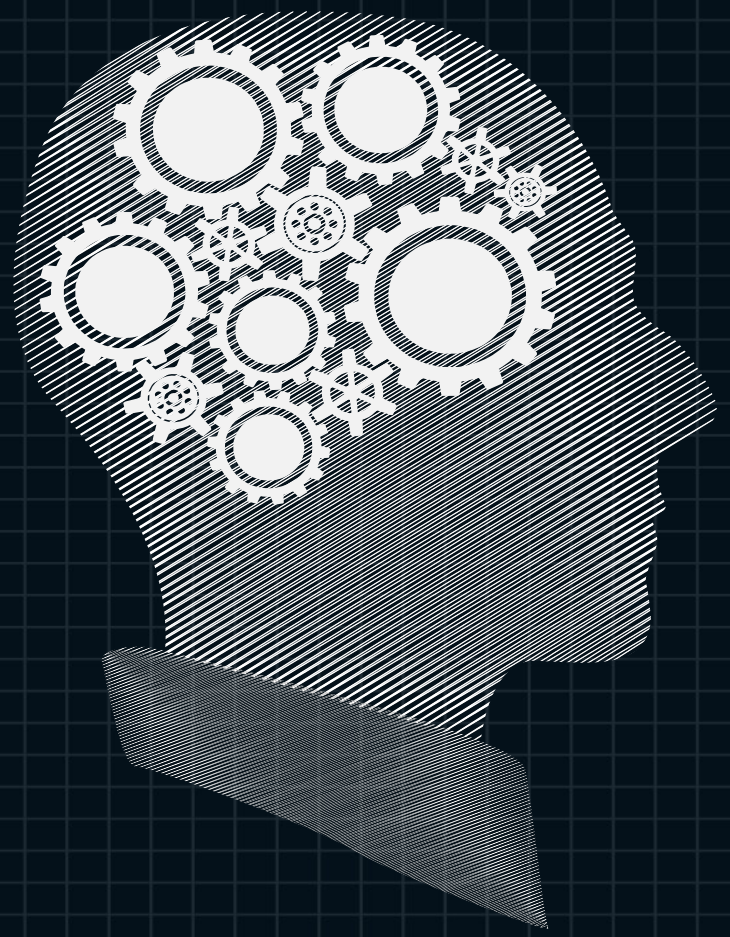
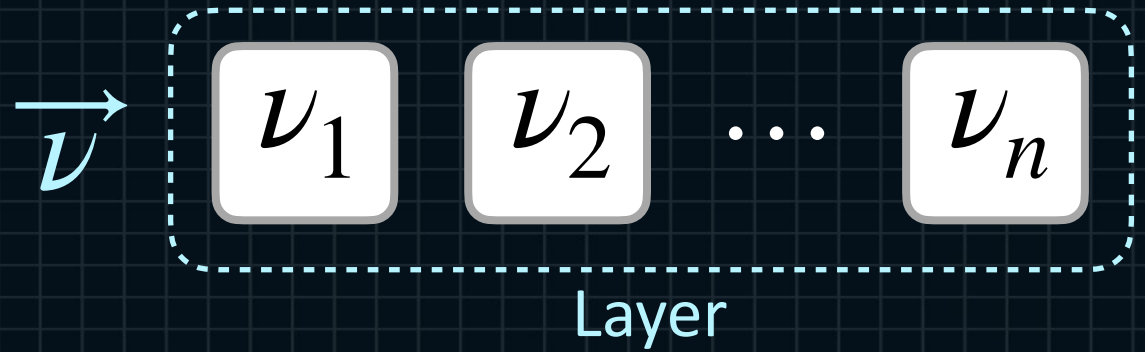


Forward Propagation of Neural Networks

Lecture.2
Dense Layers

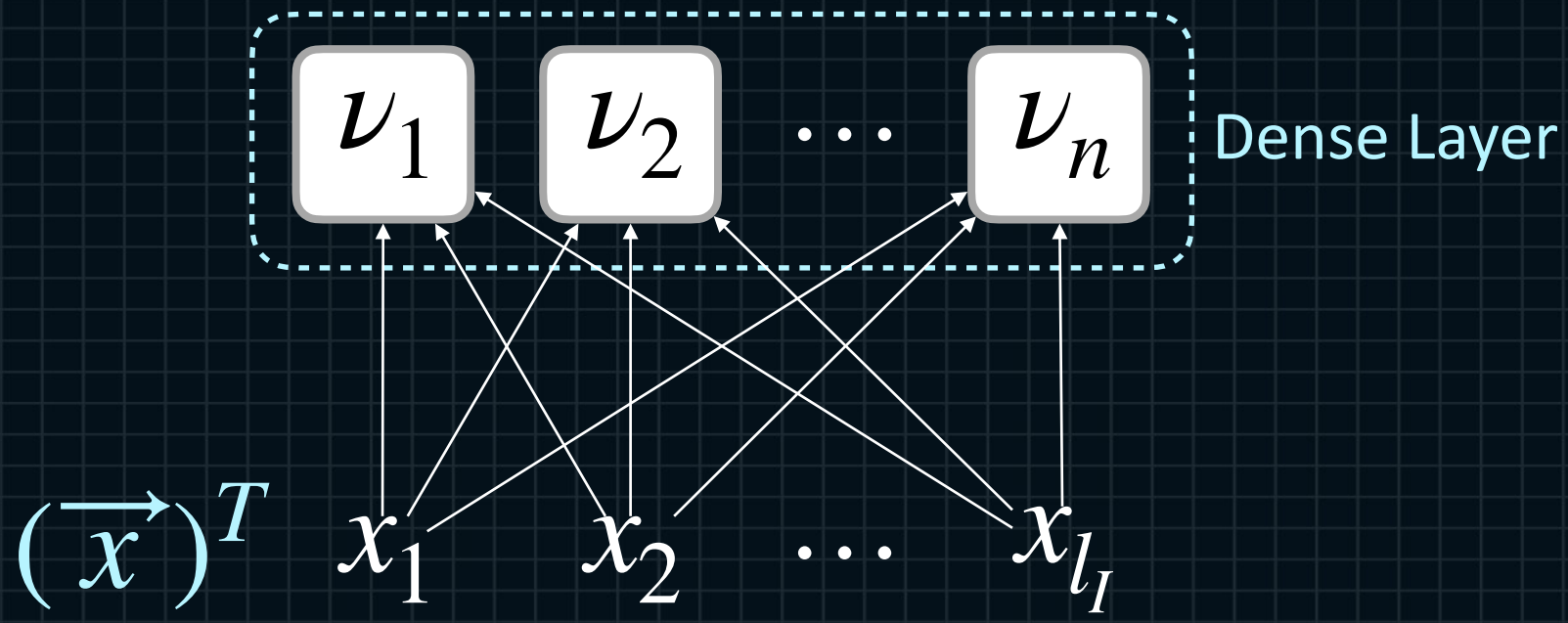
Lecture.2
Dense Layers

- Neuron Vectors and Layers



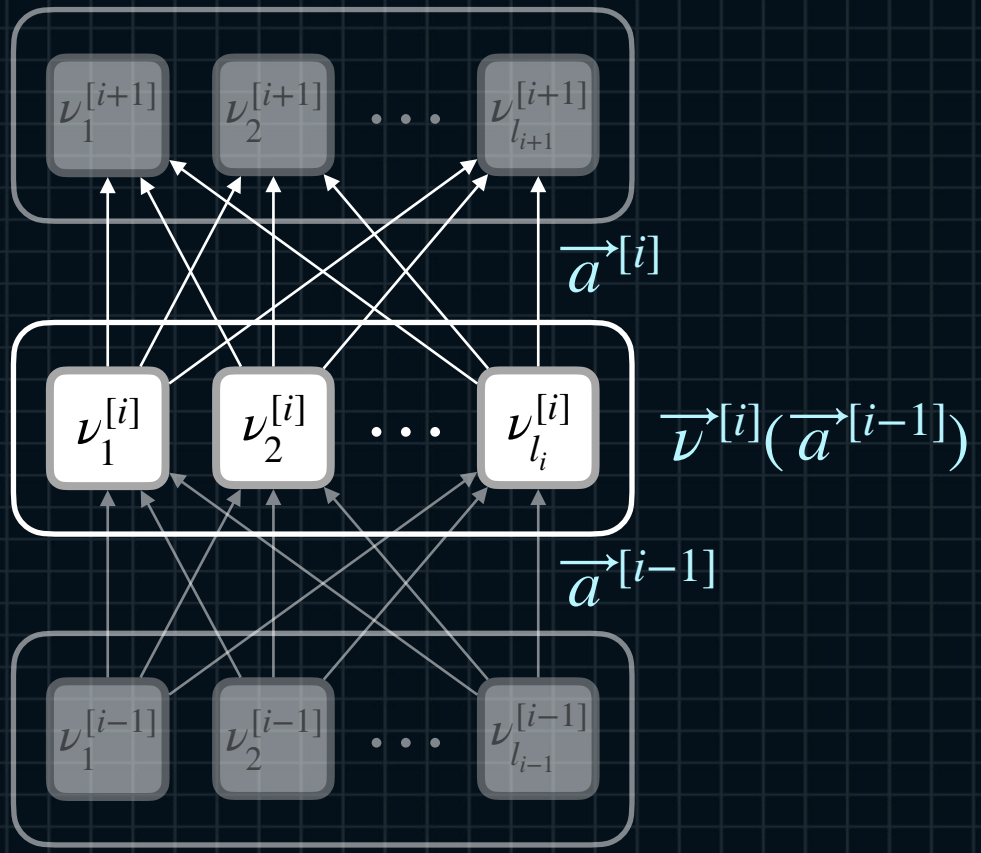
Lecture.2
Dense Layers

- Dense Layers



Lecture.2 Dense Layers

- Notation



Layer: $L^{[i]}$ (L for Layer)

Neuron Vector: $\vec{\nu}^{[i]}$

# of Neurons: l_i (l for length)

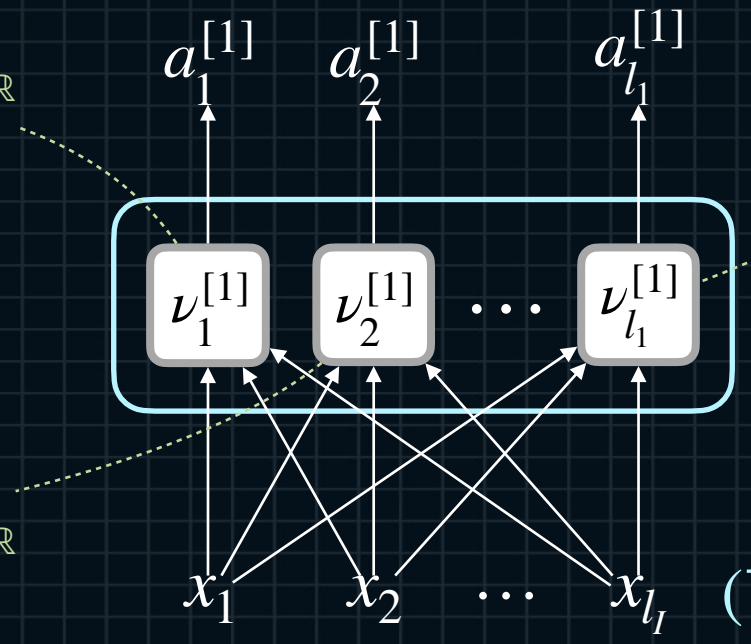
Lecture.2
Dense Layers

- Params of Dense Layer

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

$$\vec{w}_2^{[1]} = \begin{pmatrix} w_{1,2}^{[1]} \\ w_{2,2}^{[1]} \\ \vdots \\ w_{l_I,2}^{[1]} \end{pmatrix} \in \mathbb{R}^{l_I \times 1}, \quad b_2^{[1]} \in \mathbb{R}$$

$$\vec{w}_{l_1}^{[1]} = \begin{pmatrix} w_{1,l_1}^{[1]} \\ w_{2,l_1}^{[1]} \\ \vdots \\ w_{l_I,l_1}^{[1]} \end{pmatrix} \in \mathbb{R}^{l_I \times 1}, \quad b_{l_1}^{[1]} \in \mathbb{R}$$



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

$$(\vec{x})^T \in \mathbb{R}^{1 \times l_I}$$

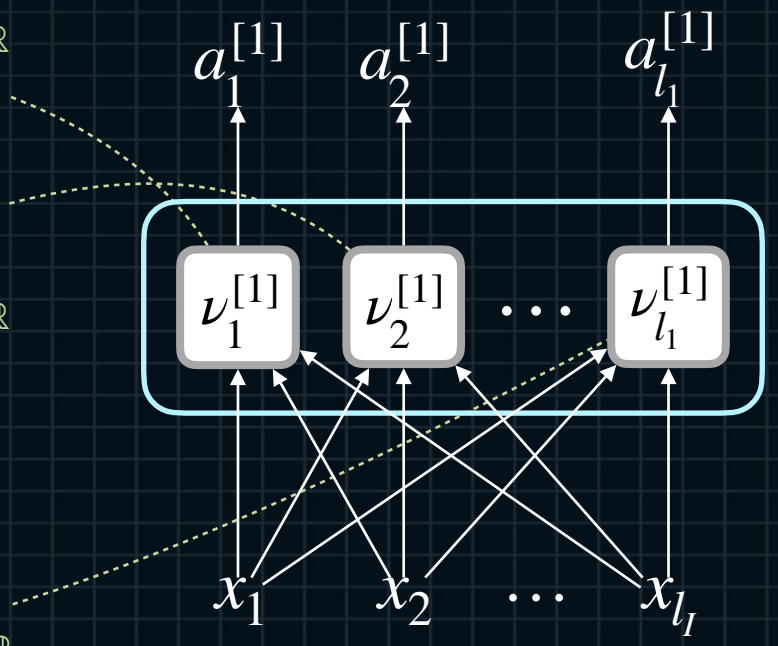
Lecture.2 Dense Layers

- Weight Matrix and Bias Vector

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

$$\vec{w}_2^{[1]} = \begin{pmatrix} w_{1,2}^{[1]} \\ w_{2,2}^{[1]} \\ \vdots \\ w_{l_I,2}^{[1]} \end{pmatrix} \in \mathbb{R}^{l_I \times 1}, \quad b_2^{[1]} \in \mathbb{R}$$

$$\vec{w}_{l_1}^{[1]} = \begin{pmatrix} w_{1,l_1}^{[1]} \\ w_{2,l_1}^{[1]} \\ \vdots \\ w_{l_I,l_1}^{[1]} \end{pmatrix} \in \mathbb{R}^{l_I \times 1}, \quad b_{l_1}^{[1]} \in \mathbb{R}$$



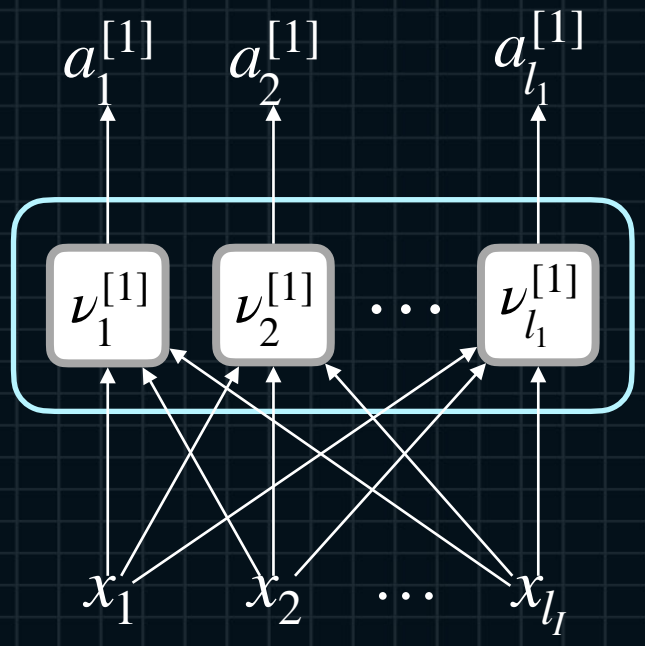
$$W^{[1]} = \begin{pmatrix} \uparrow & \uparrow & & \uparrow \\ \vec{w}_1^{[1]} & \vec{w}_2^{[1]} & \cdots & \vec{w}_{l_1}^{[1]} \\ \downarrow & \downarrow & & \downarrow \end{pmatrix}$$

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

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

Lecture.2
Dense Layers

- FP of Dense Layer



$$W^{[1]} = \begin{pmatrix} \uparrow & \uparrow & & \uparrow \\ \overrightarrow{w}_1^{[1]} & \overrightarrow{w}_2^{[1]} & \cdots & \overrightarrow{w}_{l_1}^{[1]} \\ \downarrow & \downarrow & & \downarrow \end{pmatrix} \in \mathbb{R}^{l_I \times l_1}$$

$$(\overrightarrow{b}^{[1]})^T = \begin{pmatrix} b_1^{[1]} & b_2^{[1]} & \cdots & b_{l_1}^{[1]} \end{pmatrix} \in \mathbb{R}^{1 \times l_1}$$

$$a_i^{[1]} = g\left((\overrightarrow{x})^T \overrightarrow{w}_i^{[1]} + b_i^{[1]}\right) \quad 1 \le i \le l_1$$

$$= g\left(x_1 w_{1,i}^{[1]} + x_2 w_{2,i}^{[1]} + \ldots + x_{l_I} w_{l_I,i}^{[1]} + b_i^{[1]}\right)$$

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

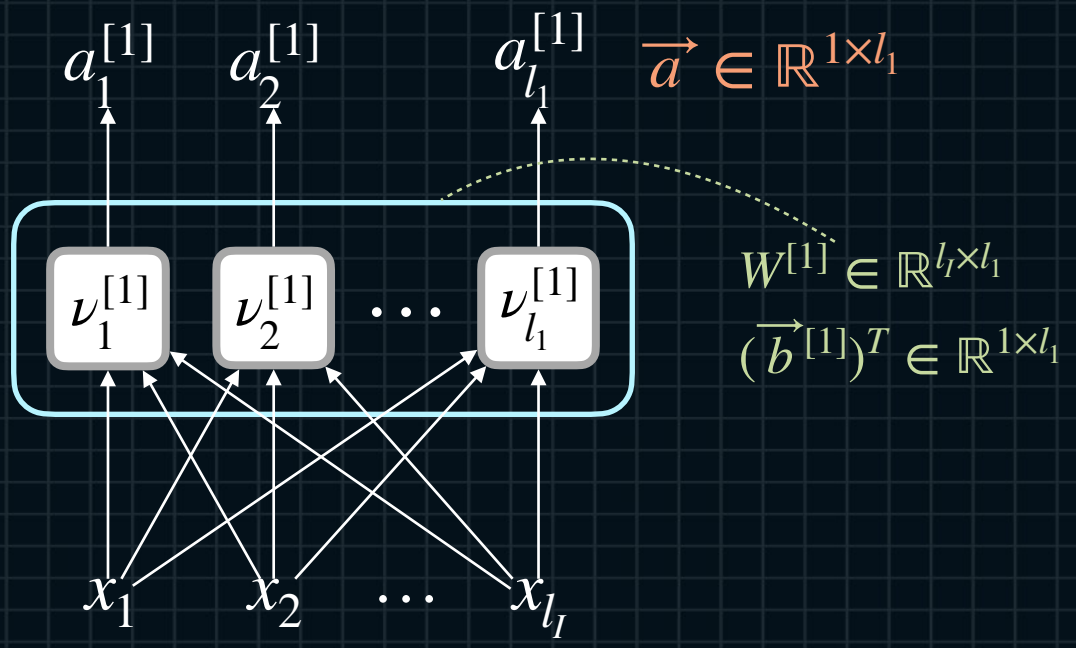
$$(\overrightarrow{a}^{[1]})^T = g\left((\overrightarrow{x})^T W^{[1]} + (\overrightarrow{b}^{[1]})^T\right)$$

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

$$= \begin{pmatrix} \nu_1^{[1]}((\overrightarrow{x})^T) & \nu_2^{[1]}((\overrightarrow{x})^T) & \ldots & \nu_{l_1}^{[1]}((\overrightarrow{x})^T) \end{pmatrix}$$

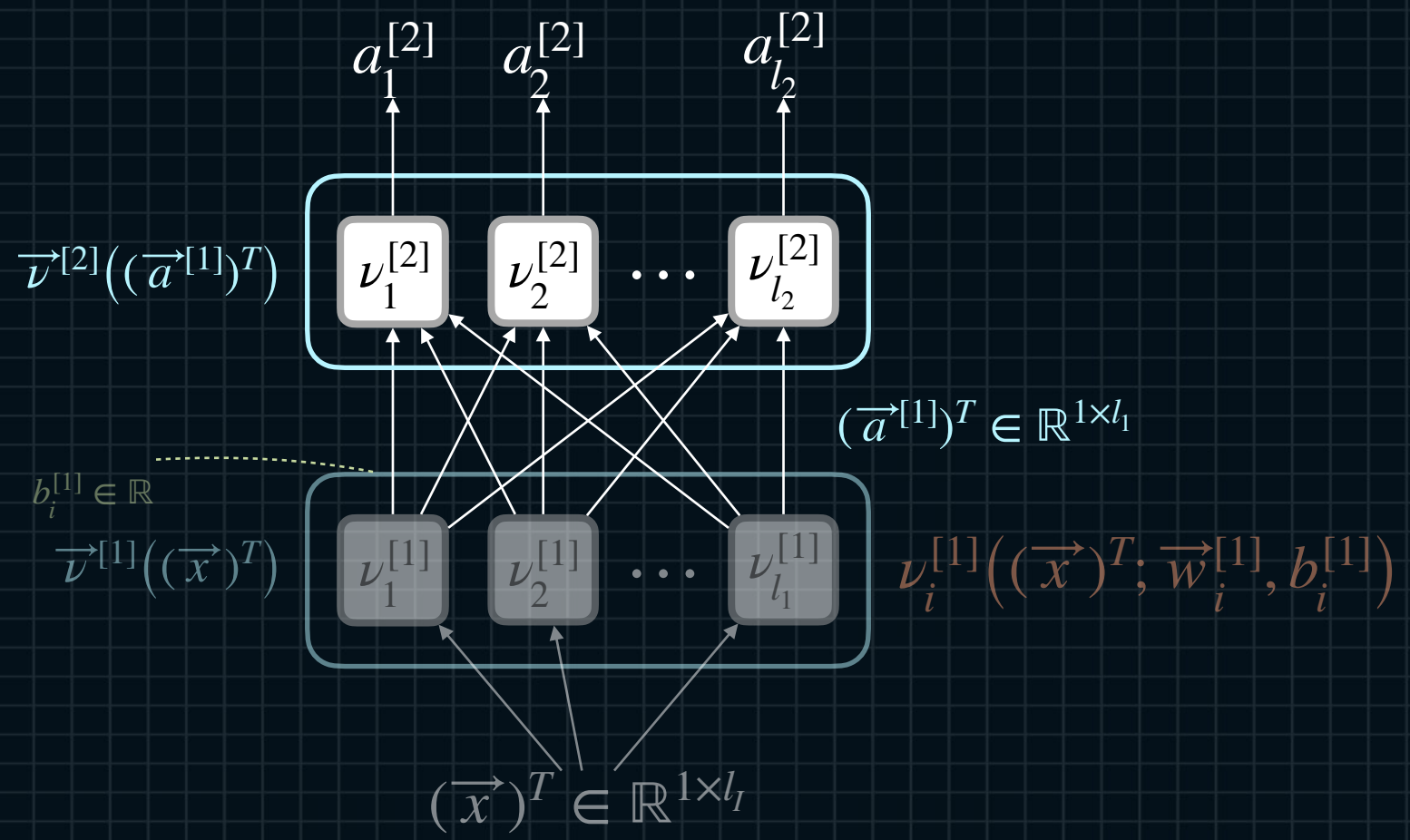
Lecture.2
Dense Layers

- Dimensions of Dense Layer



Lecture.2
Dense Layers

- The Second Dense Layer

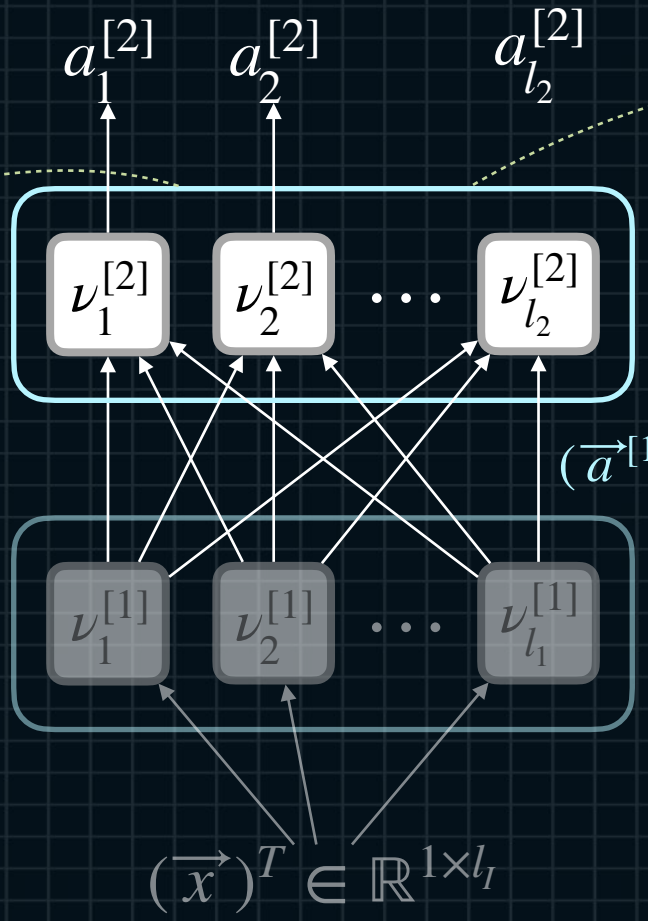


Lecture.2 Dense Layers

- The Params of The Second Dense Layer

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

$$1 \le i \le l_2$$



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

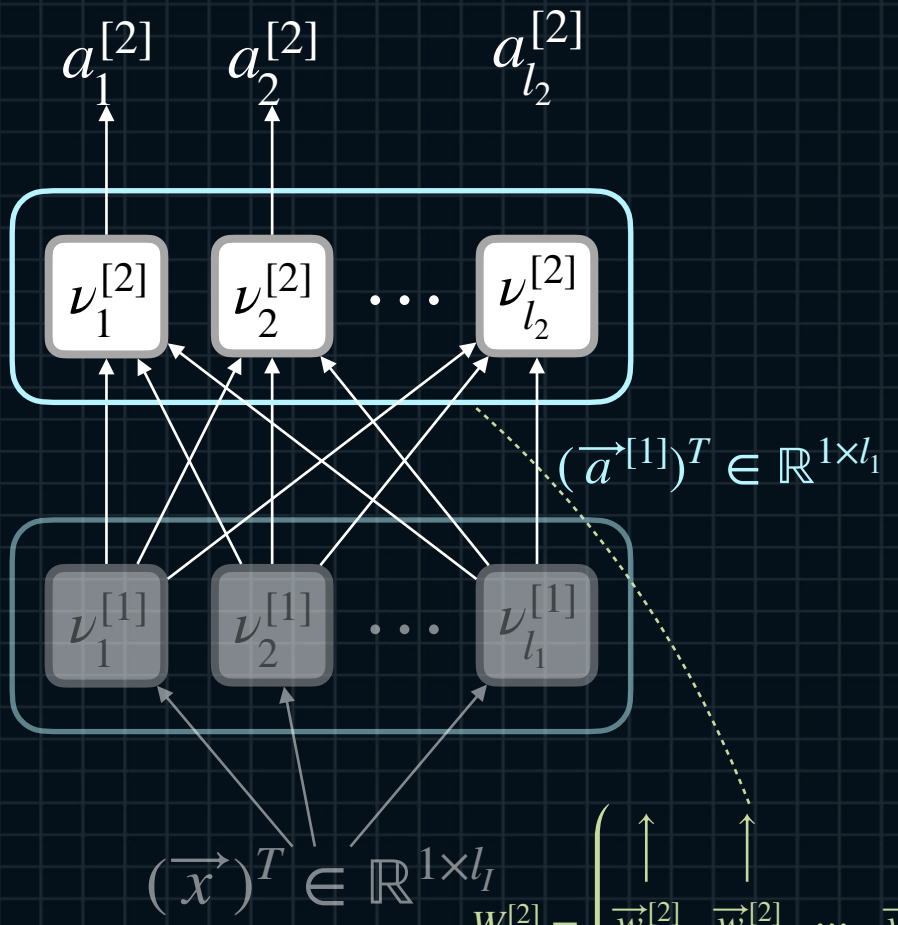
$$W^{[2]} = \begin{pmatrix} \uparrow & \uparrow & & \uparrow \\ \vec{w}_1^{[2]} & \vec{w}_2^{[2]} & \cdots & \vec{w}_{l_2}^{[2]} \\ \downarrow & \downarrow & & \downarrow \end{pmatrix}$$

$$= \begin{pmatrix} w_{1,1}^{[2]} & w_{1,2}^{[2]} & \cdots & w_{1,l_2}^{[2]} \\ w_{2,1}^{[2]} & w_{2,2}^{[2]} & \cdots & w_{2,l_2}^{[2]} \\ \vdots & \vdots & \ddots & \vdots \\ w_{l_1,1}^{[2]} & w_{l_1,2}^{[2]} & \cdots & w_{l_1,l_2}^{[2]} \end{pmatrix} \in \mathbb{R}^{l_1 \times l_2}$$

$$(\vec{b}^{[2]})^T = \begin{pmatrix} b_1^{[2]} & b_2^{[2]} & \cdots & b_{l_2}^{[2]} \end{pmatrix} \in \mathbb{R}^{1 \times l_2}$$

Lecture.2
Dense Layers

- FP of The Second Dense Layer



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

$$= g\left(a_1^{[1]} w_{1,i}^{[2]} + a_2^{[1]} w_{2,i}^{[2]} + \ldots + a_{l_1}^{[1]} w_{l_1,i}^{[2]} + b_i^{[2]}\right)$$

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

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

$$= \begin{pmatrix} \nu_1^{[2]}\left((\vec{a}^{[1]})^T\right) & \nu_2^{[2]}\left((\vec{a}^{[1]})^T\right) & \ldots & \nu_{l_2}^{[2]}\left((\vec{a}^{[1]})^T\right) \end{pmatrix}$$

$$W^{[2]} = \begin{pmatrix} \uparrow & \uparrow & & \uparrow \\ \vec{w}_1^{[2]} & \vec{w}_2^{[2]} & \cdots & \vec{w}_{l_2}^{[2]} \\ \downarrow & \downarrow & & \downarrow \end{pmatrix} \in \mathbb{R}^{l_1\times l_2}$$

$$(\vec{b}^{[2]})^T = \begin{pmatrix} b_1^{[2]} & b_2^{[2]} & \ldots & b_{l_2}^{[2]} \end{pmatrix} \in \mathbb{R}^{1\times l_2}$$

Lecture.2
Dense Layers

- The Second Dense Layer

$$W^{[2]} = \begin{pmatrix} \uparrow & \uparrow & & \uparrow \\ \vec{w}_1^{[2]} & \vec{w}_2^{[2]} & \cdots & \vec{w}_{l_2}^{[2]} \\ \downarrow & \downarrow & & \downarrow \end{pmatrix} \in \mathbb{R}^{l_1 \times l_2}$$

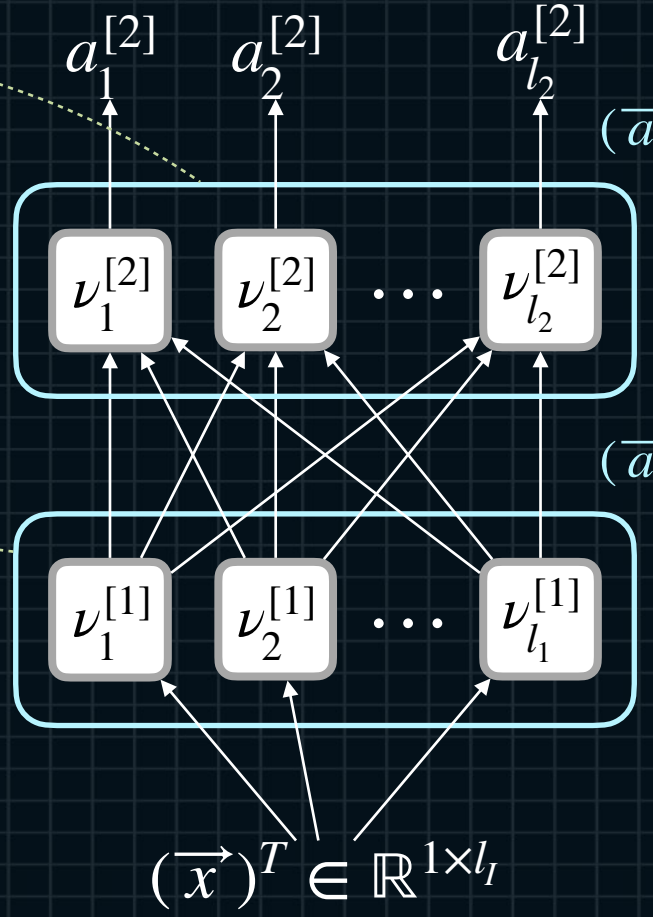
$$(\vec{b}^{[2]})^T = \begin{pmatrix} b_1^{[2]} & b_2^{[2]} & \cdots & b_{l_2}^{[2]} \end{pmatrix} \in \mathbb{R}^{1 \times l_2}$$

$$\vec{\nu}^{[2]}\left((\vec{a}^{[1]})^T\right)$$

$$W^{[1]} = \begin{pmatrix} \uparrow & \uparrow & & \uparrow \\ \vec{w}_1^{[1]} & \vec{w}_2^{[1]} & \cdots & \vec{w}_{l_1}^{[1]} \\ \downarrow & \downarrow & & \downarrow \end{pmatrix} \in \mathbb{R}^{l_I \times l_1}$$

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

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



$$(\vec{a}^{[2]})^T \in \mathbb{R}^{1 \times l_2}$$

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

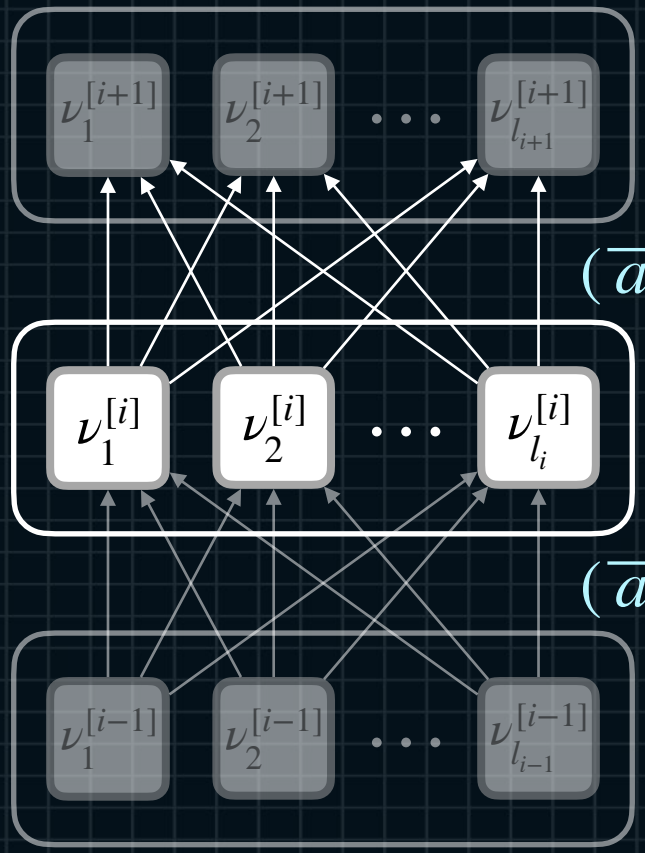
$$(\vec{a}^{[1]})^T \in \mathbb{R}^{1 \times l_1}$$

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

$$(\vec{x})^T \in \mathbb{R}^{1 \times l_I}$$

Lecture.2
Dense Layers

- Generalized Dense Layer



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

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

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

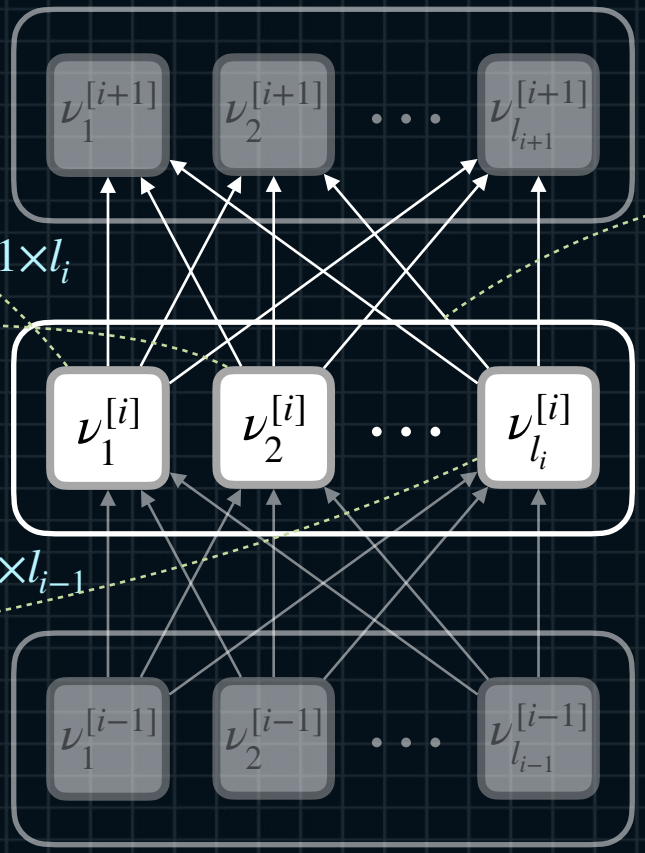
Lecture.2
Dense Layers

- Generalized Dense Layer

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

$$\vec{w}_2^{[i]} = \begin{pmatrix} w_{1,2}^{[i]} \\ w_{2,2}^{[i]} \\ \vdots \\ w_{l_{i-1},2}^{[i]} \end{pmatrix} \in \mathbb{R}^{l_{i-1}\times 1}, \quad b_2^{[i]} \in \mathbb{R}$$

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

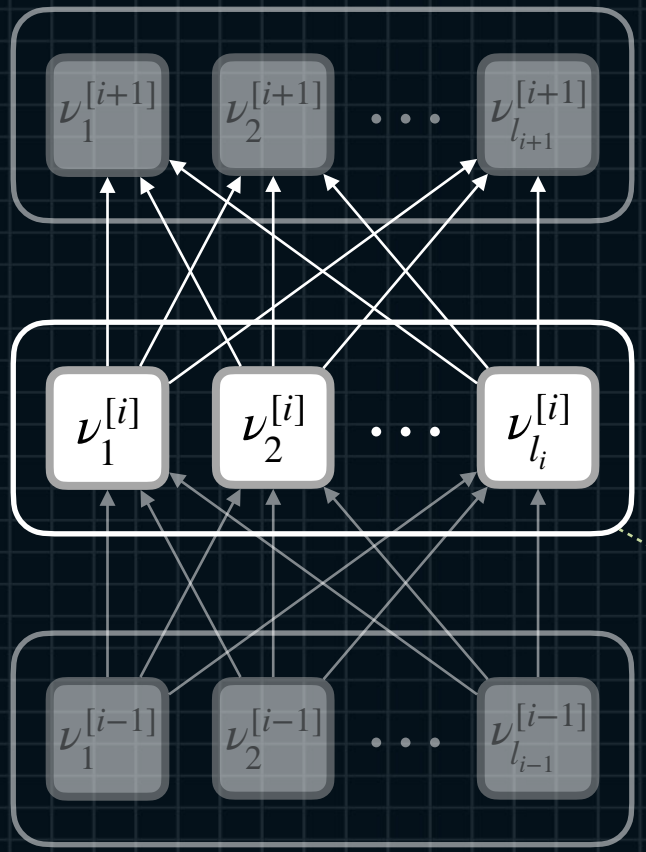


$$W^{[i]} = \begin{pmatrix} \uparrow & \uparrow & & \uparrow \\ \vec{w}_1^{[i]} & \vec{w}_2^{[i]} & \cdots & \vec{w}_{l_i}^{[i]} \\ \downarrow & \downarrow & & \downarrow \end{pmatrix} = \begin{pmatrix} w_{1,1}^{[i]} & w_{1,2}^{[i]} & \cdots & w_{1,l_i}^{[i]} \\ w_{2,1}^{[i]} & w_{2,2}^{[i]} & \cdots & w_{2,l_i}^{[i]} \\ \vdots & \vdots & \ddots & \vdots \\ w_{l_{i-1},1}^{[i]} & w_{l_{i-1},2}^{[i]} & \cdots & w_{l_{i-1},l_i}^{[i]} \end{pmatrix} \in \mathbb{R}^{l_{i-1}\times l_i}$$

$$(\vec{b}^{[i]})^T = \begin{pmatrix} b_1^{[i]} & b_2^{[i]} & \cdots & b_{l_i}^{[i]} \end{pmatrix} \in \mathbb{R}^{1\times l_i}$$

Lecture.2 Dense Layers

- Generalized Dense Layer



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

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

$$\mathbb{R}^{1\times l_i} \qquad \mathbb{R}^{1\times l_{i-1}}\ \mathbb{R}^{l_{i-1}\times l_i}\ \mathbb{R}^{1\times l_i}$$

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

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

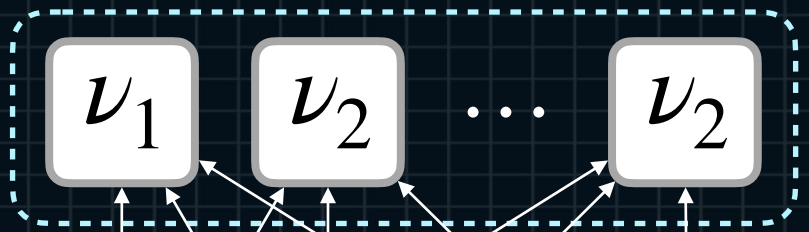
$$= \left(\nu_1^{[i]}\left((\vec{a}^{[i-1]})^T\right) \quad \nu_2^{[i]}\left((\vec{a}^{[i-1]})^T\right) \quad \ldots \quad \nu_{l_i}^{[i]}\left((\vec{a}^{[i-1]})^T\right)\right)$$

$$W^{[i]} = \begin{pmatrix} \uparrow & \uparrow & & \uparrow \\ \vec{w}_1^{[i]} & \vec{w}_2^{[i]} & \cdots & \vec{w}_{l_i}^{[i]} \\ \downarrow & \downarrow & & \downarrow \end{pmatrix} \in \mathbb{R}^{l_{i-1}\times l_i}$$

$$(\vec{b}^{[i]})^T = \begin{pmatrix} b_1^{[i]} & b_2^{[i]} & \ldots & b_{l_i}^{[i]} \end{pmatrix} \in \mathbb{R}^{1\times l_i}$$

Lecture.2
Dense Layers

- Minibatch in Dense Layers

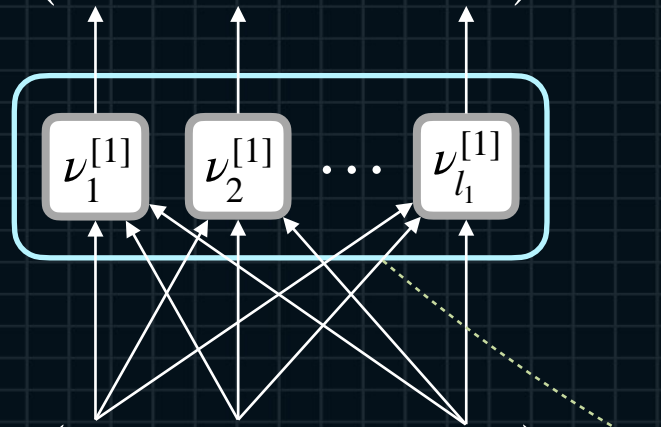


$$X^T = \begin{pmatrix} x_1^{(1)} & x_2^{(1)} & \cdots & x_{l_I}^{(1)} \\ x_1^{(2)} & x_2^{(2)} & \cdots & x_{l_I}^{(2)} \\ \vdots & \vdots & \ddots & \vdots \\ x_1^{(N)} & x_2^{(N)} & \cdots & x_{l_I}^{(N)} \end{pmatrix} \in \mathbb{R}^{N \times l_I}$$

Lecture.2 Dense Layers

- Minibatch in Dense Layers

$$(A^{[1]})^T = \begin{pmatrix} a^{[1]}_{1,1} & a^{[1]}_{1,2} & \cdots & a^{[1]}_{1,l_1} \\ a^{[1]}_{2,1} & a^{[1]}_{2,2} & \cdots & a^{[1]}_{2,l_1} \\ \vdots & \vdots & \ddots & \vdots \\ a^{[1]}_{N,1} & a^{[1]}_{N,2} & \cdots & a^{[1]}_{N,l_1} \end{pmatrix}$$



$$X^T = \begin{pmatrix} x^{(1)}_1 & x^{(1)}_2 & \cdots & x^{(1)}_{l_I} \\ x^{(2)}_1 & x^{(2)}_2 & \cdots & x^{(2)}_{l_I} \\ \vdots & \vdots & \ddots & \vdots \\ x^{(N)}_1 & x^{(N)}_2 & \cdots & x^{(N)}_{l_I} \end{pmatrix}$$

$$(Z^{[1]})^T = \begin{pmatrix} \longleftarrow & (\vec{x}^{(1)})^T & \longrightarrow \\ \longleftarrow & (\vec{x}^{(2)})^T & \longrightarrow \\ & \vdots & \\ \longleftarrow & (\vec{x}^{(N)})^T & \longrightarrow \end{pmatrix} \begin{pmatrix} \uparrow & \uparrow & & \uparrow \\ \vec{w}^{[1]}_1 & \vec{w}^{[1]}_2 & \cdots & \vec{w}^{[1]}_{l_1} \\ \downarrow & \downarrow & & \downarrow \end{pmatrix} + \begin{pmatrix} b^{[1]}_1 & b^{[1]}_2 & \cdots & b^{[1]}_{l_1} \end{pmatrix}$$

$$= \begin{pmatrix} (\vec{x}^{(1)})^T \cdot \vec{w}^{[1]}_1 + b^{[1]}_1 & (\vec{x}^{(1)})^T \cdot \vec{w}^{[1]}_2 + b^{[1]}_2 & \cdots & (\vec{x}^{(1)})^T \cdot \vec{w}^{[1]}_{l_1} + b^{[1]}_{l_1} \\ (\vec{x}^{(2)})^T \cdot \vec{w}^{[1]}_1 + b^{[1]}_1 & (\vec{x}^{(2)})^T \cdot \vec{w}^{[1]}_2 + b^{[1]}_2 & \cdots & (\vec{x}^{(2)})^T \cdot \vec{w}^{[1]}_{l_1} + b^{[1]}_{l_1} \\ \vdots & \vdots & \ddots & \vdots \\ (\vec{x}^{(N)})^T \cdot \vec{w}^{[1]}_1 + b^{[1]}_1 & (\vec{x}^{(N)})^T \cdot \vec{w}^{[1]}_2 + b^{[1]}_2 & \cdots & (\vec{x}^{(N)})^T \cdot \vec{w}^{[1]}_{l_1} + b^{[1]}_{l_1} \end{pmatrix}$$

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

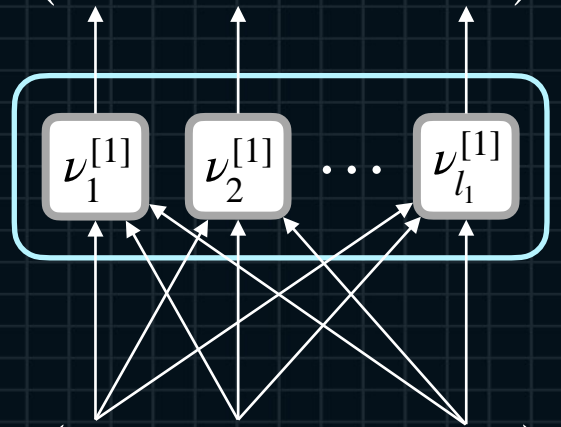
$$W^{[1]} = \begin{pmatrix} \uparrow & \uparrow & & \uparrow \\ \vec{w}^{[1]}_1 & \vec{w}^{[1]}_2 & \cdots & \vec{w}^{[1]}_{l_1} \\ \downarrow & \downarrow & & \downarrow \end{pmatrix} \in \mathbb{R}^{l_I \times l_1}$$

$$(\vec{b}^{[1]})^T = \begin{pmatrix} b^{[1]}_1 & b^{[1]}_2 & \cdots & b^{[1]}_{l_1} \end{pmatrix} \in \mathbb{R}^{1 \times l_1}$$

Lecture.2 Dense Layers

- Minibatch in Dense Layers

$$(A^{[1]})^T = \begin{pmatrix} a^{[1]}_{1,1} & a^{[1]}_{1,2} & \dots & a^{[1]}_{1,l_1} \\ a^{[1]}_{2,1} & a^{[1]}_{2,2} & \dots & a^{[1]}_{2,l_1} \\ \vdots & \vdots & \ddots & \vdots \\ a^{[1]}_{N,1} & a^{[1]}_{N,2} & \dots & a^{[1]}_{N,l_1} \end{pmatrix}$$



$$X^T = \begin{pmatrix} x^{(1)}_1 & x^{(1)}_2 & \dots & x^{(1)}_{l_I} \\ x^{(2)}_1 & x^{(2)}_2 & \dots & x^{(2)}_{l_I} \\ \vdots & \vdots & \ddots & \vdots \\ x^{(N)}_1 & x^{(N)}_2 & \dots & x^{(N)}_{l_I} \end{pmatrix}$$

$$(Z^{[1]})^T = \begin{pmatrix} (\vec{x}^{(1)})^T \cdot \vec{w}^{[1]}_1 + b^{[1]}_1 & (\vec{x}^{(1)})^T \cdot \vec{w}^{[1]}_2 + b^{[1]}_2 & \dots & (\vec{x}^{(1)})^T \cdot \vec{w}^{[1]}_{l_1} + b^{[1]}_{l_1} \\ (\vec{x}^{(2)})^T \cdot \vec{w}^{[1]}_1 + b^{[1]}_1 & (\vec{x}^{(2)})^T \cdot \vec{w}^{[1]}_2 + b^{[1]}_2 & \dots & (\vec{x}^{(2)})^T \cdot \vec{w}^{[1]}_{l_1} + b^{[1]}_{l_1} \\ \vdots & \vdots & \ddots & \vdots \\ (\vec{x}^{(N)})^T \cdot \vec{w}^{[1]}_1 + b^{[1]}_1 & (\vec{x}^{(N)})^T \cdot \vec{w}^{[1]}_2 + b^{[1]}_2 & \dots & (\vec{x}^{(N)})^T \cdot \vec{w}^{[1]}_{l_1} + b^{[1]}_{l_1} \end{pmatrix}$$

$$(A^{[1]})^T = \begin{pmatrix} g\left((\vec{x}^{(1)})^T \cdot \vec{w}^{[1]}_1 + b^{[1]}_1\right) & g\left((\vec{x}^{(1)})^T \cdot \vec{w}^{[1]}_2 + b^{[1]}_2\right) & \dots & g\left((\vec{x}^{(1)})^T \cdot \vec{w}^{[1]}_{l_1} + b^{[1]}_{l_1}\right) \\ g\left((\vec{x}^{(2)})^T \cdot \vec{w}^{[1]}_1 + b^{[1]}_1\right) & g\left((\vec{x}^{(2)})^T \cdot \vec{w}^{[1]}_2 + b^{[1]}_2\right) & \dots & g\left((\vec{x}^{(2)})^T \cdot \vec{w}^{[1]}_{l_1} + b^{[1]}_{l_1}\right) \\ \vdots & \vdots & \ddots & \vdots \\ g\left((\vec{x}^{(N)})^T \cdot \vec{w}^{[1]}_1 + b^{[1]}_1\right) & g\left((\vec{x}^{(N)})^T \cdot \vec{w}^{[1]}_2 + b^{[1]}_2\right) & \dots & g\left((\vec{x}^{(N)})^T \cdot \vec{w}^{[1]}_{l_1} + b^{[1]}_{l_1}\right) \end{pmatrix}$$

$$a^{[1]}_{i,j} = g\left(z^{[1]}_{i,j}\right)$$

Lecture.2 Dense Layers

- Minibatch in Dense Layers

$$
(A^{[i]})^T = \in \mathbb{R}^{N \times l_i}
\begin{pmatrix}
\nu_1^{[i]}\left((\overrightarrow{a}^{(1)})^T ;\ \overrightarrow{w}_1^{[i]}, b_1^{[i]}\right) & \nu_2^{[i]}\left((\overrightarrow{a}^{(1)})^T ;\ \overrightarrow{w}_2^{[i]}, b_2^{[i]}\right) & \dots & \nu_{l_i}^{[i]}\left((\overrightarrow{a}^{(1)})^T ;\ \overrightarrow{w}_{l_i}^{[i]}, b_{l_i}^{[i]}\right) \\
\nu_1^{[i]}\left((\overrightarrow{a}^{(2)})^T ;\ \overrightarrow{w}_1^{[i]}, b_1^{[i]}\right) & \nu_2^{[i]}\left((\overrightarrow{a}^{(2)})^T ;\ \overrightarrow{w}_2^{[i]}, b_2^{[i]}\right) & \dots & \nu_{l_i}^{[i]}\left((\overrightarrow{a}^{(2)})^T ;\ \overrightarrow{w}_{l_i}^{[i]}, b_{l_i}^{[i]}\right) \\
\vdots & \vdots & \ddots & \vdots \\
\nu_1^{[i]}\left((\overrightarrow{a}^{(N)})^T ;\ \overrightarrow{w}_1^{[i]}, b_1^{[i]}\right) & \nu_2^{[i]}\left((\overrightarrow{a}^{(N)})^T ;\ \overrightarrow{w}_2^{[i]}, b_2^{[i]}\right) & \dots & \nu_{l_i}^{[i]}\left((\overrightarrow{a}^{(N)})^T ;\ \overrightarrow{w}_{l_i}^{[i]}, b_{l_i}^{[i]}\right)
\end{pmatrix}
$$

Lecture.2 Dense Layers

- Minibatch in Dense Layers

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

Batch-wise

$$\begin{pmatrix} \longleftarrow & (\overrightarrow{a}^{[i](1)})^T & \longrightarrow \\ \longleftarrow & (\overrightarrow{a}^{[i](2)})^T & \longrightarrow \\ & \vdots & \\ \longleftarrow & (\overrightarrow{a}^{[i](N)})^T & \longrightarrow \end{pmatrix}$$

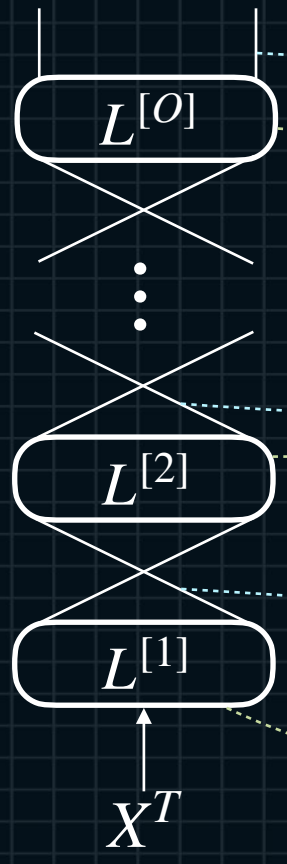
$$(A^{[i]})^T = \begin{pmatrix} a^{[i]}_{1,1} & a^{[i]}_{1,2} & \cdots & a^{[i]}_{1,l_i} \\ a^{[i]}_{2,1} & a^{[i]}_{2,2} & \cdots & a^{[i]}_{2,l_i} \\ \vdots & \vdots & \ddots & \vdots \\ a^{[i]}_{N,1} & a^{[i]}_{N,2} & \cdots & a^{[i]}_{N,l_i} \end{pmatrix}$$

Neuron-wise

$$\begin{pmatrix} \uparrow & \uparrow & & \uparrow \\ \overrightarrow{a}^{[i]}_1 & \overrightarrow{a}^{[i]}_2 & \cdots & \overrightarrow{a}^{[i]}_{l_i} \\ \downarrow & \downarrow & & \downarrow \end{pmatrix}$$

Lecture.2 Dense Layers

- Cascaded Dense Layers



$$W^{[O]} = \begin{pmatrix} \uparrow & \uparrow & & \uparrow \\ \vec{w}_1^{[O]} & \vec{w}_2^{[O]} & \cdots & \vec{w}_{l_O}^{[O]} \\ \downarrow & \downarrow & & \downarrow \end{pmatrix} \in \mathbb{R}^{l_{O-1} \times l_O}$$

$$(\vec{b}^{[O]})^T = \begin{pmatrix} b_1^{[O]} & b_2^{[O]} & \cdots & b_{l_O}^{[O]} \end{pmatrix} \in \mathbb{R}^{1 \times l_O}$$

$$(A^{[O]})^T = (A^{[O-1]})^T W^{[O]} + (\vec{b}^{[O]})^T$$

$\mathbb{R}^{N \times l_O}$ $\mathbb{R}^{N \times l_{O-1}}$ $\mathbb{R}^{l_{O-1} \times l_O}$ $\mathbb{R}^{1 \times l_O}$

$$W^{[2]} = \begin{pmatrix} \uparrow & \uparrow & & \uparrow \\ \vec{w}_1^{[2]} & \vec{w}_2^{[2]} & \cdots & \vec{w}_{l_2}^{[2]} \\ \downarrow & \downarrow & & \downarrow \end{pmatrix} \in \mathbb{R}^{l_1 \times l_2}$$

$$(\vec{b}^{[2]})^T = \begin{pmatrix} b_1^{[2]} & b_2^{[2]} & \cdots & b_{l_2}^{[2]} \end{pmatrix} \in \mathbb{R}^{1 \times l_2}$$

$$(A^{[2]})^T = (A^{[1]})^T W^{[2]} + (\vec{b}^{[2]})^T$$

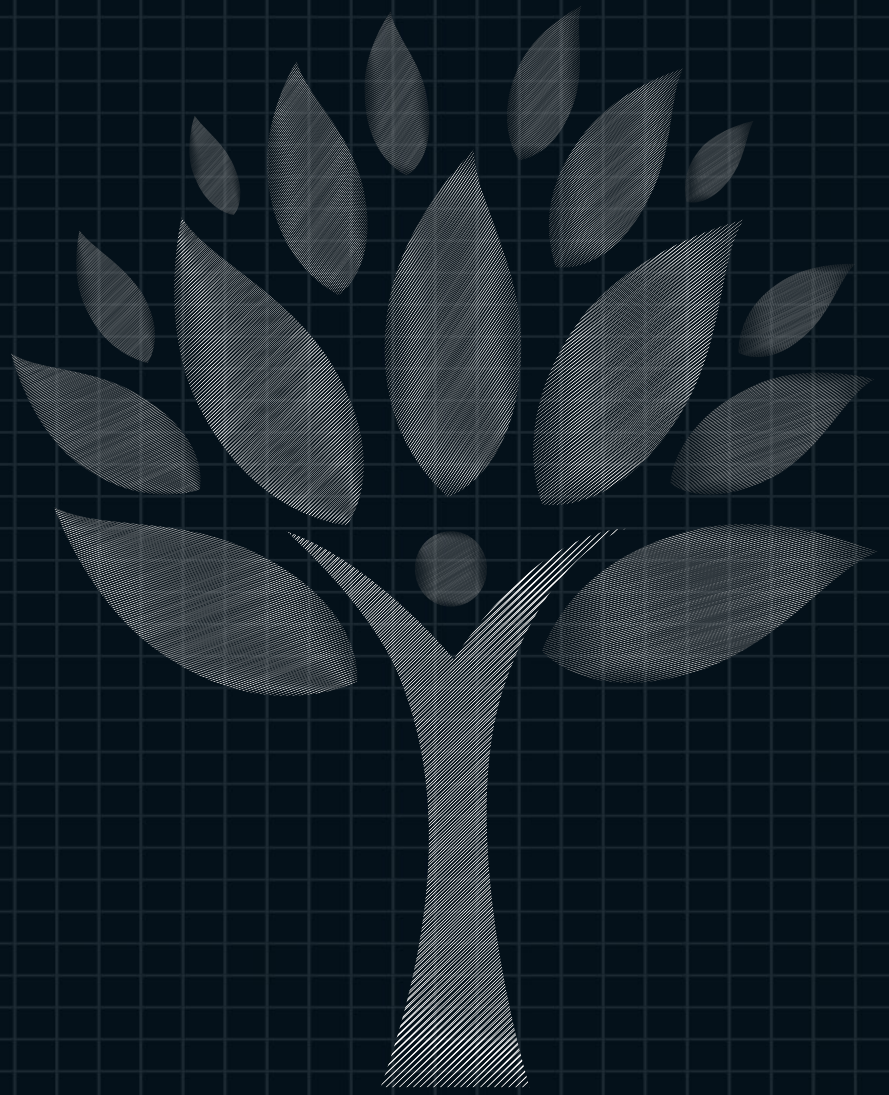
$\mathbb{R}^{N \times l_2}$ $\mathbb{R}^{N \times l_1}$ $\mathbb{R}^{l_1 \times l_2}$ $\mathbb{R}^{1 \times l_2}$

$$W^{[1]} = \begin{pmatrix} \uparrow & \uparrow & & \uparrow \\ \vec{w}_1^{[1]} & \vec{w}_2^{[1]} & \cdots & \vec{w}_{l_1}^{[1]} \\ \downarrow & \downarrow & & \downarrow \end{pmatrix} \in \mathbb{R}^{l_I \times l_1}$$

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

$$(A^{[1]})^T = X^T W^{[1]} + (\vec{b}^{[1]})^T$$

$\mathbb{R}^{N \times l_1}$ $\mathbb{R}^{N \times l_I}$ $\mathbb{R}^{l_I \times l_1}$ $\mathbb{R}^{1 \times l_1}$



Forward Propagation of Neural Networks

Lecture.2
Dense Layers