Neural Network II

Chul Min Yeum
Assistant Professor
Civil and Environmental Engineering
University of Waterloo, Canada

CIVE 497 – CIVE 700: Smart Structure Technology

UNIVERSITY OF WATERLOO
FACULTY OF ENGINEERING

Last updated: 2020-03-27

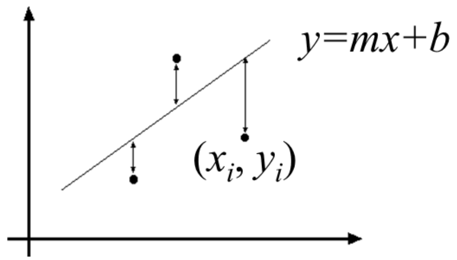
Recall: Linear Regression

Data (measurement): $(x_1, y_1), \ldots, (x_n, y_n)$

Model: Line $f(x_i, m, b) = mx_i + b$

Task: Find (m, b)

Minimize $E = J(m,b) = \sum_{i=1}^{n} (y_i - f(x_i, m, b))^2 = \sum_{i=1}^{n} (y_i - mx_i - b)^2$



$$J(\theta) = J(\theta^1, \theta^2) = \sum_{i=1}^{n} (y_i - \theta^1 x_i - \theta^2)^2$$

$$\frac{\partial J(\theta^1, \theta^2)}{\partial \theta^1} = -2 \sum_{i=1}^{n} [y_i - \theta^1 x_i - \theta^2] x_i \qquad \theta_{j+1}^1 \leftarrow \theta_j^1 - \alpha \frac{\partial}{\partial \theta_j^1} J(\theta)$$

$$\frac{\partial J(\theta^1, \theta^2)}{\partial \theta^2} = -2 \sum_{i=1}^{n} [y_i - \theta^1 x_i - \theta^2] \qquad \theta_{j+1}^2 \leftarrow \theta_j^2 - \alpha \frac{\partial}{\partial \theta_j^2} J(\theta)$$

Recall: Linear Regression (Continue)

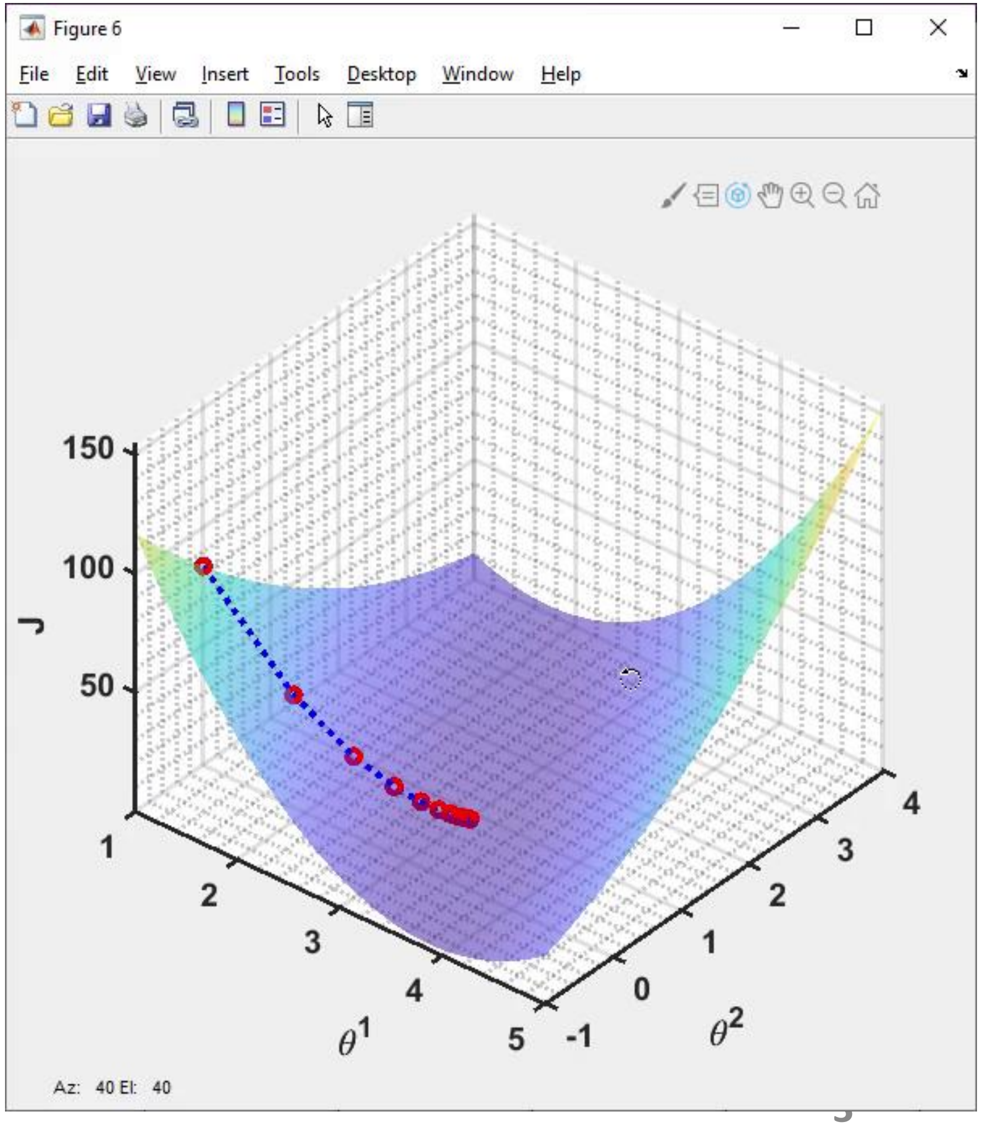
Gradient Descent

$$\theta_{j+1} \leftarrow \theta_j - \alpha \frac{\partial}{\partial \theta_j} J(\theta)$$

Repeat until convergence

$$\frac{\partial J(\theta^1, \theta^2)}{\partial \theta^1} = -2 \sum_{i=1}^{n} [y_i - \theta^1 x_i - \theta^2] x_i \qquad \theta^1_{j+1} \leftarrow \theta^1_j - \alpha \frac{\partial}{\partial \theta^1_j} J(\theta)$$

$$\frac{\partial J(\theta^1, \theta^2)}{\partial \theta^2} = -2 \sum_{i=1}^{n} [y_i - \theta^1 x_i - \theta^2] \qquad \theta^2_{j+1} \leftarrow \theta^2_j - \alpha \frac{\partial}{\partial \theta^2_j} J(\theta)$$



Backward Pass

Our goal with backpropagation is **to update each of the weights in the network** so that they cause the actual output to be closer the target output, thereby minimizing the error for each output neuron and the network as a whole.

$$\theta_{j+1} \leftarrow \theta_j - \alpha \frac{\partial}{\partial \theta_j} J(\theta)$$

How to find $\frac{\partial}{\partial \theta_j} J(\theta)$ to update the parameter θ ?

Chain Rule

Given a multivariable function $f(x, y)$, and two single variable functions $x(t)$ and $y(t)$, here's what the multivariable chain rule says:

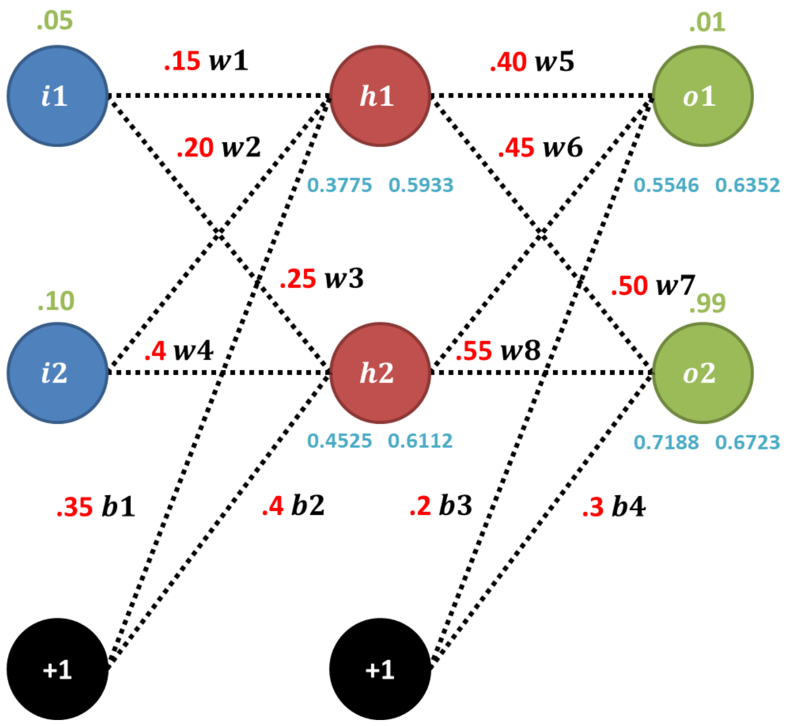
$$\underbrace{\frac{d}{dt} f(x(t), y(t))}_{\text{Derivative of composition function}} = \frac{\partial f}{\partial x}\frac{dx}{dt} + \frac{\partial f}{\partial y}\frac{dy}{dt}$$

The single variable chain rule tells you how to take the derivative of the composition of two functions:

$$\frac{d}{dt} f(g(t)) = \frac{df}{dg}\frac{dg}{dt} = f'(g(t))g'(t)$$

https://www.khanacademy.org/math/multivariable-calculus/multivariable-derivatives/differentiating-vector-valued-functions/a/multivariable-chain-rule-simple-version

Backpropagation (w_5)



$$E = J(\mathbf{w}, \mathbf{b})$$

$$= (o_1 - f(i_1, i_2, \mathbf{w}, \mathbf{b})_1)^2 + (o_2 - f(i_1, i_2, \mathbf{w}, \mathbf{b})_2)^2$$

$$= (o_1 - out_{o1})^2 + (o_2 - out_{o2})^2$$

$$out_{o1} = f(net_{o1}) = \frac{1}{1 + e^{-net_{o1}}}$$

$$net_{o1} = w_5 out_{h1} + w_6 out_{h2} + b_3$$

$$\boxed{\theta_{j+1} \leftarrow \theta_j - \alpha \frac{\partial}{\partial \theta_j} J(\theta)}$$

$$\frac{\partial E}{\partial w_5} = \frac{\partial E}{\partial out_{o1}} \frac{\partial out_{o1}}{\partial net_{o1}} \frac{\partial net_{o1}}{\partial w_5} + \frac{\partial E}{\partial out_{o2}} \frac{\partial out_{o2}}{\partial net_{o2}} \cancel{\frac{\partial net_{o2}}{\partial w_5}}$$

0

$$w_5 \leftarrow w_5 - \alpha \frac{\partial E}{\partial w_5}$$

$$= \frac{\partial E}{\partial out_{o1}} \frac{\partial out_{o1}}{\partial net_{o1}} \frac{\partial net_{o1}}{\partial w_5}$$

Backpropagation (w_5)

$$\frac{\partial E}{\partial w_5} = \frac{\partial E}{\partial out_{o1}} \frac{\partial out_{o1}}{\partial net_{o1}} \frac{\partial net_{o1}}{\partial w_5}$$

$$E = J(\mathbf{w}, \mathbf{b})$$
$$= (o_1 - f(i_1, i_2, \mathbf{w}, \mathbf{b})_1)^2$$
$$+ (o_2 - f(i_1, i_2, \mathbf{w}, \mathbf{b})_2)^2$$
$$= (o_1 - out_{o1})^2 + (o_2 - out_{o2})^2$$

$$out_{o1} = f(net_{o1}) = \frac{1}{1 + e^{-net_{o1}}}$$

$$net_{o1} = w_5 out_{h1} + w_6 out_{h2} + b_3$$

$$\frac{\partial E}{\partial out_{o1}} = -2(o_1 - out_{o1})$$

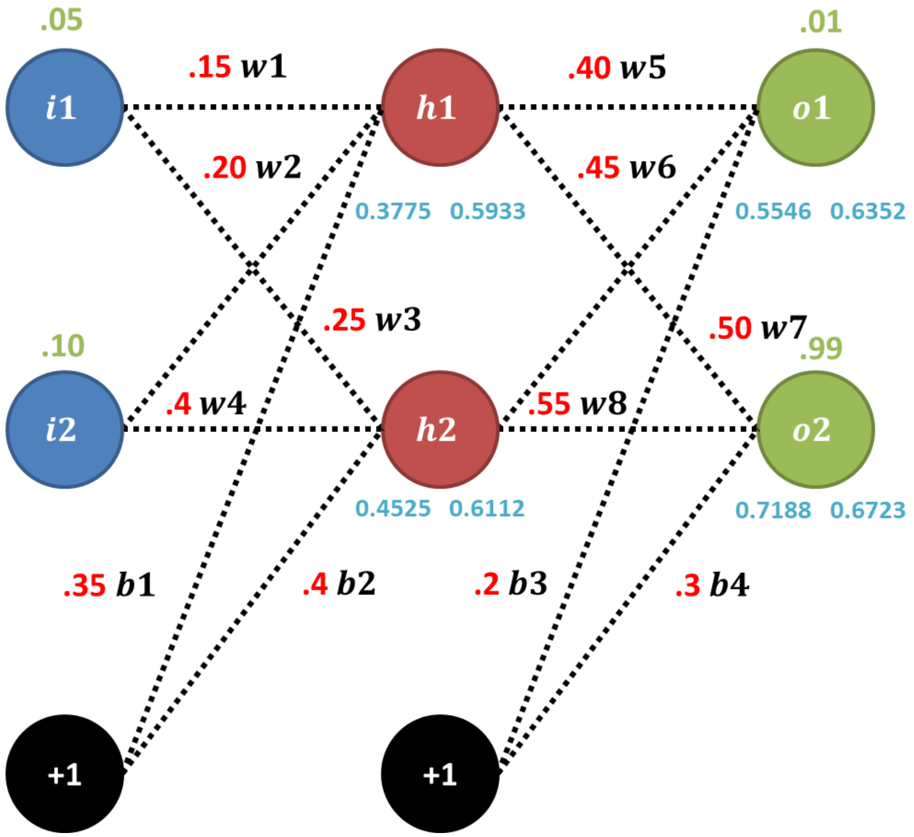
$$\frac{\partial out_{o1}}{\partial net_{o1}} = f(net_{o1}) * f(-net_{o1})$$

$$\frac{\partial net_{o1}}{\partial w_5} = out_{h1}$$

$$f(x) = \frac{1}{1 + e^{-x}} = \frac{e^x}{1 + e^x},$$

$$\frac{\mathrm{d}}{\mathrm{d}x} f(x) = \frac{e^x \cdot (1 + e^x) - e^x \cdot e^x}{(1 + e^x)^2} = \frac{e^x}{(1 + e^x)^2} = f(x)(1 - f(x)) = f(x)f(-x).$$

Backpropagation (w_5)



$$\boxed{\theta_{j+1} \leftarrow \theta_j - \alpha \frac{\partial}{\partial \theta_j} J(\theta)}$$

$$w_5 \leftarrow w_5 - \alpha \frac{\partial E}{\partial w_5}$$

$$\frac{\partial E}{\partial out_{o1}} = -2(o_1 - out_{o1}) = -2(0.01 - 0.6352)$$
$$= 1.2504$$

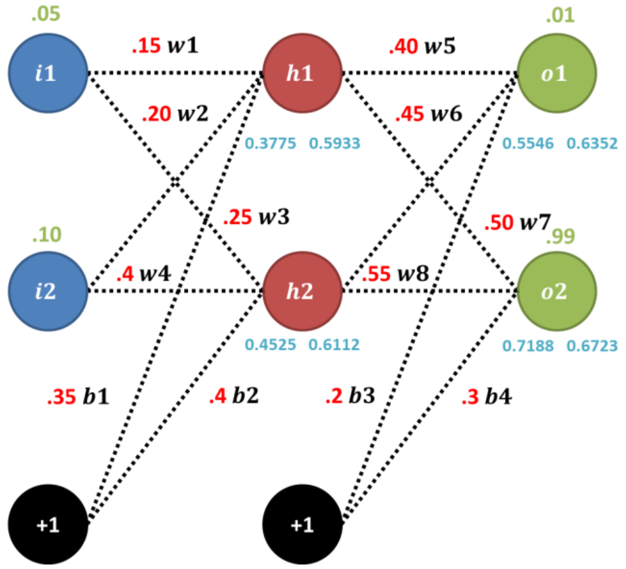
$$\frac{\partial out_{o1}}{\partial net_{o1}} = f(net_{o1}) * f(-net_{o1})$$
$$= f(0.5546) * f(-0.5546) = 0.2317$$

$$\frac{\partial net_{o1}}{\partial w_5} = out_{h1} = 0.5933$$

$$\frac{\partial E}{\partial w_5} = \frac{\partial E}{\partial out_{o1}} \frac{\partial out_{o1}}{\partial net_{o1}} \frac{\partial net_{o1}}{\partial w_5}$$

$$= 1.2504 * 0.2317 * 0.5933 = 0.1719$$

Backpropagation (w_1)



$$E = J(\mathbf{w}, \mathbf{b}) = (o_1 - out_{o1})^2 + (o_2 - out_{o2})^2$$

$$out_{o1} = f(net_{o1}) = \frac{1}{1 + e^{-net_{o1}}} \qquad net_{o2} = w_7 out_{h1} + w_8 out_{h2} + b_4$$

$$net_{o1} = w_5 out_{h1} + w_6 out_{h2} + b_3 \qquad out_{o2} = f(net_{o2}) = \frac{1}{1 + e^{-net_{o2}}}$$

$$net_{h1} = w_1 i_1 + w_2 i_2 + b_1 \qquad net_{h2} = w_3 i_1 + w_4 i_2 + b_2$$

$$out_{h2} = f(net_{h2}) = \frac{1}{1 + e^{-net_{h2}}}$$

$$out_{h1} = f(net_{h1}) = \frac{1}{1 + e^{-net_{h1}}}$$

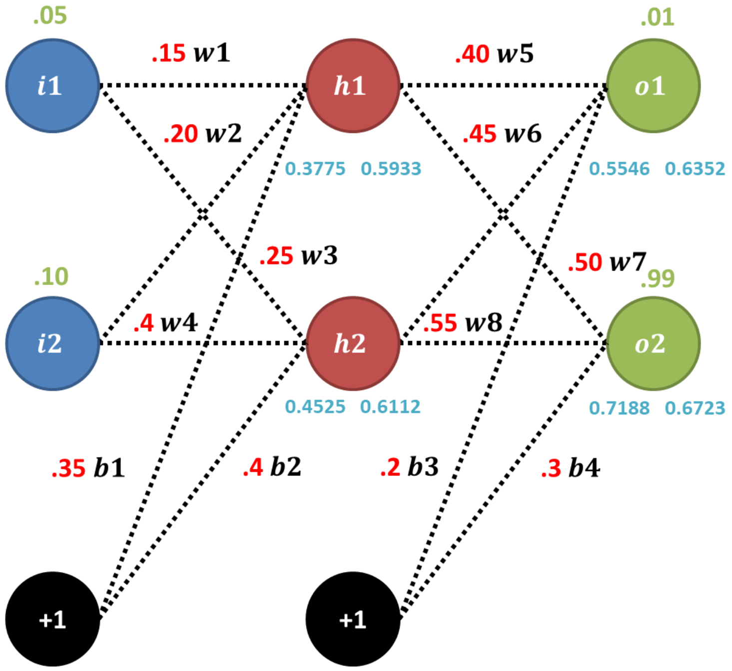
$$\frac{\partial E}{\partial w_1} = \frac{\partial E}{\partial out_{o1}} \frac{\partial out_{o1}}{\partial net_{o1}} \frac{\partial net_{o1}}{\partial w_1} + \frac{\partial E}{\partial out_{o2}} \frac{\partial out_{o2}}{\partial net_{o2}} \frac{\partial net_{o2}}{\partial w_1}$$

$$= -2(o_1 - out_{o1}) * f(net_{o1}) * f(-net_{o1}) * \frac{\partial net_{o1}}{\partial w_1} + -2(o_2 - out_{o2}) * f(net_{o2}) * f(-net_{o2}) * \frac{\partial net_{o2}}{\partial w_1}$$

$$\frac{\partial net_{o1}}{\partial w_1} = \frac{\partial net_{o1}}{\partial out_{h1}} \frac{\partial out_{h1}}{\partial net_{h1}} \frac{\partial net_{h1}}{\partial w_1} + \frac{\partial net_{o1}}{\partial out_{h2}} \frac{\partial out_{h2}}{\partial net_{h2}} \underbrace{\frac{\partial net_{h2}}{\partial w_1}}_{0} = \frac{\partial net_{o1}}{\partial out_{h1}} \frac{\partial out_{h1}}{\partial net_{h1}} \frac{\partial net_{h1}}{\partial w_1} = w_5 * f(net_{h1}) * f(-net_{h1}) * i_1$$

$$\frac{\partial net_{o2}}{\partial w_1} = \frac{\partial net_{o2}}{\partial out_{h1}} \frac{\partial out_{h1}}{\partial net_{h1}} \frac{\partial net_{h1}}{\partial w_1} + \frac{\partial net_{o2}}{\partial out_{h2}} \frac{\partial out_{h2}}{\partial net_{h2}} \underbrace{\frac{\partial net_{h2}}{\partial w_1}}_{0} = \frac{\partial net_{o2}}{\partial out_{h1}} \frac{\partial out_{h1}}{\partial net_{h1}} \frac{\partial net_{h1}}{\partial w_1} = w_7 * f(net_{h1}) * f(-net_{h1}) * i_1$$

Backpropagation (w_1)



$$\boxed{\theta^{j+1} \leftarrow \theta^{j} - \alpha \frac{\partial}{\partial \theta^{j}} J(\theta)}$$

$$w_1 \leftarrow w_1 - \alpha \frac{\partial E}{\partial w_1}$$

$$\frac{\partial E}{\partial w_1} = \frac{\partial E}{\partial out_{o1}} \frac{\partial out_{o1}}{\partial net_{o1}} \frac{\partial net_{o1}}{\partial w_1} + \frac{\partial E}{\partial out_{o2}} \frac{\partial out_{o2}}{\partial net_{o2}} \frac{\partial net_{o2}}{\partial w_1}$$

$$= -2(o_1 - out_{o1}) * f(net_{o1}) * f(-net_{o1}) * \frac{\partial net_{o1}}{\partial w_1} +$$

$$-2(o_2 - out_{o2}) * f(net_{o2}) * f(-net_{o2}) * \frac{\partial net_{o2}}{\partial w_1}$$

$$= -2(0.01 - 0.6352) * f(0.5546) * f(-0.5546) * 0.0048 - 2(0.99 - 0.6723) * f(0.7188) * f(-0.7188) * 0.006 = 5.5090e - 04$$

$$\frac{\partial net_{o1}}{\partial w_1} = \frac{\partial net_{o1}}{\partial out_{h1}} \frac{\partial out_{h1}}{\partial net_{h1}} \frac{\partial net_{h1}}{\partial w_1} + \frac{\partial net_{o1}}{\partial out_{h2}} \frac{\partial out_{h2}}{\partial net_{h2}} \frac{\partial net_{h2}}{\partial w_1} = \frac{\partial net_{o1}}{\partial out_{h1}} \frac{\partial out_{h1}}{\partial net_{h1}} \frac{\partial net_{h1}}{\partial w_1}$$

$$= w_5 * f(net_{h1}) * f(-net_{h1}) * i_1 = 0.4 * f(0.3775) * f(-0.3775) * 0.05 = 0.0048$$

$$\frac{\partial net_{o2}}{\partial w_1} = \frac{\partial net_{o2}}{\partial out_{h1}} \frac{\partial out_{h1}}{\partial net_{h1}} \frac{\partial net_{h1}}{\partial w_1} + \frac{\partial net_{o2}}{\partial out_{h2}} \frac{\partial out_{h2}}{\partial net_{h2}} \frac{\partial net_{h2}}{\partial w_1} = \frac{\partial net_{o2}}{\partial out_{h1}} \frac{\partial out_{h1}}{\partial net_{h1}} \frac{\partial net_{h1}}{\partial w_1}$$

$$= w_7 * f(net_{h1}) * f(-net_{h1}) * i_1 = 0.5 * f(0.3775) * f(-0.3775) * 0.05 = 0.006$$

Efficient Computation Forward and Backward Propagation

$$\frac{\partial E}{\partial w_5} = \frac{\partial E}{\partial out_{o1}} \frac{\partial out_{o1}}{\partial net_{o1}} \frac{\partial net_{o1}}{\partial w_5} + \frac{\partial E}{\partial out_{o2}} \frac{\partial out_{o2}}{\partial net_{o2}} \frac{\partial net_{o2}}{\partial w_5} = \frac{\partial E}{\partial out_{o1}} \frac{\partial out_{o1}}{\partial net_{o1}} \frac{\partial net_{o1}}{\partial w_5}$$

$$\frac{\partial E}{\partial w_7} = \frac{\partial E}{\partial out_{o1}} \frac{\partial out_{o1}}{\partial net_{o1}} \frac{\partial net_{o1}}{\partial w_7} + \frac{\partial E}{\partial out_{o2}} \frac{\partial out_{o2}}{\partial net_{o2}} \frac{\partial net_{o2}}{\partial w_7} = \frac{\partial E}{\partial out_{o2}} \frac{\partial out_{o2}}{\partial net_{o2}} \frac{\partial net_{o2}}{\partial w_7}$$

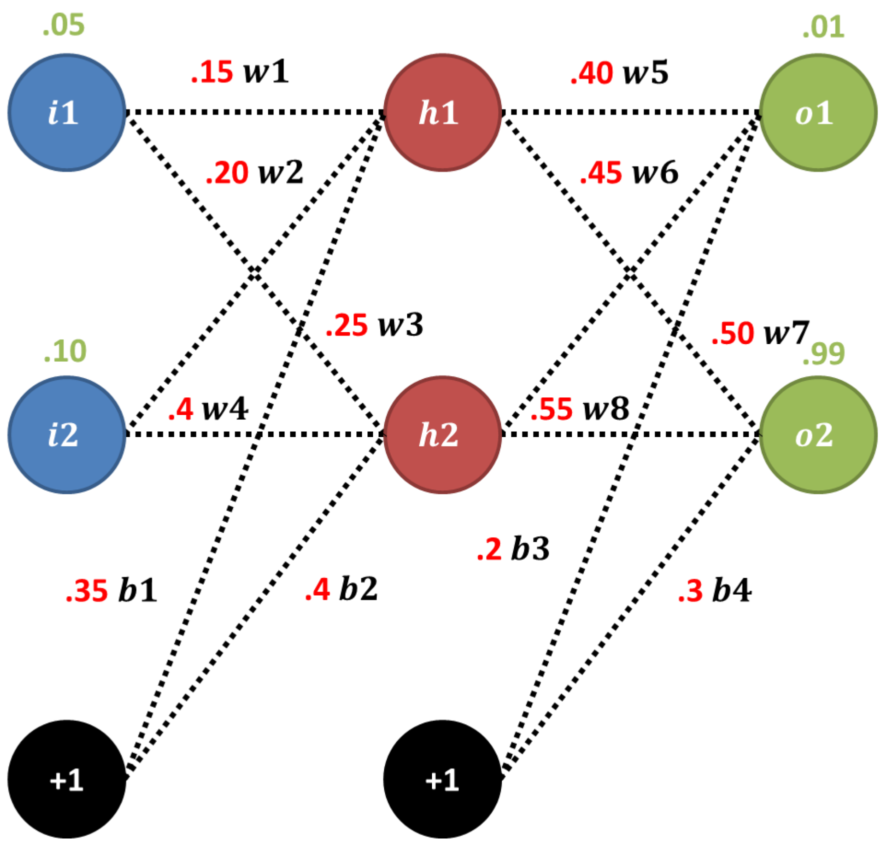
$$\frac{\partial E}{\partial w_1} = \frac{\partial E}{\partial out_{o1}} \frac{\partial out_{o1}}{\partial net_{o1}} \frac{\partial net_{o1}}{\partial w_1} + \frac{\partial E}{\partial out_{o2}} \frac{\partial out_{o2}}{\partial net_{o2}} \frac{\partial net_{o2}}{\partial w_1}$$

$$= \frac{\partial E}{\partial out_{o1}} \frac{\partial out_{o1}}{\partial net_{o1}} \frac{\partial net_{o1}}{\partial out_{h1}} \frac{\partial out_{h1}}{\partial net_{h1}} \frac{\partial net_{h1}}{\partial w_1} + \frac{\partial E}{\partial out_{o2}} \frac{\partial out_{o2}}{\partial net_{o2}} \frac{\partial net_{o2}}{\partial out_{h1}} \frac{\partial out_{h1}}{\partial net_{h1}} \frac{\partial net_{h1}}{\partial w_1}$$

$$\frac{\partial E}{\partial w_4} = \frac{\partial E}{\partial out_{o1}} \frac{\partial out_{o1}}{\partial net_{o1}} \frac{\partial net_{o1}}{\partial w_4} + \frac{\partial E}{\partial out_{o2}} \frac{\partial out_{o2}}{\partial net_{o2}} \frac{\partial net_{o2}}{\partial w_4}$$

$$= \frac{\partial E}{\partial out_{o1}} \frac{\partial out_{o1}}{\partial net_{o1}} \frac{\partial net_{o1}}{\partial out_{h2}} \frac{\partial out_{h2}}{\partial net_{h2}} \frac{\partial net_{h2}}{\partial w_4} + \frac{\partial E}{\partial out_{o2}} \frac{\partial out_{o2}}{\partial net_{o2}} \frac{\partial net_{o2}}{\partial out_{h2}} \frac{\partial out_{h2}}{\partial net_{h2}} \frac{\partial net_{h2}}{\partial w_4}$$

Matrix Representation of Network Parameters



$$Input = \begin{bmatrix} i_1 & i_2 \end{bmatrix} \quad Ouptut = \begin{bmatrix} o_1 \\ o_2 \end{bmatrix}$$

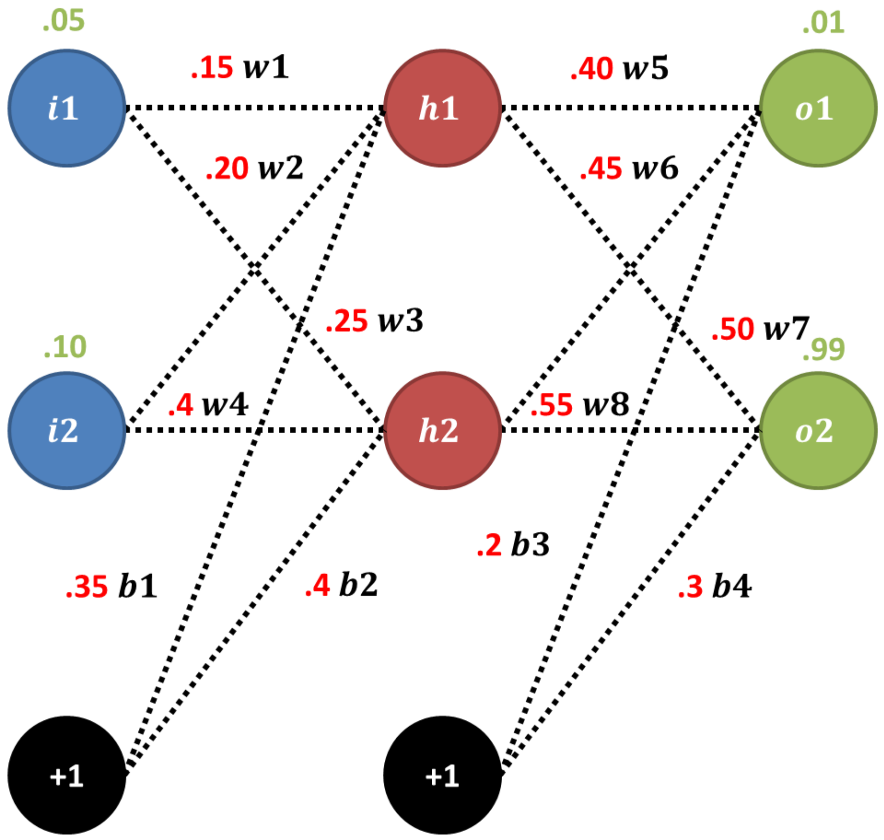
$$W_{ih} = \begin{bmatrix} w_1 & w_3 \\ w_2 & w_4 \end{bmatrix} \quad b_{ih} = \begin{bmatrix} b_1 \\ b_2 \end{bmatrix}$$

$$W_{ho} = \begin{bmatrix} w_5 & w_7 \\ w_6 & w_8 \end{bmatrix} \quad b_{oh} = \begin{bmatrix} b_3 \\ b_4 \end{bmatrix}$$

$$h_{net} = \begin{bmatrix} net_{h1} \\ net_{h2} \end{bmatrix} \quad h_{out} = \begin{bmatrix} out_{h1} \\ out_{h2} \end{bmatrix}$$

$$O_{net} = \begin{bmatrix} net_{o1} \\ net_{o2} \end{bmatrix} \quad O_{out} = \begin{bmatrix} out_{o1} \\ out_{o2} \end{bmatrix}$$

Forward Propagation (Matrix Operation)



$$h_{net} = Input * W_{ih} + b_{ih} = \begin{bmatrix} i_1 & i_2 \end{bmatrix} \begin{bmatrix} w_1 & w_3 \\ w_2 & w_4 \end{bmatrix} + \begin{bmatrix} b_1 \\ b_2 \end{bmatrix} = \begin{bmatrix} net_{h1} \\ net_{h2} \end{bmatrix}$$

$$h_{out} = f\left(\begin{bmatrix} net_{h1} \\ net_{h2} \end{bmatrix}\right) = \begin{bmatrix} out_{h1} \\ out_{h2} \end{bmatrix}$$

$$o_{net} = h_{out} * W_{ho} + b_{ho} = \begin{bmatrix} out_{h1} \\ out_{h2} \end{bmatrix} \begin{bmatrix} w_5 & w_7 \\ w_6 & w_8 \end{bmatrix} + \begin{bmatrix} b_3 \\ b_4 \end{bmatrix} = \begin{bmatrix} net_{o1} \\ net_{o2} \end{bmatrix}$$

$$o_{out} = f\left(\begin{bmatrix} net_{o1} \\ net_{o2} \end{bmatrix}\right) = \begin{bmatrix} out_{o1} \\ out_{o2} \end{bmatrix}$$

$$E = \left[\begin{bmatrix} out_{o1} \\ out_{o2} \end{bmatrix} - Ouptut\right]^{\mathrm{T}} \left[\begin{bmatrix} out_{o1} \\ out_{o2} \end{bmatrix} - Ouptut\right]$$

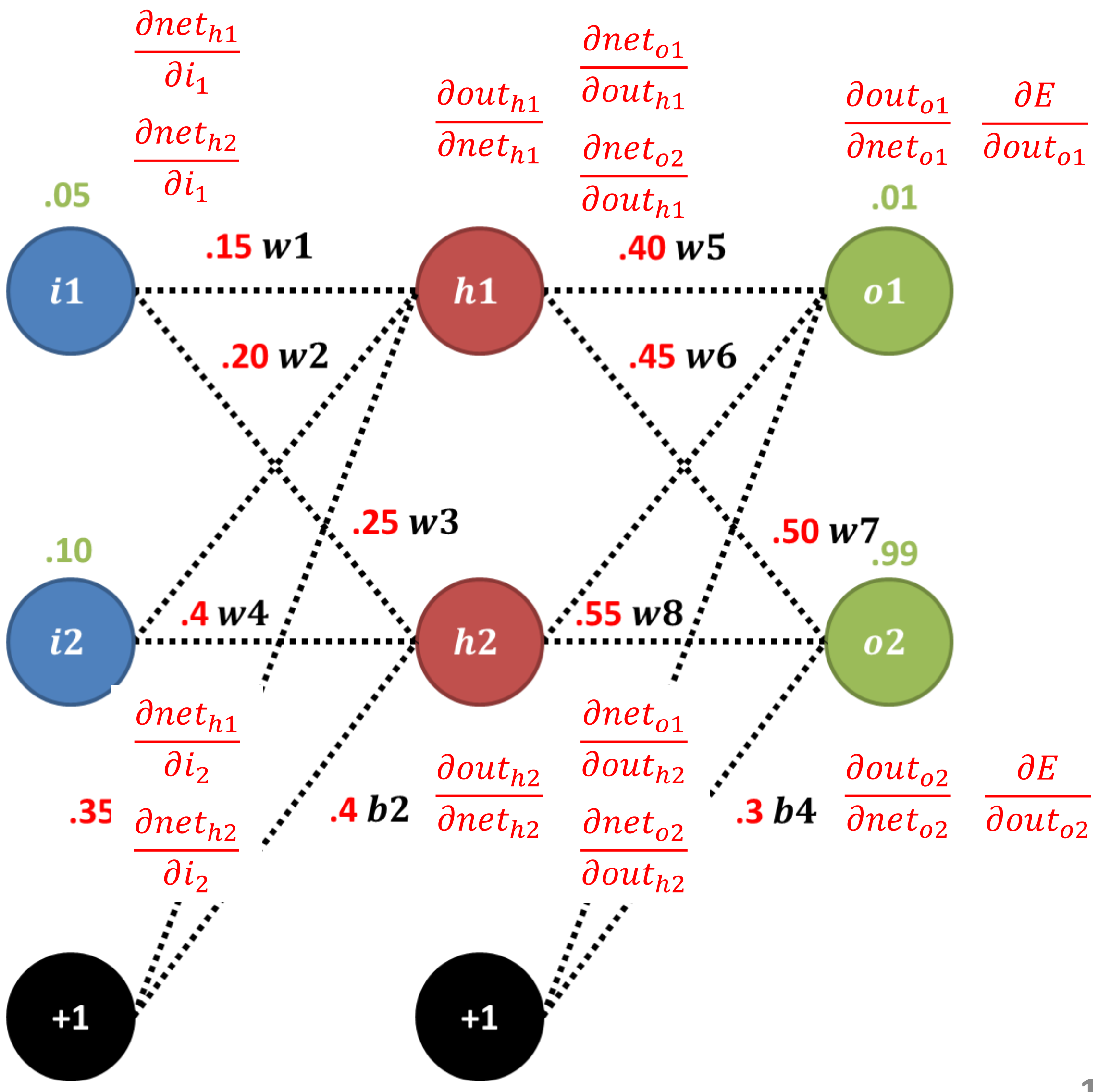
Backward Propagation (Matrix Operation)

$$O_{ON} = \begin{bmatrix} \frac{\partial out_{o1}}{\partial net_{o1}} & \frac{\partial out_{o1}}{\partial net_{o1}} \\ \frac{\partial out_{o2}}{\partial net_{o2}} & \frac{\partial out_{o2}}{\partial net_{o2}} \end{bmatrix} \qquad O_{EO} = \begin{bmatrix} \frac{\partial E}{\partial out_{o1}} & \frac{\partial E}{\partial out_{o1}} \\ \frac{\partial E}{\partial out_{o2}} & \frac{\partial E}{\partial out_{o2}} \end{bmatrix}$$

$$O_{NW} = \begin{bmatrix} \frac{\partial net_{o1}}{\partial w_5} & \frac{\partial net_{o2}}{\partial w_7} \\ \frac{\partial net_{o1}}{\partial w_6} & \frac{\partial net_{o2}}{\partial w_5} \end{bmatrix}$$

$$H_{ON} = \begin{bmatrix} \frac{\partial out_{h1}}{\partial net_{h1}} & \frac{\partial out_{h1}}{\partial net_{h1}} \\ \frac{\partial out_{h2}}{\partial net_{h2}} & \frac{\partial out_{h2}}{\partial net_{h2}} \end{bmatrix} \qquad H_{EO} = \begin{bmatrix} \frac{\partial net_{o1}}{\partial out_{h1}} & \frac{\partial net_{o1}}{\partial out_{h2}} \\ \frac{\partial net_{o2}}{\partial out_{h1}} & \frac{\partial net_{o2}}{\partial out_{h2}} \end{bmatrix}$$

$$H_{NW} = \begin{bmatrix} \frac{\partial net_{h1}}{\partial w_1} & \frac{\partial net_{h1}}{\partial w_3} \\ \frac{\partial net_{h1}}{\partial w_2} & \frac{\partial net_{h1}}{\partial w_4} \end{bmatrix} \qquad I_{EO} = \begin{bmatrix} \frac{\partial net_{h1}}{\partial i_1} & \frac{\partial net_{h1}}{\partial i_2} \\ \frac{\partial net_{h2}}{\partial i_1} & \frac{\partial net_{h2}}{\partial i_2} \end{bmatrix}$$



Backward Propagation (Matrix Operation)

$$O_{ON} = \begin{bmatrix} \frac{\partial out_{o1}}{\partial net_{o1}} & \frac{\partial out_{o1}}{\partial net_{o1}} \\ \frac{\partial out_{o2}}{\partial net_{o2}} & \frac{\partial out_{o2}}{\partial net_{o2}} \end{bmatrix}$$

$$O_{NW} = \begin{bmatrix} \frac{\partial net_{o1}}{\partial w_5} & \frac{\partial net_{o2}}{\partial w_7} \\ \frac{\partial net_{o1}}{\partial w_6} & \frac{\partial net_{o2}}{\partial w_8} \end{bmatrix}$$

$$O_{ON} = \begin{bmatrix} f(net_{o1})f(-net_{o1}) & f(net_{o1})f(-net_{o1}) \\ f(net_{o2})f(-net_{o2}) & f(net_{o2})f(-net_{o2}) \end{bmatrix}$$

$$O_{NW} = \begin{bmatrix} out_{h1} & out_{h1} \\ out_{h2} & out_{h2} \end{bmatrix}$$

$$E = J(\mathbf{w}, \mathbf{b}) = (o_1 - out_{o1})^2 + (o_2 - out_{o2})^2$$

$$out_{o1} = f(net_{o1}) = \frac{1}{1 + e^{-net_{o1}}} \qquad net_{o2} = w_7 out_{h1} + w_8 out_{h2} + b_4$$

$$net_{o1} = w_5 out_{h1} + w_6 out_{h2} + b_3 \qquad out_{o2} = f(net_{o2}) = \frac{1}{1 + e^{-net_{o2}}}$$

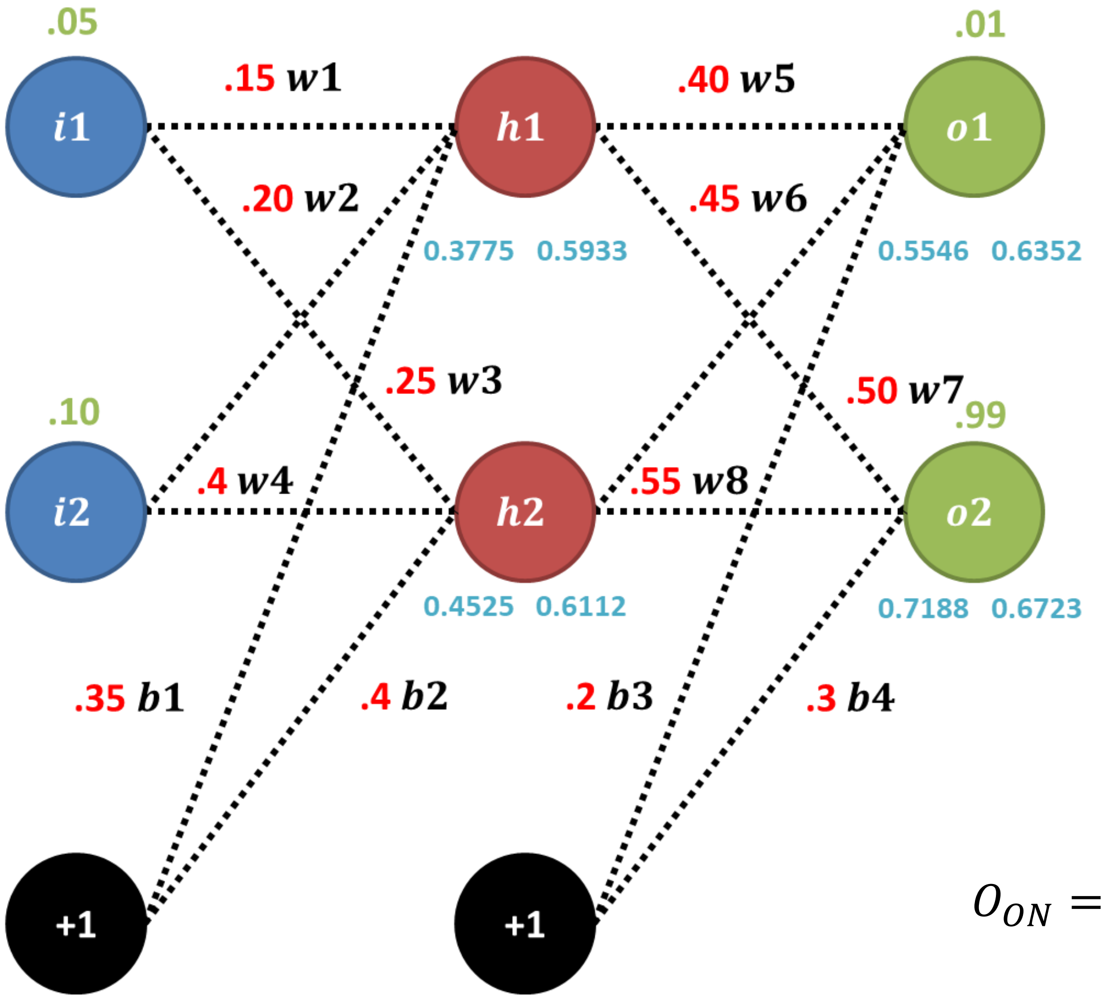
$$net_{h1} = w_1 i_1 + w_2 i_2 + b_1 \qquad net_{h2} = w_3 i_1 + w_4 i_2 + b_2$$

$$out_{h1} = f(net_{h1}) = \frac{1}{1 + e^{-net_{h1}}} \qquad out_{h2} = f(net_{h2}) = \frac{1}{1 + e^{-net_{h2}}}$$

$$O_{EO} = \begin{bmatrix} \frac{\partial E}{\partial out_{o1}} & \frac{\partial E}{\partial out_{o1}} \\ \frac{\partial E}{\partial out_{o2}} & \frac{\partial E}{\partial out_{o2}} \end{bmatrix}$$

$$O_{EO} = \begin{bmatrix} -2(o_1 - out_{o1}) & -2(o_1 - out_{o1}) \\ -2(o_2 - out_{o2}) & -2(o_2 - out_{o2}) \end{bmatrix}$$

Backward Propagation (Matrix Operation)



$$O_{EO} = \begin{bmatrix} \frac{\partial E}{\partial out_{o1}} & \frac{\partial E}{\partial out_{o1}} \\ \frac{\partial E}{\partial out_{o2}} & \frac{\partial E}{\partial out_{o2}} \end{bmatrix} = \begin{bmatrix} -2(o_1 - out_{o1}) & -2(o_1 - out_{o1}) \\ -2(o_2 - out_{o2}) & -2(o_2 - out_{o2}) \end{bmatrix}$$

$$O_{NW} = \begin{bmatrix} \frac{\partial net_{o1}}{\partial w_5} & \frac{\partial net_{o2}}{\partial w_7} \\ \frac{\partial net_{o1}}{\partial w_6} & \frac{\partial net_{o2}}{\partial w_8} \end{bmatrix} = \begin{bmatrix} out_{h1} & out_{h1} \\ out_{h2} & out_{h2} \end{bmatrix}$$

$$O_{ON} = \begin{bmatrix} \frac{\partial out_{o1}}{\partial net_{o1}} & \frac{\partial out_{o1}}{\partial net_{o1}} \\ \frac{\partial out_{o2}}{\partial net_{o2}} & \frac{\partial out_{o2}}{\partial net_{o2}} \end{bmatrix} = \begin{bmatrix} f(net_{o1})f(-net_{o1}) & f(net_{o1})f(-net_{o1}) \\ f(net_{o2})f(-net_{o2}) & f(net_{o2})f(-net_{o2}) \end{bmatrix}$$

Backward Propagation (Matrix Operation)

$$H_{ON} = \begin{bmatrix} \frac{\partial out_{h1}}{\partial net_{h1}} & \frac{\partial out_{h1}}{\partial net_{h1}} \\ \frac{\partial out_{h2}}{\partial net_{h2}} & \frac{\partial out_{h2}}{\partial net_{h2}} \end{bmatrix} = \begin{bmatrix} f(net_{h1})f(-net_{h1}) & f(net_{h1})f(-net_{h1}) \\ f(net_{h2})f(-net_{h2}) & f(net_{h2})f(-net_{h2}) \end{bmatrix}$$

$$H_{EO} = \begin{bmatrix} \frac{\partial net_{o1}}{\partial out_{h1}} & \frac{\partial net_{o1}}{\partial out_{h2}} \\ \frac{\partial net_{o2}}{\partial out_{h1}} & \frac{\partial net_{o2}}{\partial out_{h2}} \end{bmatrix} = \begin{bmatrix} w_5 & w_6 \\ w_7 & w_8 \end{bmatrix}$$

$$H_{NW} = \begin{bmatrix} \frac{\partial net_{h1}}{\partial w_1} & \frac{\partial net_{h1}}{\partial w_3} \\ \frac{\partial net_{h1}}{\partial w_2} & \frac{\partial net_{h1}}{\partial w_4} \end{bmatrix} = \begin{bmatrix} i_1 & i_1 \\ i_2 & i_2 \end{bmatrix}$$

$$E = J(\mathbf{w}, \mathbf{b}) = (o_1 - out_{o1})^2 + (o_2 - out_{o2})^2$$

$$out_{o1} = f(net_{o1}) = \frac{1}{1 + e^{-net_{o1}}} \qquad net_{o2} = w_7 out_{h1} + w_8 out_{h2} + b_4$$

$$net_{o1} = w_5 out_{h1} + w_6 out_{h2} + b_3 \qquad out_{o2} = f(net_{o2}) = \frac{1}{1 + e^{-net_{o2}}}$$

$$net_{h1} = w_1 i_1 + w_2 i_2 + b_1 \qquad net_{h2} = w_3 i_1 + w_4 i_2 + b_2$$

$$out_{h1} = f(net_{h1}) = \frac{1}{1 + e^{-net_{h1}}} \qquad out_{h2} = f(net_{h2}) = \frac{1}{1 + e^{-net_{h2}}}$$

$$I_{EO} = \begin{bmatrix} \frac{\partial net_{h1}}{\partial i_1} & \frac{\partial net_{h1}}{\partial i_2} \\ \frac{\partial net_{h2}}{\partial i_1} & \frac{\partial net_{h2}}{\partial i_2} \end{bmatrix} = \begin{bmatrix} w_1 & w_2 \\ w_3 & w_4 \end{bmatrix}$$

Backward Propagation (Matrix Operation)

$$\frac{\partial E}{\partial w_5} = \frac{\partial E}{\partial out_{o1}}\frac{\partial out_{o1}}{\partial net_{o1}}\frac{\partial net_{o1}}{\partial w_5} + \frac{\partial E}{\partial out_{o2}}\frac{\partial out_{o2}}{\partial net_{o2}}\frac{\partial net_{o2}}{\partial w_5} = \frac{\partial E}{\partial out_{o1}}\frac{\partial out_{o1}}{\partial net_{o1}}\frac{\partial net_{o1}}{\partial w_5}$$

$$W_{ih} = \begin{bmatrix} w_1 & w_3 \\ w_2 & w_4 \end{bmatrix}$$

$$\frac{\partial E}{\partial w_7} = \frac{\partial E}{\partial out_{o1}}\frac{\partial out_{o1}}{\partial net_{o1}}\frac{\partial net_{o1}}{\partial w_7} + \frac{\partial E}{\partial out_{o2}}\frac{\partial out_{o2}}{\partial net_{o2}}\frac{\partial net_{o2}}{\partial w_7} = \frac{\partial E}{\partial out_{o2}}\frac{\partial out_{o2}}{\partial net_{o2}}\frac{\partial net_{o2}}{\partial w_7}$$

$$\frac{\partial E}{\partial w_{ih}} = \begin{bmatrix} \frac{\partial E}{\partial w_5} & \frac{\partial E}{\partial w_7} \\ \frac{\partial E}{\partial w_6} & \frac{\partial E}{\partial w_8} \end{bmatrix} = \begin{bmatrix} \frac{\partial E}{\partial out_{o1}} & \frac{\partial E}{\partial out_{o1}} \\ \frac{\partial E}{\partial out_{o2}} & \frac{\partial E}{\partial out_{o2}} \end{bmatrix} .* \begin{bmatrix} \frac{\partial out_{o1}}{\partial net_{o1}} & \frac{\partial out_{o1}}{\partial net_{o1}} \\ \frac{\partial out_{o2}}{\partial net_{o2}} & \frac{\partial out_{o2}}{\partial net_{o2}} \end{bmatrix} .* \begin{bmatrix} \frac{\partial net_{o1}}{\partial w_5} & \frac{\partial net_{o2}}{\partial w_7} \\ \frac{\partial net_{o1}}{\partial w_6} & \frac{\partial net_{o2}}{\partial w_5} \end{bmatrix} = O_{EO} .* O_{ON} .* O_{NW}$$

$.*$: Element-wise multiplication

Backward Propagation (Matrix Operation)

$$\frac{\partial E}{\partial w_1} = \frac{\partial E}{\partial out_{o1}}\frac{\partial out_{o1}}{\partial net_{o1}}\frac{\partial net_{o1}}{\partial w_1} + \frac{\partial E}{\partial out_{o2}}\frac{\partial out_{o2}}{\partial net_{o2}}\frac{\partial net_{o2}}{\partial w_1}$$

$$W_{ho} = \begin{bmatrix} w_5 & w_7 \\ w_6 & w_8 \end{bmatrix}$$

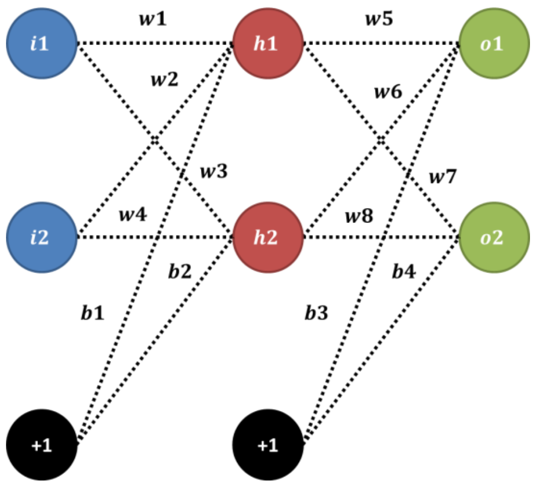
$$= \frac{\partial E}{\partial out_{o1}}\frac{\partial out_{o1}}{\partial net_{o1}}\frac{\partial net_{o1}}{\partial out_{h1}}\frac{\partial out_{h1}}{\partial net_{h1}}\frac{\partial net_{h1}}{\partial w_1} + \frac{\partial E}{\partial out_{o2}}\frac{\partial out_{o2}}{\partial net_{o2}}\frac{\partial net_{o2}}{\partial out_{h1}}\frac{\partial out_{h1}}{\partial net_{h1}}\frac{\partial net_{h1}}{\partial w_1}$$

$$\frac{\partial E}{\partial w_{ho}} = \begin{bmatrix} \frac{\partial E}{\partial w_1} & \frac{\partial E}{\partial w_3} \\ \frac{\partial E}{\partial w_2} & \frac{\partial E}{\partial w_4} \end{bmatrix} = (O_{EO}.* O_{ON} * H_{EO} + flipud(O_{EO}.* O_{ON} * H_{EO})).* H_{ON}.* H_{NW}$$

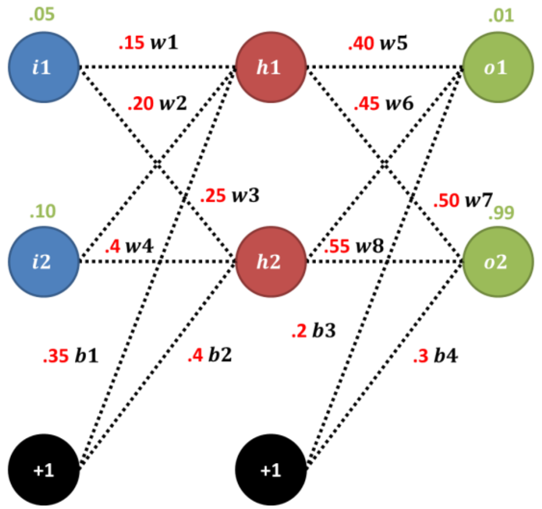
$$O_{EO}.* O_{ON} * H_{EO} = \begin{bmatrix} \frac{\partial E}{\partial out_{o1}} & \frac{\partial E}{\partial out_{o1}} \\ \frac{\partial E}{\partial out_{o2}} & \frac{\partial E}{\partial out_{o2}} \end{bmatrix} .* \begin{bmatrix} \frac{\partial out_{o1}}{\partial net_{o1}} & \frac{\partial out_{o1}}{\partial net_{o1}} \\ \frac{\partial out_{o2}}{\partial net_{o2}} & \frac{\partial out_{o2}}{\partial net_{o2}} \end{bmatrix} .* \begin{bmatrix} \frac{\partial net_{o1}}{\partial out_{h1}} & \frac{\partial net_{o1}}{\partial out_{h2}} \\ \frac{\partial net_{o2}}{\partial out_{h1}} & \frac{\partial net_{o2}}{\partial out_{h2}} \end{bmatrix} = \begin{bmatrix} \frac{\partial E}{\partial out_{o1}}\frac{\partial out_{o1}}{\partial net_{o1}}\frac{\partial net_{o1}}{\partial out_{h1}} & \frac{\partial E}{\partial out_{o1}}\frac{\partial out_{o1}}{\partial net_{o1}}\frac{\partial net_{o1}}{\partial out_{h2}} \\ \frac{\partial E}{\partial out_{o2}}\frac{\partial out_{o2}}{\partial net_{o2}}\frac{\partial net_{o2}}{\partial out_{h1}} & \frac{\partial E}{\partial out_{o2}}\frac{\partial out_{o2}}{\partial net_{o2}}\frac{\partial net_{o2}}{\partial out_{h2}} \end{bmatrix}$$

$$H_{ON}.* H_{NW} = \begin{bmatrix} \frac{\partial out_{h1}}{\partial net_{h1}} & \frac{\partial out_{h1}}{\partial net_{h1}} \\ \frac{\partial out_{h2}}{\partial net_{h2}} & \frac{\partial out_{h2}}{\partial net_{h2}} \end{bmatrix} .* \begin{bmatrix} \frac{\partial net_{h1}}{\partial w_1} & \frac{\partial net_{h1}}{\partial w_3} \\ \frac{\partial net_{h1}}{\partial w_2} & \frac{\partial net_{h1}}{\partial w_4} \end{bmatrix} = \begin{bmatrix} \frac{\partial out_{h1}}{\partial net_{h1}}\frac{\partial net_{h1}}{\partial w_1} & \frac{\partial out_{h1}}{\partial net_{h1}}\frac{\partial net_{h1}}{\partial w_3} \\ \frac{\partial out_{h2}}{\partial net_{h2}}\frac{\partial net_{h1}}{\partial w_2} & \frac{\partial out_{h2}}{\partial net_{h2}}\frac{\partial net_{h1}}{\partial w_4} \end{bmatrix}$$

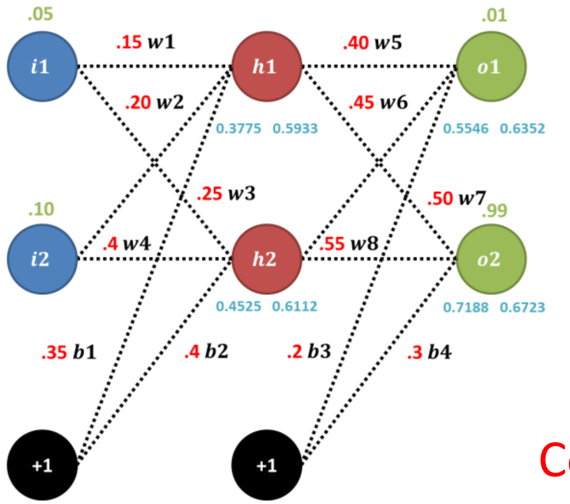
Summary of Neural Network



S1. Design Neural Network

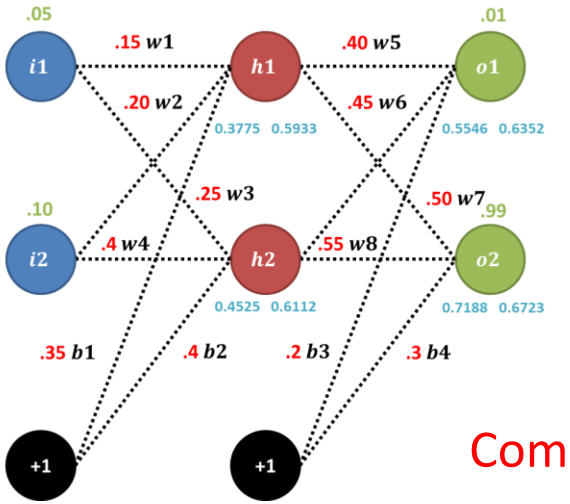


S2. Initialization of NN



Compute $J(\theta)$

S3. Forward Propagation



Compute $\frac{\partial}{\partial \theta_j} J(\theta)$

S4. Backward Propagation

$$\theta_{j+1} \leftarrow \theta_j - \alpha \frac{\partial}{\partial \theta_j} J(\theta)$$

S5. Update NN

$$\theta_{j+1} \leftarrow \theta_j - \alpha \frac{\partial}{\partial \theta_j} J(\theta)$$

Batch (Vanilla) Gradient Descent

$$\theta_{j+1} \leftarrow \theta_j - \alpha \frac{\partial}{\partial \theta_j} J(x^i, y^j; \theta)$$

Stochastic Gradient Descent

$$\theta_{j+1} \leftarrow \theta_j - \alpha \frac{\partial}{\partial \theta_j} J(x^{\{i:i+b\}}, y^{\{i:i+b\}}; \theta)$$

Mini-batch Gradient Descent