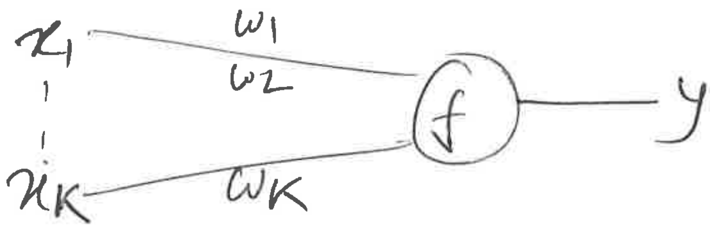
# Single unit neural network

0




$$u = \sum_{i=1}^{K} w_i x_i$$

$$u = W^T x$$

$$w^T = [w_1 \cdots w_K]$$

$$x = \begin{bmatrix} x_1 \\ \vdots \\ x_K \end{bmatrix}$$

$$y = f(u).$$

---

2 choices of $f$:

(1) unit step function (discrete).

$$f(u) = \begin{cases} 1 & \text{if } u > 0 \\ 0 & \text{otherwise.} \end{cases}$$

Perceptron

Perceptron algorithm

$$w^{(new)} = w^{(old)} - \eta (y - t) x$$

(2) $f(u)$ is the logistic function.

$$f(u) = \sigma(u) = \frac{1}{1 + e^{-u}}.$$

$E = \frac{1}{2}(t - y)^2$

$E = \frac{1}{2}(t - f(u))^2$

$= \frac{1}{2}(t - f(W^T x))^2$

~~Neural~~ (1-level Back propagation) ①

$$\boxed{\sigma(-u) = \frac{1}{1+e^{-(-u)}} = \frac{1}{1+e^{u}}} \qquad \sigma(u) = \frac{1}{1+e^{-u}}$$

$$1-\sigma(u) = 1 - \frac{1}{1+e^{-u}} = \frac{1+e^{-u}-1}{1+e^{-u}} = \frac{e^{-u}}{1+e^{-u}}$$

$$= \frac{1}{\frac{1}{e^{-u}}+1} = \frac{1}{1+e^{u}}$$

---

$$\sigma(u)\,\sigma(-u) = \frac{1}{(1+e^{-u})} \cdot \frac{1}{(1+e^{u})}$$

$$= \frac{1}{1+1/e^{u}} \cdot \frac{1}{1+e^{u}}$$

$$= \frac{e^{u}}{1+e^{u}} \cdot \frac{1}{1+e^{u}} = \frac{e^{u}}{(1+e^{u})^2}$$

$$= \frac{e^{u}/e^{2u}}{\frac{(1+e^{u})}{e^{u}}\frac{(1+e^{u})}{e^{u}}} = \frac{e^{-u}}{(1+e^{-u})(1+e^{-u})}$$

---

$$\frac{d\sigma(u)}{du} = \frac{d(1+e^{-u})^{-1}}{du} = \frac{-(1+e^{-u})^{-2}}{\cancel{\phi}}\frac{de^{-u}}{du}$$

$$= \cancel{-e^{u}} - (1+e^{-u})^{-2}(-)e^{-u}$$

$$= \frac{e^{-u}}{(1+e^{-u})^2}$$

$$= \sigma(u)\,\sigma(-u) = \boxed{\sigma(u)\,(1-\sigma(u))}$$

t – gold standard.

y – current output value.

Goal: Minimize $E = \frac{1}{2}(t-y)^2$

Note. $y = \sigma(u)$.

and $u = \sum_{i=0}^{K} w_i x_i$

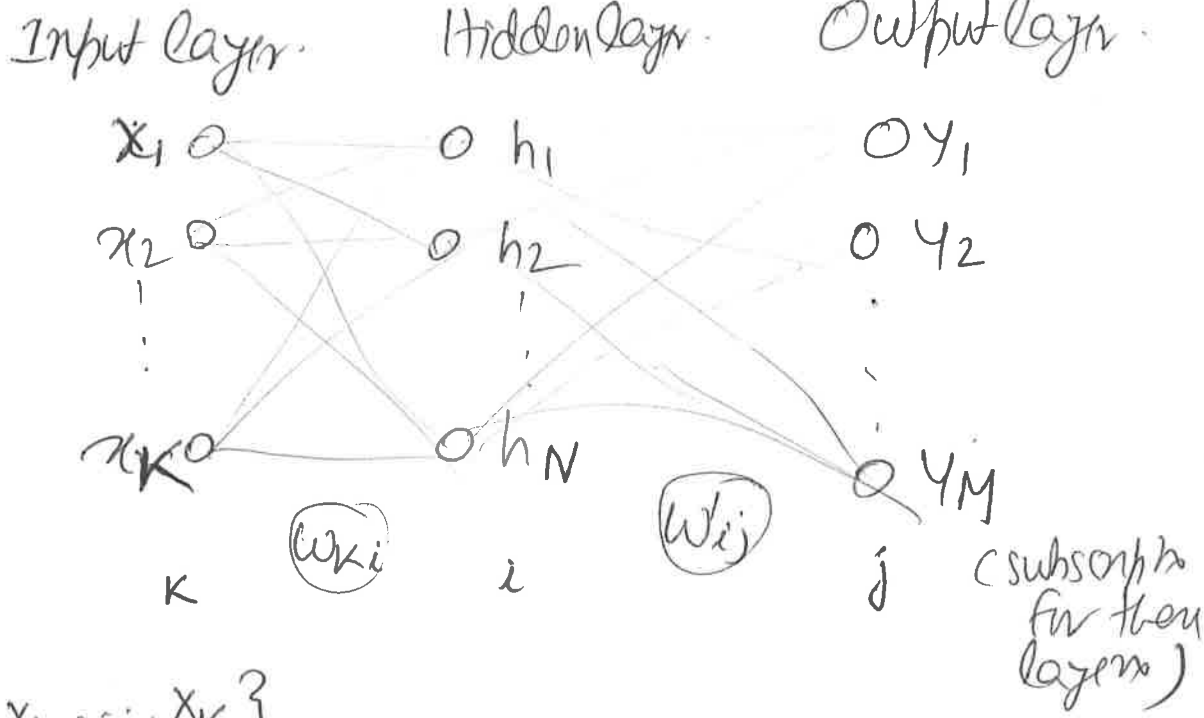
$$\frac{\partial E}{\partial w_i} = \frac{\partial E}{\partial y} \cdot \frac{\partial y}{\partial u} \frac{\partial u}{\partial w_i}$$

See Page 1

$$= (y-t) \cdot \boxed{y(1-y)} \cdot x_i$$

$$w^{(new)} = w^{(old)} - \eta (y-t) y (1-y) x$$

# Back propagation with multi-layer network




$\{x_k\} = \{x_1, \cdots, x_K\}$

$\{h_i\} = \{h_1, \cdots h_N\}$

$\{y_j\} = \{y_1, \cdots, y_M\}$

$u_i$ — net input of hidden layer units.

$u'_j$ — net input of output layer units.

---

$$h_i = \sigma(u_i) = \sigma\left(\sum_{k=1}^{K} w_{ki} x_k\right)$$

$$x \quad \begin{bmatrix} x_1 \\ \vdots \\ x_K \end{bmatrix} \qquad W = \begin{bmatrix} w_{11} & \cdots & w_{1N} \\ \vdots & & \\ w_{K1} & \cdots & w_{KN} \end{bmatrix}$$

$K \times N$

$$W^T = \begin{bmatrix} w_{11} & \cdots & w_{K1} \\ \vdots & & \vdots \\ w_{1N} & \cdots & w_{KN} \end{bmatrix} \qquad \begin{bmatrix} x_1 \\ \vdots \\ x_K \end{bmatrix}$$

$$= \begin{bmatrix} w_{11}x_1 + \cdots & w_{K1}x_K \\ \vdots & \\ w_{1i}x_1 + \cdots & w_{Ki}x_K \\ \vdots & \\ w_{1N}x_1 + \cdots & w_{KN}x_K \end{bmatrix} = \begin{bmatrix} u_1 \\ \vdots \\ u_i \\ \vdots \\ u_N \end{bmatrix}$$

$$\cancel{h_i} \quad u_i = \sum_{k=1}^{K} w_{ki} x_k$$

---

$$h \begin{bmatrix} h_1 \\ \vdots \\ h_N \end{bmatrix} \qquad W' = \begin{bmatrix} w'_{11} & \cdots & w'_{1M} \\ \vdots & & \\ w'_{N1} & \cdots & w'_{NM} \end{bmatrix}$$

$$N \times M$$

(5)

$$W'^{T} = \begin{bmatrix} w'_{11} & \cdots & w'_{N1} \\ \vdots & & \\ w'_{1M} & \cdots & w'_{NM} \end{bmatrix} \begin{bmatrix} h_1 \\ \vdots \\ h_N \end{bmatrix}$$

$$= \begin{bmatrix} w'_{11} h_1 + \cdots + w'_{N1} h_N \\ \vdots \\ w'_{1j} h_1 + \cdots + w'_{Nj} h_N \\ \vdots \\ w'_{1M} h_1 + \cdots + w'_{NM} h_N \end{bmatrix} = \begin{bmatrix} u'_1 \\ \vdots \\ u'_j \\ \vdots \\ u_M \end{bmatrix}$$

$$u'_j = w'_{1j} h_1 + \cdots + w'_{Nj} h_N$$

$$= \sum_{i=1}^{N} w'_{ij} h_i$$

$$y_j = 6(u'_j) = 6\left( \sum_{i=1}^{N} w'_{ij} h_i \right)$$

Squared sum error function

$$E(x, t, W, W') = \frac{1}{2}\sum_{j=1}^{M}(y_j - t_j)^2$$

$x$ input. $\{x_1, \cdots x_K\}$

$t = \{t_1 \cdots t_M\}$ gold standard labels of output.

$W = \{w_{ki}\}$ $K \times N$ matrix (input to hidden)

$W' = \{w'_{ij}\}$ $N \times M$ matrix. (hidden to output).

---

Need to find update equation for $w_{ki}$ & $w'_{ij}$

---

$$\frac{\partial E}{\partial w_{ki}} = \frac{\partial E}{\partial u_i}\frac{\partial u_i}{\partial w_{ki}}$$

$$= \boxed{\frac{\partial E}{\partial h_i}}\frac{\partial h_i}{\partial u_i}\frac{\partial u_i}{\partial w_{ki}}$$

$$= \boxed{\sum_{j=1}^{M} EI'_j \cdot w_{ij}} \times h_i(1-h_i) \times x_k$$

See Page 5. See Page 1. See Page 4.

$EI_i$

$= EI_i \cdot x_k.$

$$w_{ki}^{(new)} = w_{ki}^{(old)} - \eta EI_i \cdot x_k.$$

$$\frac{\partial E}{\partial y_j} = \frac{1}{2}\times 2(y_j - t_j) = y_j - t_j$$

$$\frac{\partial E}{\partial u_j'} = \frac{\partial E}{\partial y_j}\frac{\partial y_j}{\partial u_j'}$$

$$= \underbrace{(y_j - t_j)\, y_j(1 - y_j)}_{EI'_j}$$

From below.

$$\frac{\partial E}{\partial w'_{ij}} = \frac{\partial E}{\partial u_j'}\frac{\partial u_j'}{\partial w'_{ij}} = EI'_j \times h_i$$

$$w'^{(new)}_{ij} = w'^{(old)}_{ij} - \eta \frac{\partial E}{\partial w'_{ij}}$$

Computing $\frac{\partial E}{\partial h_i}$

$$= \frac{\partial \left( \frac{1}{2} \sum_{j=1}^{M} (y_j - t_j)^2 \right)}{\partial h_i}$$

$$= \frac{\partial \left( \frac{1}{2} \sum_{j=1}^{M} ( \sigma(u_j') - t_j)^2 \right)}{\partial h_i}$$

$$= \frac{\partial \left( \frac{1}{2} \sum_{j=1}^{M} \left( \sigma\left( \sum_{l=1}^{N} w'_{lj} h_l \right) - t_j \right)^2 \right)}{\partial h_i}$$

$$= \frac{\partial}{\partial h_i} \left( \frac{1}{2} (\sigma(u_1') - t_1)^2 + \frac{1}{2} (\sigma(u_2') - t_2)^2 + \dots \frac{1}{2} (\sigma(u_M') - t_M)^2 \right)$$

Because of this

$$\boxed{\frac{\partial E}{\partial h_i}} = \sum_{j=1}^{M} \frac{\partial E}{\partial u_j'} \frac{\partial u_j'}{\partial h_j}$$

$$= \sum_{j=1}^{M} EI_j' \cdot w'_{ij}$$