Multilayer Networks

INST 808: Jordan Boyd-Graber

University of Maryland

Hands-On Demo

Data and Model

Data

x_1	x_2	y
1.00	1.00	0.00
1.00	0.00	1.00
0.00	0.00	0.00
0.00	1.00	1.00

Using ReLU as non-linearity

First Layer

$$w^{(1)} = \begin{bmatrix} 1.00 & 1.00 \\ 1.00 & 1.00 \end{bmatrix} \quad (1)$$

$$b^{(1)} = \begin{bmatrix} -1.00 & 0.00 \end{bmatrix} \quad (2)$$

Second Layer

$$w^{(2)} = \begin{bmatrix} -2.00 & 1.00 \end{bmatrix} \quad (3)$$

$$b^{(2)} = 0.00 \quad (4)$$

Prediction for $x_0 = (1.00, 1.00)$

Prediction for $x_0 = (1.00, 1.00)$

- Hidden Computation

$$a_{0,0}^{(1)} = f(w_{0,0}^{(1)} \cdot 1.00 + w_{0,1}^{(1)} \cdot 1.00 + b_0) \tag{5}$$

$$= f(1.00 \cdot 1.00 + 1.00 \cdot 1.00 + -1.00) \tag{6}$$

Prediction for $x_0 = (1.00, 1.00)$

- Hidden Computation

$$a_{0,0}^{(1)} = f(w_{0,0}^{(1)} \cdot 1.00 + w_{0,1}^{(1)} \cdot 1.00 + b_0) \tag{5}$$

$$= f(1.00 \cdot 1.00 + 1.00 \cdot 1.00 + -1.00) \tag{6}$$

$$a_{0,1}^{(1)} = f(w_{1,0}^{(1)} \cdot 1.00 + w_{1,1}^{(1)} \cdot 1.00 + b_1) \tag{7}$$

$$= f(1.00 \cdot 1.00 + 1.00 \cdot 1.00 + 0.00) \tag{8}$$

Prediction for $x_0 = (1.00, 1.00)$

- Hidden Computation

$$a_{0,0}^{(1)} = f(w_{0,0}^{(1)} \cdot 1.00 + w_{0,1}^{(1)} \cdot 1.00 + b_0) \tag{5}$$
$$= f(1.00 \cdot 1.00 + 1.00 \cdot 1.00 + -1.00) \tag{6}$$

$$a_{0,1}^{(1)} = f(w_{1,0}^{(1)} \cdot 1.00 + w_{1,1}^{(1)} \cdot 1.00 + b_1) \tag{7}$$
$$= f(1.00 \cdot 1.00 + 1.00 \cdot 1.00 + 0.00) \tag{8}$$

- Hidden Layer: [1 2]
- Output Answer

$$a_{0,0}^{(3)} = f(w_{0,0}^{(2)} \cdot 1.00 + w_{0,1}^{(2)} \cdot 2.00 + b_0) \tag{9}$$
$$= f(-2.00 \cdot 1.00 + 1.00 \cdot 2.00 + 0.00) \tag{10}$$

- Prediction: 0.00, Error: 0.00

Prediction for $x_0 = (1.00, 1.00)$

- Hidden Computation

$$a_{0,0}^{(1)} = f(w_{0,0}^{(1)} \cdot 1.00 + w_{0,1}^{(1)} \cdot 1.00 + b_0) \tag{5}$$
$$= f(1.00 \cdot 1.00 + 1.00 \cdot 1.00 + -1.00) \tag{6}$$

$$a_{0,1}^{(1)} = f(w_{1,0}^{(1)} \cdot 1.00 + w_{1,1}^{(1)} \cdot 1.00 + b_1) \tag{7}$$
$$= f(1.00 \cdot 1.00 + 1.00 \cdot 1.00 + 0.00) \tag{8}$$

- Hidden Layer: [1 2]
- Output Answer

$$a_{0,0}^{(3)} = f(w_{0,0}^{(2)} \cdot 1.00 + w_{0,1}^{(2)} \cdot 2.00 + b_0) \tag{9}$$
$$= f(-2.00 \cdot 1.00 + 1.00 \cdot 2.00 + 0.00) \tag{10}$$

Prediction for $x_0 = (1.00, 1.00)$

- Hidden Computation

$$a_{0,0}^{(1)} = f(w_{0,0}^{(1)} \cdot 1.00 + w_{0,1}^{(1)} \cdot 1.00 + b_0) \tag{5}$$
$$= f(1.00 \cdot 1.00 + 1.00 \cdot 1.00 + -1.00) \tag{6}$$

$$a_{0,1}^{(1)} = f(w_{1,0}^{(1)} \cdot 1.00 + w_{1,1}^{(1)} \cdot 1.00 + b_1) \tag{7}$$
$$= f(1.00 \cdot 1.00 + 1.00 \cdot 1.00 + 0.00) \tag{8}$$

- Hidden Layer: [1 2]
- Output Answer

$$a_{0,0}^{(3)} = f(w_{0,0}^{(2)} \cdot 1.00 + w_{0,1}^{(2)} \cdot 2.00 + b_0) \tag{9}$$
$$= f(-2.00 \cdot 1.00 + 1.00 \cdot 2.00 + 0.00) \tag{10}$$

- Prediction: 0.00, Error: 0.00

Prediction for $x_1 = (1.00, 0.00)$

Prediction for $x_1 = (1.00, 0.00)$

- Hidden Computation

$$a_{1,0}^{(1)} = f(w_{0,0}^{(1)} \cdot 1.00 + w_{0,1}^{(1)} \cdot 0.00 + b_0) \tag{11}$$

$$= f(1.00 \cdot 1.00 + 1.00 \cdot 0.00 + -1.00) \tag{12}$$

Prediction for $x_1 = (1.00, 0.00)$

- Hidden Computation

$$a_{1,0}^{(1)} = f\left(w_{0,0}^{(1)} \cdot 1.00 + w_{0,1}^{(1)} \cdot 0.00 + b_0\right) \tag{11}$$

$$= f\left(1.00 \cdot 1.00 + 1.00 \cdot 0.00 + -1.00\right) \tag{12}$$

$$a_{1,1}^{(1)} = f\left(w_{1,0}^{(1)} \cdot 1.00 + w_{1,1}^{(1)} \cdot 0.00 + b_1\right) \tag{13}$$

$$= f\left(1.00 \cdot 1.00 + 1.00 \cdot 0.00 + 0.00\right) \tag{14}$$

Prediction for $x_1 = (1.00, 0.00)$

- Hidden Computation

$$a_{1,0}^{(1)} = f(w_{0,0}^{(1)} \cdot 1.00 + w_{0,1}^{(1)} \cdot 0.00 + b_0) \tag{11}$$
$$= f(1.00 \cdot 1.00 + 1.00 \cdot 0.00 + -1.00) \tag{12}$$

$$a_{1,1}^{(1)} = f(w_{1,0}^{(1)} \cdot 1.00 + w_{1,1}^{(1)} \cdot 0.00 + b_1) \tag{13}$$
$$= f(1.00 \cdot 1.00 + 1.00 \cdot 0.00 + 0.00) \tag{14}$$

- Hidden Layer: [0 1]
- Output Answer

$$a_{1,0}^{(3)} = f(w_{0,0}^{(2)} \cdot 0.00 + w_{0,1}^{(2)} \cdot 1.00 + b_0) \tag{15}$$
$$= f(-2.00 \cdot 0.00 + 1.00 \cdot 1.00 + 0.00) \tag{16}$$

- Prediction: 1.00, Error: 0.00

Prediction for $x_1 = (1.00, 0.00)$

- Hidden Computation

$$a_{1,0}^{(1)} = f(w_{0,0}^{(1)} \cdot 1.00 + w_{0,1}^{(1)} \cdot 0.00 + b_0) \tag{11}$$
$$= f(1.00 \cdot 1.00 + 1.00 \cdot 0.00 + -1.00) \tag{12}$$

$$a_{1,1}^{(1)} = f(w_{1,0}^{(1)} \cdot 1.00 + w_{1,1}^{(1)} \cdot 0.00 + b_1) \tag{13}$$
$$= f(1.00 \cdot 1.00 + 1.00 \cdot 0.00 + 0.00) \tag{14}$$

- Hidden Layer: [0 1]
- Output Answer

$$a_{1,0}^{(3)} = f(w_{0,0}^{(2)} \cdot 0.00 + w_{0,1}^{(2)} \cdot 1.00 + b_0) \tag{15}$$
$$= f(-2.00 \cdot 0.00 + 1.00 \cdot 1.00 + 0.00) \tag{16}$$

Prediction for $x_1 = (1.00, 0.00)$

- Hidden Computation

$$a_{1,0}^{(1)} = f(w_{0,0}^{(1)} \cdot 1.00 + w_{0,1}^{(1)} \cdot 0.00 + b_0) \tag{11}$$
$$= f(1.00 \cdot 1.00 + 1.00 \cdot 0.00 + -1.00) \tag{12}$$

$$a_{1,1}^{(1)} = f(w_{1,0}^{(1)} \cdot 1.00 + w_{1,1}^{(1)} \cdot 0.00 + b_1) \tag{13}$$
$$= f(1.00 \cdot 1.00 + 1.00 \cdot 0.00 + 0.00) \tag{14}$$

- Hidden Layer: [0 1]
- Output Answer

$$a_{1,0}^{(3)} = f(w_{0,0}^{(2)} \cdot 0.00 + w_{0,1}^{(2)} \cdot 1.00 + b_0) \tag{15}$$
$$= f(-2.00 \cdot 0.00 + 1.00 \cdot 1.00 + 0.00) \tag{16}$$

- Prediction: 1.00, Error: 0.00

Prediction for $x_2 = (0.00, 0.00)$

Prediction for $x_2 = (0.00, 0.00)$

- Hidden Computation

$$a_{2,0}^{(1)} = f\left(w_{0,0}^{(1)} \cdot 0.00 + w_{0,1}^{(1)} \cdot 0.00 + b_0\right) \tag{17}$$

$$= f\left(1.00 \cdot 0.00 + 1.00 \cdot 0.00 + -1.00\right) \tag{18}$$

Prediction for $x_2 = (0.00, 0.00)$

- Hidden Computation

$$a_{2,0}^{(1)} = f\left(w_{0,0}^{(1)} \cdot 0.00 + w_{0,1}^{(1)} \cdot 0.00 + b_0\right) \tag{17}$$

$$= f\left(1.00 \cdot 0.00 + 1.00 \cdot 0.00 + -1.00\right) \tag{18}$$

$$a_{2,1}^{(1)} = f\left(w_{1,0}^{(1)} \cdot 0.00 + w_{1,1}^{(1)} \cdot 0.00 + b_1\right) \tag{19}$$

$$= f\left(1.00 \cdot 0.00 + 1.00 \cdot 0.00 + 0.00\right) \tag{20}$$

Prediction for $x_2 = (0.00, 0.00)$

- Hidden Computation

$$a_{2,0}^{(1)} = f(w_{0,0}^{(1)} \cdot 0.00 + w_{0,1}^{(1)} \cdot 0.00 + b_0) \tag{17}$$
$$= f(1.00 \cdot 0.00 + 1.00 \cdot 0.00 + -1.00) \tag{18}$$

$$a_{2,1}^{(1)} = f(w_{1,0}^{(1)} \cdot 0.00 + w_{1,1}^{(1)} \cdot 0.00 + b_1) \tag{19}$$
$$= f(1.00 \cdot 0.00 + 1.00 \cdot 0.00 + 0.00) \tag{20}$$

- Hidden Layer: [0. 0.]
- Output Answer

$$a_{2,0}^{(3)} = f(w_{0,0}^{(2)} \cdot 0.00 + w_{0,1}^{(2)} \cdot 0.00 + b_0) \tag{21}$$
$$= f(-2.00 \cdot 0.00 + 1.00 \cdot 0.00 + 0.00) \tag{22}$$

- Prediction: 0.00, Error: 0.00

Prediction for $x_2 = (0.00, 0.00)$

- Hidden Computation

$$a_{2,0}^{(1)} = f(w_{0,0}^{(1)} \cdot 0.00 + w_{0,1}^{(1)} \cdot 0.00 + b_0) \tag{17}$$
$$= f(1.00 \cdot 0.00 + 1.00 \cdot 0.00 + -1.00) \tag{18}$$

$$a_{2,1}^{(1)} = f(w_{1,0}^{(1)} \cdot 0.00 + w_{1,1}^{(1)} \cdot 0.00 + b_1) \tag{19}$$
$$= f(1.00 \cdot 0.00 + 1.00 \cdot 0.00 + 0.00) \tag{20}$$

- Hidden Layer: [0. 0.]
- Output Answer

$$a_{2,0}^{(3)} = f(w_{0,0}^{(2)} \cdot 0.00 + w_{0,1}^{(2)} \cdot 0.00 + b_0) \tag{21}$$
$$= f(-2.00 \cdot 0.00 + 1.00 \cdot 0.00 + 0.00) \tag{22}$$

Prediction for $x_2 = (0.00, 0.00)$

- Hidden Computation

$$a_{2,0}^{(1)} = f(w_{0,0}^{(1)} \cdot 0.00 + w_{0,1}^{(1)} \cdot 0.00 + b_0) \tag{17}$$
$$= f(1.00 \cdot 0.00 + 1.00 \cdot 0.00 + -1.00) \tag{18}$$

$$a_{2,1}^{(1)} = f(w_{1,0}^{(1)} \cdot 0.00 + w_{1,1}^{(1)} \cdot 0.00 + b_1) \tag{19}$$
$$= f(1.00 \cdot 0.00 + 1.00 \cdot 0.00 + 0.00) \tag{20}$$

- Hidden Layer: [0. 0.]
- Output Answer

$$a_{2,0}^{(3)} = f(w_{0,0}^{(2)} \cdot 0.00 + w_{0,1}^{(2)} \cdot 0.00 + b_0) \tag{21}$$
$$= f(-2.00 \cdot 0.00 + 1.00 \cdot 0.00 + 0.00) \tag{22}$$

- Prediction: 0.00, Error: 0.00

Prediction for $x_3 = (0.00, 1.00)$

Prediction for $x_3 = (0.00, 1.00)$

- Hidden Computation

$$a_{3,0}^{(1)} = f\left(w_{0,0}^{(1)} \cdot 0.00 + w_{0,1}^{(1)} \cdot 1.00 + b_0\right) \tag{23}$$

$$= f\left(1.00 \cdot 0.00 + 1.00 \cdot 1.00 + -1.00\right) \tag{24}$$

Prediction for $x_3 = (0.00, 1.00)$

- Hidden Computation

$$a_{3,0}^{(1)} = f\left(w_{0,0}^{(1)} \cdot 0.00 + w_{0,1}^{(1)} \cdot 1.00 + b_0\right) \tag{23}$$

$$= f\left(1.00 \cdot 0.00 + 1.00 \cdot 1.00 + -1.00\right) \tag{24}$$

$$a_{3,1}^{(1)} = f\left(w_{1,0}^{(1)} \cdot 0.00 + w_{1,1}^{(1)} \cdot 1.00 + b_1\right) \tag{25}$$

$$= f\left(1.00 \cdot 0.00 + 1.00 \cdot 1.00 + 0.00\right) \tag{26}$$

Prediction for $x_3 = (0.00, 1.00)$

- Hidden Computation

$$a_{3,0}^{(1)} = f(w_{0,0}^{(1)} \cdot 0.00 + w_{0,1}^{(1)} \cdot 1.00 + b_0) \tag{23}$$
$$= f(1.00 \cdot 0.00 + 1.00 \cdot 1.00 + -1.00) \tag{24}$$

$$a_{3,1}^{(1)} = f(w_{1,0}^{(1)} \cdot 0.00 + w_{1,1}^{(1)} \cdot 1.00 + b_1) \tag{25}$$
$$= f(1.00 \cdot 0.00 + 1.00 \cdot 1.00 + 0.00) \tag{26}$$

- Hidden Layer: [0 1]
- Output Answer

$$a_{3,0}^{(3)} = f(w_{0,0}^{(2)} \cdot 0.00 + w_{0,1}^{(2)} \cdot 1.00 + b_0) \tag{27}$$
$$= f(-2.00 \cdot 0.00 + 1.00 \cdot 1.00 + 0.00) \tag{28}$$

- Prediction: 1.00, Error: 0.00

Prediction for $x_3 = (0.00, 1.00)$

- Hidden Computation

$$a_{3,0}^{(1)} = f\left(w_{0,0}^{(1)} \cdot 0.00 + w_{0,1}^{(1)} \cdot 1.00 + b_0\right) \tag{23}$$

$$= f\left(1.00 \cdot 0.00 + 1.00 \cdot 1.00 + -1.00\right) \tag{24}$$

$$a_{3,1}^{(1)} = f\left(w_{1,0}^{(1)} \cdot 0.00 + w_{1,1}^{(1)} \cdot 1.00 + b_1\right) \tag{25}$$

$$= f\left(1.00 \cdot 0.00 + 1.00 \cdot 1.00 + 0.00\right) \tag{26}$$

- Hidden Layer: [0 1]
- Output Answer

$$a_{3,0}^{(3)} = f\left(w_{0,0}^{(2)} \cdot 0.00 + w_{0,1}^{(2)} \cdot 1.00 + b_0\right) \tag{27}$$

$$= f\left(-2.00 \cdot 0.00 + 1.00 \cdot 1.00 + 0.00\right) \tag{28}$$

Prediction for $x_3 = (0.00, 1.00)$

- Hidden Computation

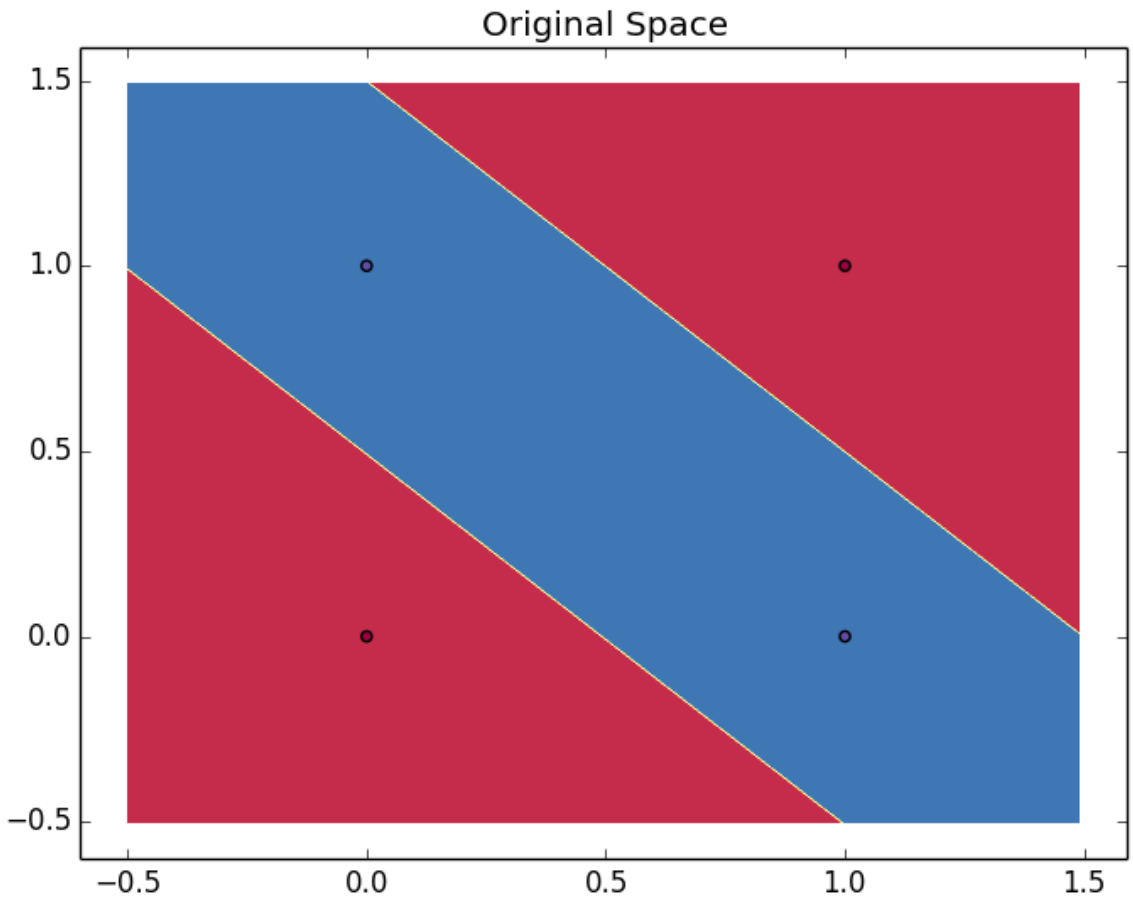
$$a_{3,0}^{(1)} = f(w_{0,0}^{(1)} \cdot 0.00 + w_{0,1}^{(1)} \cdot 1.00 + b_0) \tag{23}$$
$$= f(1.00 \cdot 0.00 + 1.00 \cdot 1.00 + -1.00) \tag{24}$$

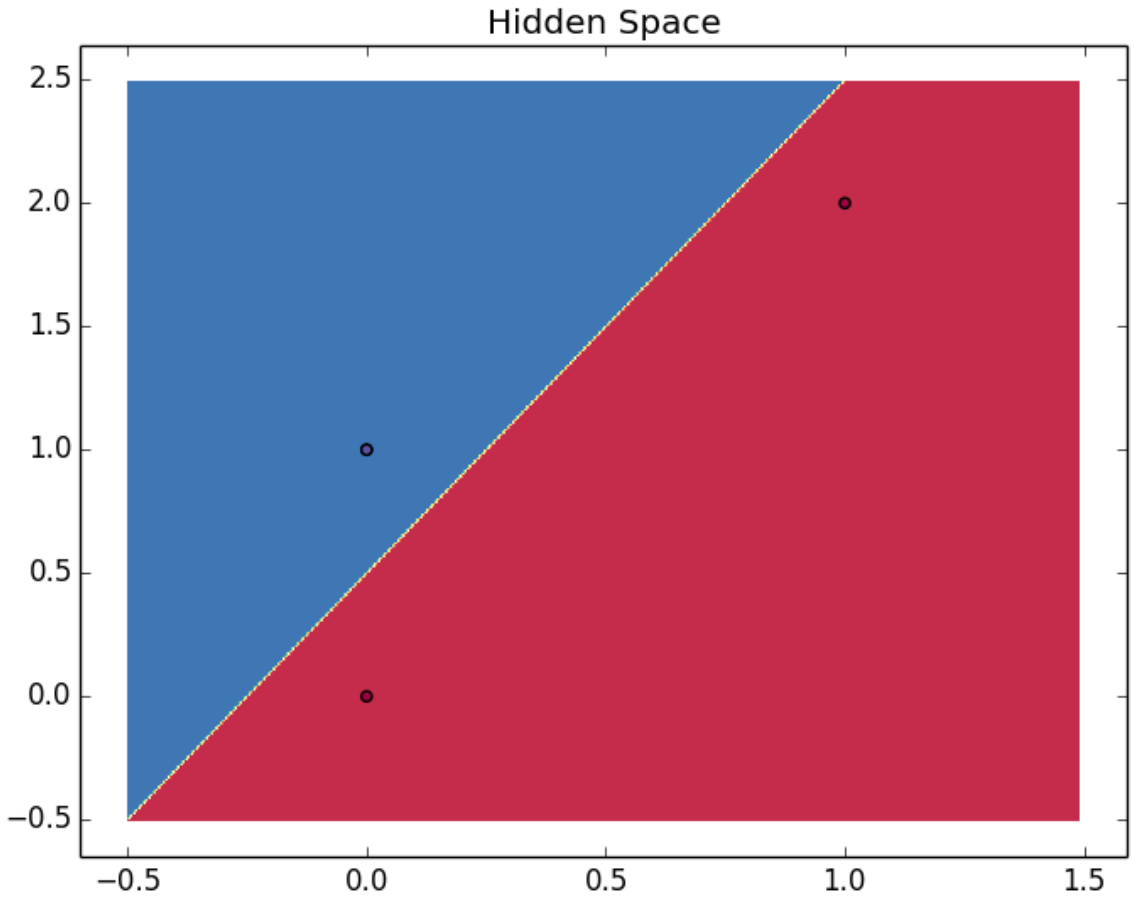
$$a_{3,1}^{(1)} = f(w_{1,0}^{(1)} \cdot 0.00 + w_{1,1}^{(1)} \cdot 1.00 + b_1) \tag{25}$$
$$= f(1.00 \cdot 0.00 + 1.00 \cdot 1.00 + 0.00) \tag{26}$$

- Hidden Layer: [0 1]
- Output Answer

$$a_{3,0}^{(3)} = f(w_{0,0}^{(2)} \cdot 0.00 + w_{0,1}^{(2)} \cdot 1.00 + b_0) \tag{27}$$
$$= f(-2.00 \cdot 0.00 + 1.00 \cdot 1.00 + 0.00) \tag{28}$$

- Prediction: 1.00, Error: 0.00


Original Space
1.5
1.0
0.5
0.0
−0.5
−0.5
0.0
0.5
1.0
1.5


Hidden Space
2.5
2.0
1.5
1.0
0.5
0.0
−0.5
−0.5
0.0
0.5
1.0
1.5