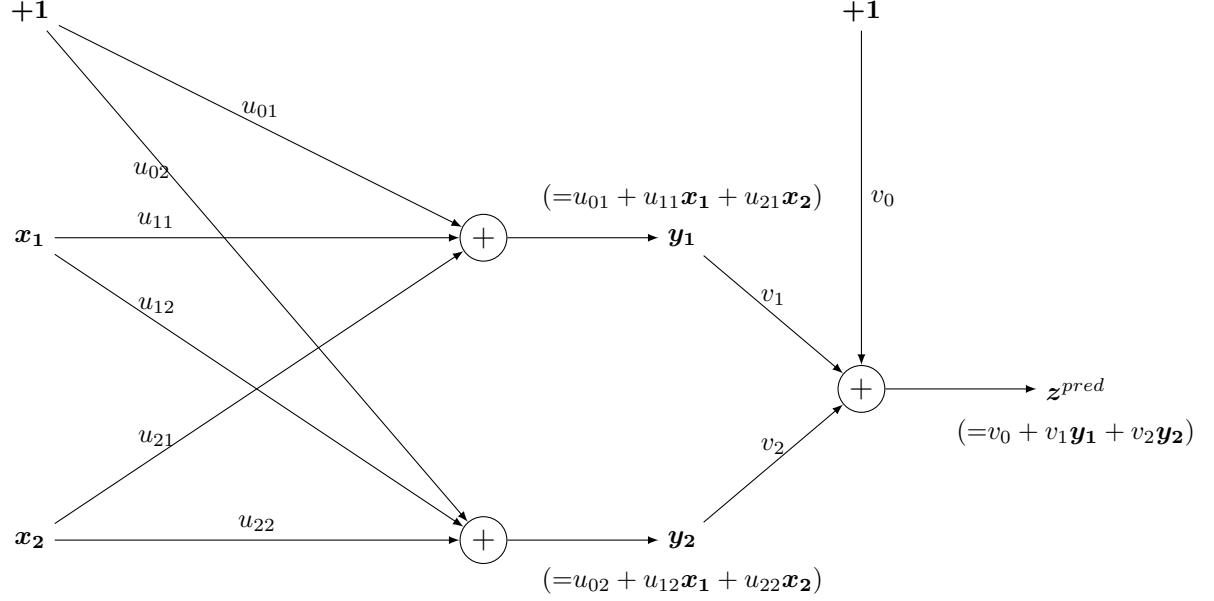


Let us suppose we want to use the following linear neural network architecture, with no activation function:



Let us define the following matrices, which model the above network:

$$\mathbf{X}_0 = \begin{bmatrix} x_{11} & x_{12} \\ x_{21} & x_{22} \\ x_{31} & x_{32} \\ x_{41} & x_{42} \end{bmatrix} \quad \mathbf{X} = \begin{bmatrix} 1 & \mathbf{X}_0 \end{bmatrix} = \begin{bmatrix} 1 & x_{11} & x_{12} \\ 1 & x_{21} & x_{22} \\ 1 & x_{31} & x_{32} \\ 1 & x_{41} & x_{42} \end{bmatrix} \quad \mathbf{u} = \begin{bmatrix} u_{01} & u_{02} \\ u_{11} & u_{12} \\ u_{21} & u_{22} \end{bmatrix}$$

$$\mathbf{Y}_0 = \mathbf{X}\mathbf{u} = \begin{bmatrix} y_{11} & y_{12} \\ y_{21} & y_{22} \\ y_{31} & y_{32} \\ y_{41} & y_{42} \end{bmatrix} \quad \mathbf{Y} = \begin{bmatrix} 1 & \mathbf{Y}_0 \end{bmatrix} = \begin{bmatrix} 1 & y_{11} & y_{12} \\ 1 & y_{21} & y_{22} \\ 1 & y_{31} & y_{32} \\ 1 & y_{41} & y_{42} \end{bmatrix} \quad \mathbf{v} = \begin{bmatrix} v_0 \\ v_1 \\ v_2 \end{bmatrix}$$

$$\mathbf{z}^{pred} = \mathbf{Y}\mathbf{v} = \begin{bmatrix} z_1 \\ z_2 \\ z_3 \\ z_4 \end{bmatrix}$$

The final output of the network can also be written as:

$$z_i^{pred} = v_0 + v_1y_{i1} + v_2y_{i2} \quad (1)$$

$$= u_0 + v_1(u_{01} + u_{11}x_{i1} + u_{21}x_{i2}) + v_2(u_{02} + u_{12}x_{i1} + u_{22}x_{i2}) \quad (2)$$

$$= v_0 + u_{01}v_1 + u_{02}v_2 + u_{11}v_1x_{i1} + u_{21}v_1x_{i2} + u_{12}v_2x_{i1} + u_{22}v_2x_{i2} \quad (3)$$

Let us use the following cost function, which we want to minimize:

$$C = \frac{1}{2} \sum_{i=1}^n (z_i^{reality} - z_i^{pred})^2$$

The derivative of C with respect to any coefficient w is:

$$\frac{\partial C}{\partial w} = - \sum_{i=1}^n \frac{\partial z_i^{pred}}{\partial w} (z_i^{reality} - z_i^{pred})$$

Using (3) we can calculate the derivative of C with respect to each weight u_{ij} and v_i of the network and write them in vector form:

$$\begin{aligned} \frac{\partial z_i^{pred}}{\partial \mathbf{u}} &= \begin{bmatrix} \frac{\partial z_i^{pred}}{\partial u_{01}} & \frac{\partial z_i^{pred}}{\partial u_{02}} \\ \frac{\partial z_i^{pred}}{\partial u_{11}} & \frac{\partial z_i^{pred}}{\partial u_{12}} \\ \frac{\partial z_i^{pred}}{\partial u_{21}} & \frac{\partial z_i^{pred}}{\partial u_{22}} \end{bmatrix} = \begin{bmatrix} v_1 & v_2 \\ v_1 x_{i1} & v_2 x_{i1} \\ v_1 x_{i2} & v_2 x_{i2} \end{bmatrix} \\ \frac{\partial z_i^{pred}}{\partial \mathbf{v}} &= \begin{bmatrix} \frac{\partial z_i^{pred}}{\partial v_0} \\ \frac{\partial z_i^{pred}}{\partial v_1} \\ \frac{\partial z_i^{pred}}{\partial v_2} \end{bmatrix} = \begin{bmatrix} 1 \\ u_{01} + u_{11}x_{i1} + u_{21}x_{i2} \\ u_{02} + u_{12}x_{i1} + u_{22}x_{i2} \end{bmatrix} \end{aligned}$$

Let us define d_i such that

$$d_i = -(z_i^{reality} - z_i^{pred})$$

By plugging in the corresponding derivatives found above in to (dC), we obtain:

$$\begin{aligned} \frac{\partial C}{\partial v_0} &= \sum_{i=1}^n d_i \\ \frac{\partial C}{\partial v_1} &= \sum_{i=1}^n (u_{01} + u_{11}x_{i1} + u_{21}x_{i2})d_i \\ \frac{\partial C}{\partial v_2} &= \sum_{i=1}^n (u_{02} + u_{12}x_{i1} + u_{22}x_{i2})d_i \end{aligned}$$

$$\begin{aligned} \frac{\partial C}{\partial u_{01}} &= v_1 \sum_{i=1}^n d_i & \frac{\partial C}{\partial u_{02}} &= v_2 \sum_{i=1}^n d_i \\ \frac{\partial C}{\partial u_{11}} &= v_1 \sum_{i=1}^n x_{i1} d_i & \frac{\partial C}{\partial u_{12}} &= v_2 \sum_{i=1}^n x_{i1} d_i \\ \frac{\partial C}{\partial u_{21}} &= v_1 \sum_{i=1}^n x_{i2} d_i & \frac{\partial C}{\partial u_{22}} &= v_2 \sum_{i=1}^n x_{i2} d_i \end{aligned}$$

Which can be cast in matrix form as:

$$\begin{aligned}\frac{\partial C}{\partial \mathbf{u}} &= \mathbf{X}^T \mathbf{d} \begin{bmatrix} v_1 & v_2 \end{bmatrix} \\ \frac{\partial C}{\partial \mathbf{v}} &= \mathbf{Y}^T \mathbf{d}\end{aligned}$$

with

$$\mathbf{d} = \begin{bmatrix} d_1 \\ d_2 \\ d_3 \\ d_4 \end{bmatrix}$$