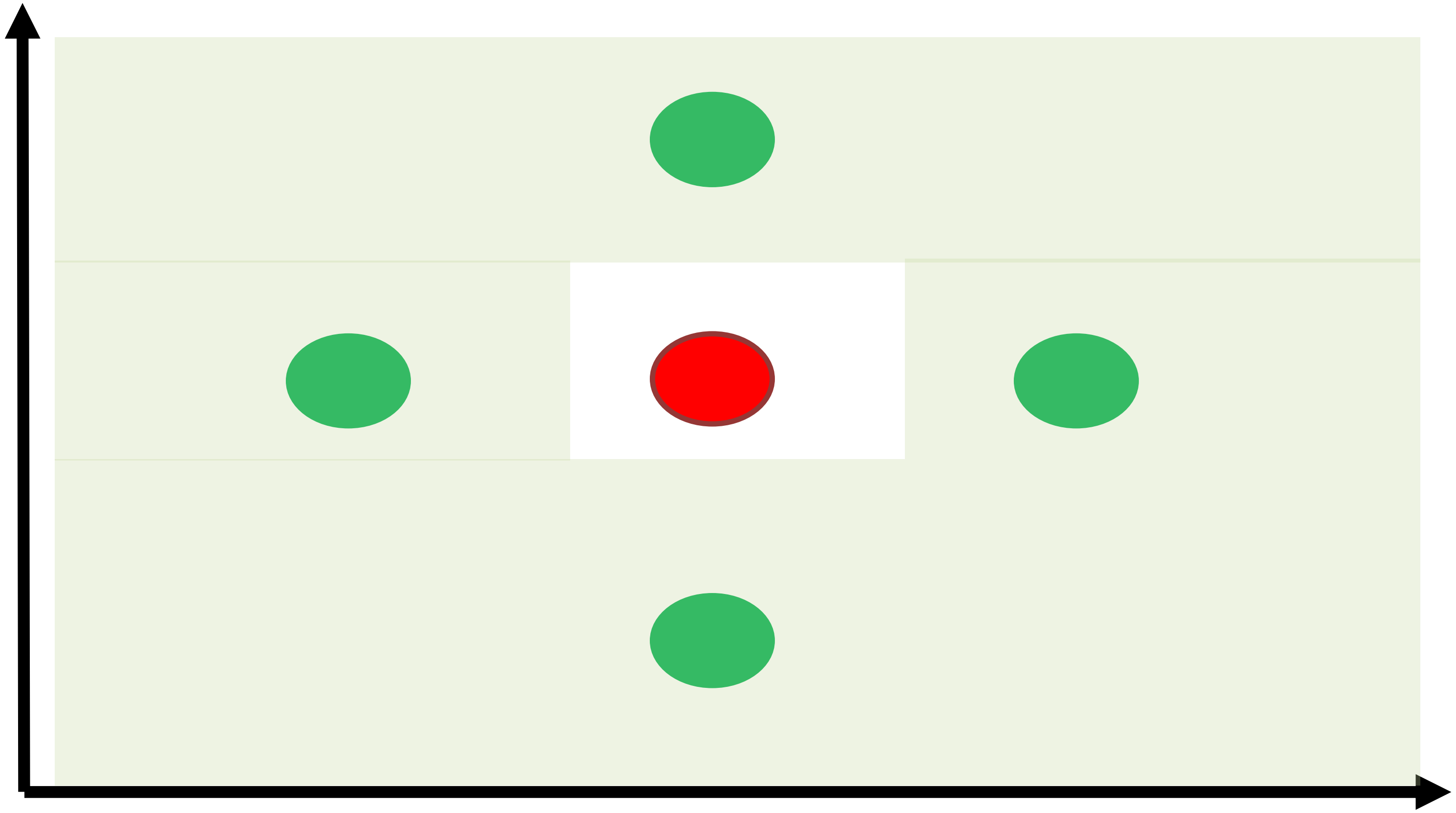
Deep Learning: Question

Given labelled data:

$$D = \{(1,3) \rightarrow 1\,, (3,1) \rightarrow 1, (3,5) \rightarrow 1, (5,3) \rightarrow 1, (3,3) \rightarrow 0\}$$

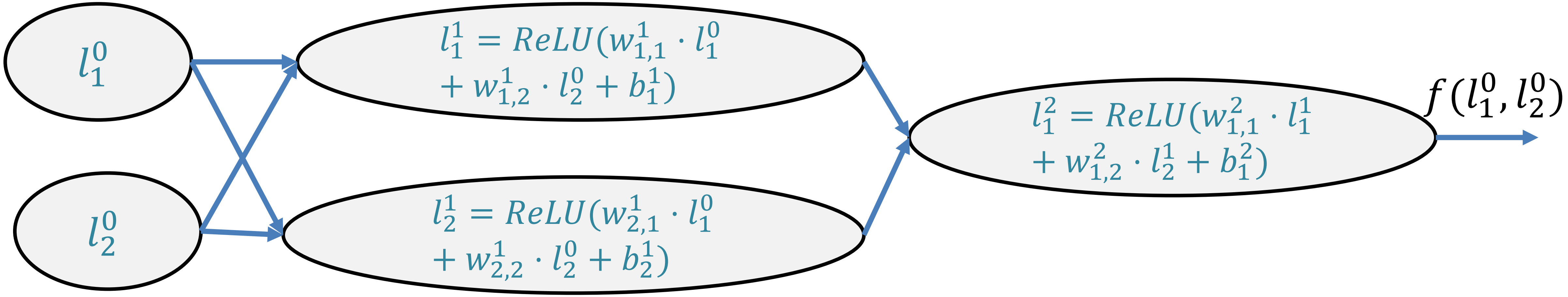


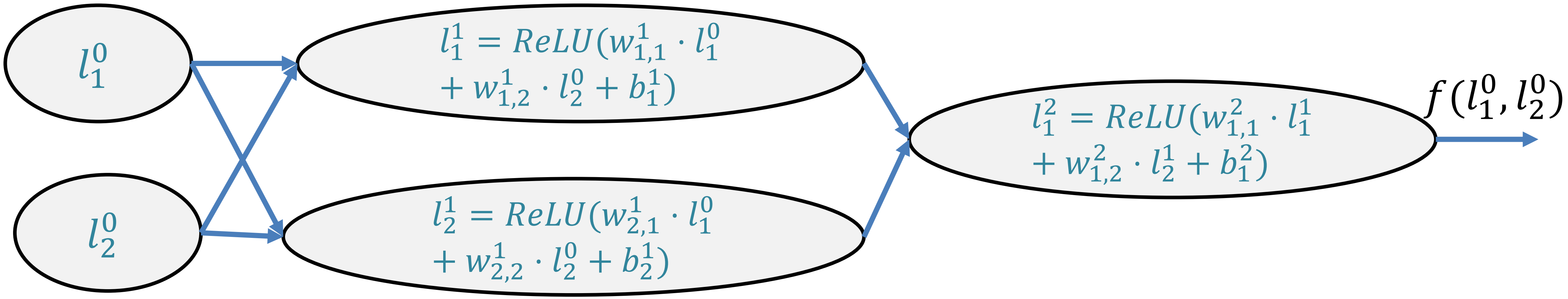
Deep Learning: Question

Given labelled data:

$$D = \{(1,3) \to 1, (3,1) \to 1, (3,5) \to 1, (5,3) \to 1, (3,3) \to 0\}$$

Consider a deep learning model with one hidden layer with two neurons.





Define:

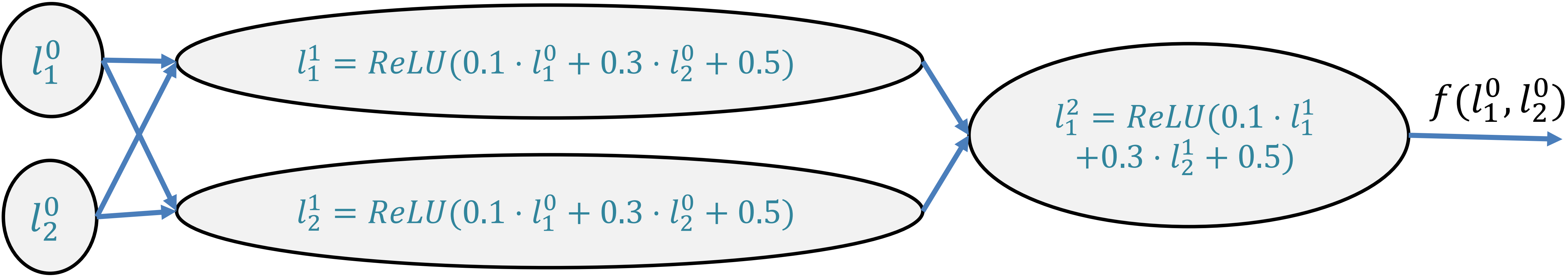
$$w_{1,1}^1 = w_{2,1}^1 = w_{1,1}^2 = 0.1$$

$$w_{1,2}^1 = w_{2,2}^1 = w_{1,2}^2 = 0.3$$

$$b_1^1 = b_2^1 = b_1^2 = 0.5$$

Compute a single iteration of backpropagation.

Initial Model



l_1^0	l_2^0	f^*	l_1^1	l_2^1	$l_1^2 = f(l_1^0, l_2^0)$
1	3	1	$ReLU(0.1 \cdot 1$	$ReLU(0.1 \cdot 1$	$ReLU(0.1 \cdot 1.5$ $+ 0.3 \cdot 1.5$
3	1	1	1.1	1.1	0.94
3	5	1	2.3	2.3	1.42
5	3	1	1.9	1.9	1.26
3	3	0	1.7	1.7	1.18

Initial MSE

$$\text{MSE} = \Sigma_{(l_1^0, l_2^0) \in D} \left(f^*(l_1^0, l_2^0) - f(l_1^0, l_2^0) \right)^2$$

$$(1 - 1.1)^2 + (1 - 0.94)^2 + (1 - 1.42)^2 + (1 - 1.26)^2 + (0 - 1.18)^2 = 1.65$$

Finding Minima via Gradient Descent

1. **Initialize** randomly with certain values for weights/bias a_0 , set $\mathrm{n} = 0$

2. Compute the **gradient** of the MSE at a_n

3. The **next point** is the one maximizing the decrease in MSE

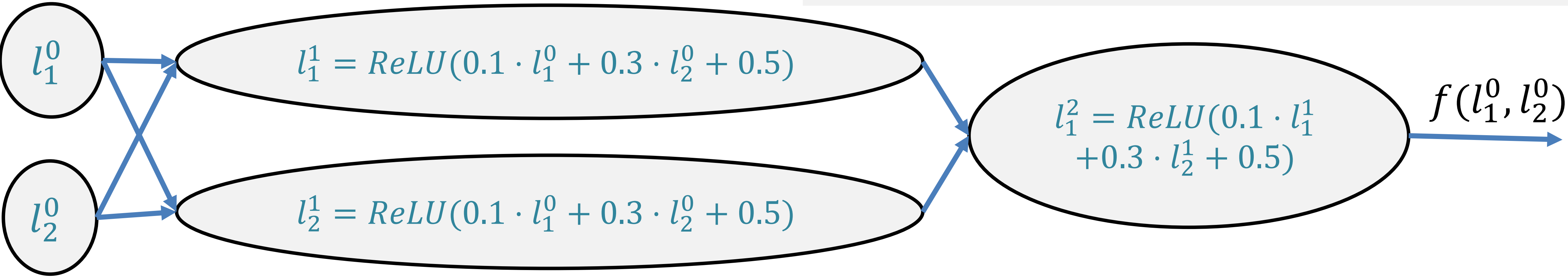
   $a_{n+1} = a_n - \gamma \nabla MSE(a_n)$ (γ is called the learning rate)

   n = n + 1

4. If **loss is small enough**, complete; otherwise, repeat 2

Gradients

$$\text{MSE} = \Sigma_{(l_1^0,l_2^0)\in D}\left(f^*(l_1^0,l_2^0) - f(l_1^0,l_2^0)\right)^2$$

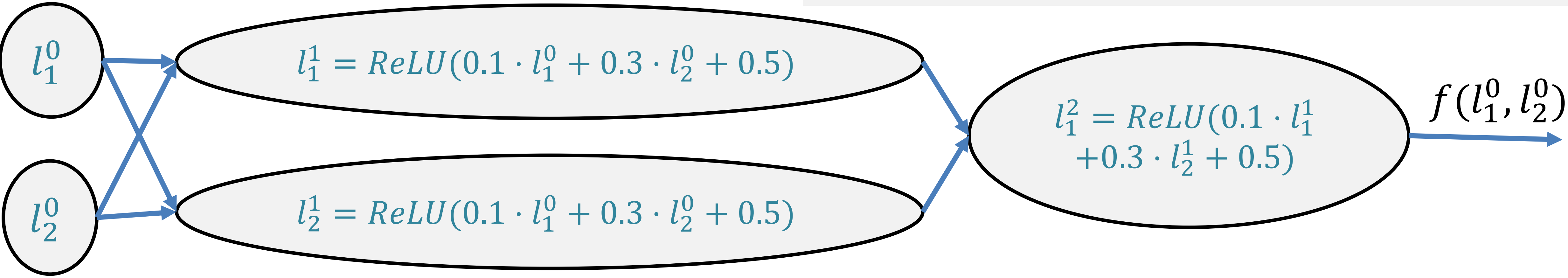


$$\frac{\partial MSE}{\partial w_{1,1}^2} = \Sigma_{(l_1^0,l_2^0)\in D_{Tr}} \frac{\partial MSE}{\partial l_1^2}\cdot\frac{\partial l_1^2}{\partial p_1^2}\cdot\frac{\partial p_1^2}{\partial w_{1,1}^2} = \Sigma_{(l_1^0,l_2^0)\in D_{Tr}} 2(f^*(l_1^0,l_2^0) - l_1^2)\cdot(-1)\cdot l_1^1$$

$$= -2[(1-1.1)\cdot 1.5 + (1-0.94)\cdot 1.1 + (1-1.42)\cdot 2.3 + (1-1.26)\cdot 1.9$$

Gradients

$$\text{MSE} = \Sigma_{(l_1^0, l_2^0) \in D} \left(f^*(l_1^0, l_2^0) - f(l_1^0, l_2^0) \right)^2$$

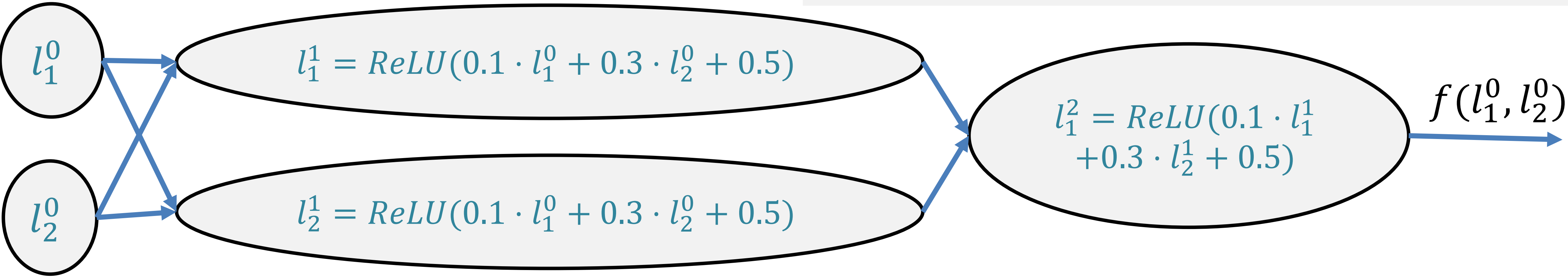


$$\frac{\partial MSE}{\partial w_{1,2}^2} = \Sigma_{(l_1^0, l_2^0) \in D_{Tr}} \frac{\partial MSE}{\partial l_1^2} \cdot \frac{\partial l_1^2}{\partial p_1^2} \cdot \frac{\partial p_1^2}{\partial w_{1,2}^2} = \Sigma_{(l_1^0, l_2^0) \in D_{Tr}} 2(f^*(l_1^0, l_2^0) - l_1^2) \cdot (-1) \cdot l_2^1$$

$$= -2[(1 - 1.1) \cdot 1.5 + (1 - 0.94) \cdot 1.1 + (1 - 1.42) \cdot 2.3 + (1 - 1.26) \cdot 1.9$$

Gradients

$$\text{MSE} = \Sigma_{(l_1^0, l_2^0) \in D} \left(f^*(l_1^0, l_2^0) - f(l_1^0, l_2^0) \right)^2$$



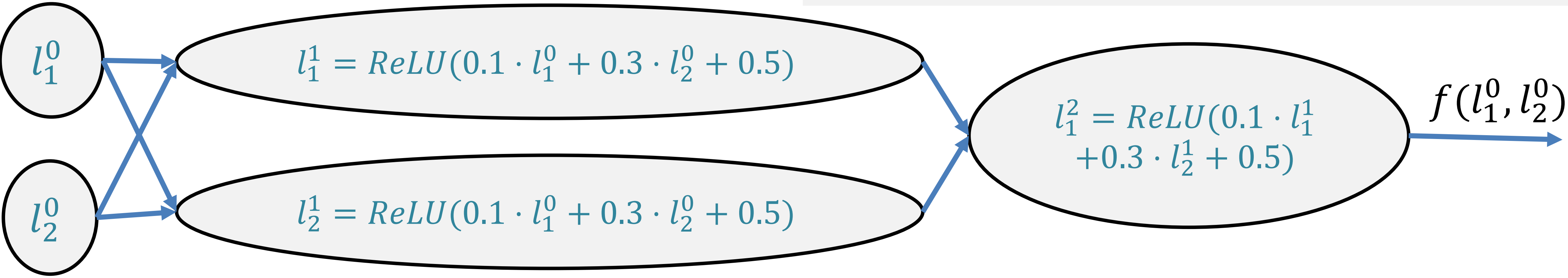
$$\frac{\partial MSE}{\partial b_1^2} = \Sigma_{(l_1^0, l_2^0) \in D_{Tr}} \frac{\partial MSE}{\partial l_1^2} \cdot \frac{\partial l_1^2}{\partial p_1^2} \cdot \frac{\partial p_1^2}{\partial b_1^2}$$

$$= \Sigma_{(l_1^0, l_2^0) \in D_{Tr}} 2(f^*(l_1^0, l_2^0) - l_1^2) \cdot (-1) \cdot 1$$

$$= -2[(1 - 1.1) + (1 - 0.94) + (1 - 1.42) + (1 - 1.26)$$

Gradients

$$\mathrm{MSE} = \Sigma_{(l_1^0, l_2^0) \in D} \left(f^*(l_1^0, l_2^0) - f(l_1^0, l_2^0) \right)^2$$



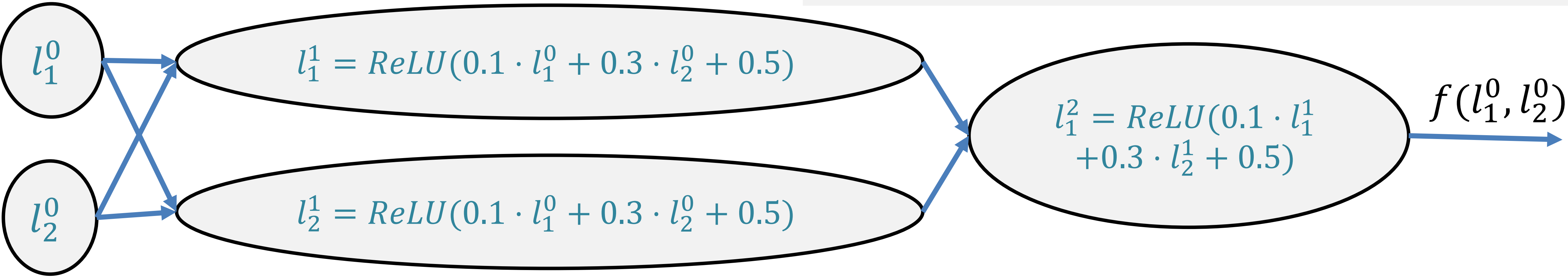
$$\frac{\partial MSE}{\partial w_{1,1}^1} = \Sigma_{(l_1^0, l_2^0) \in D_{Tr}} \frac{\partial MSE}{\partial l_1^2} \cdot \frac{\partial l_1^2}{\partial p_1^2} \cdot \frac{\partial p_1^2}{\partial l_1^1} \cdot \frac{\partial l_1^1}{\partial p_1^1} \cdot \frac{\partial p_1^1}{\partial w_{1,1}^1}$$

$$= \Sigma_{(l_1^0, l_2^0) \in D_{Tr}} \frac{\partial MSE}{\partial l_1^2} \cdot \frac{\partial l_1^2}{\partial p_1^2} \cdot w_{1,1}^2 \cdot 1 \cdot l_1^0$$

$$= 0.1 \cdot [0.2 \cdot 1 - 0.12 \cdot 3 + 0.84 \cdot 3 + 0.52 \cdot 5 + 2.36 \cdot 3] = 1.204$$

Gradients

$$\text{MSE} = \Sigma_{(l_1^0, l_2^0) \in D} \left(f^*(l_1^0, l_2^0) - f(l_1^0, l_2^0) \right)^2$$



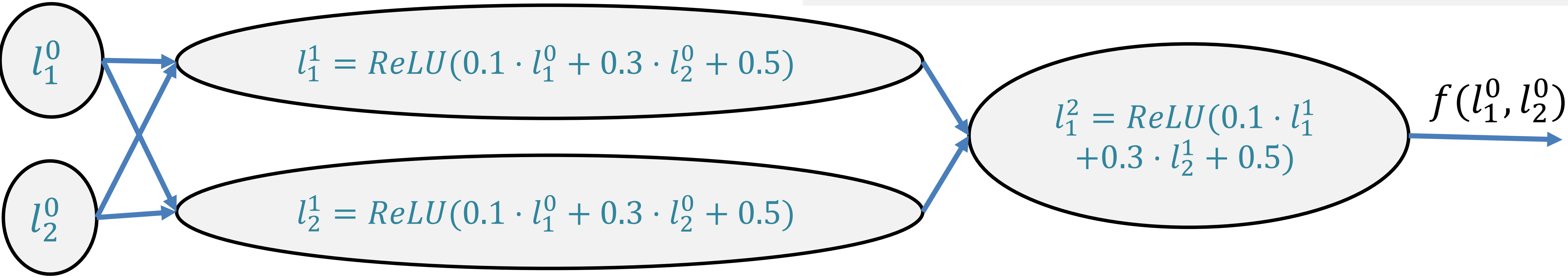
$$\frac{\partial MSE}{\partial w_{1,2}^1} = \Sigma_{(l_1^0, l_2^0) \in D_{Tr}} \frac{\partial MSE}{\partial l_1^2} \cdot \frac{\partial l_1^2}{\partial p_1^2} \cdot \frac{\partial p_1^2}{\partial l_1^1} \cdot \frac{\partial l_1^1}{\partial p_1^1} \cdot \frac{\partial p_1^1}{\partial w_{1,1}^1} = \Sigma_{(l_1^0, l_2^0) \in D_{Tr}} \frac{\partial MSE}{\partial l_1^2} \cdot \frac{\partial l_1^2}{\partial p_1^2} \cdot w_{1,1}^2 \cdot 1 \cdot l_2^0$$

$$= 0.1 \cdot [0.2 \cdot 1 - 0.12 \cdot 3 + 0.84 \cdot 3 + 0.52 \cdot 5 + 2.36 \cdot 3] = 1.204$$

$$\frac{\partial MSE}{\partial b_1^1} = \frac{\partial MSE}{\partial l_1^2} \cdot \frac{\partial l_1^2}{\partial p_1^2} \cdot \frac{\partial p_1^2}{\partial l_1^1} \cdot \frac{\partial l_1^1}{\partial p_1^1} \cdot \frac{\partial p_1^1}{\partial w_{1,1}^1} = \Sigma_{(l_1^0, l_2^0) \in D_{Tr}} \frac{\partial MSE}{\partial l_1^2} \cdot \frac{\partial l_1^2}{\partial p_1^2} \cdot w_{1,1}^2 \cdot 1 = 3.8 \cdot 0.1 = 0.38$$

Gradients

$$\text{MSE} = \Sigma_{(l_1^0, l_2^0)\in D}\left(f^*(l_1^0, l_2^0) - f(l_1^0, l_2^0)\right)^2$$



$$\frac{\partial MSE}{\partial w_{2,1}^1} = \Sigma_{(l_1^0, l_2^0)\in D_{Tr}} \frac{\partial MSE}{\partial l_1^2} \cdot \frac{\partial l_1^2}{\partial p_1^2} \cdot \frac{\partial p_1^2}{\partial l_1^1} \cdot \frac{\partial l_1^1}{\partial p_1^1} \cdot \frac{\partial p_1^1}{\partial w_{1,1}^1} = \Sigma_{(l_1^0, l_2^0)\in D_{Tr}} \frac{\partial MSE}{\partial l_1^2} \cdot \frac{\partial l_1^2}{\partial p_1^2} \cdot w_{2,1}^2 \cdot 1 \cdot l_1^0$$

$$= 0.3 \cdot [0.2 \cdot 1 - 0.12 \cdot 3 + 0.84 \cdot 3 + 0.52 \cdot 5 + 2.36 \cdot 3] = 3.612$$

$$\frac{\partial MSE}{\partial b_2^1} = \Sigma_{(l_1^0, l_2^0)\in D_{Tr}} \frac{\partial MSE}{\partial l_1^2} \cdot \frac{\partial l_1^2}{\partial p_1^2} \cdot \frac{\partial p_1^2}{\partial l_1^1} \cdot \frac{\partial l_1^1}{\partial p_1^1} \cdot \frac{\partial p_1^1}{\partial w_{1,1}^1} = \Sigma_{(l_1^0, l_2^0)\in D_{Tr}} \frac{\partial MSE}{\partial l_1^2} \cdot \frac{\partial l_1^2}{\partial p_1^2} \cdot w_{2,1}^2 \cdot 1 = 3.8 \cdot 0.3 = 1.14$$

Finding Minima via Gradient Descent

The **next point** is the one maximizing the decrease in MSE

$$a_{n+1} = a_n - \gamma \nabla MSE(a_n)$$ (γ is called the learning rate)

Pick $\gamma = 0.01$

$w_{1,1}^2 = 0.1 - 0.01 \cdot 7.1 = 0.029$ $b_1^2 = 0.5 - 0.01 \cdot 3.8 = 0.462$

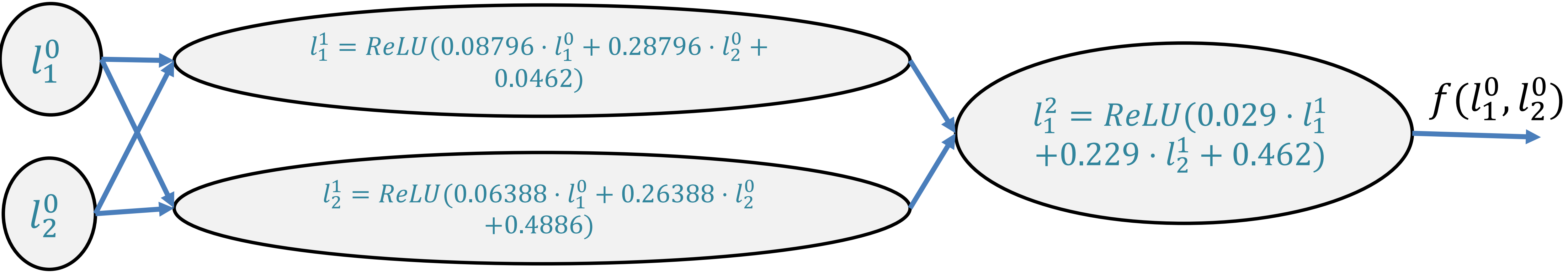
$w_{1,2}^2 = 0.3 - 0.01 \cdot 7.1 = 0.229$ $w_{1,1}^1 = 0.1 - 0.01 \cdot 1.204 = 0.08796$

$w_{1,2}^1 = 0.3 - 0.01 \cdot 1.204 = 0.28796$ $b_1^1 = 0.5 - 0.01 \cdot 0.38 = 0.0462$

$w_{2,1}^1 = 0.1 - 0.01 \cdot 3.612 = 0.06388$ $w_{2,2}^1 = 0.3 - 0.01 \cdot 3.612 = 0.26388$

$b_2^1 = 0.5 - 0.01 \cdot 1.14 = 0.4886$

Updated Model



l_1^0	l_2^0	f^*	l_1^1	l_2^1	$l_1^2 = f(l_1^0, l_2^0)$
1	3	1	0.99804	1.34412	0.79875
3	1	1	0.59804	0.94412	0.69555
3	5	1	1.74988	1.99964	0.97066
5	3	1	1.34988	1.59964	0.86746
3	3	0	1.17396	1.47188	0.83311

MSE

$$\text{MSE} = \Sigma_{(l_1^0, l_2^0) \in D} \left(f^*(l_1^0, l_2^0) - f(l_1^0, l_2^0) \right)^2$$

$$(1 - 0.79875)^2 + (1 - 0.69555)^2 + (1 - 0.97066)^2 + (1 - 0.86746)^2 + (0 - 0.83311)^2 = 0.84569$$