

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. $\mathbf{U} = \begin{pmatrix} u_{11} & u_{12} \\ u_{21} & u_{22} \\ u_{31} & u_{32} \end{pmatrix} \in \mathbb{R}^{3 \times 2}$, $\mathbf{b} = \begin{pmatrix} b_1 \\ b_2 \\ b_3 \end{pmatrix} \in \mathbb{R}^3$, $\mathbf{W} = (w_1 \ w_2 \ w_3) \in \mathbb{R}^{1 \times 3}$, $c \in \mathbb{R}$ are neural network parameters.

Forward Equations (Vectorized form)

$$\mathbf{G} = \mathbf{X}\mathbf{U}^T + \mathbf{1}\mathbf{b}^T$$

$$\mathbf{H} = \tanh(\mathbf{G})$$

$$\mathbf{z} = \mathbf{H}\mathbf{W}^T + \mathbf{1}c$$

$$\mathbf{y} = \sigma(\mathbf{z})$$

Forward equations (Scalar form)

$$g_{ij} = u_{j1}x_{i1} + u_{j2}x_{i2} + b_j$$

$$h_{ij} = \tanh(g_{ij})$$

$$z_i = w_1h_{i1} + w_2h_{i2} + w_3h_{i3} + c$$

$$y_i = \sigma(z_i)$$

Here, i indexes data points and j indexes hidden units, so $i \in \{1, \dots, N\}$ 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)

$$\bar{\mathcal{E}} = 1$$

$$\bar{z}_i = \bar{\mathcal{E}} \frac{\partial \mathcal{E}}{\partial z_i} = \frac{1}{N} (y_i - t_i) \text{ (see Lecture 4 notes)}$$

$$\bar{w}_j = \sum_{i=1}^N \bar{z}_i \frac{\partial z_i}{\partial w_j} = \sum_{i=1}^N \bar{z}_i h_{ij}$$

$$\bar{c} = \sum_{i=1}^N \bar{z}_i \frac{\partial z_i}{\partial c} = \sum_{i=1}^N \bar{z}_i$$

$$\bar{h}_{ij} = \bar{z}_i \frac{\partial z_i}{\partial h_{ij}} = \bar{z}_i w_j$$

$$\bar{g}_{ij} = \bar{h}_{ij} \frac{\partial h_{ij}}{\partial g_{ij}} = \bar{h}_{ij} (1 - \tanh^2(g_{ij})) \text{ (check derivative of tanh)}$$

$$\bar{u}_{jk} = \sum_{i=1}^N \bar{g}_{ij} \frac{\partial g_{ij}}{\partial u_{jk}} = \sum_{i=1}^N \bar{g}_{ij} x_{ik}$$

$$\bar{b}_j = \sum_{i=1}^N \bar{g}_{ij} \frac{\partial g_{ij}}{\partial b_j} = \sum_{i=1}^N \bar{g}_{ij}$$

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

Backward Equations (Vectorized form)

$$\bar{\mathcal{E}} = 1$$

$$\bar{\mathbf{z}} = \frac{1}{N} (\mathbf{y} - \mathbf{t})$$

$$\bar{\mathbf{W}} = \mathbf{H}^T \bar{\mathbf{z}}$$

$$\bar{c} = \mathbf{z}^T \mathbf{1} \text{ zer bar } ^T * \mathbf{1}$$

$$\bar{\mathbf{H}} = \bar{\mathbf{z}} \mathbf{W}$$

$$\bar{\mathbf{G}} = \bar{\mathbf{H}} \odot (1 - \tanh^2(\mathbf{G}))$$

$$\bar{\mathbf{U}} = \bar{\mathbf{G}}^T \mathbf{X}$$

$$\bar{\mathbf{b}} = \bar{\mathbf{G}}^T \mathbf{1}$$