To find the equations for params, we take log likelihood of the distribution:

$$LL = \sum_{n=1}^{N} \left\{ -\frac{1}{2}\log 2\pi\sigma_u^2 - \frac{1}{2\sigma_u^2}\left(x_{u,n} - \left(\sum_{c\in x_{\pi_u}} w_{uc}x_{uc,n} + w_{u0}\right)\right)^2 \right\} \tag{1}$$

We'll proceed to find the parameters for each node independently, and so dropping the u subscript from here on. Further, to avoid confusion, let's call the output node y, and the parents of y (which will our inputs) as x_c for c^{th} parent/input. y_n is the n^{th} output observation for the node in consideration, and $x_{c,n}$ is the n^{th} observation for the c^{th} input.

$$LL = \sum_{n=1}^{N} \left\{ -\frac{1}{2}\log(2\pi\sigma^2) - \frac{1}{2\sigma^2}\left(y_n - \left(\sum_{c\in x_{\pi}} w_c x_{c,n} + w_0\right)\right)^2 \right\} \tag{2}$$

Since we will be working with matrices, let's also define the required matrices for these variables:

$W = \begin{bmatrix} w_0 \\ w_1 \\ w_2 \\ \vdots \\ w_c \end{bmatrix}$, $dim(W) = (c+1)\times 1$ is our weights matrix including w_0 . $Y = \begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ \vdots \\ y_n \end{bmatrix}$, $dim(Y) = n \times 1$ where n is the total number of samples.

$X = \begin{bmatrix} 1 & x_{1,1} & x_{2,1} & x_{3,1} & \dots & x_{c,1} \\ 1 & x_{1,2} & x_{2,2} & x_{3,2} & \dots & x_{c,2} \\ 1 & x_{1,3} & x_{2,3} & x_{3,3} & \dots & x_{c,3} \\ \vdots & & & & & \\ 1 & x_{1,n} & x_{2,n} & x_{3,n} & \dots & x_{c,n} \end{bmatrix}$, $dim(X) = n \times (c+1)$, and $x_{i,j}$ is j^{th} observation sample for i^{th} input variable/parent. For notational convenience, we define $x_{0,i} = 1$ for all i .

Going back to our LL equation, for MLE, we take partial derivatives $w.r.t.$ our weight parameters $\{w_0, w_1, w_2, \cdots, w_c\}$ and setting them to 0 one by one. That yields $(c+1)$ equations. For w_i, the equation is:

$$\frac{\partial(LL)}{\partial w_i} = \sum_{n=1}^{N}(y_n - (w_0x_{0,n} + w_1x_{1,n} + \ldots + w_cx_{c,n}))x_{i,n} = 0$$

Solving this further, and moving the summations in:

$$w_0\sum_{n=1}^{N} x_{0,n}x_{i,n} + w_1\sum_{n=1}^{N} x_{1,n}x_{i,n} + w_2\sum_{n=1}^{N} x_{2,n}x_{i,n} + w_3\sum_{n=1}^{N} x_{3,n}x_{i,n} + \cdots + w_c\sum_{n=1}^{N} x_{c,n}x_{i,n} = \sum_{n=1}^{N} y_nx_{i,n}$$

We have $(c+1)$ such equations for each w_i with $i \in [0, c]$. This is a system of $(c+1)$ equations in $(c+1)$ variables which can be uniquely solved. With the above matrix definitions, we can write the combined matrix replacement for the above equations as:

$$(X^T \times X) \times W = X^T \times Y$$

which is what we have implemented in code.

To solve for variance σ^2, we again take partial derivative of $LL\ w.r.t\ \sigma^2$ and set it to 0, which after some simplification gives:

$$\sigma^2 = \frac{1}{N}\left(\sum_{n=1}^{N}(y_n - (w_0x_{0,n} + w_1x_{1,n} + w_2x_{2,n}\cdots w_cx_{c,n}))^2\right)$$

which can be rewritten in matrix form as:

$$\sigma^2 = \frac{1}{N}[(XW - Y)^T \times (XW - Y)]$$