$$z = \mathrm{softmax}(\hat{z})$$

$$z_i = \frac{e^{\hat{z}_i}}{\sum_\ell e^{\hat{z}_\ell}} = e^{\hat{z}_i}\left(\sum e^{\hat{z}_\ell}\right)^{-1}$$

$$\frac{\partial z_i}{\partial \hat{z}_j} = -1\left(\sum e^{\hat{z}_\ell}\right)^{-2} \cdot e^{\hat{z}_j} e^{\hat{z}_i} + \delta_{ij}\, e^{\hat{z}_i}\left(\sum -\right)^{-1}$$

$$= -z_i z_j + \delta_{ij} z_i$$

$$-E = \sum_\ell r_\ell \log z_\ell$$

$$-\frac{\partial E}{\partial z_i} = \frac{r_i}{z_i}$$

$$-\frac{\partial E}{\partial \hat{z}_j} = \sum_\ell \frac{\partial E}{\partial z_\ell}\frac{\partial z_\ell}{\partial \hat{z}_j} = \sum_\ell \frac{r_\ell}{z_\ell}\cdot\left(\delta_{\ell j} z_\ell - z_\ell z_j\right)$$

$$= \sum_\ell r_\ell\left(\delta_{\ell j} - z_j\right) = \underbrace{\sum_\ell r_\ell \delta_{\ell j}}_{\text{selects one term}} - z_j \underbrace{\sum_\ell r_\ell}_{1}$$

$$= r_j - z_j$$

/tmp/IMG_2974.jpeg

$$\hat{z}_i = w_{i0} + w_{i1} y_1 + w_{i2} y_2 + \dots w_{iN} y_N$$
$$= \underline{w}_i \cdot \begin{pmatrix} 1 \\ y \end{pmatrix}$$

$$\frac{\partial z_k}{\partial w_{ij}} = 0 \text{ if } k \neq i$$

$$\frac{\partial \hat{z}_i}{\partial y_j} = w_{ij} \qquad \frac{\partial \hat{z}_i}{\partial w_{ij}} = y_j \text{ w/ } y_0 \equiv 1$$

$$-\frac{\partial E}{\partial y_j} = \sum_{l=1}^{N} \frac{\partial E}{\partial \hat{z}_l} \frac{\partial \hat{z}_l}{\partial y_j}$$
$$= \sum_l (r_l - z_l) w_{lj}$$
$$-\nabla_y E = (r - z)^T W$$

$$-\frac{\partial E}{\partial w_{ij}} = \sum_{l=0}^{N} \frac{\partial E}{\partial \hat{z}_l} \frac{\partial \hat{z}_l}{\partial w_{ij}}$$
$$= \sum (r_l - z_l) \cdot y_j \delta_{il}$$
$$= (r_i - z_i) \cdot y_j$$
$$-\nabla_W E = (\underline{r} - \underline{z}) \cdot [1, y^T]$$

CLASS

$$\sigma(x) = \begin{cases} \dfrac{1}{1+e^{-x}} & \text{sigmoid} \\ \dfrac{e^{x}-e^{-x}}{e^{x}+e^{-x}} & \text{tanh} = \dfrac{1-e^{-2x}}{1+e^{-2x}} \\ \max(x, \alpha x) & \text{ReLU} \end{cases} \quad \sigma' = \begin{cases} \sigma(1-\sigma) \\ 1-\sigma^2 \\ = \begin{cases} 1 & \text{if } x>0 \\ \alpha & \text{if } x<0 \end{cases} \end{cases}$$

$$\frac{\partial y'_i}{\partial y_r} = \sigma'(\hat{y}_i)\,\delta_{ij} = \sigma'(y_i)|_{i=r}$$

$$\hat{y}_i = V_{i0} + V_{i1}x_1 + V_{i2}x_2 \cdots V_{im}x_m$$

$$\frac{\partial y'_i}{\partial v_{ij}} = \frac{\partial y_i}{\partial \hat{y}_i} \cdot \frac{\partial \hat{y}_i}{\partial V_{ij}}$$

$$= \sigma'(\hat{y}_i) \cdot x_j$$

$$-\frac{\partial E}{\partial v_{ij}} = \frac{\partial E}{\partial y'_i}\frac{\partial \hat{y}_i}{\partial v_{ij}} =$$

$$= \left[\sum_l (r_l - z_l) w_{li}\right] \cdot \sigma'(y_i) \cdot x_j$$

$$= \left[\nabla_y E\right]_i \otimes \sigma'(y_i) \cdot x_j$$

$$-\nabla E = (\nabla_y E) \otimes \sigma'(y) \cdot [1, x^T]$$

/tmp/IMG_2976.jpeg