In the equations below, $\mathbf{X} = \begin{pmatrix} x_{11} & x_{12} \\ \vdots & \vdots \\ x_{N1} & x_{N2} \end{pmatrix}$ is a matrix of data points. $\mathbf{U} = \begin{pmatrix} u_{11} & u_{12} \\ u_{21} & u_{22} \\ u_{31} & u_{32} \end{pmatrix} \in \mathbb{R}^{3\times 2}$, $\mathbf{b} = \begin{pmatrix} b_1 \\ b_2 \\ b_3 \end{pmatrix} \in \mathbb{R}^3$, $\mathbf{W} = \begin{pmatrix} w_1 & w_2 & w_3 \end{pmatrix} \in \mathbb{R}^{1\times 3}$, $c \in \mathbb{R}$ are neural network parameters.

Forward Equations (Vectorized form)

$$\begin{aligned}
\mathbf{G} &= \mathbf{X}\mathbf{U}^T + \mathbf{1}\mathbf{b}^T \\
\mathbf{H} &= \tanh(\mathbf{G}) \\
\mathbf{z} &= \mathbf{H}\mathbf{W}^T + \mathbf{1}c \\
\mathbf{y} &= \sigma(\mathbf{z})
\end{aligned}$$

Forward equations (Scalar form)

$$\begin{aligned}
g_{ij} &= u_{j1}x_{i1} + u_{j2}x_{i2} + b_j \\
h_{ij} &= \tanh(g_{ij}) \\
z_i &= w_1 h_{i1} + w_2 h_{i2} + w_3 h_{i3} + c \\
y_i &= \sigma(z_i)
\end{aligned}$$

Here, i indexes data points and j indexes hidden units, so $i \in \{1, \ldots, N\}$ and $j \in \{1, 2, 3\}$.

Cost function

$$\mathcal{E}(\mathbf{z}, \mathbf{t}) = \frac{1}{N}\left[\sum_{i=1}^{N} \mathcal{L}(z_i, t_i)\right]$$

$$\mathcal{L}(z, t) = t\log(1 + \exp(-z)) + (1 - t)\log(1 + \exp(z))$$

Backward Equations (Scalar form)

$$\begin{aligned}
\overline{\mathcal{E}} &= 1 \\
\overline{z_i} &= \overline{\mathcal{E}}\frac{\partial \mathcal{E}}{\partial z_i} = \boxed{\frac{1}{N}(y_i - t_i)} \text{ (see Lecture 4 notes)} \\
\overline{w_j} &= \sum_{i=1}^{N} \overline{z_i}\frac{\partial z_i}{\partial w_j} = \sum_{i=1}^{N} \overline{z_i} h_{ij} \\
\overline{c} &= \sum_{i=1}^{N} \overline{z_i}\frac{\partial z_i}{\partial c} = \sum_{i=1}^{N} \overline{z_i} \\
\overline{h_{ij}} &= \overline{z_i}\frac{\partial z_i}{\partial h_{ij}} = \overline{z_i} w_j \\
\overline{g_{ij}} &= \overline{h_{ij}}\frac{\partial h_{ij}}{\partial g_{ij}} = \overline{h_{ij}}(1 - \tanh^2(g_{ij})) \text{ (check derivative of tanh)} \\
\overline{u_{jk}} &= \sum_{i=1}^{N} \overline{g_{ij}}\frac{\partial g_{ij}}{\partial u_{jk}} = \sum_{i=1}^{N} \overline{g_{ij}} x_{ik} \\
\overline{b_j} &= \sum_{i=1}^{N} \overline{g_{ij}}\frac{\partial g_{ij}}{\partial b_j} = \sum_{i=1}^{N} \overline{g_{ij}}
\end{aligned}$$

y is logistic y = 1/(1+e^{-z})

summation over all the samples
by multivariate chain rule

As above, i indexes data points and j indexes hidden units, so $i \in \{1, \ldots, N\}$ and $j \in \{1, 2, 3\}$. In addition, k indexes the data dimension so $k \in 1, 2$.

Backward Equations (Vectorized form)

$$\begin{aligned}
\overline{\mathcal{E}} &= 1 \\
\overline{\mathbf{z}} &= \frac{1}{N}(\mathbf{y} - \mathbf{t}) \\
\overline{\mathbf{W}} &= \mathbf{H}^T \overline{\mathbf{z}} \\
\overline{c} &= \mathbf{z}^T \mathbf{1} \\
\overline{\mathbf{H}} &= \overline{\mathbf{z}} \mathbf{W}^{T} \\
\overline{\mathbf{G}} &= \overline{\mathbf{H}} \odot (1 - \tanh^2(\mathbf{G})) \\
\overline{\mathbf{U}} &= \overline{\mathbf{G}}^T \mathbf{X} \\
\overline{\mathbf{b}} &= \overline{\mathbf{G}}^T \mathbf{1}
\end{aligned}$$

inner product, where H_{ij} = z_{i} dot W_{j}

hadamard element-wise product