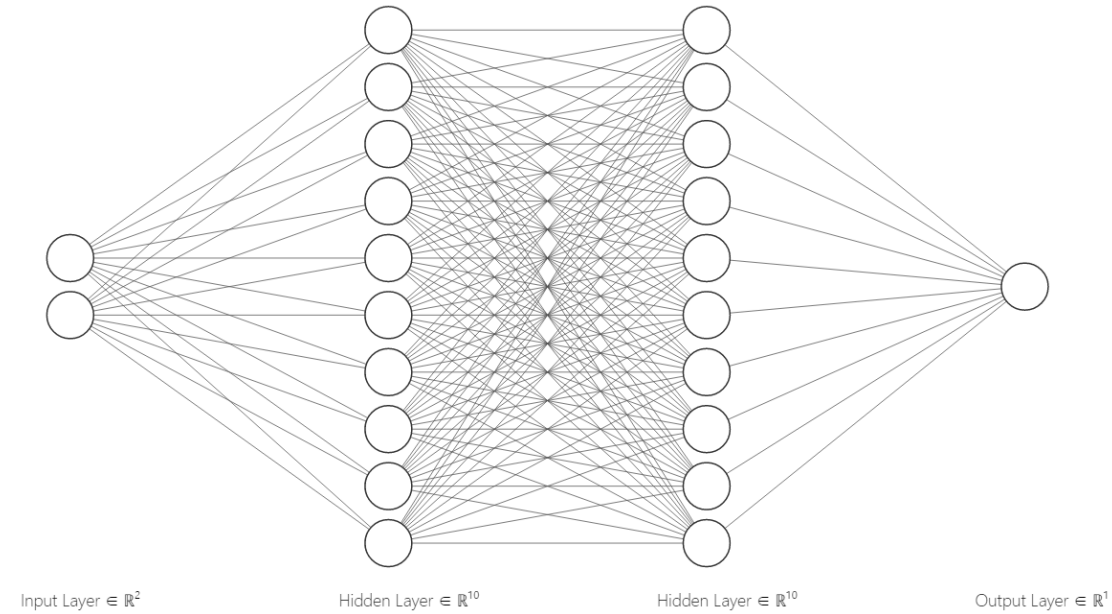


The structured fully connected neural network shows below, input layer contains 2 nodes, the first hidden layer contains 10 nodes, the second hidden layer contains 10 nodes, and the last output layer contains one node.



In this case the shape of b_1 , w_1 between input and first hidden layer are $(1,10)$, $(10,2)$, and shape of b_2 , w_2 between first and second hidden layer are $(1,10)$, $(10,10)$, respectively. And shape of b_3 , w_3 between second and output hidden layer are $(1,1)$ and $(1,10)$, respectively.

Forward propagation:

$$z_1 = xw_1^T + b_1$$

$$y_1 = \text{sigmoid}(z_1)$$

$$z_2 = y_1w_2^T + b_2$$

$$y_2 = \text{sigmoid}(z_2)$$

$$z_3 = y_2w_3^T + b_3$$

$$y_3 = z_3$$

Loss Function:

$$\text{loss} = (y_3 - y)^2$$

Backward Propagation:

$$\frac{\partial \text{loss}}{\partial w_3} = \frac{\partial \text{loss}}{\partial y_3} \frac{\partial y_3}{\partial w_3} = 2(y_3 - y) \cdot y_2$$

$$\frac{\partial \text{loss}}{\partial b_3} = \frac{\partial \text{loss}}{\partial y_3} \frac{\partial y_3}{\partial b_3} = 2(y_3 - y)$$

$$\frac{\partial \text{loss}}{\partial w_2} = \frac{\partial \text{loss}}{\partial y_3} \frac{\partial y_3}{\partial y_2} \frac{\partial y_2}{\partial z_2} \frac{\partial z_2}{\partial w_2} = 2(y_3 - y) \cdot w_3 \cdot \text{sigmoid}(z_2)(1 - \text{sigmoid}(z_2)) \cdot y_1$$

$$\frac{\partial \text{loss}}{\partial b_2} = \frac{\partial \text{loss}}{\partial y_3} \frac{\partial y_3}{\partial y_2} \frac{\partial y_2}{\partial z_2} \frac{\partial z_2}{\partial b_2} = 2(y_3 - y) \cdot w_3 \cdot \text{sigmoid}(z_2)(1 - \text{sigmoid}(z_2))$$

$$\begin{aligned} \frac{\partial \text{loss}}{\partial w_1} &= \frac{\partial \text{loss}}{\partial y_3} \frac{\partial y_3}{\partial y_2} \frac{\partial y_2}{\partial z_2} \frac{\partial z_2}{\partial y_1} \frac{\partial y_1}{\partial z_1} \frac{\partial z_1}{\partial w_1} \\ &= 2(y_3 - y) \cdot w_3 \cdot \text{sigmoid}(z_2)(1 - \text{sigmoid}(z_2)) \cdot w_2 \cdot \text{sigmoid}(z_1)(1 - \text{sigmoid}(z_1)) \cdot x \end{aligned}$$

$$\begin{aligned}\frac{\partial loss}{\partial b_1} &= \frac{\partial loss}{\partial y_3} \frac{\partial y_3}{\partial y_2} \frac{\partial y_2}{\partial z_2} \frac{\partial z_2}{\partial y_1} \frac{\partial y_1}{\partial z_1} \frac{\partial z_1}{\partial b_1} = \\ &= 2(y_3 - y) \cdot w_3 \cdot \text{sigmoid}(z_2)(1 - \text{sigmoid}(z_2)) \cdot w_2 \cdot \text{sigmoid}(z_1)(1 \\ &\quad - \text{sigmoid}(z_1))\end{aligned}$$

After this backward propagation, the file can be generated by `torch.autograd()` and by my self-backward-propagation code. Please see the code in "my_autograd.dat" and "torch_autograd.dat" files.