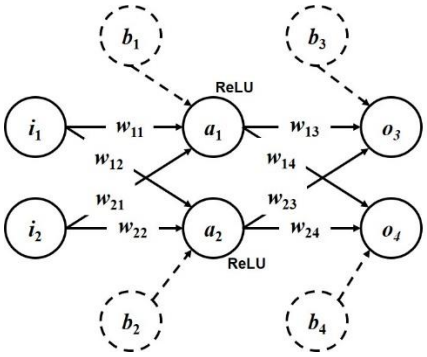
Question & Answer

You are given the toy neural network on the right with an input layer, a single hidden layer, and an output layer. The input layer has two units (i1, i2), which feed the fully connected layer's two hidden units (a1, a2) that utilize ReLU activation functions. The output of the activated units is then fed to the final layer with two output units (o3, o4) that don't have any activation functions associated with. The pertinent weights and bias terms as well as the values of variables are presented in the figure below.



Variable	i_1	i_2	w_{11}	w_{12}	w_{21}	w_{22}	w_{13}	w_{14}	w_{23}	w_{24}	b_1	b_2	b_3	b_4	t_3	t_4
Value	2.0	-1.0	1.0	-0.5	0.5	-1.0	0.5	-1.0	-0.5	1.0	0.5	-0.5	-1.0	0.5	1.0	0.5

(i) (10 pts.) Compute the output (o3, o4) with the input (i1, i2), and network parameters as specified above. Write down all calculations, including the hidden layer results.

h1 = i1 × w11 + i2 × w21 + b1 = 2.0 × 1.0 − 1.0 × 0.5 + 0.5 = 2.0

h2 = i1 × w12 + i2 × w22 + b2 = 2.0 × −0.5 + −1.0 × −1.0 − 0.5 = −0.5

a1 = max(0, h1) = 2

a2 = max(0, h2) = 0

o3 = a1 × w13 + a2 × w23 + b3 = 2 × 0.5 + 0 × −0.5 − 1.0 = 0

o4 = a1 × w14 + a2 × w24 + b4 = 2 × −1.0 + 0 × 1.0 + 0.5 = −1.5

(ii) (5 pts.) Compute the mean squared error (MSE) of the output (o3, o4) calculated above and the target (t3, t4).

MSE = $\frac{1}{2}$ x $(t3 - o3)^2 + \frac{1}{2}$ x $(t4 - o4)^2$ = 0.5 × 1.0 + 0.5 × 4.0 = 2.5

(iii) (10 pts.) Update the weight w21 using gradient descent with learning rate 0.1 as well as the loss computed previously. (Please write down all your computations).

$$\frac{\partial MSE}{\partial w_{21}} = \frac{\partial \frac{1}{2}(t_3 - o_3)^2}{\partial o_3} * \frac{\partial o_3}{\partial a_1} * \frac{\partial a_1}{\partial w_{21}} + \frac{\partial \frac{1}{2}(t_4 - o_4)^2}{\partial o_4} * \frac{\partial o_4}{\partial a_1} * \frac{\partial a_1}{\partial w_{21}}$$

= (o3 – t3) × w13 x i2 + (o4 – t4) × w14 x i2

= (0 – 1.0) x 0.5 x -1.0 + (-1.5 – 0.5) x -1.0 x -1.0

= 0.5 + -2.0 = -1.5

Update using gradient descent:

$$w_{21}^{+} = w_{21} - lr * \frac{\partial MSE}{\partial w_{21}}$$ = 0.5 – 0.1 x -1.5 = 0.65