \in R$

1) Find $\dfrac{\partial L}{\partial Z_3}$

$Z_3 = (z_1, \dots, z_k)$

$\text{Softmax}(Z_3) = (\hat{y}_1, \dots, \hat{y}_k)$

$y = (y_1, \dots y_k), \begin{cases} y_c = 1, c \text{ is true class} \\ y_c = 0, c \text{ is false} \end{cases}$

$$L = -\sum_{i=1}^{k} y_i \log \hat{y}_i$$

$$\frac{\partial L}{\partial Z_3} = \left(\frac{\partial L}{\partial Z_r} \right)_{r \in \{1, \dots, k\}}$$

find it

L depends on all $\hat{y}_i$

all $\hat{y}_i$ depends on all Z_r

$$\frac{\partial L}{\partial Z_r} = \sum_{i=1}^{K} \frac{\partial L}{\partial \hat{y}_i} \frac{\partial \hat{y}_i}{\partial Z_r}$$

$$\frac{\partial L}{\partial y_i} = \frac{\partial}{\partial \hat{y}_i}\left(-\sum_{j=1} y_j \log \hat{y}_j\right)$$

if $i \neq j$ it's const

$$= -\frac{y_i}{\hat{y}_i}$$

$$\frac{\partial \hat{y}_i}{\partial z_r} = \frac{\partial}{\partial z_r}\left(\frac{\exp(z_r)}{\sum_{j=1}^{K} \exp(z_j)}\right)$$

case 1 ; $r = i$:

$$\frac{\partial \hat{y}_i}{\partial z_r} = \frac{\exp(z_r)\sum_{j=1}^{K} \exp(z_j) - \exp(2z_r)}{\left(\sum_{j=1}^{K} \exp(z_j)\right)^2}$$

$$= \frac{\exp(z_r) \cancel{\sum_{j=1}^{K} \exp(z_j)}}{\left(\sum_{j=1}^{K} \exp(z_j)\right)^{\cancel{2}}} - \frac{\exp(z_r) \cdot \exp(z_r)}{\left(\sum_{j=1}^{K} \exp(z_j)\right)^2}$$

$$= \frac{\exp(z_r)}{\sum_{j=1}^{K} \exp(z_j)} - \left(\frac{\exp(z_r)}{\sum_{j=1}^{K} \exp(z_j)}\right)^2 =$$

$$= \frac{\exp(z_r)}{\sum_{j=1}^{k} \exp(z_j)} \left(1 - \frac{\exp(z_r)}{\sum_{j=1}^{k} \exp(z_j)} \right)$$

$$= \underline{\hat{y}_i (1 - \hat{y}_i)}$$

Case 2: $i \neq r$ $\quad \frac{\partial \hat{y}_i}{\partial z_r} =$

$$= \frac{0 \cdot \sum_{j=1}^{k} \exp(z_j) - \exp(z_r) \cdot \exp(z_i)}{\left(\sum_{j=1}^{k} \exp(z_j) \right)^2}$$

$$= \frac{- \exp(z_r) \cdot \exp(z_i)}{\left(\sum_{j=1}^{k} \exp(z_j) \right)^2} =$$

$$= - \hat{y}_r \cdot \hat{y}_i$$

Thus:

Thus:

$$\frac{\partial L}{\partial z_r} = \sum_{i=1}^{k} \frac{\partial L}{\partial \hat{y}_i}\cdot\frac{\partial \hat{y}_i}{\partial z_r} =$$

$$= \left(\sum_{\substack{i=1,\\ i\neq r}}^{k} -\frac{y_i}{\hat{y}_i}\cdot(-\hat{y}_r\cdot\hat{y}_i)\right) +$$

$$+\left(-\frac{y_r}{\hat{y}_r}\right)\cdot\hat{y}_r\cdot(1-\hat{y}_r) =$$

$$= \left(\sum_{\substack{i=1\\ i\neq r}}^{k} y_i\hat{y}_r\right) + \left(-y_r + y_r\hat{y}_r\right)$$

$$= \left(\hat{y}_r\sum_{\substack{i=1\\ i\neq r}}^{k} y_i\right) - y_r + y_r\hat{y}_r =$$

$$= -y_r + \hat{y}_r\sum_{\substack{i=1\\ i\neq r}}^{k} y_i + \hat{y}_r y_r$$

$$= -y_r + \hat{y}_r\left(\underbrace{\sum_{\substack{i=1\\ i\neq r}}^{k} y_i}_{0\ \text{(false classes)}} + \underbrace{y_r}_{1\ \text{(true class)}}\right)$$

$$= \hat{y}_r - y_r$$

$$\frac{\partial L}{\partial Z_3} = \hat{y}_r - y_r$$

Remember: $Z_3 = A_2 W_3 + b_3$

$$\frac{\partial L}{\partial W_3} = \frac{\partial L}{\partial Z_3} \cdot \frac{\partial Z_3}{\partial W_3} = A_2^T \cdot \frac{\partial Z_3}{\partial W_3}$$

$$\frac{\partial L}{\partial A_2} = \frac{\partial Z_3}{\partial W_3} \cdot W_3^T$$

$$\frac{\partial L}{\partial b_3} = \sum_i \frac{\partial L}{\partial Z_{3ik}}$$

$$A_2 = \mathrm{ReLU}(Z_2) = \max(Z_2, 0)$$

$$Z_2 = A_1 W_2 + b_2$$

$$A_2 = \mathrm{ReLU}(Z_2) = \max(Z_2, 0)$$

$$\frac{\partial L}{\partial Z_2} = \frac{\partial L}{\partial A_2} \cdot \frac{\partial A_2}{\partial Z_2} = \frac{\partial L}{\partial A_2} \cdot \begin{cases} 1, & Z_2 > 0 \\ 0, & Z_2 < 0 \end{cases}$$

$$\frac{\partial \mathrm{ReLU}(Z_2)}{Z_2} = \begin{cases} 1, & Z_2 > 0 \\ 0, & Z_2 < 0 \end{cases}$$