c) $\ell(z, t) = -\sum_{i=1}^{N} t_i \ln(z_i)$, $t \in [0, 1]^{1\times N}$

$$D_z \ell(z,t) = \left[\frac{\partial \ell(z,t)}{\partial z_1}, \ldots\ldots, \frac{\partial \ell(z,t)}{\partial z_N} \right]$$

$$\frac{\partial \ell(z,t)}{\partial z_i} = \frac{\partial}{\partial z_i}\left(-t_i \ln(z_i)\right)$$

$$= \frac{-t_i}{z_i}$$

d) division by zero if $z_i \approx 0$

QUESTION 3:

a) $\tanh(x) = \dfrac{e^{x} - e^{-x}}{e^{x} + e^{-x}}$

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

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

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

$$= \cancel{1 - \tan}\quad 1 - \tanh^2 x$$

QUESTION 4:

a) $\text{softmax}(x) = \text{softmax}(x+c)$

$$\text{softmax}_i(x+c) = \frac{e^{x_i+c}}{\sum_{j=1}^{N} e^{x_j+c}} = \frac{\cancel{e^{c}}\, e^{x_i}}{\cancel{e^{c}} \sum_{j=1}^{N} e^{x_j}}$$

$$= \text{softmax}_i(x)$$

b) $D_x \log(\sigma(x)) \in \mathbb{R}^{N \times N}$

$\{\log = \ln\}$

$$\log \sigma_i(x) = \log \frac{e^{x_i}}{\sum_{j=1}^{N} e^{x_j}} = \log e^{x_i} - \log\left(\sum_{j=1}^{N} e^{x_j}\right)$$

$$= x_i - \log\left(\sum_{j=1}^{N} e^{x_j}\right)$$

diagonal:

$$\frac{\partial}{\partial x_i} \log(\sigma_i(x)) = 1 - \frac{e^{x_i}}{\sum_{j=1}^{N} e^{x_j}} \star = \boxed{1 - \sigma_i(x)}$$

$$\star \quad \frac{\partial \log(f(x))}{\partial x} = \frac{1}{f(x)} \cdot f'(x)$$

off-diagonal:

$$\frac{\partial}{\partial x_j} \log(\sigma_i(x)) = \cancel{\frac{\partial x_i}{\partial x_j}}^{\,0} - \frac{\partial}{\partial x_j} \log\left(\sum_{k=1}^{N} e^{x_k}\right)$$

$$= \frac{-e^{x_j}}{\sum_{k=1}^{N} e^{x_k}} = \boxed{-\sigma_j(x)}$$

Combining:

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

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

$\cancel{z_1} = \quad v = [v_1 \ldots\ldots v_N]$

$$z_1 = v_1(1-\sigma_1(x)) - v_2\sigma_1(x) - \ldots\ldots v_N\sigma_1(x)$$

$$= v_1 - \sigma_1(x)\sum_{j=1}^{N} v_j$$

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

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

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

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

$$D_z l(z,t) = [-t_1, \ldots, -t_N] = -t$$