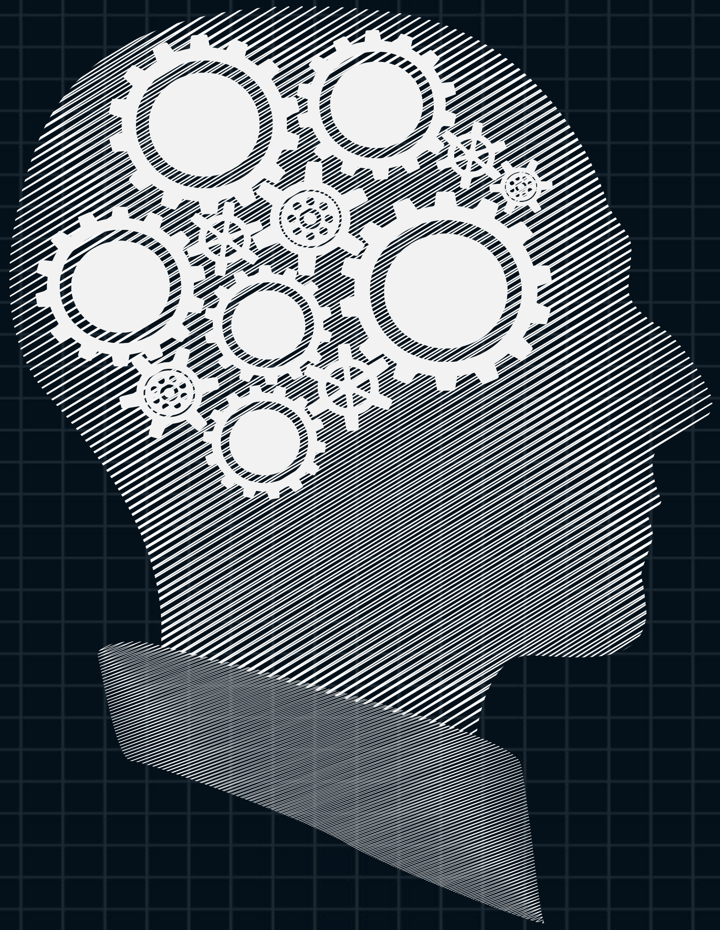


Forward Propagation of Neural Networks

Lecture.1 Artificial Neurons

- Parametric Functions

$$x \in \mathbb{R} \rightarrow f(x) \rightarrow z \in \mathbb{R}$$

$$z = e^x$$
$$z = log(x)$$
$$\vdots$$

$$x \in \mathbb{R} \rightarrow f(x;\theta) \rightarrow z \in \mathbb{R}$$

$$z = x + \theta$$
$$z = \theta x$$
$$\vdots$$

Lecture.1 Artificial Neurons - Hierarchy of Tensor Computations

Zeroth-order Tensor Operations

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

$$a + b : \mathbb{R} \times \mathbb{R} \to \mathbb{R}$$
$$a \cdot b : \mathbb{R} \times \mathbb{R} \to \mathbb{R}$$

First-order Tensor Operations

$$a \in \mathbb{R}$$
$$\overrightarrow{u}, \overrightarrow{v} \in \mathbb{R}^n$$

$$a\overrightarrow{u} : \mathbb{R} \times \mathbb{R}^n \to \mathbb{R}^n$$
$$\overrightarrow{u} + \overrightarrow{v} : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}^n$$
$$\overrightarrow{u} \bigcirc \overrightarrow{v} : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}^n$$
$$(\overrightarrow{u})^T \cdot \overrightarrow{v} : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$$

Second-order Tensor Operations

$$a \in \mathbb{R} \quad \overrightarrow{u} \in \mathbb{R}^n$$
$$M, N \in \mathbb{R}^{m \times n}, O \in \mathbb{R}^{n \times o}$$

$$aM : \mathbb{R} \times \mathbb{R}^{m \times n} \to \mathbb{R}^{m \times n}$$
$$M + N : \mathbb{R}^{m \times n} \times \mathbb{R}^{m \times n} \to \mathbb{R}^{m \times n}$$
$$M \bigcirc N : \mathbb{R}^{m \times n} \times \mathbb{R}^{m \times n} \to \mathbb{R}^{m \times n}$$
$$M \cdot \overrightarrow{u} : \mathbb{R}^{m \times n} \times \mathbb{R}^n \to \mathbb{R}^m$$
$$MO : \mathbb{R}^{m \times n} \times \mathbb{R}^{n \times o} \to \mathbb{R}^{m \times o}$$

Third-order Tensor Operations

$$a \in \mathbb{R}$$
$$M, N \in \mathbb{R}^{m \times n \times o}$$

$$aM : \mathbb{R} \times \mathbb{R}^{m \times n \times o} \to \mathbb{R}^{m \times n \times o}$$
$$M + N : \mathbb{R}^{m \times n \times o} \times \mathbb{R}^{m \times n \times o} \to \mathbb{R}^{m \times n \times o}$$
$$M \bigcirc N : \mathbb{R}^{m \times n \times o} \times \mathbb{R}^{m \times n \times o} \to \mathbb{R}^{m \times n \times o}$$

Lecture.1
Artificial Neurons

- Dataset(X Data)

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

$$\left(\overrightarrow{x}^{(1)}\right)^T = \left(x_1^{(1)} \quad x_2^{(1)} \quad \cdots \quad x_{l_I}^{(1)}\right)$$

$$\left(\overrightarrow{x}^{(2)}\right)^T = \left(x_1^{(2)} \quad x_2^{(2)} \quad \cdots \quad x_{l_I}^{(2)}\right)$$

$$\vdots$$

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

Lecture.1
Artificial Neurons

- Dataset(X Data)

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

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

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

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

Lecture.1 Artificial Neurons

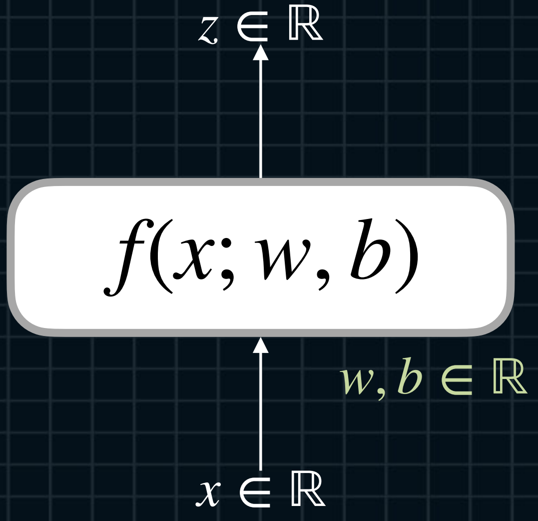
- Affine Functions with One Feature

Weighted Sum

$$z = xw$$

Affine Transformation

$$z = xw + b$$



- Affine Functions with n Features

Weighted Sum

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

Affine Transformation

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

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

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

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

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

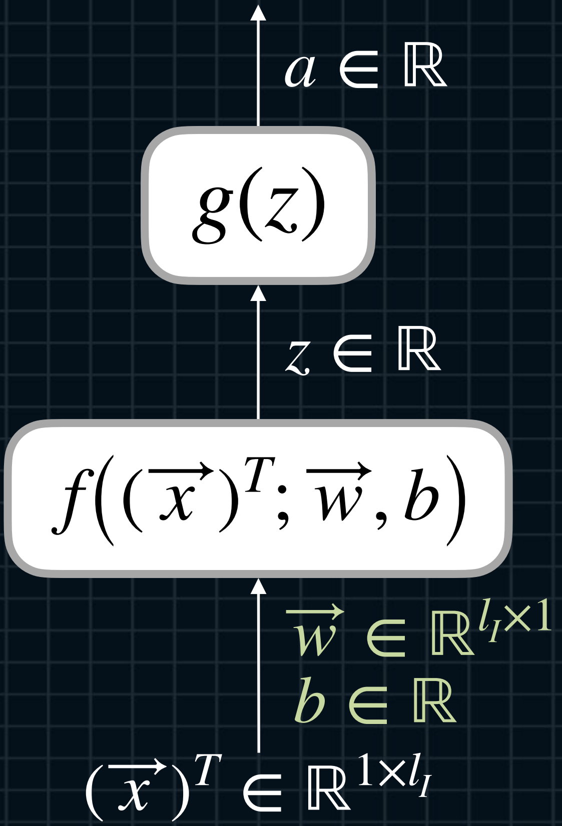
- Activation Functions

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

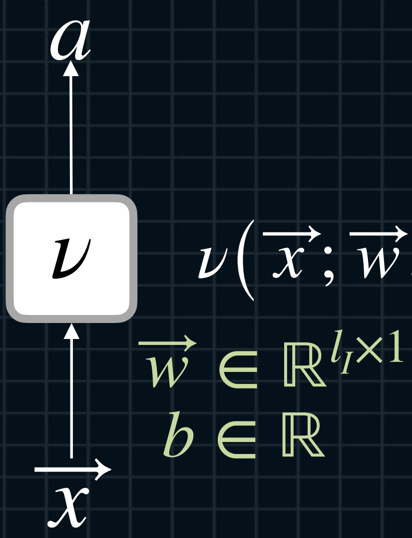
- Artificial Neurons



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

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

$$\Longleftrightarrow$$



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

Lecture.1 Artificial Neurons

- Minibatch in Artificial Neurons

$$(\vec{x}^{(1)})^T, \ (\vec{x}^{(2)})^T, \ \vdots, \ (\vec{x}^{(N)})^T \longrightarrow \boxed{f\big((\vec{x})^T; \vec{w}, b\big)}$$

$$a^{(1)} = g\big((\vec{x}^{(1)})^T \cdot \vec{w} + b\big)$$

$$a^{(2)} = g\big((\vec{x}^{(2)})^T \cdot \vec{w} + b\big)$$

$$\vdots$$

$$a^{(N)} = g\big((\vec{x}^{(N)})^T \cdot \vec{w} + b\big)$$

$$\underbrace{X^T = \begin{pmatrix} \longleftarrow & (\vec{x}^{(1)})^T & \longrightarrow \\ \longleftarrow & (\vec{x}^{(2)})^T & \longrightarrow \\ & \vdots & \\ \longleftarrow & (\vec{x}^{(N)})^T & \longrightarrow \end{pmatrix}}_{\in \mathbb{R}^{N \times l_I}} \longrightarrow \boxed{f\big(X^T; \vec{w}, b\big)} \longrightarrow \underbrace{A = \begin{pmatrix} a^{(1)} \\ a^{(2)} \\ \vdots \\ a^{(N)} \end{pmatrix}}_{\in \mathbb{R}^{N \times 1}}$$

Lecture.1 Artificial Neurons

- Minibatch in Artificial Neurons

Minibatch Input	Weight/Bias	Affine Function	Activation Function
$(\vec{x}^{(1)})^T \in \mathbb{R}^{1 \times l_I}$		$z^{(1)} = (\vec{x}^{(1)})^T \cdot \vec{w} + b$	$a^{(1)} = g\left((\vec{x}^{(1)})^T \cdot \vec{w} + b\right)$
$(\vec{x}^{(2)})^T \in \mathbb{R}^{1 \times l_I}$	$\vec{w} = \begin{pmatrix} w_1 \\ w_2 \\ \vdots \\ w_{l_I} \end{pmatrix} \in \mathbb{R}^{l_I \times 1}$	$z^{(2)} = (\vec{x}^{(2)})^T \cdot \vec{w} + b$	$a^{(2)} = g\left((\vec{x}^{(2)})^T \cdot \vec{w} + b\right)$
$\vdots$		$\vdots$	$\vdots$
$(\vec{x}^{(N)})^T \in \mathbb{R}^{1 \times l_I}$	$b \in \mathbb{R}$	$z^{(N)} = (\vec{x}^{(N)})^T \cdot \vec{w} + b$	$a^{(N)} = g\left((\vec{x}^{(N)})^T \cdot \vec{w} + b\right)$

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

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

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

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

Lecture.1 Artificial Neurons

- Minibatch in Artificial Neurons

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

$g(z)$

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

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

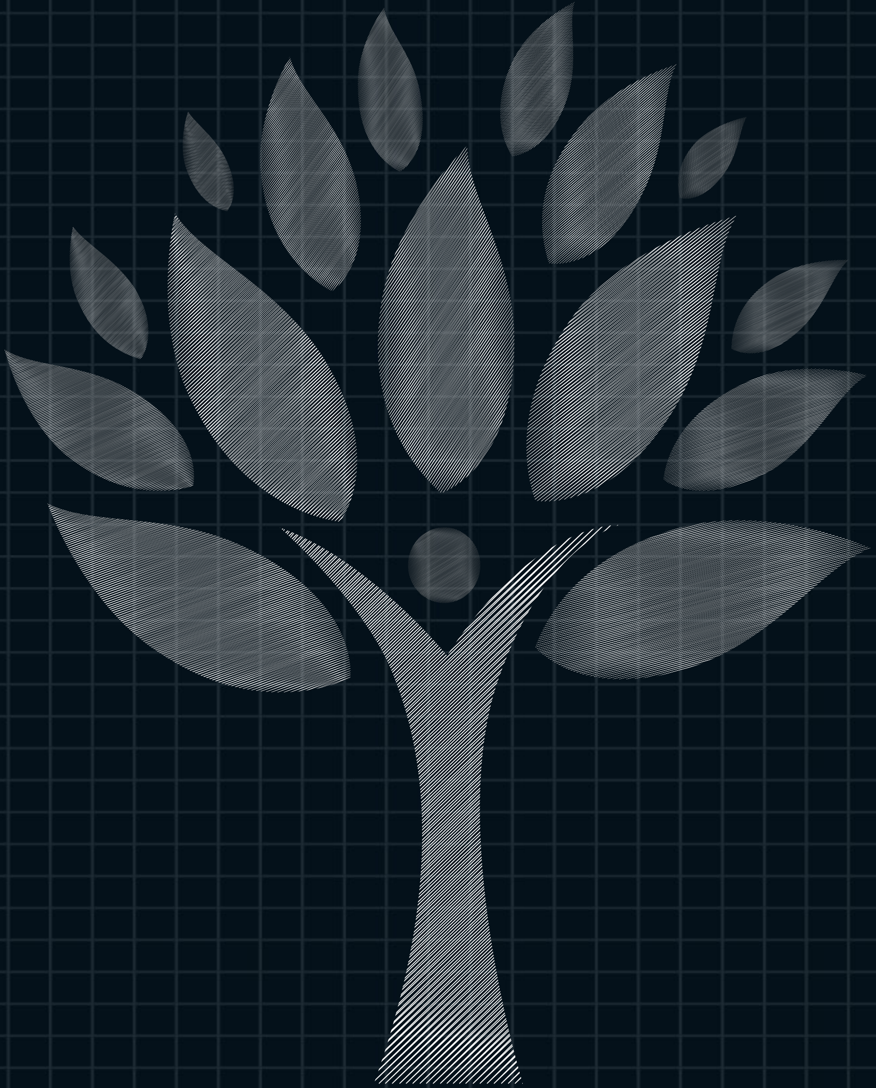
$\overrightarrow{w} \in \mathbb{R}^{l_I \times 1}$

$b \in \mathbb{R}$

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

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

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



Forward Propagation of Neural Networks

Lecture.1
Artificial Neurons