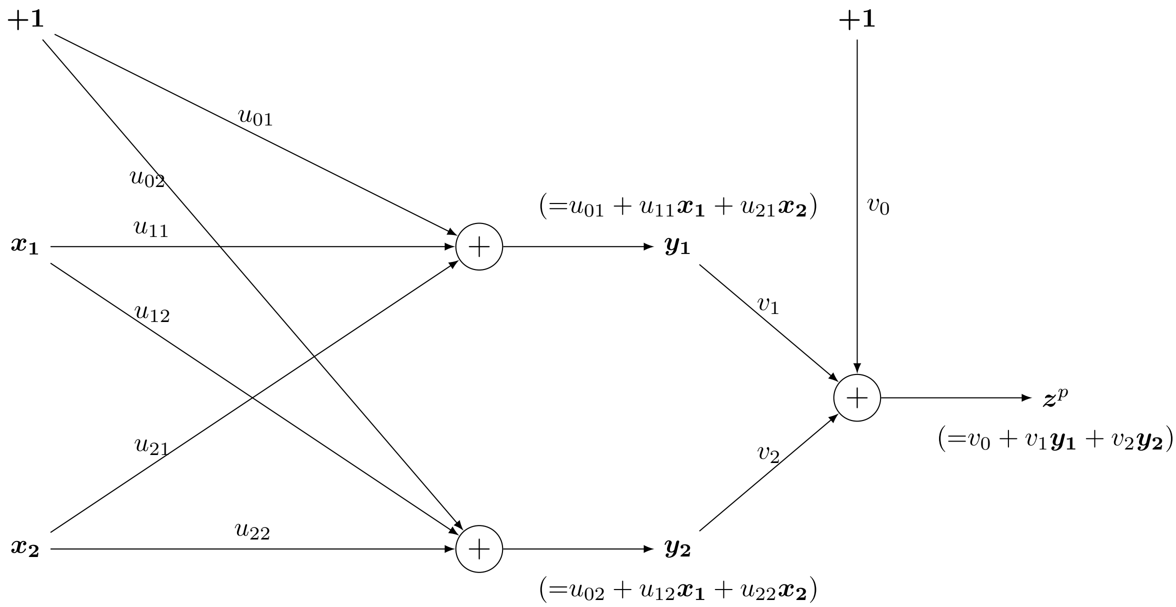
1 Model outline

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



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

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

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

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

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

$$\begin{aligned} z_i^p &= v_0 + v_1 y_{i1} + v_2 y_{i2} \\ &= u_0 + v_1(u_{01} + u_{11}x_{i1} + u_{21}x_{i2}) + v_2(u_{02} + u_{12}x_{i1} + u_{22}x_{i2}) \\ &= v_0 + u_{01}v_1 + u_{02}v_2 + (u_{11}v_1 + u_{12}v_2)x_{i1} + (u_{21}v_1 + u_{22}v_2)x_{i2} \end{aligned} \tag{1}$$

2 Training the model

Let us use the following quadratic cost function which in this case we want to maximize:

$$C = -\frac{1}{2}\sum_{i=1}^{n}(z_i^r - z_i^p)^2 \tag{2}$$

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

$$\frac{\partial C}{\partial w} = \sum_{i=1}^{n} \frac{\partial z_i^p}{\partial w}(z_i^r - z_i^p) \tag{3}$$

Using (1) we can calculate the derivative of C with respect to each weight u_{ij} and v_i of the network and arrange them in matrix form:

$$\frac{\partial z_i^p}{\partial \mathbf{u}} = \begin{bmatrix} \frac{\partial z_i^p}{\partial u_{01}} & \frac{\partial z_i^p}{\partial u_{02}} \\ \frac{\partial z_i^p}{\partial u_{11}} & \frac{\partial z_i^p}{\partial u_{12}} \\ \frac{\partial z_i^p}{\partial u_{21}} & \frac{\partial z_i^p}{\partial u_{22}} \end{bmatrix} = \begin{bmatrix} v_1 & v_2 \\ v_1 x_{i1} & v_2 x_{i1} \\ v_1 x_{i2} & v_2 x_{i2} \end{bmatrix} \tag{4}$$

$$\frac{\partial z_i^p}{\partial \mathbf{v}} = \begin{bmatrix} \frac{\partial z_i^p}{\partial v_0} \\ \frac{\partial z_i^p}{\partial v_1} \\ \frac{\partial z_i^p}{\partial v_2} \end{bmatrix} = \begin{bmatrix} 1 \\ u_{01} + u_{11} x_{i1} + u_{21} x_{i2} \\ u_{02} + u_{12} x_{i1} + u_{22} x_{i2} \end{bmatrix} \tag{5}$$

Let us define d_i such that

$$d_i = z_i^r - z_i^p \tag{6}$$

By plugging the corresponding derivatives found above in (3), we obtain:

$$\frac{\partial C}{\partial u_{01}} = v_1 \sum_{i=1}^{n} d_i \tag{7}$$

$$\frac{\partial C}{\partial u_{11}} = v_1 \sum_{i=1}^{n} x_{i1} d_i \tag{8}$$

$$\frac{\partial C}{\partial u_{21}} = v_1 \sum_{i=1}^{n} x_{i2} d_i \tag{9}$$

$$\frac{\partial C}{\partial u_{02}} = v_2 \sum_{i=1}^{n} d_i \tag{10}$$

$$\frac{\partial C}{\partial u_{12}} = v_2 \sum_{i=1}^{n} x_{i1} d_i \tag{11}$$

$$\frac{\partial C}{\partial u_{22}} = v_2 \sum_{i=1}^{n} x_{i2} d_i \tag{12}$$

$$\frac{\partial C}{\partial v_0} = \sum_{i=1}^{n} d_i \tag{13}$$

$$\frac{\partial C}{\partial v_1} = \sum_{i=1}^{n} (u_{01} + u_{11} x_{i1} + u_{21} x_{i2}) d_i \tag{14}$$

$$\frac{\partial C}{\partial v_2} = \sum_{i=1}^{n} (u_{02} + u_{12} x_{i1} + u_{22} x_{i2}) d_i \tag{15}$$

Which can be cast in matrix form as:

$$\frac{\partial C}{\partial \mathbf{u}} = \mathbf{X}^T \mathbf{d} \begin{bmatrix} v_1 & v_2 \end{bmatrix} \tag{16}$$

$$\frac{\partial C}{\partial \mathbf{v}} = \mathbf{Y}^T \mathbf{d} \tag{17}$$

with

$$\mathbf{d} = \begin{bmatrix} d_1 \\ d_2 \\ d_3 \\ d_4 \end{bmatrix} \tag{18}$$

3 Equivalence with the linear least squares regression model

To find the optimal $\mathbf{u}$ and $\mathbf{v}$, we set derivatives (7) to (15) to 0 and solve for the u_{ij} and v_j coefficients. (7) to (12) reduce to:

$$\sum_{i=1}^{n} d_i = 0 \tag{19}$$

$$\sum_{i=1}^{n} x_{i1} d_i = 0 \tag{20}$$

$$\sum_{i=1}^{n} x_{i2} d_i = 0 \tag{21}$$

Moreover, we can rewrite (14) as:

$$u_{01}\sum_{i=1}^{n} d_i + u_{11}\sum_{i=1}^{n} x_{i1}d_i + u_{21}\sum_{i=1}^{n} x_{i2}d_i = 0 \tag{22}$$

which is always true if equations (19)-(20) are satisfied, and similarly for equation (15). So only equations (19)-(20) need to be solved. Those equations can be rewritten as:

$$\sum_{i=1}^{n} z_i^r - w_0 - w_1 x_{i1} - w_2 x_{i2} = 0 \tag{23}$$

$$\sum_{i=1}^{n} x_{i1}(z_i^r - w_0 - w_1 x_{i1} - w_2 x_{i2}) = 0 \tag{24}$$

$$\sum_{i=1}^{n} x_{i2}(z_i^r - w_0 - w_1 x_{i1} - w_2 x_{i2}) = 0 \tag{25}$$

where

$$w_0 = v_0 + u_{01}v_1 + u_{02}v_2 \tag{26}$$

$$w_1 = u_{11}v_1 + u_{12}v_2 \tag{27}$$

$$w_2 = u_{21}v_1 + u_{22}v_2 \tag{28}$$

These are the normal equations of the linear least squares regression model, so the coefficients w_k need to satisfy the normal equations. There are possibly multiple solutions for u_{ij} and v_j, but they are constrained by the normal equations, and eventually the optimal solutions is the same as that of the least squares.