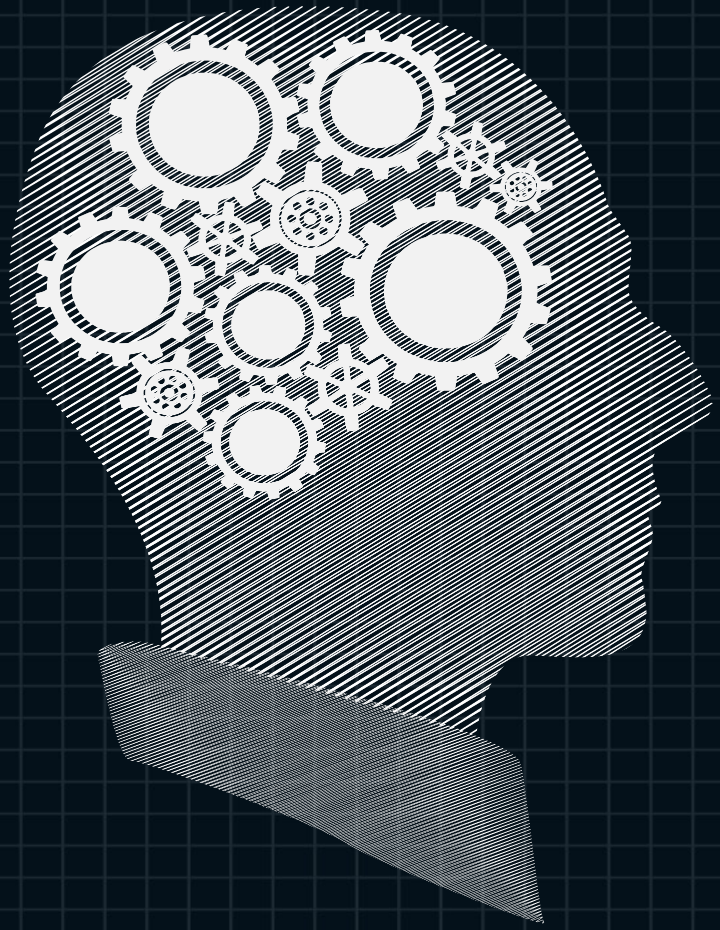


Forward Propagation of Neural Networks

Lecture.1 Artificial Neurons

Lecture.1
Artificial Neurons

- Parametric Functions

output을 만들어내는 데 사용되는 변수를 가지고 있는 함수

$z \in \mathbb{R}$ output

$\uparrow$

$f(x)$

$\uparrow$

$x \in \mathbb{R}$ input

$z \in \mathbb{R}$

$\uparrow$

$f(\underbrace{x}_{\text{입력}}; \underbrace{\theta}_{\text{파라미터}})$

$\uparrow$

$x \in \mathbb{R}$

ex) $\theta = 5$

ex)

$$z = e^x$$

$$z = log(x)$$

$$\vdots$$

$$z = \underline{x} + \theta \quad z = x + 5$$

$$z = \theta x \quad z = 5x$$

$$\vdots$$

Lecture.1 Artificial Neurons

- Hierarchy of Tensor Computations

Scalar, vector, matrix, ...

Zeroth-order Tensor Operations (= Scalar)	First-order Tensor Operations (= Vector)	Second-order Tensor Operations (= Matrix)	Third-order Tensor Operations
$a, b \in \mathbb{R}$	$a \in \mathbb{R}$ $\vec{u}, \vec{v} \in \mathbb{R}^n$	$a \in \mathbb{R}$ $\vec{u} \in \mathbb{R}^n$ $M, N \in \mathbb{R}^{m \times n}, O \in \mathbb{R}^{n \times o}$	$a \in \mathbb{R}$ $M, N \in \mathbb{R}^{m \times n \times o}$
$a + b : \mathbb{R} \times \mathbb{R} \to \mathbb{R}$ $a \cdot b : \mathbb{R} \times \mathbb{R} \to \mathbb{R}$	$a\vec{u} : \mathbb{R} \times \mathbb{R}^n \to \mathbb{R}^n$ $\vec{u} + \vec{v} : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}^n$ $\vec{u} \bigcirc \vec{v} : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}^n$ $(\vec{u})^T \cdot \vec{v} : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$	$aM : \mathbb{R} \times \mathbb{R}^{m \times n} \to \mathbb{R}^{m \times n}$ $M + N : \mathbb{R}^{m \times n} \times \mathbb{R}^{m \times n} \to \mathbb{R}^{m \times n}$ $M \bigcirc N : \mathbb{R}^{m \times n} \times \mathbb{R}^{m \times n} \to \mathbb{R}^{m \times n}$ $M \cdot \vec{u} : \mathbb{R}^{m \times n} \times \mathbb{R}^n \to \mathbb{R}^m$ $MO : \mathbb{R}^{m \times n} \times \mathbb{R}^{n \times o} \to \mathbb{R}^{m \times o}$	$aM : \mathbb{R} \times \mathbb{R}^{m \times n \times o} \to \mathbb{R}^{m \times n \times o}$ $M + N : \mathbb{R}^{m \times n \times o} \times \mathbb{R}^{m \times n \times o} \to \mathbb{R}^{m \times n \times o}$ $M \bigcirc N : \mathbb{R}^{m \times n \times o} \times \mathbb{R}^{m \times n \times o} \to \mathbb{R}^{m \times n \times o}$

$$\begin{pmatrix} 2 \\ 3 \\ 4 \end{pmatrix} \circ \begin{pmatrix} 10 \\ 20 \\ 30 \end{pmatrix} = \begin{pmatrix} 20 \\ 60 \\ 120 \end{pmatrix}$$

Lecture.1
Artificial Neurons

- Dataset(X Data)

$$\vec{x}^T = \boxed{\begin{pmatrix} x_1 & x_2 & \dots & x_{l_I} \end{pmatrix}}$$

$$\left(\vec{x}^{(1)}\right)^T = \begin{pmatrix} x_1^{(1)} & x_2^{(1)} & \dots & x_{l_I}^{(1)} \end{pmatrix}$$

$$\left(\vec{x}^{(2)}\right)^T = \begin{pmatrix} x_1^{(2)} & x_2^{(2)} & \dots & x_{l_I}^{(2)} \end{pmatrix}$$

$$\vdots$$

$$\left(\vec{x}^{(N)}\right)^T = \begin{pmatrix} x_1^{(N)} & x_2^{(N)} & \dots & x_{l_I}^{(N)} \end{pmatrix}$$

수학성적 ← 공부 출퇴근시간 수면시간

$\vec{x}$

$$\vec{x} = \begin{pmatrix} \text{공부} \\ \text{출퇴근시간} \\ \text{수면시간} \end{pmatrix}$$

	공부	수면	용돈
학생 1	a	a	a
2	a	a	a
3	a	a	a
4	a	a	a

Lecture.1
Artificial Neurons

- Dataset(X Data)

$$\vec{x}^T = \begin{pmatrix} x_1 & x_2 & \cdots & x_{l_I} \end{pmatrix}$$

$$X^T = \begin{pmatrix} x^{(1)} \\ x^{(2)} \\ \vdots \\ x^{(N)} \end{pmatrix} \in \mathbb{R}^{N \times 1}$$

$$X^T = \begin{pmatrix} \longleftarrow & \left(\vec{x}^{(1)}\right)^T & \longrightarrow \\ \longleftarrow & \left(\vec{x}^{(2)}\right)^T & \longrightarrow \\ & \vdots & \\ \longleftarrow & \left(\vec{x}^{(N)}\right)^T & \longrightarrow \end{pmatrix} = \begin{pmatrix} x_1^{(1)} & x_2^{(1)} & \cdots & x_{n_I}^{(1)} \\ x_1^{(2)} & x_2^{(2)} & \cdots & x_{n_I}^{(2)} \\ \vdots & \vdots & \ddots & \vdots \\ x_1^{(N)} & x_2^{(N)} & \cdots & x_{n_I}^{(N)} \end{pmatrix}$$

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

$$(\vec{x})^T = (x_1, x_2, x_3, \cdots x_{l_I})$$

$$(\vec{x}^{(1)})^T = (x_1^{(1)} \quad x_2^{(1)} \quad x_3^{(1)} \quad \cdots \quad x_{l_I}^{(1)})$$

$$(\vec{x}^{(2)})^T = (x_1^{(2)} \quad x_2^{(2)} \quad x_3^{(2)} \quad \cdots \quad x_{l_I}^{(2)})$$

$$\vdots$$

$$(\vec{x}^{(N)})^T = (x_1^{(N)} \quad x_2^{(N)} \quad \cdots \quad x_{l_I}^{(N)})$$

$$X^T = \begin{pmatrix} x^{(1)} \\ x^{(2)} \\ \vdots \\ x^{(N)} \end{pmatrix} \qquad X^T = \begin{pmatrix} \leftarrow & (\vec{x}^{(1)})^T & \rightarrow \\ \leftarrow & (\vec{x}^{(2)})^T & \rightarrow \\ & \vdots & \\ \leftarrow & (x^{(N)})^T & \rightarrow \end{pmatrix}$$

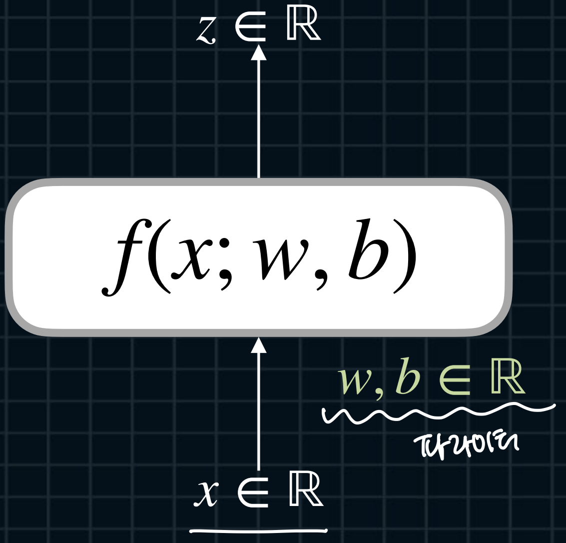
- Affine Functions with One Feature

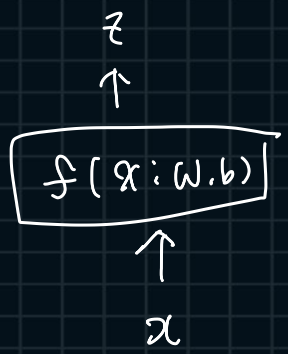
Weighted Sum

$$z = xw$$

Affine Transformation

$$z = xw + b$$





$z = x \cdot w + b$

Lecture.1
Artificial Neurons

- Affine Functions with n Features

Weighted Sum

$$z = \underset{10}{w_1}\underset{키}{x_1} + \underset{100}{w_2}\underset{몸무게}{x_2} + \ldots + \underset{-20}{w_n}\underset{나이}{x_n} = (\vec{w})^T\vec{x} = (\vec{x})^T\vec{w}$$

$$\vec{x} = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} \qquad \vec{w} = \begin{pmatrix} w_1 \\ w_2 \\ \vdots \\ w_n \end{pmatrix}$$

Affine Transformation

$$z = w_1x_1 + w_2x_2 + \ldots + w_nx_n + b = (\vec{x})^T\vec{w} + b$$

$z \in \mathbb{R}$

$\uparrow$

$$\boxed{f((\vec{x})^T; \vec{w}, b)}$$

low vector

$\uparrow$

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

$$\vec{w} = \begin{pmatrix} w_1 \\ w_2 \\ \vdots \\ w_{l_I} \end{pmatrix} \in \mathbb{R}^{l_I \times 1}, \quad b \in \mathbb{R}$$

$(\vec{x})^T$의 차원 = $\vec{w}$의 차원

Lecture.1 Artificial Neurons

- Activation Functions

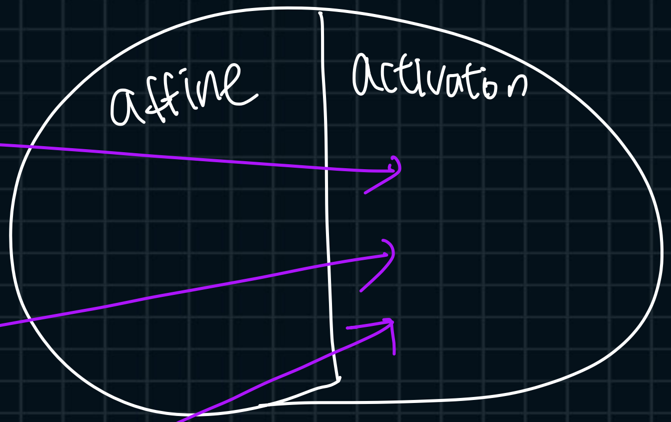
#activation function

Sigmoid $$g(x) = \sigma(x) = \frac{1}{1 + e^{-x}}$$

Tanh $$g(x) = tanh(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}}$$

ReLU $$g(x) = ReLU(x) = max(0, x)$$

기본적으로 [illegible] 사용!



Lecture.1 Artificial Neurons

- Artificial Neurons

$$a \in \mathbb{R}$$

$$g(z)$$

$$z \in \mathbb{R}$$

$$f\left((\vec{x})^T; \vec{w}, b\right)$$

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

activation value

$$a = g(z) = \frac{1}{1 + e^{-z}}$$

$$z = (\vec{x})^T \vec{w} + b$$

$$\Longleftrightarrow$$

$$a$$

$$\nu$$

$$\vec{x}$$

$$\nu(\vec{x}; \vec{w}, b) = g\left((\vec{x})^T \vec{w} + b\right)$$

$$\vec{w} \in \mathbb{R}^{l_I \times 1}$$
$$b \in \mathbb{R}$$

Lecture.1 Artificial Neurons

- Minibatch in Artificial Neurons

$(\vec{x}^{(1)})^T \rightarrow f((\vec{x})^T; \vec{w}, b) \rightarrow a^{(1)} = g\left((\vec{x}^{(1)})^T \cdot \vec{w} + b\right)$

$(\vec{x}^{(2)})^T \rightarrow f((\vec{x})^T; \vec{w}, b) \rightarrow a^{(2)} = g\left((\vec{x}^{(2)})^T \cdot \vec{w} + b\right)$

$\vdots$

$(\vec{x}^{(N)})^T \rightarrow f((\vec{x})^T; \vec{w}, b) \rightarrow a^{(N)} = g\left((\vec{x}^{(N)})^T \cdot \vec{w} + b\right)$

Weight & Bias 는 바뀌지 X

$$X^T = \begin{pmatrix} \longleftarrow & (\vec{x}^{(1)})^T & \longrightarrow \\ \longleftarrow & (\vec{x}^{(2)})^T & \longrightarrow \\ & \vdots & \\ \longleftarrow & (\vec{x}^{(N)})^T & \longrightarrow \end{pmatrix} \longrightarrow f(X^T; \vec{w}, b) \longrightarrow A = \begin{pmatrix} a^{(1)} \\ a^{(2)} \\ \vdots \\ a^{(N)} \end{pmatrix}$$

$X^T \in \mathbb{R}^{N \times l_I}$ $\qquad A \in \mathbb{R}^{N \times 1}$

$$\begin{pmatrix} \leftarrow (x^{(1)})^T \rightarrow \\ \leftarrow (x^{(2)})^T \rightarrow \\ \vdots \\ \leftarrow (x^{(N)})^T \rightarrow \end{pmatrix} \begin{pmatrix} w_1 \\ w_2 \\ \vdots \\ w_n \end{pmatrix} + b = \begin{pmatrix} a \\ a \\ \vdots \\ a \end{pmatrix} = a$$

Lecture.1 Artificial Neurons

- Minibatch in Artificial Neurons

Minibatch Input

$$(\vec{x}^{(1)})^T \in \mathbb{R}^{1\times l_I}$$
$$(\vec{x}^{(2)})^T \in \mathbb{R}^{1\times l_I}$$
$$\vdots$$
$$(\vec{x}^{(N)})^T \in \mathbb{R}^{1\times l_I}$$

Weight/Bias

$$\vec{w} = \begin{pmatrix} w_1 \\ w_2 \\ \vdots \\ w_{l_I} \end{pmatrix} \in \mathbb{R}^{l_I\times 1}$$
$$b \in \mathbb{R}$$

Affine Function

$$z^{(1)} = (\vec{x}^{(1)})^T \cdot \vec{w} + b$$
$$z^{(2)} = (\vec{x}^{(2)})^T \cdot \vec{w} + b$$
$$\vdots$$
$$z^{(N)} = (\vec{x}^{(N)})^T \cdot \vec{w} + b$$

Activation Function

$$a^{(1)} = g\left((\vec{x}^{(1)})^T \cdot \vec{w} + b\right)$$
$$a^{(2)} = g\left((\vec{x}^{(2)})^T \cdot \vec{w} + b\right)$$
$$\vdots$$
$$a^{(N)} = g\left((\vec{x}^{(N)})^T \cdot \vec{w} + b\right)$$

$$X^T = \begin{pmatrix} \longleftarrow & (\vec{x}^{(1)})^T & \longrightarrow \\ \longleftarrow & (\vec{x}^{(2)})^T & \longrightarrow \\ & \vdots & \\ \longleftarrow & (\vec{x}^{(N)})^T & \longrightarrow \end{pmatrix} \in \mathbb{R}^{N\times l_I}$$

$$\vec{z} = X^T\vec{w} + b$$

$$Z = \begin{pmatrix} z^{(1)} \\ z^{(2)} \\ \vdots \\ z^{(N)} \end{pmatrix} \in \mathbb{R}^{N\times 1}$$

$$A = \begin{pmatrix} a^{(1)} \\ a^{(2)} \\ \vdots \\ a^{(N)} \end{pmatrix} \in \mathbb{R}^{N\times 1}$$

Lecture.1
Artificial Neurons

- Minibatch in Artificial Neurons

$A \in \mathbb{R}^{N\times 1}$

$g(z)$

$Z \in \mathbb{R}^{N\times 1}$

$f(X^T; \vec{w}, b)$

$\vec{w} \in \mathbb{R}^{l_I \times 1}$

$b \in \mathbb{R}$

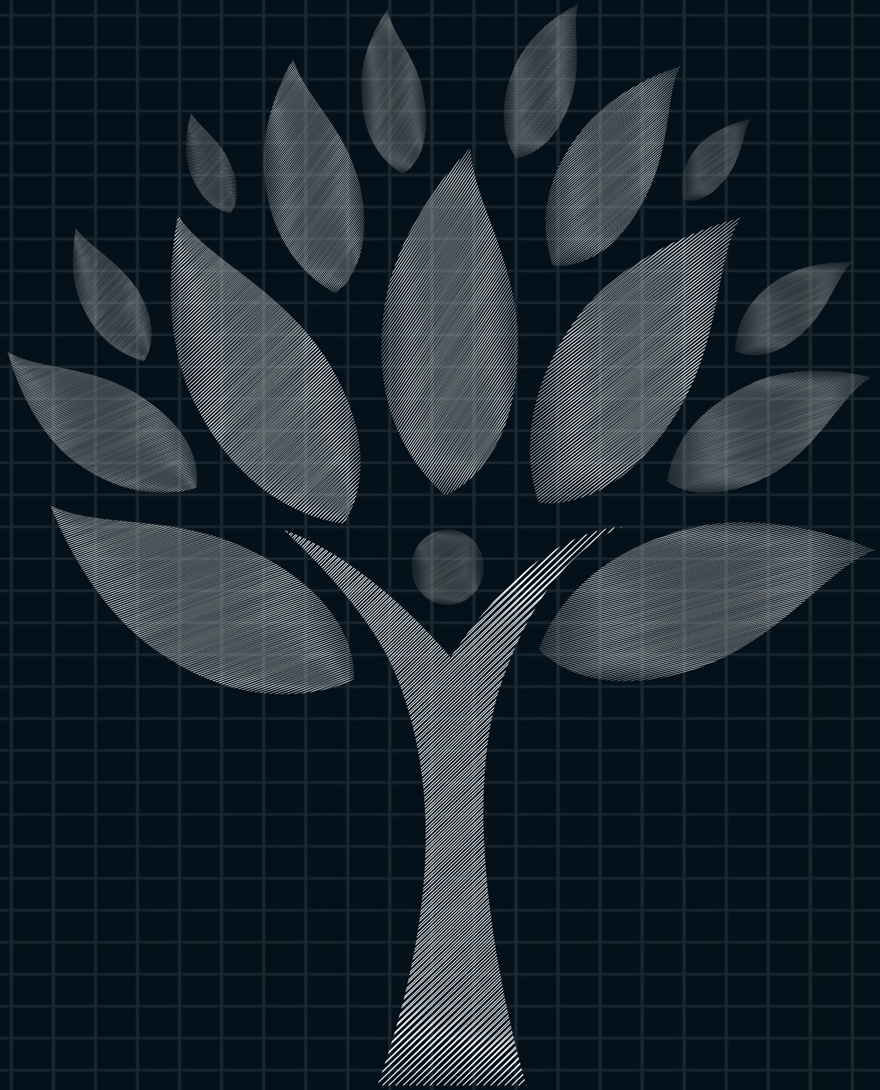
$X^T \in \mathbb{R}^{N\times l_I}$

$$A = \begin{pmatrix} g\left((\vec{x}^{(1)})^T \cdot \vec{w} + b\right) \\ g\left((\vec{x}^{(2)})^T \cdot \vec{w} + b\right) \\ \vdots \\ g\left((\vec{x}^{(N)})^T \cdot \vec{w} + b\right) \end{pmatrix}$$

$$Z = \begin{pmatrix} (\vec{x}^{(1)})^T \cdot \vec{w} + b \\ (\vec{x}^{(2)})^T \cdot \vec{w} + b \\ \vdots \\ (\vec{x}^{(N)})^T \cdot \vec{w} + b \end{pmatrix}$$

$$\underbrace{(\vec{x})^T \cdot \vec{w}}_{\downarrow} + b$$

$$X^T \cdot \vec{w} + b$$



Forward Propagation of Neural Networks

Lecture.1
Artificial Neurons