

Combining:

$$\begin{bmatrix} 1 - \sigma_1(x) & -\sigma_2(x) & \dots & -\sigma_N(x) \\ -\sigma_1(x) & & & \\ \vdots & \ddots & & \\ -\sigma_1(x) & & & 1 - \sigma_N(x) \end{bmatrix}$$

c) $z = v \cdot D_x \log(\sigma(x))$

~~z_1~~ $v = [v_1, \dots, v_N]$

$$\begin{aligned} z_1 &= v_1(1 - \sigma_1(x)) - v_2 \sigma_1(x) - \dots - v_N \sigma_1(x) \\ &= v_1 - \sigma_1(x) \sum_{j=1}^N v_j \end{aligned}$$

$$z_i = v_i - \sigma_i(x) \sum_{j=1}^N v_j$$

d) with softmax: $l(z, t) = - \sum_{j=1}^N t_j \ln(z_j)$

with log-softmax: $l(z, t) = - \sum_{j=1}^N t_j z_j$

$$\frac{\partial l(z, t)}{\partial z_i} = - \frac{\partial}{\partial z_i} t_i z_i = -t_i$$

$$\nabla_z l(z, t) = [-t_1, \dots, -t_N] = \textcircled{-t}$$