Equivalent Networks:

1. Background on multi layer perceptron:

A multi layer perceptron MLP is a network which maps some network to produce some outputs. The network consist of layers of nodes connected by weights. The input layer is transformed to get the output layer.

The MLP with 2 input nodes and 2 output nodes are as shown. assumption of linear activation we have the values given as

$$a_0^1 = w_{0,0}^{(1)}a_0^{(0)} + w_{0,1}^{(1)}a_1^{(0)} + b_0^{(1)}$$
$$a_1^1 = w_{1,0}^{(1)}a_0^{(0)} + w_{1,1}^{(1)}a_1^{(0)} + b_1^{(1)}$$

where
$a_i^1$: equation of i node at layer l
$b_i^l$: bias action on i th node at layer l
$w_{i,j}$: weight connecting nth node at layer l-1 to th node at layer l

The equation $\overrightarrow{a}^1 = W^1\overrightarrow{a}^0 + \overrightarrow{b}^1$ where $\overrightarrow{a}^{(0)} = \begin{pmatrix} a_0^0 \\ a_1^0 \end{pmatrix}$ $\overrightarrow{a}^1 = \begin{pmatrix} a_0^1 \\ a_1^1 \end{pmatrix}$ $\overrightarrow{b}^1 = \begin{pmatrix} b_0^1 \\ b_1^1 \end{pmatrix}$

$$W^1 = \begin{pmatrix} w_{0,0}^1 & w_{0,1}^1 \\ w_{1,0}^1 & w_{1,1}^1 \end{pmatrix}$$

There can be multiple hidden layers and output layers where the above transformation equation is used

$$\overrightarrow{a}^2 = W^2\overrightarrow{a}^1 + \overrightarrow{b}^2$$
$$\overrightarrow{a}^3 = W^3\overrightarrow{a}^2 + \overrightarrow{b}^3$$

For the given Network 1 (MLP with multiple hidden layers)

The representation of the first hidden layer is

$\overrightarrow{a}^1 = W^1\overrightarrow{a}^0 + \overrightarrow{b}^1$ ….. (i)

Second Hidden layer is

$\overrightarrow{a}^2 = W^2\overrightarrow{a}^1 + \overrightarrow{b}^2$ ….. (ii)

Final Layer being

$\overrightarrow{a}^3 = W^3\overrightarrow{a}^2 + \overrightarrow{b}^3$ …. (iii)

Where $\overrightarrow{a}^{(0)} = \begin{pmatrix} a_0^0 \\ a_1^0 \\ a_2^0 \\ a_3^0 \\ a_4^0 \\ a_5^0 \end{pmatrix}$, $\overrightarrow{a}^{(1)} = \begin{pmatrix} a_0^1 \\ a_1^1 \\ a_2^1 \\ a_3^1 \\ a_4^1 \\ a_5^1 \end{pmatrix}$ $\overrightarrow{a}^{(2)} = \begin{pmatrix} a_0^2 \\ a_1^2 \\ a_2^2 \\ a_3^2 \\ a_4^2 \\ a_5^2 \end{pmatrix}$ ,$W^{(1)} = \begin{pmatrix} w_{0,0}^1 & w_{0,1}^1 & w_{0,2}^1 & w_{0,3}^1 & w_{0,4}^1 \\ w_{1,0}^1 & w_{1,1}^1 & w_{1,2}^1 & w_{1,3}^1 & w_{1,4}^1 \\ w_{2,0}^1 & w_{2,1}^1 & w_{2,2}^1 & w_{2,3}^1 & w_{2,4}^1 \\ w_{3,0}^1 & w_{3,1}^1 & w_{3,2}^1 & w_{3,3}^1 & w_{3,4}^1 \\ w_{4,0}^1 & w_{4,1}^1 & w_{4,2}^1 & w_{4,3}^1 & w_{4,4}^1 \end{pmatrix}$

$W^{(2)} = \begin{pmatrix} w_{0,0}^2 & w_{0,1}^2 & w_{0,2}^2 & w_{0,3}^2 & w_{0,4}^2 \\ w_{1,0}^2 & w_{1,1}^2 & w_{1,2}^2 & w_{1,3}^2 & w_{1,4}^2 \\ w_{2,0}^2 & w_{2,1}^2 & w_{2,2}^2 & w_{2,3}^2 & w_{2,4}^2 \\ w_{3,0}^2 & w_{3,1}^2 & w_{3,2}^2 & w_{3,3}^2 & w_{3,4}^2 \\ w_{4,0}^2 & w_{4,1}^2 & w_{4,2}^2 & w_{4,3}^2 & w_{4,4}^2 \end{pmatrix}$, $W^{(3)} = \begin{pmatrix} w_{0,0}^3 & w_{0,1}^3 & w_{0,2}^3 & w_{0,3}^3 & w_{0,4}^3 \\ w_{1,0}^3 & w_{1,1}^3 & w_{1,2}^3 & w_{1,3}^3 & w_{1,4}^3 \\ w_{2,0}^3 & w_{2,1}^3 & w_{2,2}^3 & w_{2,3}^3 & w_{2,4}^3 \\ w_{3,0}^3 & w_{3,1}^3 & w_{3,2}^3 & w_{3,3}^3 & w_{3,4}^3 \\ w_{4,0}^3 & w_{4,1}^3 & w_{4,2}^3 & w_{4,3}^3 & w_{4,4}^3 \end{pmatrix}$,

$\overrightarrow{b}^{(1)} = \begin{pmatrix} b_0^1 \\ b_1^1 \\ b_2^1 \\ b_3^1 \\ b_4^1 \\ b_5^1 \end{pmatrix}$, $\overrightarrow{b}^{(2)} = \begin{pmatrix} b_0^2 \\ b_1^2 \\ b_2^2 \\ b_3^2 \\ b_4^2 \\ b_5^2 \end{pmatrix}$, $\overrightarrow{b}^{(3)} = \begin{pmatrix} b_0^3 \\ b_1^3 \\ b_2^3 \\ b_3^3 \\ b_4^3 \\ b_5^3 \end{pmatrix}$

The equivalent representation of a Neural network with 2 hidden layers n inputs (5 in this case) and n outputs (5 in this case) as a single equivalent representation as a single layer neural network requires representation of the layers as the function of input layers with respect to the output

Backtracking from the last layer representation we have the following:

From equation (iii)

$$\overrightarrow{a}^3 = W^3\overrightarrow{a}^2 + \overrightarrow{b}^3$$

where

$$\overrightarrow{a}^2 = W^2\overrightarrow{a}^1 + \overrightarrow{b}^2$$

Thus

$$\overrightarrow{a}^3 = W^3(W^2\overrightarrow{a}^1 + \overrightarrow{b}^2) + \overrightarrow{b}^3$$
$$\overrightarrow{a}^3 = W^3 \,.\, W^2\overrightarrow{a}^2 + (W^3\overrightarrow{b}^2 + \overrightarrow{b}^3)$$

where

$$\overrightarrow{a}^1 = W^1\overrightarrow{a}^0 + \overrightarrow{b}^1$$

Substituting the same we obtain:

$$\overrightarrow{a}^3 = W^3 \,.\, W^2(W^1\overrightarrow{a}^0 + \overrightarrow{b}^1) + (W^3\overrightarrow{b}^2 + \overrightarrow{b}^3)$$

$$\overrightarrow{a}^3 = W^3 \,.\, W^2 \,.\, W^1\overrightarrow{a}^0 + (W^3 \,.\, W^2 \,.\, \overrightarrow{b}^1 + W^3\overrightarrow{b}^2 + \overrightarrow{b}^3) \;\ldots \text{(iv)}$$

Where . represents the matrix dot product.

Based on the dimensionality representation (current scenario)

1. $W^3 \,.\, W^2 \,.\, W^1$ is a 5 x 5 matrix
2. $W^2 \,.\, W^2$ is a 5 x 5 matrix
3. $W^3, W^2, W^1$ is a 5x 5 matrix
4. $\overrightarrow{b}^1, \overrightarrow{b}^2, \overrightarrow{b}^3$ is a 5 x 1 matrix

Thus

$W^3 \,.\, W^2 \,.\, \overrightarrow{b}^1, W^3\overrightarrow{b}^2, \overrightarrow{b}^3$ represents a 5x1 matrix

The linear combination of the above matrices also results in a 5x1 matrix

Thus

$W^3 \,.\, W^2 \,.\, \overrightarrow{b}^1 + W^3\overrightarrow{b}^2 + \overrightarrow{b}^3$ is a 5 x1 matrix and can be represented as a substituted matrix

$$b^{\sim} = W^3 \,.\, W^2 \,.\, \overrightarrow{b}^1 + W^3\overrightarrow{b}^2 + \overrightarrow{b}^3$$

$$W^{\sim} = W^3 \,.\, W^2 \,.\, W^1$$

The equation iv can thus be reduced as

$$\overrightarrow{a}^3 = W^{\sim}\overrightarrow{a}^0 + b^{\sim} \ldots \text{(vi)}$$

Given the MLP network with no hidden layers the network representation can be given as

$\overrightarrow{a}^1 = W\overrightarrow{a} + \overrightarrow{b}$ … (vii)

The equation (vi) and (vii) are equivalent if the vector size and values of $\overrightarrow{a}^1, \overrightarrow{a}^0$ of equation (vii) is equal to vector size $\overrightarrow{a}^3, \overrightarrow{a}^0$ of equation vi

The above representation canoe generalised for n layered network with m inputs an k outputs with following representation as

$\overrightarrow{a}^{out} = W_{single}\overrightarrow{a}^{in} + \overrightarrow{b}^{bias}$

where

$W_{single} = \prod_{i=1}^{n} W^i$ ($W^i$ represents the weight at each layer of the multi layer perceptron)

$\overrightarrow{b}^{bias} = \overrightarrow{b}^n + W^n . b^{n-1} + W^n . W^{n-1} . b^{n-2} + .... + (\prod_{i=2}^{n} W^i)\overrightarrow{b}^1$