lecture 10. Feedforward Neural Networks

target.

$$x \rightarrow \boxed{f} \rightarrow \underline{\hat{y}} \longleftrightarrow y$$

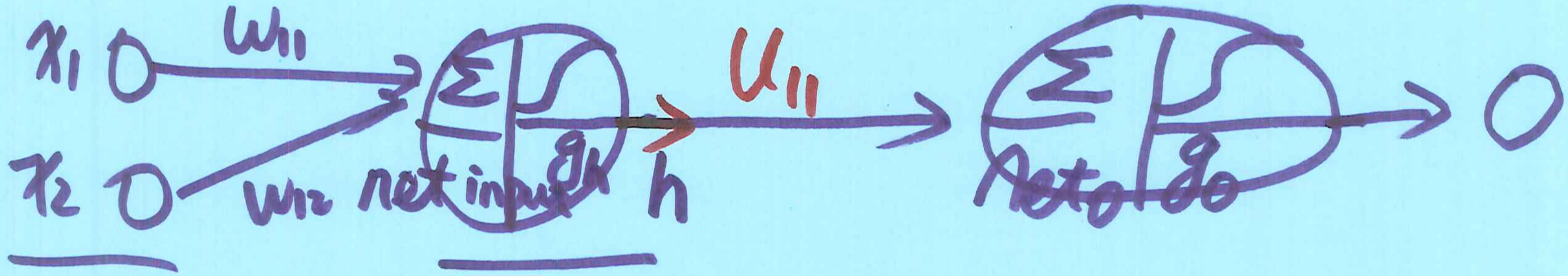
① Linear model : $f = w^T x$

② model : $f = w^T x$

③ Loss function : $L(w) = \sum_{i=1}^{m} (\hat{y}_i - y_i)^2$

④ Optimization : find out / learn w that can minimize $L(w)$

1. Model of Neural Network.



input

hidden layer

output layer

g_h: g_o: activation function

1.1. Hidden layer / Units / Nodes.

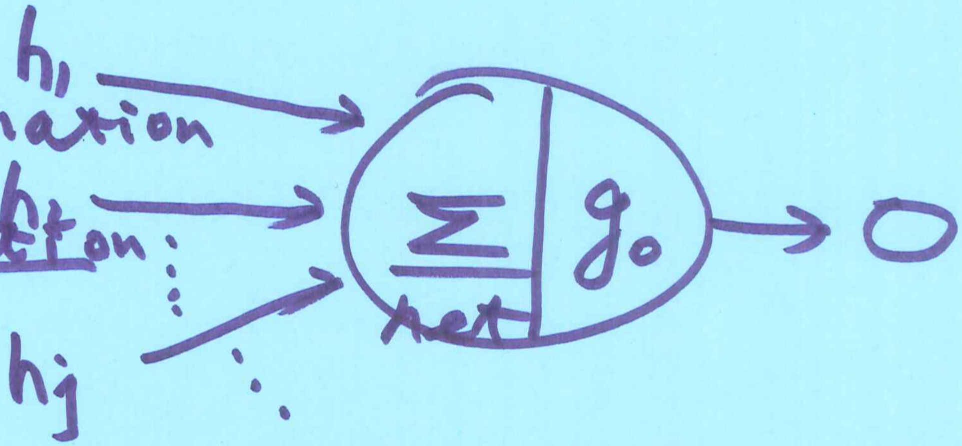
1) provides / extracts a set of features from data

Output of ~~the~~ a hidden node:

$h = \underline{g_h(net)}$ $\underline{net = x_1 w_{11} + x_2 \cdot w_{12}}$

1.2 Output Nodes.

1) provide transformation to the approximation.



$$O = g_o(net), \quad net = u_{11} h_1 + \cdots + u_{1j} h_j \cdots$$

$$= \sum_{j=1}^{n_h} u_{1j} \cdot h_j$$

n_h is the # of hidden nodes.

2). activation function.

① $g_o(x) = \text{sigm}(x)$

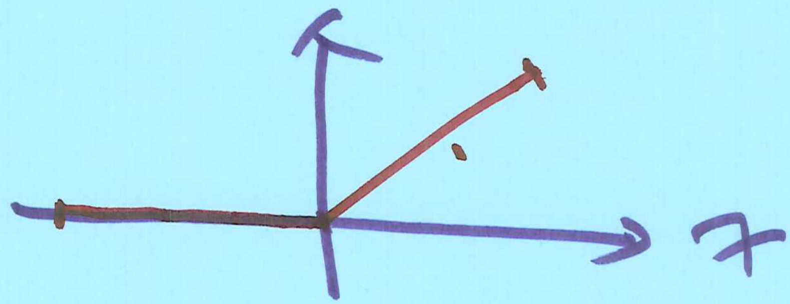
② linear units.

③ $\text{softmax}(x_i) = \dfrac{\exp(x_i)}{\sum_j \exp(x_j)}$

2) Activation function.

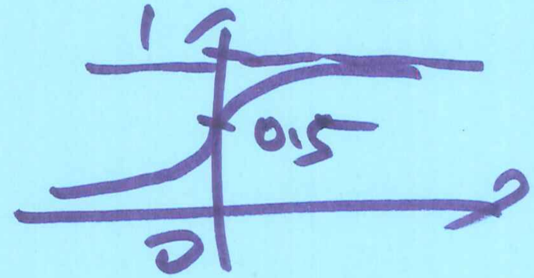
① Rectified linear units. (Relu)

$$g(\underline{x}) = \max\{0, x\} \quad x \in R.$$

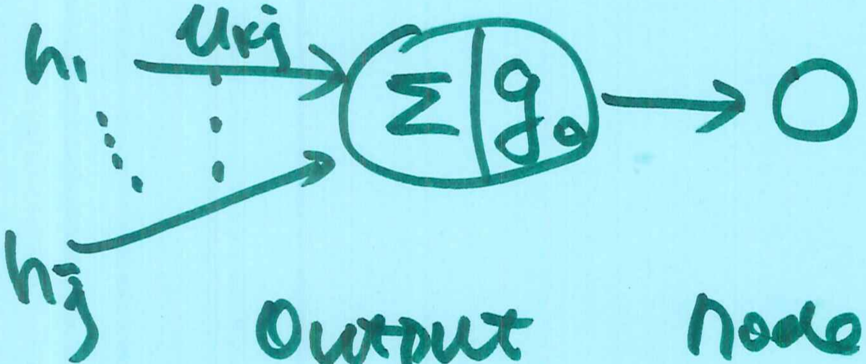


② sigmoid.

$$\operatorname{sigm}(x) = \frac{1}{1+e^{-x}}$$



(3) Hidden - to - output functions



Output node 1:

$$O_1 = g_o(net)$$

$$net = \sum_j u_{1j} \cdot h_j$$

Output node k:

$$O_k = g_o(net)$$

$$net = \sum_j (u_{kj} \cdot h_j)$$

Model for the NN (1st page)

$$O = g_o(net_o)$$
$$= g_o(U_{11} \cdot \underline{h})$$
$$= g_o(U_{11} \cdot g_h(net_h))$$
$$\boxed{= g_o(U_{11} \cdot g_h(x_1 \cdot w_{11} + x_2 \cdot w_{12}))}$$

Model 1.

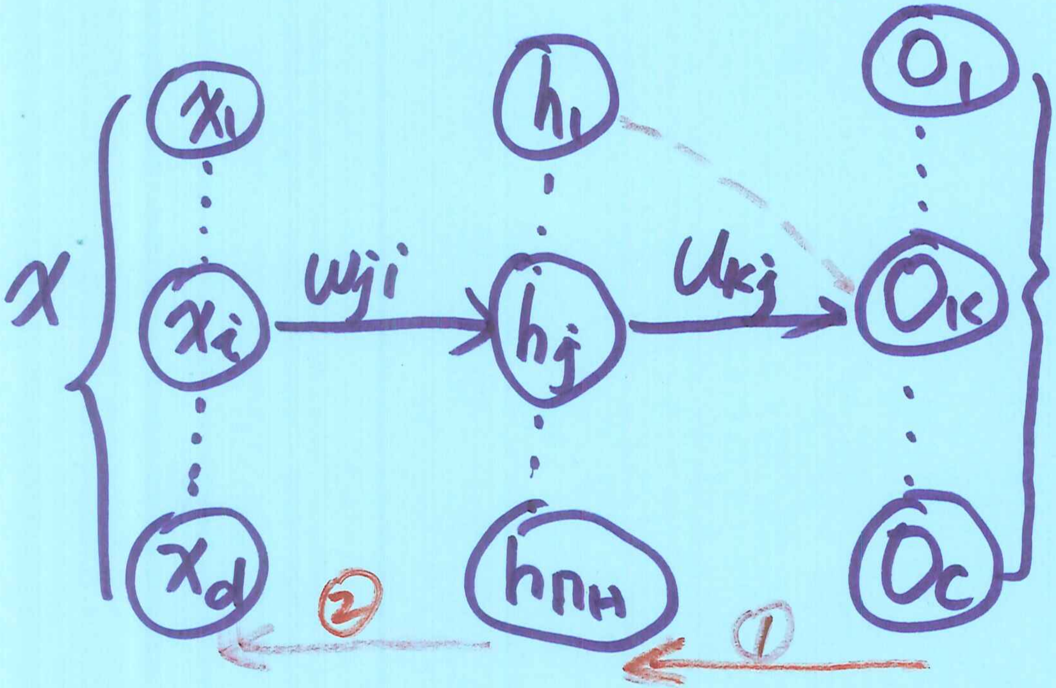
2. Loss function

$$x \xrightarrow{W} \bigcirc \xrightarrow{h_1 \ U} \bigcirc \rightarrow \boxed{O \longleftrightarrow y}$$

$$Loss(w, u) = \frac{1}{2} \sum_{i=1}^{m} (O_i - y_i)^2$$

Lecture 11.

Neural Network Optimization



$h_j = g_H(net_j)$; $net_j = \sum_{i=1}^{d} w_{ji} x_i$

$O_k = g_o(net_k)$; $net_k = \sum_{j=1}^{n_H} u_{kj} \cdot h_j$

$O \leftrightarrow y$ model:

target

Loss on a single sample:

$L(w, u) = \frac{1}{2} \sum_{k=1}^{c} (y_k - O_k)^2$

1. SGD to learn w. and u.

for ep in range(max_epoch):

for a single sample $(x, y) \in (X, Y)$:

$\forall$. $i \in [1, d]$, $j \in [1, n_H]$, $k \in [1, c]$

$w_{ji} = w_{ji} - \varepsilon \cdot \boxed{\nabla_{w_{ji}} L(w, u)}$

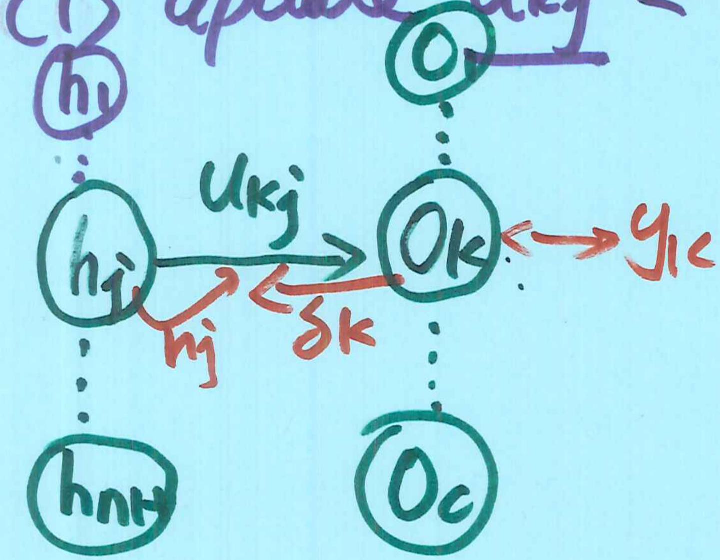
$u_{kj} = u_{kj} - \varepsilon \boxed{\nabla_{u_{kj}} L(w, u)}$

1.1 Backpropagation Algorithm
(BP)
: compute the two gradients.

→ update Hidden-to-output weights
$k=1,\cdots c$. $u_{k1}, \cdots u_{kj}, \cdots, u_{kn_H}$ } ①

→ update input-to-Hidden weights
$j=1,\cdots,n_H$: $w_{j1},\cdots w_{ji}, \cdots w_{jd}$ } ②

(1) update $u_{kj} \leftarrow \nabla_{u_{kj}} L(w.u)$



$$L(w,u) = \frac{1}{2}\sum_{k=1}^{c}(y_k - O_k)^2$$

$$O_k = g_o(net_k);\quad net_k = \sum_{j=1}^{n_H} u_{kj}\cdot h_j$$

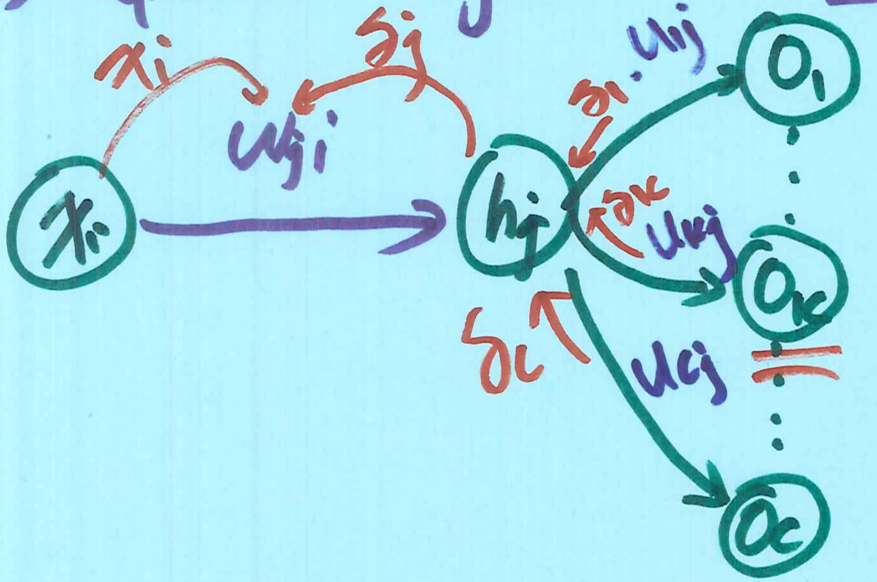
$$\nabla_{u_{kj}} L(w.u) = \frac{\partial L}{\partial O_k}\cdot\frac{\partial O_k}{\partial net_k}\cdot\frac{\partial net_k}{\partial u_{kj}}$$

$$= \underbrace{-(y_k - O_k)\cdot g_o'}_{\delta_k}\cdot h_j$$

$$= \delta_k \cdot h_j$$

δ_k: feed back from k-th output

(2) update $w_{ji} \leftarrow \nabla_{w_{ji}} L(w,u)$



$$L(w,u) = \frac{1}{2}\sum_{k=1}^{c}(y_k - O_k)^2$$

$$O_k = g_O(net_k); \quad net_k = \sum_{j=1}^{n_H} u_{kj}\cdot h_j$$

$$h_j = g_H(net_j); \quad net_j = \sum_{i=1}^{d} w_{ji}\cdot x_i$$

$$\nabla_{w_{ji}} L(w,u) = \overbrace{\sum_{k=1}^{c}\left[\frac{\partial L}{\partial O_k}\cdot\frac{\partial O_k}{\partial net_k}\cdot\frac{\partial net_k}{\partial h_j}\right]}^{\partial L/\partial h_j}\cdot\frac{\partial h_j}{\partial net_j}\cdot\frac{\partial net_j}{\partial w_{ji}}$$

$$= \frac{\partial L}{\partial h_j}\cdot\frac{\partial h_j}{\partial net_j}\cdot\frac{\partial net_j}{\partial w_{ji}}$$

$$= \underbrace{\left[\sum_{k=1}^{c}\underbrace{-(y_k - O_k)\cdot g_O'}_{\delta_k}\cdot u_{kj}\right]\cdot g_H'}_{\delta_j}\cdot x_i$$

$$= \delta_j \cdot x_i$$

$$\nabla_{u_{kj}} L(w, u) = \delta_k \cdot h_j \;:\; \delta_k = -(y_k - o_k)\cdot g_o'$$

$$\nabla_{w_{ji}} L(w, u) = \delta_j \cdot x_i \;:\; \delta_j = \left[\sum_{k=1}^{c} -(y_k - o_k)\cdot g_o' \cdot u_{kj}\right] \cdot g_H'$$