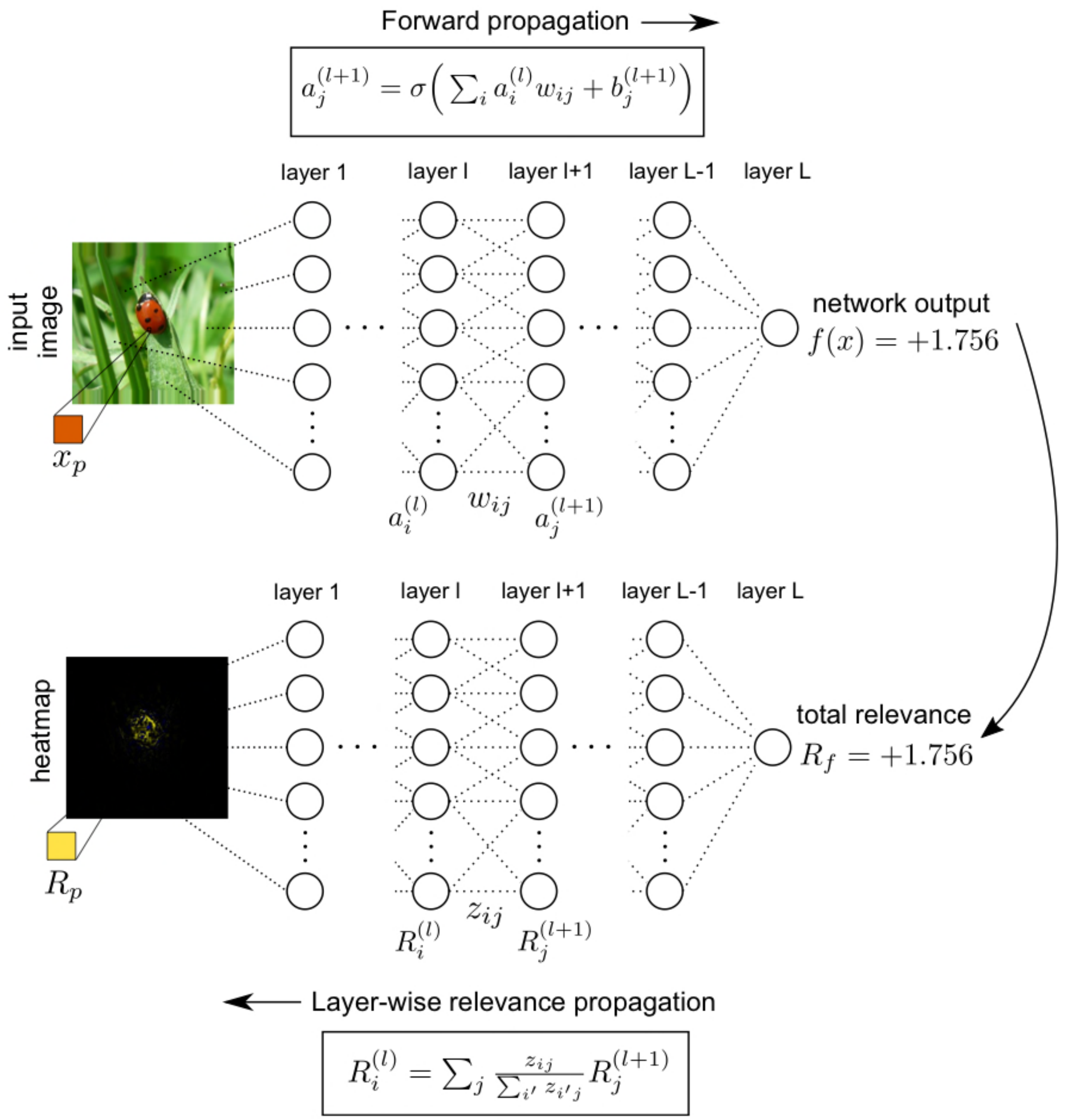
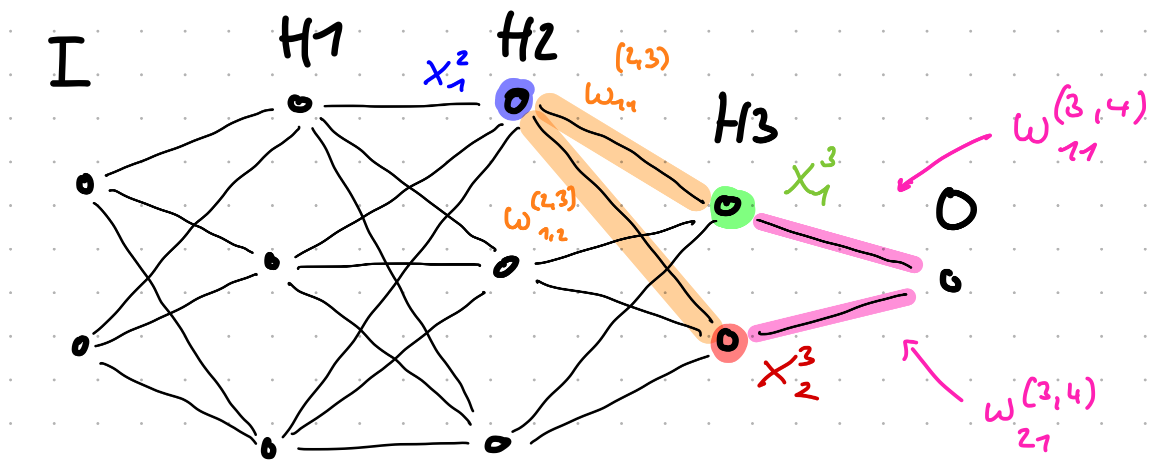
$$R_i^{(l)} = \sum_j \frac{z_{ij}}{\sum_{i'} z_{i'j}} R_j^{(l+1)} \quad \text{with} \quad \mathrm{z_{ij}} = \mathrm{x_i^{(l)} w_{ij}^{(l,l+1)}}$$

$$\sum_p R_p^{(1)} = f(\boldsymbol{x}).$$



$$R_i^{(l)} = \sum_j \frac{z_{ij}}{\sum_{i'} z_{i'j}} R_j^{(l+1)} \quad \text{with} \quad \mathrm{z_{ij} = x_i^{(l)} w_{ij}^{(l,l+1)}}$$



$$R_1^3 = \sum_{j=1}^{1} \frac{z_{1j}}{\sum_{i'=1}^{2} z_{i'j}} = \sum_{j=1}^{1} \frac{x_1^3 \cdot w_{1j}^3}{\sum_{i'=1}^{2} x_{i'}^3 \cdot w_{i'j}^3}$$

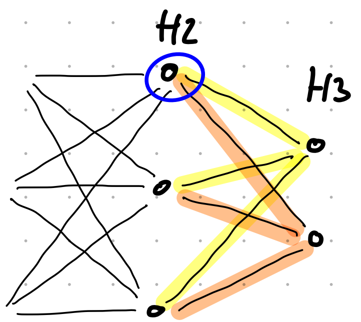
$$= \frac{x_1^3 \cdot w_{11}^{(3,4)}}{x_1^3 \cdot w_{11}^{(3,4)} + x_2^3 \cdot w_{21}^{(3,4)}}$$

D.h. R_1^3 ist der relative Anteil von x_1^3 am Output der der Schicht H3.

$$R^1_2 = \sum_{j=1}^{2} \frac{z_{1j}}{\sum_{i'=1}^{3} z_{i'j}} = \frac{z_{11}}{z_{11} + z_{21} + z_{31}} + \frac{z_{12}}{z_{12} + z_{22} + z_{32}}$$

$$= \frac{x_1^2 \cdot w_{11}^{(2,3)}}{x_1 \cdot w_{11} + x_2 \cdot w_{21} + x_3 \cdot w_{31}}$$

$$+ \frac{x_1^2 \cdot w_{12}^{(2,3)}}{x_1 \cdot w_{12} + x_2 \cdot w_{22} + x_3 \cdot w_{32}}$$



$$R_i^{(l)} = \sum_j \frac{z_{ij}}{\sum_{i'} z_{i'j}} R_j^{(l+1)} \quad \text{with} \quad \mathrm{z_{ij} = x_i^{(l)} w_{ij}^{(l,l+1)}}$$

nach

$$\begin{pmatrix} R_1^{l} \\ R_2^{l} \\ \vdots \\ R_n^{l} \end{pmatrix} = \begin{pmatrix} \sum_j \frac{z_{1j}}{\sum_{i'} z_{i'j}} R_j^{(l+1)} \\ \vdots \\ \sum_j \frac{z_{nj}}{\sum_{i'} z_{i'j}} R_j^{(l+1)} \end{pmatrix}$$

$$\Longleftrightarrow \begin{pmatrix} R_1^{l} \\ R_2^{l} \\ \vdots \\ R_n^{l} \end{pmatrix} = \begin{pmatrix} \boxed{\sum_j \frac{x_1^{(l)} w_{1j}^{(l,l+1)}}{\sum_{i=1}^{n} x_i^{(l)} \cdot w_{ij}^{(l,l+1)}} \cdot R_j^{(l+1)}} \\ \vdots \\ \sum_j \frac{x_n^{(l)} w_{nj}^{(l,l+1)}}{\sum_{i=1}^{n} x_i^{(l)} \cdot w_{ij}^{(l,l+1)}} \cdot R_j^{(l+1)} \end{pmatrix} \qquad (1)$$

Schicht $l+1$ habe m Neuronen

$$\boxed{\sum_{j=1}^{m} \frac{x_1^{(l)} w_{1j}^{(l,l+1)}}{\sum_{i=1}^{n} x_i^{(l)} \cdot w_{ij}^{(l,l+1)}} \cdot R_j^{(l+1)}} = \qquad (2)$$

$$\text{matmul}\left(\boxed{\begin{pmatrix} \frac{x_1^{(l)} w_{11}^{(l,l+1)}}{\sum_{i=1}^{n} x_i^{(l)} \cdot w_{i1}^{(l,l+1)}} \\ \vdots \\ \frac{x_1^{(l)} w_{1m}^{(l,l+1)}}{\sum_{i=1}^{n} x_i^{(l)} \cdot w_{im}^{(l,l+1)}} \end{pmatrix}^T} , \underbrace{\begin{pmatrix} R_1^{(l+1)} \\ \vdots \\ R_m^{(l+1)} \end{pmatrix}}_{=: R^{(l+1)}} \right)$$

ELEMENTWEISE

$$\boxed{\begin{pmatrix} \dfrac{x_1^{(l)} w_{11}^{(l,l+1)}}{\sum_{i=1}^{n} x_i^{(l)} \cdot w_{i1}^{(l,l+1)}} \\ \vdots \\ \dfrac{x_1^{(l)} w_{1m}^{(l,l+1)}}{\sum_{i=1}^{n} x_i^{(l)} \cdot w_{im}^{(l,l+1)}} \end{pmatrix}} = \begin{pmatrix} x_1^{(l)} \cdot w_{11}^{(l,l+1)} \\ \vdots \\ x_1^{(l)} w_{1m}^{(l,l+1)} \end{pmatrix} : \begin{pmatrix} \sum_{i=1}^{n} x_i^{(l)} \cdot w_{i1}^{(l,l+1)} \\ \vdots \\ \sum_{i=1}^{n} x_i^{(l)} \cdot w_{im}^{(l,l+1)} \end{pmatrix} \quad (3)$$

$$\sum_{i=1}^{n} x_i^{(l)} \cdot w_{i1}^{(l,l+1)} = \begin{pmatrix} w_{11}^{(l,l+1)} \\ \vdots \\ w_{n1}^{(l,l+1)} \end{pmatrix}^T \cdot \begin{pmatrix} x_1^{(l)} \\ \vdots \\ x_n^{(l)} \end{pmatrix} \quad (4)$$

1. Spalte

$$= W_{:,1}^{(l,l+1)^T} \cdot \text{output } H_l$$

$$= \text{matmul}\left(W_{:,1}^{(l,l+1)^T}, X^{(l)}\right)$$

1. Zeile

$$\begin{pmatrix} x_1^{(l)} \cdot w_{11}^{(l,l+1)} \\ \vdots \\ x_1^{(l)} w_{1m}^{(l,l+1)} \end{pmatrix} = x_1^{(l)} \cdot \left(W_{1,:}^{(l,l+1)}\right)^T \quad (5)$$

$$\boxed{\begin{pmatrix} \dfrac{x_1^{(l)} w_{11}^{(l,l+1)}}{\sum_{i=1}^{n} x_i^{(l)} \cdot w_{i1}^{(l,l+1)}} \\ \vdots \\ \dfrac{x_1^{(l)} w_{1m}^{(l,l+1)}}{\sum_{i=1}^{n} x_i^{(l)} \cdot w_{im}^{(l,l+1)}} \end{pmatrix}} = x_1^{(l)} \cdot W_{1,:}^{(l,l+1)^T} : \text{matmul}\left(W^{(l,l+1)^T}, X^{(l)}\right) \quad (6)$$

$$\begin{pmatrix} \dfrac{x_1^{(\ell)} w_{11}^{(\ell,\ell+1)}}{\sum_{i=1}^{n} x_i^{(\ell)} \cdot w_{i1}^{(\ell,\ell+1)}} \\ \vdots \\ \dfrac{x_1^{(\ell)} w_{1m}^{(\ell,\ell+1)}}{\sum_{i=1}^{n} x_i^{(\ell)} \cdot w_{im}^{(\ell,\ell+1)}} \end{pmatrix} = X_1^{(\ell)} \cdot W_{1,:}^{(\ell,\ell+1)^T} : \mathrm{matmul}\left(W^{(\ell,\ell+1)^T}, X^{(\ell)}\right)$$

(6) + (2)

$$\boxed{\sum_{j=1}^{m} \frac{x_1^{(\ell)} w_{1j}^{(\ell,\ell+1)}}{\sum_{i=1}^{n} x_i^{(\ell)} \cdot w_{ij}^{(\ell,\ell+1)}} \cdot R_j^{(\ell+1)}} =$$

(7)

$$\mathrm{matmul}\left(\left(\underset{\substack{\uparrow \\ \text{Spaltenvektor}}}{X_1^{(\ell)} \cdot W_{1,:}^{(\ell,\ell+1)^T}} : \mathrm{matmul}\underset{\substack{\uparrow \\ \text{Spaltenvektor}}}{\left(W^{(\ell,\ell+1)^T}, X^{(\ell)}\right)}\right)^T, \underset{\substack{\uparrow \\ \text{Spaltenvektor}}}{R^{(\ell+1)}} \right)$$

$$\boxed{R^{\ell} = \mathrm{matmul}\left(X^{(\ell)^T} \cdot W^{(\ell,\ell+1)^T} : \mathrm{matmul}\left(\left(W^{(\ell,\ell+1)^T}, X^{(\ell)}\right)\right)^T, R^{(\ell+1)}\right)}$$

$$R^{\ell} = \text{matmul}\Big(\underbrace{\underbrace{X^{(\ell)T} \cdot W^{(\ell,\ell+1)^T}}_{①} : \underbrace{\text{matmul}\left(W^{(\ell,\ell+1)^T}, X^{(\ell)}\right)}_{②}\Big)^T}_{③}, R^{(\ell+1)}\Big)$$

①

$$X^{(\ell)T} \cdot W^{(\ell,\ell+1)^T} = (x_1, \ldots, x_n) \cdot \begin{pmatrix} w_{11} & w_{21} & \cdots & w_{n1} \\ w_{12} & w_{22} & \cdots & w_{n2} \\ \vdots & & \ddots & \\ w_{1m} & w_{2m} & \cdots & w_{nm} \end{pmatrix}$$

$$= \begin{pmatrix} x_1 w_{11} & \cdots & x_n w_{n1} \\ \vdots & & \vdots \\ x_1 w_{1m} & \cdots & x_n w_{nm} \end{pmatrix}$$

②

$$\text{matmul}\left(W^{(\ell,\ell+1)^T}, X^{(\ell)}\right) = \begin{pmatrix} w_{11} \cdot x_1 + w_{21} + w_{31} \cdot x_3 \cdots \\ \vdots \\ w_{1m} \cdot x_1 + w_{2m} \cdot x_2 + w_{3m} \cdot x_3 \cdots \end{pmatrix}$$

$$= \begin{pmatrix} \sum_{i=1}^{n} x_i \cdot w_{i1} \\ \vdots \\ \sum_{i=1}^{n} x_i \cdot w_{im} \end{pmatrix}$$

③

$$X^{(\ell)T} \cdot W^{(\ell,\ell+1)} : \text{matmul}\left(W^{(\ell,\ell+1)}, X^{(\ell)}\right)$$

$$= \begin{pmatrix} \dfrac{x_1 w_{11}}{\sum_{i=1}^{n} x_i \cdot w_{i1}} & \cdots & \dfrac{x_n w_{n1}}{\sum_{i=1}^{n} x_i \cdot w_{i1}} \\ \vdots & & \vdots \\ \dfrac{x_1 w_{1m}}{\sum_{i=1}^{n} x_i \cdot w_{im}} & \cdots & \dfrac{x_n w_{nm}}{\sum_{i=1}^{n} x_i \cdot w_{im}} \end{pmatrix}$$

$$\mathrm{matmul}\left(X^{(\ell)T} \cdot W^{(\ell,\ell+1)T} : \mathrm{matmul}\left(W^{(\ell,\ell+1)}, X^{(\ell)}\right)^T, R^{(\ell+1)}\right)$$

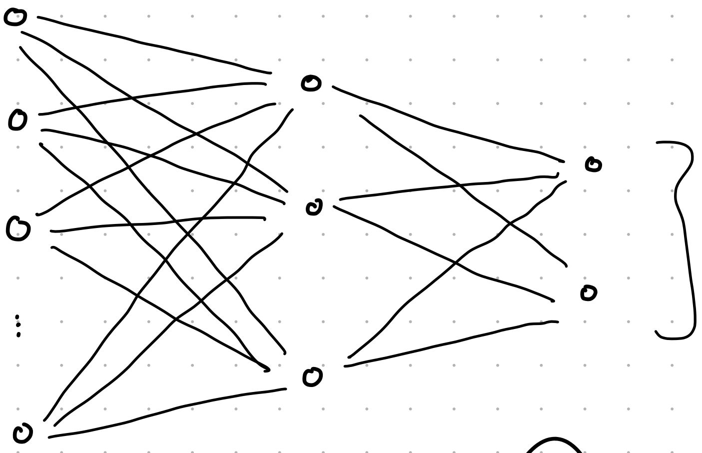
$$= \mathrm{matmul}\left(\begin{pmatrix} \frac{x_1 w_{11}}{\sum_{i=1}^{n} x_i \cdot w_{i1}} & \cdots & \frac{x_1 w_{1m}}{\sum_{i=1}^{n} x_i \cdot w_{im}} \\ \vdots & & \vdots \\ \frac{x_n w_{n1}}{\sum_{i=1}^{n} x_i \cdot w_{i1}} & \cdots & \frac{x_n w_{nm}}{\sum_{i=1}^{n} x_i \cdot w_{im}} \end{pmatrix}, \begin{pmatrix} R_1^{(\ell+1)} \\ \vdots \\ R_m^{(\ell+1)} \end{pmatrix}\right)$$

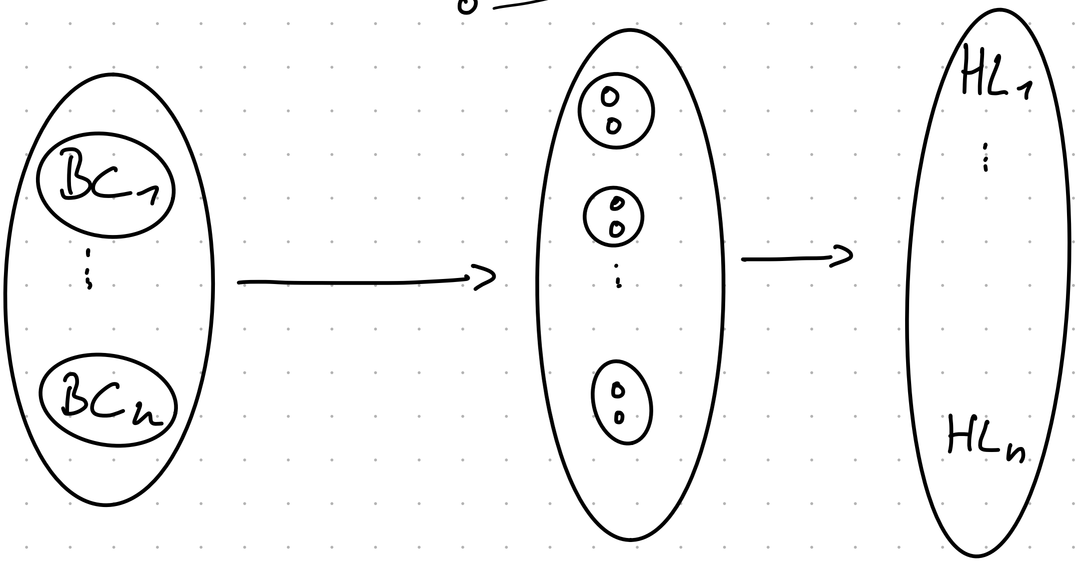
$$= \begin{pmatrix} \frac{x_1 w_{11}}{\sum_{i=1}^{n} x_i \cdot w_{i1}} \cdot R_1^{(\ell+1)} + \dots + \frac{x_1 w_{1m}}{\sum_{i=1}^{n} x_i \cdot w_{im}} \cdot R_m^{(\ell+1)} \\ \\ \frac{x_n w_{n1}}{\sum_{i=1}^{n} x_i \cdot w_{i1}} + R_1^{(\ell+1)} + \dots + \frac{x_n w_{nm}}{\sum_{i=1}^{n} x_i \cdot w_{im}} \cdot R_m^{(\ell+1)} \end{pmatrix}$$

$$= \begin{pmatrix} \sum_j \frac{x_1 w_{1j}}{\sum_i x_i \cdot w_{ij}} \cdot R_j^{(\ell+1)} \\ \\ \sum_j \frac{x_n w_{nj}}{\sum_i x_i \cdot w_{ij}} \cdot R_j^{(\ell+1)} \end{pmatrix}$$

Classification task :

Binary Classifier
BC





$\rightarrow$ Loss $= \sum_{i=1}^{n} HL_i$