



Department of Computer Science

UNIVERSITY OF COLORADO **BOULDER**



Deep Learning

Advanced Machine Learning for NLP

Jordan Boyd-Graber

HANDS-ON DEMO

Data and Model

Data

| x_1 | x_2 | y |
|-------|-------|------|
| 1.00 | 1.00 | 0.00 |
| 1.00 | 0.00 | 0.00 |
| 0.00 | 0.00 | 0.00 |
| 0.00 | 1.00 | 0.00 |

Using ReLU as non-linearity

First Layer

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

$$b^{(1)} = [-1.00 \quad 0.00] \quad (2)$$

Second Layer

$$w^{(2)} = [-2.00 \quad 1.00] \quad (3)$$

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

Prediction for $x_0 = (1.00, 1.00)$

Prediction for $x_0 = (1.00, 1.00)$

- Hidden Computation

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

$$= f(1.00 \cdot 1.00 + 1.00 \cdot 1.00 + -1.00) \quad (6)$$

Prediction for $x_0 = (1.00, 1.00)$

- Hidden Computation

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

$$= f(1.00 \cdot 1.00 + 1.00 \cdot 1.00 + -1.00) \quad (6)$$

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

$$= f(1.00 \cdot 1.00 + 1.00 \cdot 1.00 + 0.00) \quad (8)$$

Prediction for $x_0 = (1.00, 1.00)$

- Hidden Computation

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

$$= f(1.00 \cdot 1.00 + 1.00 \cdot 1.00 + -1.00) \quad (6)$$

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

$$= f(1.00 \cdot 1.00 + 1.00 \cdot 1.00 + 0.00) \quad (8)$$

- Hidden Layer: [1 2]
- Output Answer

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

$$= f(-2.00 \cdot 1.00 + 1.00 \cdot 2.00 + 0.00) \quad (10)$$

- Prediction: 0.00. Error: 0.00

Prediction for $x_0 = (1.00, 1.00)$

- Hidden Computation

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

$$= f(1.00 \cdot 1.00 + 1.00 \cdot 1.00 + -1.00) \quad (6)$$

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

$$= f(1.00 \cdot 1.00 + 1.00 \cdot 1.00 + 0.00) \quad (8)$$

- Hidden Layer: [1 2]
- Output Answer

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

$$= f(-2.00 \cdot 1.00 + 1.00 \cdot 2.00 + 0.00) \quad (10)$$

Prediction for $x_0 = (1.00, 1.00)$

- Hidden Computation

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

$$= f(1.00 \cdot 1.00 + 1.00 \cdot 1.00 + -1.00) \quad (6)$$

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

$$= f(1.00 \cdot 1.00 + 1.00 \cdot 1.00 + 0.00) \quad (8)$$

- Hidden Layer: [1 2]
- Output Answer

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

$$= f(-2.00 \cdot 1.00 + 1.00 \cdot 2.00 + 0.00) \quad (10)$$

Prediction for $x_1 = (1.00, 0.00)$

Prediction for $x_1 = (1.00, 0.00)$

- Hidden Computation

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

$$= f(1.00 \cdot 1.00 + 1.00 \cdot 0.00 + -1.00) \quad (12)$$

Prediction for $x_1 = (1.00, 0.00)$

- Hidden Computation

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

$$= f(1.00 \cdot 1.00 + 1.00 \cdot 0.00 + -1.00) \quad (12)$$

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

$$= f(1.00 \cdot 1.00 + 1.00 \cdot 0.00 + 0.00) \quad (14)$$

Prediction for $x_1 = (1.00, 0.00)$

- Hidden Computation

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

$$= f(1.00 \cdot 1.00 + 1.00 \cdot 0.00 + -1.00) \quad (12)$$

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

$$= f(1.00 \cdot 1.00 + 1.00 \cdot 0.00 + 0.00) \quad (14)$$

- Hidden Layer: [0 1]
- Output Answer

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

$$= f(-2.00 \cdot 0.00 + 1.00 \cdot 1.00 + 0.00) \quad (16)$$

- Prediction: 1.00. Error: 0.00

Prediction for $x_1 = (1.00, 0.00)$

- Hidden Computation

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

$$= f(1.00 \cdot 1.00 + 1.00 \cdot 0.00 + -1.00) \quad (12)$$

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

$$= f(1.00 \cdot 1.00 + 1.00 \cdot 0.00 + 0.00) \quad (14)$$

- Hidden Layer: [0 1]
- Output Answer

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

$$= f(-2.00 \cdot 0.00 + 1.00 \cdot 1.00 + 0.00) \quad (16)$$

Prediction for $x_1 = (1.00, 0.00)$

- Hidden Computation

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

$$= f(1.00 \cdot 1.00 + 1.00 \cdot 0.00 + -1.00) \quad (12)$$

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

$$= f(1.00 \cdot 1.00 + 1.00 \cdot 0.00 + 0.00) \quad (14)$$

- Hidden Layer: [0 1]
- Output Answer

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

$$= f(-2.00 \cdot 0.00 + 1.00 \cdot 1.00 + 0.00) \quad (16)$$

Prediction for $x_2 = (0.00, 0.00)$

Prediction for $x_2 = (0.00, 0.00)$

- Hidden Computation

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

$$= f(1.00 \cdot 0.00 + 1.00 \cdot 0.00 + -1.00) \quad (18)$$

Prediction for $x_2 = (0.00, 0.00)$

- Hidden Computation

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

$$= f(1.00 \cdot 0.00 + 1.00 \cdot 0.00 + -1.00) \quad (18)$$

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

$$= f(1.00 \cdot 0.00 + 1.00 \cdot 0.00 + 0.00) \quad (20)$$

Prediction for $x_2 = (0.00, 0.00)$

- Hidden Computation

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

$$= f(1.00 \cdot 0.00 + 1.00 \cdot 0.00 + -1.00) \quad (18)$$

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

$$= f(1.00 \cdot 0.00 + 1.00 \cdot 0.00 + 0.00) \quad (20)$$

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

$$a_{2,0}^{(3)} = f(w_{0,0}^{(2)} \cdot 0.00 + w_{0,1}^{(2)} \cdot 0.00 + b_0) \quad (21)$$

$$= f(-2.00 \cdot 0.00 + 1.00 \cdot 0.00 + 0.00) \quad (22)$$

- Prediction: 0.00. Error: 0.00

Prediction for $x_2 = (0.00, 0.00)$

- Hidden Computation

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

$$= f(1.00 \cdot 0.00 + 1.00 \cdot 0.00 + -1.00) \quad (18)$$

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

$$= f(1.00 \cdot 0.00 + 1.00 \cdot 0.00 + 0.00) \quad (20)$$

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

$$a_{2,0}^{(3)} = f(w_{0,0}^{(2)} \cdot 0.00 + w_{0,1}^{(2)} \cdot 0.00 + b_0) \quad (21)$$

$$= f(-2.00 \cdot 0.00 + 1.00 \cdot 0.00 + 0.00) \quad (22)$$

Prediction for $x_2 = (0.00, 0.00)$

- Hidden Computation

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

$$= f(1.00 \cdot 0.00 + 1.00 \cdot 0.00 + -1.00) \quad (18)$$

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

$$= f(1.00 \cdot 0.00 + 1.00 \cdot 0.00 + 0.00) \quad (20)$$

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

$$a_{2,0}^{(3)} = f(w_{0,0}^{(2)} \cdot 0.00 + w_{0,1}^{(2)} \cdot 0.00 + b_0) \quad (21)$$

$$= f(-2.00 \cdot 0.00 + 1.00 \cdot 0.00 + 0.00) \quad (22)$$

Prediction for $x_3 = (0.00, 1.00)$

Prediction for $x_3 = (0.00, 1.00)$

- Hidden Computation

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

$$= f(1.00 \cdot 0.00 + 1.00 \cdot 1.00 + -1.00) \quad (24)$$

Prediction for $x_3 = (0.00, 1.00)$

- Hidden Computation

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

$$= f(1.00 \cdot 0.00 + 1.00 \cdot 1.00 + -1.00) \quad (24)$$

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

$$= f(1.00 \cdot 0.00 + 1.00 \cdot 1.00 + 0.00) \quad (26)$$

Prediction for $x_3 = (0.00, 1.00)$

- Hidden Computation

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

$$= f(1.00 \cdot 0.00 + 1.00 \cdot 1.00 + -1.00) \quad (24)$$

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

$$= f(1.00 \cdot 0.00 + 1.00 \cdot 1.00 + 0.00) \quad (26)$$

- Hidden Layer: [0 1]
- Output Answer

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

$$= f(-2.00 \cdot 0.00 + 1.00 \cdot 1.00 + 0.00) \quad (28)$$

- Prediction: 1.00. Error: 0.00

Prediction for $x_3 = (0.00, 1.00)$

- Hidden Computation

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

$$= f(1.00 \cdot 0.00 + 1.00 \cdot 1.00 + -1.00) \quad (24)$$

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

$$= f(1.00 \cdot 0.00 + 1.00 \cdot 1.00 + 0.00) \quad (26)$$

- Hidden Layer: [0 1]
- Output Answer

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

$$= f(-2.00 \cdot 0.00 + 1.00 \cdot 1.00 + 0.00) \quad (28)$$

Prediction for $x_3 = (0.00, 1.00)$

- Hidden Computation

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

$$= f(1.00 \cdot 0.00 + 1.00 \cdot 1.00 + -1.00) \quad (24)$$

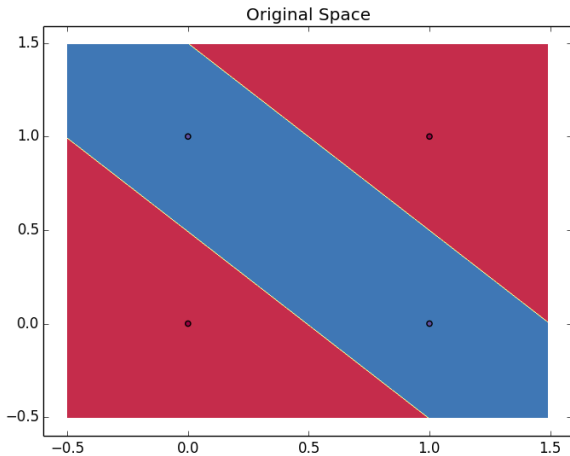
$$a_{3,1}^{(1)} = f(w_{1,0}^{(1)} \cdot 0.00 + w_{1,1}^{(1)} \cdot 1.00 + b_1) \quad (25)$$

$$= f(1.00 \cdot 0.00 + 1.00 \cdot 1.00 + 0.00) \quad (26)$$

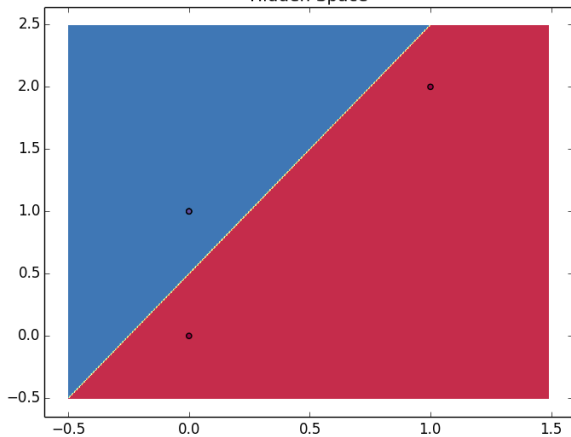
- Hidden Layer: [0 1]
- Output Answer

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

$$= f(-2.00 \cdot 0.00 + 1.00 \cdot 1.00 + 0.00) \quad (28)$$



Hidden Space



Next Time

- Representing Words
- Updating representations
- Comparing with contextual information