<Part 2>

[Ch01.]

[CH02.]

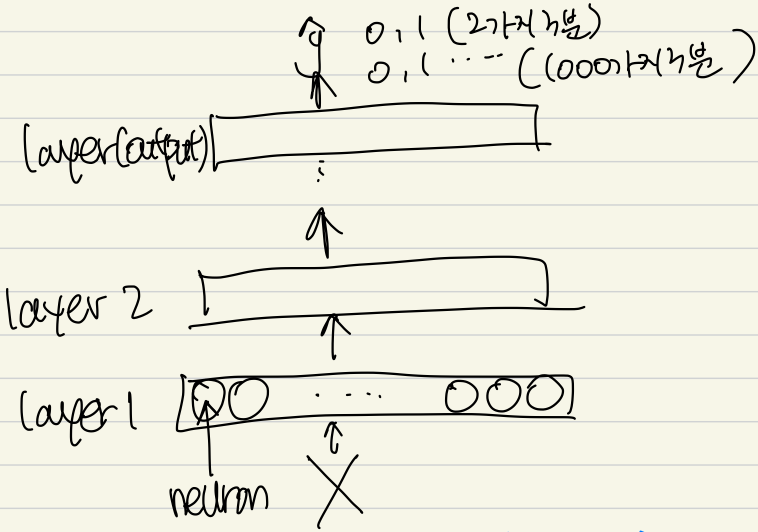
01, 02, 03, 04

CH02_01

• Dense Layers

- <Neuron Vectors and Layers>

A.N : $(\vec{x})^T \rightarrow$ (affine | activation) $\rightarrow a$



• 뉴런 묶음 (집합) : Layer : $\vec{\nu}$: (ν_1 ν_2 ⋯ ν_n) ⇒ filter들의 묶음 (각각 다른 작용)

$\vec{w}_1, b_1$ $\vec{w}_2, b_2$ $\vec{w}_n, b_n$

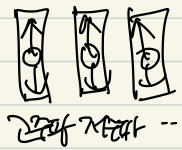
parametric function

• $(\vec{x})^T \cdot \vec{w} + b$: 하나의 필터링이라 생각가능.

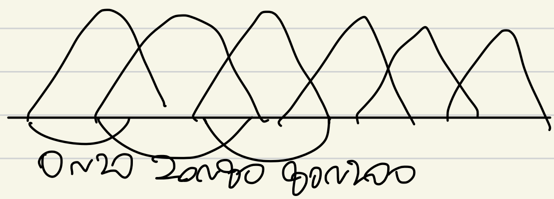
• Layer 속 AN들은 각각 다른값을 가지고있는 parametric function 이기 때문에, 같은 input: $(\vec{x})^T$를 받아도 다른 a들을 뽑아내게 된다.

• filter bank

- e.g) 음악들을 때 equalizer 필터 조절.



- e.g) MFCC

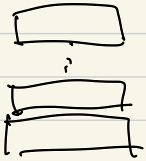


→ 7개의 필터

↓

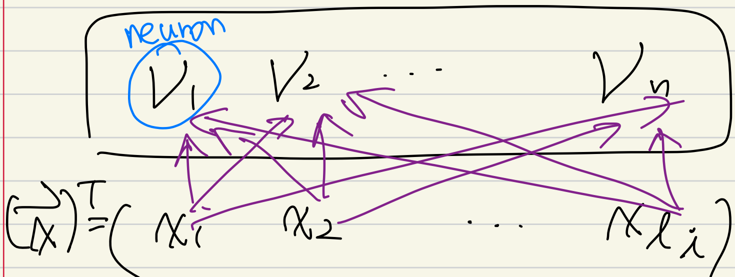
↳ Filterbank : 필터 묶음.

∴ → 하나의 레이어를 filter bank 라고 생각가능.

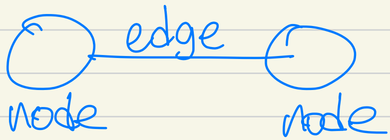


Deep Learning 구조 → correlation filter를 묶어서 만든 cascaded filter bank 구조라고 볼수있다.
(병렬)

• Dense Layer : 모든 뉴런에 모든 x가 연결된 것.



조밀하다 Dense Layer
→ 전부연결.

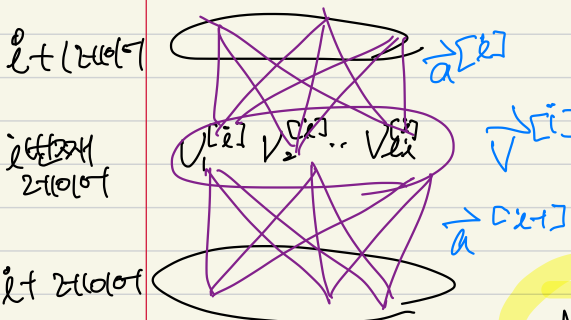


: Network 구조.

• Layer를 parametric 해야만 layer라고 부르는 것은 아님.
(사바사)

<Notation>

• i번째 layer가 dense layer라고 가정해 보자.



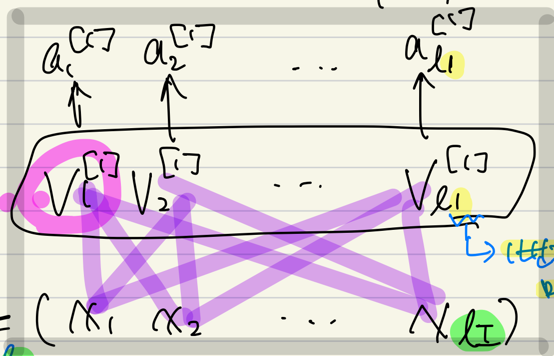
$\vec{v}^{[i]}(\vec{a}^{[i-1]})$ → 전후 layer의 뉴런들과 전부연결

Layer : $L^{[i]}$
Neuron Vector : $\vec{v}^{[i]}$
# of Neurons : l_i (l for length)

CH02_02

- The First Dense Layer

- <Params of Dense layer>



$\vec{\nu}^{[1]}((\vec{x})^T)$

l_1 → 1번째 layer의 마지막 뉴런이라는 표기.

$(\vec{x})^T = (x_1 \; x_2 \; \cdots \; x_{l_I}) \in \mathbb{R}^{1 \times l_I}$

$$\vec{w}_1^{[1]} = \begin{pmatrix} w_{1,1}^{[1]} \\ w_{2,1}^{[1]} \\ \vdots \\ w_{l_I,1}^{[1]} \end{pmatrix} \in \mathbb{R}^{l_I \times 1}, \quad b_1^{[1]} \in \mathbb{R}$$

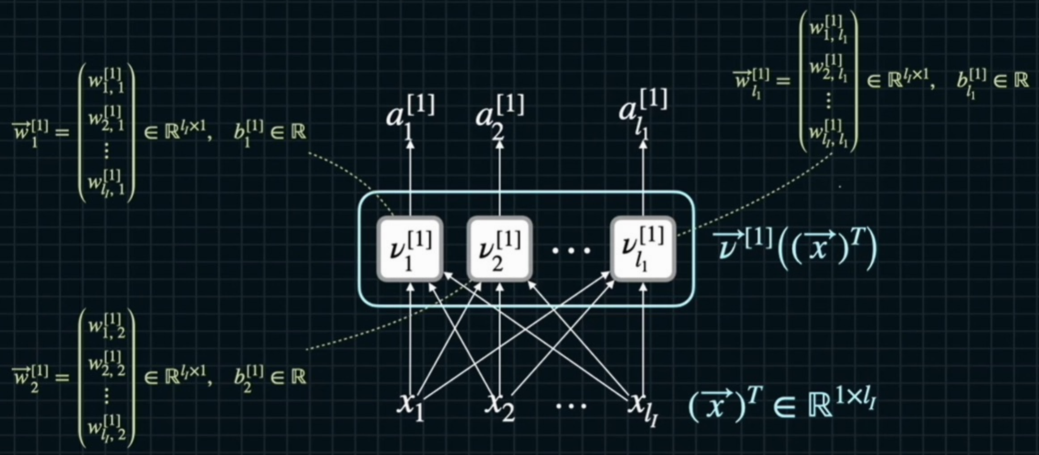
$$\rightarrow \nu_1^{[1]}\left((\vec{x})^T; \vec{w}_1^{[1]}, b_1^{[1]}\right)$$

☆ Input이 l_I개만큼 있기때문에 레이어속 한 뉴런은 weight를 l_I개만큼 가지게 된다.

☆ 한 레이어에 뉴런이 l_1개 만큼 있으면 $\vec{w}$ 자체도 l_1개 있다.

Lecture.2
Dense Layers

- Params of Dense Layer



⇒ 머리에 넣어두기 ☆☆

$$\cdot\ W^{[1]} = \overbrace{\begin{pmatrix} \uparrow & \uparrow & & \uparrow \\ \vec{w}_1^{[1]} & \vec{w}_2^{[1]} & \cdots & \vec{w}_{l_1}^{[1]} \\ \downarrow & \downarrow & & \downarrow \end{pmatrix}}^{l_1} \Big\} l_I$$

$W^{[1]} \rightarrow \mathbb{R}^{l_I \times l_1}$

$$= \begin{pmatrix} w_{1,1}^{[1]} & w_{1,2}^{[1]} & \cdots & w_{1,l_1}^{[1]} \\ w_{2,1}^{[1]} & w_{2,2}^{[1]} & \cdots & w_{2,l_1}^{[1]} \\ \vdots & \vdots & \ddots & \vdots \\ w_{l_I,1}^{[1]} & w_{l_I,2}^{[1]} & \cdots & w_{l_I,l_1}^{[1]} \end{pmatrix} \in \mathbb{R}^{l_I \times l_1}$$

→ 차원데이터 (input 수)

→ filter bank 안 filter 수.

$$\cdot\ (\vec{b}^{[1]})^T = (b_1^{[1]}\ b_2^{[1]}\ \cdots\ b_{l_1}^{[1]}) \in \mathbb{R}^{1 \times l_1}$$

→ Row vector : broadcasting 될거이여서

- <FP of Dense Layer>

하나 : $a_i^{[1]} = g((\vec{x})^T \vec{w}_i^{[1]} + b_i^{[1]}) \quad 1 \le i < l_1$

$\hookrightarrow \nu_i^{[1]}((\vec{x})^T; \vec{w}_i^{[1]}, b_i^{[1]})$

$= g(x_1 w_{1,i}^{[1]} + x_2 w_{2,i}^{[1]} \cdots + x_I w_{I,i}^{[1]} + b_i^{[1]})$

$\rightarrow (\vec{x})^T \cdot \vec{w}_i^{[1]} + b_i^{[1]}$

$\mathbb{R}^{1\times l_1} \qquad \mathbb{R}^{1\times l_I} \cdot \mathbb{R}^{l_I \times l_1} + \mathbb{R}^{1\times l_1} \rightarrow \mathbb{R}^{1\times l_1}$

전체 : $(\vec{a}^{[1]})^T = g((\vec{x})^T \cdot W^{[1]} + (\vec{b}^{[1]})^T)$

$= g\left((\vec{x})^T \begin{bmatrix} \uparrow & \uparrow & & \uparrow \\ \vec{w}_1^{[1]} & \vec{w}_2^{[1]} & \cdots & \vec{w}_{l_1}^{[1]} \\ \downarrow & \downarrow & & \downarrow \end{bmatrix} + (b_1^{[1]} \ b_2^{[1]} \cdots b_{l_1}^{[1]}) \right)$

$(x_1 \ x_2 \cdots x_{l_I})$

$= \left(g((\vec{x})^T \cdot \vec{w}_1^{[1]} + b_1^{[1]}) \quad g((\vec{x})^T \cdot \vec{w}_2^{[1]} + b_2^{[1]}) \cdots \quad g((\vec{x})^T \cdot \vec{w}_{l_1}^{[1]} + b_{l_1}^{[1]}) \right)$

$= \left(\nu_1^{[1]}((\vec{x})^T) \quad \nu_2^{[1]}((\vec{x})^T) \cdots \nu_{l_1}^{[1]}((\vec{x})^T) \right)$

☆ Row입력 → Row vector 출력.

$(\vec{x})^T \qquad \vec{a}^{[1]}$

04.02_07.

- ## Generalized_Dense_Layers.

<Dimensions of Dense Layer>

other layer

$\uparrow$ input

$a_1^{[1]} \quad a_2^{[1]} \quad \cdots \quad a_{l_1}^{[1]}$ $\qquad (\vec{a})^T \in \mathbb{R}^{1\times l_1}$

$\uparrow$ output

$\nu_1^{[1]} \quad \nu_2^{[1]} \quad \cdots \quad \nu_{l_1}^{[1]}$ $\qquad W^{[1]} \in \mathbb{R}^{l_I \times l_1}, (\vec{b}^{[1]})^T \in \mathbb{R}^{1\times l_1}$

$\uparrow$ input

$x_1 \quad x_2 \quad \cdots \quad x_{l_I}$ $\qquad (\vec{x})^T \in \mathbb{R}^{1\times l_I}$

$\Rightarrow l_1(l_I+1)$ 개의 param 생성

<The Second Dense Layer>

$a_1^{[2]} \quad a_2^{[2]} \quad \cdots \quad a_{l_2}^{[2]}$ $\qquad (\vec{a}^{[2]})^T \in \mathbb{R}^{1\times l_2}$

$\vec{\nu}^{[2]}((\vec{a}^{[1]})^T)$ $\quad \nu_1^{[2]} \quad \nu_2^{[2]} \quad \cdots \quad \nu_{l_2}^{[2]}$ $\qquad W^{[2]} \in \mathbb{R}^{l_1\times l_2}, (\vec{b}^{[2]})^T \in \mathbb{R}^{1\times l_2}$

$\uparrow$

$(\vec{a}^{[1]})^T \in \mathbb{R}^{1\times l_1}$

$\Rightarrow l_2(l_1+1)$개의 param 생성.

$$\vec{w}_i^{[2]} = \begin{pmatrix} w_{1,i} \\ w_{2,i} \\ \vdots \\ w_{l_1,i} \end{pmatrix} \in \mathbb{R}^{l_1\times 1}, \; b_i^{[2]} \in \mathbb{R}$$

$1 \le i \le l_2$

<FP of The Second Dense Layer>

$$a_i^{[2]} = g\left((\vec{a}^{[1]})^T \vec{w}_i^{[2]} + b_i^{[2]} \right) \qquad 1 \le i \le l_2$$

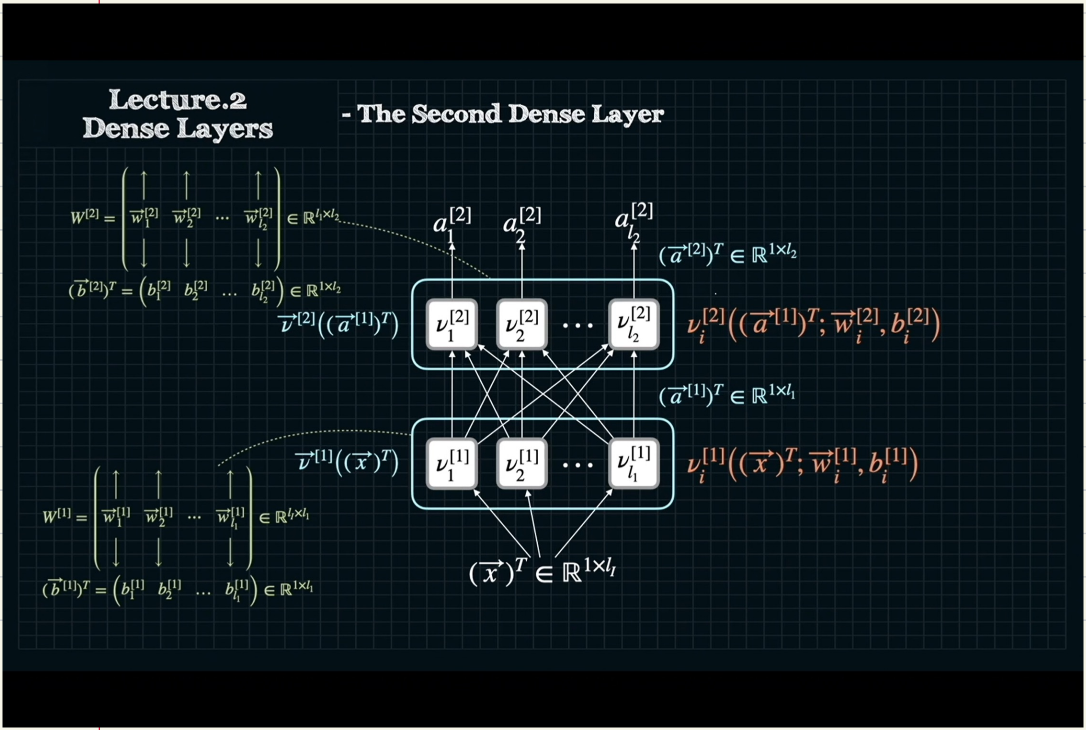
$$\nu_i^{[2]}\left((\vec{a}^{[1]})^T ; \vec{w}_i^{[2]}, b_i^{[2]} \right)$$

$$\overset{\mathbb{R}^{1\times l_2}}{(\vec{a}^{[2]})^T} = g\left(\overset{\mathbb{R}^{1\times l_1}}{(\vec{a}^{[1]})^T} \overset{\mathbb{R}^{l_1\times l_2}}{W^{[2]}} + \overset{\mathbb{R}^{1\times l_2}}{(\vec{b}^{[2]})^T} \right)$$

$$= g\left((\vec{a}^{[1]})^T \begin{pmatrix} \uparrow & \uparrow & & \uparrow \\ \vec{w}_1^{[2]} & \vec{w}_2^{[2]} & \cdots & \vec{w}_{l_2} \\ \downarrow & \downarrow & & \downarrow \end{pmatrix} + \begin{pmatrix} b_1^{[2]} & b_2^{[2]} & \cdots & b_{l_2}^{[2]} \end{pmatrix} \right)$$

$$= \left(\nu_1^{[2]}((\vec{a}^{[1]})^T) \quad \nu_2^{[2]}((\vec{a}^{[1]})^T) \quad \cdots \quad \nu_{l_2}^{[2]}((\vec{a}^{[1]})^T) \right)$$

$$= \left(a_1^{[2]} \ a_2^{[2]} \ \cdots \ a_{l_2}^{[2]} \right) \ \mathbb{R}^{1\times l_2}$$



< Generalized Dense Layer >

$$\nu_1^{[i+1]} \quad \nu_2^{[i+1]} \quad \cdots \quad \nu_{l_{i+1}}^{[i+1]}$$

$\uparrow \left(\vec{a}^{[i]}\right)^T \in R^{1 \times l_i}$

$$\nu_1^{[i]} \quad \nu_2^{[i]} \quad \cdots \quad \nu_{l_i}^{[i]}$$

$\vec{\nu}^{[i]}\left(\left(\vec{a}^{[i-1]}\right)^T\right)$

$\uparrow \left(\vec{a}^{[i-1]}\right)^T \in R^{1 \times l_{i-1}}$

$$\nu_1^{[i-1]} \quad \nu_2^{[i-1]} \quad \cdots \quad \nu_{l_{i-1}}^{[i-1]}$$

$$\vec{w}_1^{[i]} = \begin{pmatrix} w_{1,1} \\ w_{2,1} \\ \vdots \\ w_{l_{i-1},1} \end{pmatrix} \in R^{l_{i-1} \times 1}, \quad b_1^{[i]} \in R$$

$\vdots$

$$\vec{w}_{l_i}^{[i]} = \begin{pmatrix} w_{1,l_i} \\ w_{2,l_i} \\ \vdots \\ w_{l_{i-1},l_i} \end{pmatrix} \in R^{l_{i-1} \times 1}, \quad b_{l_i}^{[i]} \in R$$

앞 Data의 차원

$W^{[i]} \in R^{l_{i-1} \times l_i}$ → i번째 layer의 뉴런수

$\left(\vec{b}^{[i]}\right)^T \in R^{1 \times l_i}$ → 뉴런이어야 한다.

$$a_j^{[i]} = g\left(\left(\vec{a}^{[i-1]}\right)^T \vec{w}_j^{[i]} + b_j^{[i]}\right) \quad 1 \le j \le l_i$$

$$= g\left(a_1^{[i-1]} w_{1,j}^{[i]} + a_2^{[i-1]} w_{2,j}^{[i]} + \cdots a_{l_{i-1}}^{[i-1]} w_{l_{i-1},j}^{[i]} + b_j^{[i]}\right)$$

$$\left(\vec{a}^{[i]}\right)^T = g\left(\left(\vec{a}^{[i-1]}\right)^T W^{[i]} + \left(\vec{b}^{[i]}\right)^T\right)$$

$$= g\left(\left(\vec{a}^{[i-1]}\right) \begin{pmatrix} \uparrow & \uparrow & & \uparrow \\ \vec{w}_1^{[i]} & \vec{w}_2^{[i]} & \cdots & \vec{w}_{l_i}^{[i]} \\ \downarrow & \downarrow & & \downarrow \end{pmatrix} + \left(b_1^{[i]} \ b_2^{[i]} \cdots b_{l_i}^{[i]}\right)\right)$$

$$= \left(\nu_1^{[i]}\left(\left(\vec{a}^{[i-1]}\right)^T\right) \ \nu_2^{[i]}\left(\left(\vec{a}^{[i-1]}\right)^T\right) \cdots \nu_{l_i}^{[i]}\left(\left(\vec{a}^{[i-1]}\right)^T\right)\right)$$

CH02_04.

- Minibatches_In_Dense_Layers.

<Minibatch in Dense Layers>

$\vec{x}$마다 달라지는게 아니다.

$$(\vec{x})^T \rightarrow z = f(\vec{x}; \vec{w}, b) \rightarrow a = g(z) \rightarrow a$$

affine fcn　　　　　activation fcn

※입력값이 달라진다고 뉴런이 달라지는게 아니다. → $\vec{w}$, b의 개수, 값은 그대로임.

$(\vec{x}^{(1)})^T$

$(\vec{x}^{(2)})^T$

$\vdots$

$$X^T = \begin{pmatrix} \leftarrow (\vec{x}^{(1)})^T \rightarrow \\ \leftarrow (\vec{x}^{(2)})^T \rightarrow \\ \vdots \\ \leftarrow (\vec{x}^{(N)})^T \rightarrow \end{pmatrix} = \begin{pmatrix} x_{1,1}^{(1)} & x_{1,2}^{(1)} & \cdots & x_{1,l_I}^{(1)} \\ x_{2,1}^{(2)} & x_{2,2}^{(2)} & \cdots & x_{2,l_I}^{(2)} \\ \vdots & & \ddots & \\ x_{N,1}^{(N)} & x_{N,2}^{(N)} & & x_{N,l_I}^{(N)} \end{pmatrix}$$

$$\in \mathbb{R}^{N \times l_I}$$

$$(\vec{x}; \vec{w}, b) \rightarrow (\vec{x}; W, b) \rightarrow (X; W, b)$$

$$\underset{N\times l_1}{(Z^{[1]})^T} = \underset{N\times \cancel{l_I}}{\begin{pmatrix} \leftarrow (\vec{x}^{(1)})^T \rightarrow \\ \leftarrow (\vec{x}^{(2)})^T \rightarrow \\ \vdots \\ \leftarrow (\vec{x}^{(N)})^T \rightarrow \end{pmatrix}} \underset{\cancel{l_I}\times l_1}{\begin{pmatrix} \uparrow & \uparrow & & \uparrow \\ \vec{w}_1^{[1]} & \vec{w}_2^{[1]} & \cdots & \vec{w}_{l_1}^{[1]} \\ \downarrow & \downarrow & & \downarrow \end{pmatrix}} + \underset{l_1}{\left(b_1^{[1]} \ b_2^{[1]} \cdots b_{l_1}^{[1]}\right)}$$

$$= \begin{array}{c} \\ \begin{pmatrix} \bigcirc & \bigcirc & \cdots & \bigcirc \\ \bigcirc & \bigcirc & \cdots & \bigcirc \\ \vdots & & & \vdots \\ \bigcirc & \bigcirc & \cdots & \bigcirc \end{pmatrix} \\ \text{뉴런1 통과} \quad \text{뉴런2 통과} \quad \text{뉴런}l_1\text{ 통과} \end{array} \begin{array}{l} \vec{x}^{(1)}\text{이 들어간 값} \\ \vec{x}^{(2)}\text{이 들어간 값} \\ \\ \vec{x}^{(N)}\text{이 들어간 값.} \end{array}$$

$\mathbb{R}^{N\times l_1}$

입력 벡터수가 N개이면 결과의 행수도 N개로 바뀌지 않는다.

→ input : $N\times l_I$
output : $N\times l_1$

$$= N\left\{\begin{pmatrix} (\vec{x}^{(1)})^T\cdot\vec{w}_1^{[1]} + b_1^{[1]} & (\vec{x}^{(1)})^T\cdot\vec{w}_2^{[1]} + b_2^{[1]} & \cdots & (\vec{x}^{(1)})^T\vec{w}_{l_1}^{[1]} + b_{l_1}^{[1]} \\ (\vec{x}^{(2)})^T\cdot\vec{w}_1^{[1]} + b_1^{[1]} & (\vec{x}^{(2)})^T\cdot\vec{w}_2^{[1]} + b_2^{[1]} & \cdots & (\vec{x}^{(2)})^T\cdot\vec{w}_{l_1} + b_{l_1}^{[1]} \\ \vdots & & & \\ (\vec{x}^{(N)})^T\cdot\vec{w}_1^{[1]} + b_1^{[1]} & (\vec{x}^{(N)})^T\cdot\vec{w}_2^{[1]} + b_2^{[1]} & \cdots & (\vec{x}^{(N)})^T\cdot\vec{w}_{l_1}^{[1]} + b_{l_1}^{[1]} \end{pmatrix}\right.$$
$$\underbrace{\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad}_{l_1}$$

$$z_{i,j}^{[1]} = (\vec{x}^{(i)})^T \vec{w}_j^{[1]} + \vec{b}_j^{[1]}$$

z 값의 위치 : i 번째 input (x) 행 , j 번째 뉴런통과 (w_j, b_j) 열 .

일반화: $a^{[i]}_{i,j} = g(z^{[i]}_{i,j})$

☆ $R^{N\times l_I} \rightarrow R^{N\times l_i} \rightarrow R^{N\times l_i}$ ☆

input z a

$$(A^{[i]})^T = \begin{pmatrix} a^{[i]}_{1,1} & a^{[i]}_{1,2} & \cdots & a^{[i]}_{1,l_i} \\ a^{[i]}_{2,1} & a^{[i]}_{2,2} & & a^{[i]}_{2,l_i} \\ \vdots & \vdots & & \vdots \\ a^{[i]}_{N,1} & a^{[i]}_{N,2} & & a^{[i]}_{N,l_i} \end{pmatrix} = \begin{pmatrix} \leftarrow (\vec{a}^{[i](1)})^T \rightarrow \\ \leftarrow (\vec{a}^{[i](2)})^T \rightarrow \\ \vdots \\ \leftarrow (\vec{a}^{[i](N)})^T \rightarrow \end{pmatrix}$$

Batch-wise

1~N번째 Data들의 피처를 생성.

$$(A^{[i]})^T = \begin{pmatrix} a^{[i]}_{1,1} & a^{[i]}_{1,2} & \cdots & a^{[i]}_{1,l_i} \\ a^{[i]}_{2,1} & a^{[i]}_{2,2} & & a^{[i]}_{2,l_i} \\ \vdots & \vdots & & \vdots \\ a^{[i]}_{N,1} & a^{[i]}_{N,2} & & a^{[i]}_{N,l_i} \end{pmatrix} = \begin{pmatrix} \uparrow & \uparrow & & \uparrow \\ \vec{a}^{[i]}_1 & \vec{a}^{[i]}_2 & \cdots & \vec{a}^{[i]}_{l_i} \\ \downarrow & \downarrow & & \downarrow \end{pmatrix}$$

Neuron-wise

1~l_i번째 뉴런을 거친 값들.

- <Cascaded Dense Layers>

$(A^{[0]})^T : R^{N\times l_0}$

$L^{[0]}$ $W^{[0]} : R^{l_{0-1}\times l_0}$, $(b^{[0]})^T : R^{1\times l_0}$

⋮ $(A^{[2]})^T : R^{N\times l_2}$

$L^{[2]}$ $W^{[2]} : R^{l_1\times l_2}$, $(b^{[2]})^T : R^{1\times l_2}$

↑ $(A^{[1]})^T : R^{N\times l_1}$

$L^{[1]}$ $W^{[1]} : R^{l_I\times l_1}$, $(b^{[1]})^T : R^{1\times l_1}$

↑

$X^T : R^{N\times l_I}$