교과목: 딥러닝 수학

[Mathematics for Machine Learning]

© KOPO-Dept.
AI Engineering Limited

Daegu Campus of
KOREA POLYTECHNIC

목표

- 딥러닝에 국한된 것이 아니라 머신러닝에 대한 전반적인 이해를 위해 필요한 수학 지식 습득
  - **선형대수, 미분, 확률 통계**와 머신러닝과(딥러닝 포함)의 관계
  - 가능한 예제들을 바탕으로 설명

- 수학 문제를 잘 풀기 위한 것이 아니라 이해하고 익숙해지자

John von Neumann

"Young man, in mathematics you don't understand things, you just get used to them."

in reply to Felix Smith who had said "I'm afraid I don't understand the method of characteristics."

© KOPO-Dept.
AI Engineering Limited

AI(Artificial Intelligence) 6개 분야

- **Knowledge representation** to store what it knows or hears
- **Automated reasoning** to use the stored information to answer questions and to draw new conclusions

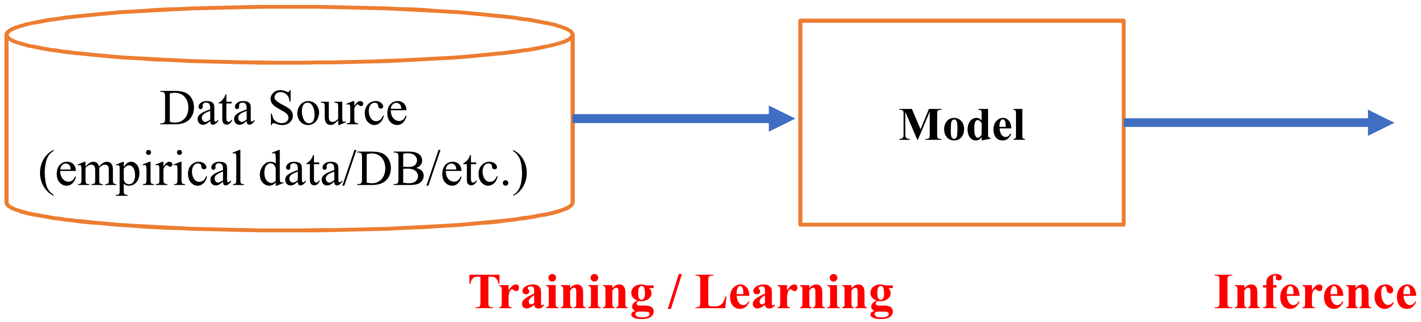
---

- **Natural language processing** to enable it to communicate successfully
- **Computer vision** to perceive objects
- **Robotics** to manipulate objects and move about
- **Machine learning** to adapt to new circumstances and to detect and extrapolate patterns

Stuart Russell and Peter Norvig, "Artificial Intelligence: A Modern Approach," Pearson.

© KOPO-Dept.
AI Engineering Limited

Machine Learning



- Machine Learning: A scientific discipline that is concerned with the design and development of algorithms that allow computers **to learn(train)** from empirical data (sensor data or database) and to **make predictions**.
- [컴퓨터가 경험적인 데이터(정제된 데이터, 활용 가능한 데이터[센서 데이터, DB])로 부터 **학습하고 예측**할 수 있도록 <알고리즘을 설계하고 개발하는 과학 분야>]

© KOPO-Dept.
AI Engineering Limited

Machine Learning

- **3가지 중요한 요소**

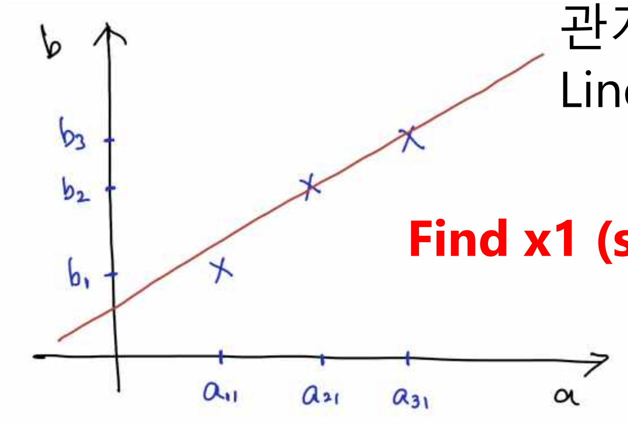
ML		Relevant Maths.
Data - Images, text, languages, time series, etc.	데이터를 효과적으로 표현	**Data** - **Vectors, matrix and tensors**
Models - Linear models - Deep neural networks - Kernel machine - Probabilistic models	모델을 체계적으로 공부	**Models** - **Linear algebra** - **Probability**
Training & Inference - Optimization - Regularization - Batch or online - Generalization		**Training & Inference** - **Optimization (Gradient Descent algorithm, backpropagation: 미분)** - Estimation - Information theory - Matrix factorization

Linear Algebra

- Motivation Example
  - **집의 크기와 가격과의 관계 도출**

$$
\begin{aligned}
a_{11}x_1 + x_2 &= b_1, \\
a_{21}x_1 + x_2 &= b_2, \\
a_{31}x_1 + x_2 &= b_3,
\end{aligned}
\quad \Rightarrow \quad
\begin{bmatrix} a_{11} & 1 \\ a_{21} & 1 \\ a_{31} & 1 \end{bmatrix}
\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}
=
\begin{bmatrix} b_1 \\ b_2 \\ b_3 \end{bmatrix}.
$$

$\mathbf{a_{i1}}$ represents the **size** of house i and $\mathbf{b_i}$ corresponds to the **price** of house i.



관계를 선으로 잘 설명할 수 있는가?
Linear Regression

Find x1 (slope) and x2 (intercept)

Linear Algebra

- Motivation Example
  - **집의 크기 + α (주변 상점 수)와 가격과의 관계 도출**

$$\begin{bmatrix} a_{11} & a_{12} & 1 \\ a_{21} & a_{22} & 1 \\ a_{31} & a_{32} & 1 \\ a_{41} & a_{42} & 1 \\ a_{51} & a_{52} & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} b_1 \\ b_2 \\ b_3 \\ b_4 \\ b_5 \end{bmatrix}$$

Modeling: Linear Equation 만드는 것

ML: Model Parameter

모델링(Linear equation from data[matrix a, vector b]) → solve/find vector x

- **Linear regression involves solving the linear equation Ax = b.**

© KOPO-Dept.
AI Engineering Limited

Linear Algebra

- **Two Important Equations in Linear Algebra**

$$\boxed{\mathbf{Ax} = \mathbf{b}} \text{ (linear equation)}$$

$$\boxed{\mathbf{Ax} = \lambda \mathbf{x}} \text{ (eigenvalue equation)}$$

- **Why Linear Algebra for ML?**
  - **Data are presented in the form of vectors and matrices**
  - **Model relates input with output, which is often represented by linear equations, which are presented in the form of vectors and matrices.**
    - 관심 가지는 것: **Linear model: Linear transformation**
      - 선형함수

Daegu Campus of KOREA POLYTECHNIC

Linear Algebra

- **Objects in Linear Algebra**

Scalar ($\mathbb{R}$): x

Vector ($\mathbb{R}^n$): $\mathbf{x}$ (x_i)

Matrix ($\mathbb{R}^{m \times n}$): $\mathbf{X}$ ($X_{i,j}$)

Tensor ($\mathbb{R}^{m \times n \times k \times \cdots}$): $\mathbf{X}$ ($X_{i,j,k}$)

- **Outline**

Vectors, vector spaces, and matrices

Inverse and transpose

Linear independence and rank

Linear equations (range space and null space)

Norm and inner product

Orthogonality and rotation

Projection

진행 예정

© KOPO-Dept.
AI Engineering Limited

Vectors

© ERAE AMS Limited

Linear Algebra - Vector

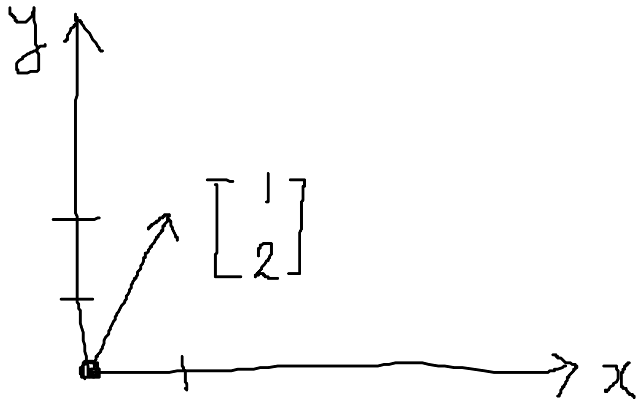
- **Physicist**
  - ▶ Arrow (with direction and length)
- **Computer scientist**
  - ▶ List of numbers

사람을 표현하는
3-dimensional vector

$$\begin{bmatrix} 175 \\ 72 \\ \text{남자} \end{bmatrix}$$

$$\begin{bmatrix} 185 \\ 85 \\ \text{남자} \end{bmatrix}$$

- **Mathematician**
  - ▶ Arrow rooted at the origin



Physicist + Computer scientist

© KOPO-Dept.
AI Engineering Limited

Linear Algebra - Vector

- **Basis Vector**

$$\begin{bmatrix} 1 \\ 2 \end{bmatrix} = 1 \underbrace{\boxed{\begin{bmatrix} 1 \\ 0 \end{bmatrix}} + 2 \boxed{\begin{bmatrix} 0 \\ 1 \end{bmatrix}}}_{\text{Basis vector}}$$

- **Column / Row Vector and Transpose**

The **transpose** of a vector $\mathbf{x}$ is denoted by $\mathbf{x}^\top$.

Consider the **column vector** $\mathbf{x}$:

$$\mathbf{x} = \begin{bmatrix} 1 \\ 3 \\ -2 \end{bmatrix}.$$

Then, $\mathbf{x}^\top$ is **row vecotor**, given by

$$\mathbf{x}^\top = [1, \ 3, \ -2].$$

$(x^T)^T = x$

Basis vector: 그 벡터 공간을 선형생성(span)하는 선형독립(linear independence)인 벡터

Daegu Campus of KOREA POLYTECHNIC

Linear Algebra – Vector Space

- **Definition**

A **vector space**, denoted by $\mathcal{V}$, is a collection of objects, called **vectors**, which is **closed** under **vector addition (+)** and **scalar multiplication ($\cdot$)**, such that the following axioms hold for any $\mathbf{x}, \mathbf{y}, \mathbf{z} \in \mathcal{V}$ and $\alpha, \beta \in \mathbb{R}$:

$$\boxed{\mathbf{x} + \mathbf{y} \in \mathcal{V} \text{ and } \alpha\mathbf{x} \in \mathcal{V} \text{ (closed under vector addition and scalar multiplication)}}$$

$\mathbf{x} + \mathbf{y} = \mathbf{y} + \mathbf{x}$ (commutativity) 교환

$\mathbf{x} + (\mathbf{y} + \mathbf{z}) = (\mathbf{x} + \mathbf{y}) + \mathbf{z}$ (associativity) 결합

$\alpha(\mathbf{x} + \mathbf{y}) = \alpha\mathbf{x} + \alpha\mathbf{y}$ and $(\alpha + \beta)\mathbf{x} = \alpha\mathbf{x} + \beta\mathbf{x}$ (distributivity) 분배

$\exists \mathbf{0} \in \mathcal{V}$ such that $\mathbf{0} + \mathbf{x} = \mathbf{x}$ and $1\mathbf{x} = \mathbf{x}$ (identity element)

for any $\mathbf{x} \in \mathcal{V}$, $\exists(-\mathbf{x}) \in \mathcal{V}$ such that $\mathbf{x} + (-\mathbf{x}) = 0$ (inverse element)

$\alpha(\beta\mathbf{x}) = (\alpha\beta)\mathbf{x}$

vector space: 어떤 vector들을 모아 놓은 set인데,
“+”연산과 “scalar multiplication”에 닫혀 있다.

© KOPO-Dept.
AI Engineering Limited

Linear Algebra – Vector Space

- **Polynomials are vectors?**
  - **A polynomial of degree n is described as:**

$$f(x) = a_0 + a_1x + a_2x^2 + \cdots + a_nx^n.$$

$$\begin{bmatrix} a_0 \\ a_1 \\ a_2 \\ \vdots \\ a_n \end{bmatrix}^T \times \begin{bmatrix} 1 \\ x \\ x^2 \\ \vdots \\ x^n \end{bmatrix} = b$$

Linear Algebra – Vector Space

- **Polynomials are vectors?**

For $f, g \in \mathcal{V}$, $\underline{f + g \in \mathcal{V}}$?

$$
\begin{aligned}
& f(x) + g(x) \\
& \quad = \left(a_0 + a_1 x + a_2 x^2 + \cdots + a_n x^n\right) + \left(b_0 + b_1 x + b_2 x^2 + \cdots + b_n x^n\right) \\
& \quad = (a_0 + b_0) + (a_1 + b_1)x + (a_2 + b_2)x^2 + \cdots + (a_n + b_n)x^n \in \mathcal{V}.
\end{aligned}
$$

For $f \in \mathcal{V}$ and $\alpha \in \mathbb{R}$, $\underline{\alpha f \in \mathcal{V}}$?

$$
\begin{aligned}
\alpha f(x) &= \alpha\left(a_0 + a_1 x + a_2 x^2 + \cdots + a_n x^n\right) \\
&= \alpha a_0 + \alpha a_1 x + \alpha a_2 x^2 + \cdots + \alpha a_n x^n \in \mathcal{V}.
\end{aligned}
$$

© KOPO-Dept.
AI Engineering Limited

Linear Algebra – Vector Space

- **R^3 is a vector space?**

Consider two (column) vectors $\mathbf{x}$ and $\mathbf{y}$:

$$\mathbf{x} = \begin{bmatrix} 1 \\ 3 \\ -2 \end{bmatrix} \in \mathbb{R}^3, \quad \mathbf{y} = \begin{bmatrix} -1 \\ 1 \\ 0 \end{bmatrix} \in \mathbb{R}^3.$$

Add two vectors with each scaled:

$$2\mathbf{x} + \mathbf{y} = \begin{bmatrix} 2 \\ 6 \\ -4 \end{bmatrix} + \begin{bmatrix} -1 \\ 1 \\ 0 \end{bmatrix} = \begin{bmatrix} 1 \\ 7 \\ -4 \end{bmatrix} \in \mathbb{R}^3.$$

© KOPO-Dept.
AI Engineering Limited

Matrices

Linear Algebra – Matrix

- **Rows and Columns**

$$\mathbf{A} \in \mathbb{R}^{m\times n} = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix} = [\mathbf{a}_1, \mathbf{a}_2, \ldots, \mathbf{a}_n],$$

where $\mathbf{a}_i \in \mathbb{R}^m$ for $i = 1, \ldots, n$.

The **transpose** of the matrix $\mathbf{A}$, denoted by $\mathbf{A}^\top \in \mathbb{R}^{n\times m}$, is given by

$$\mathbf{A}^\top = \begin{bmatrix} a_{11} & a_{21} & \cdots & a_{m1} \\ a_{12} & a_{22} & \cdots & a_{m2} \\ \vdots & \vdots & & \vdots \\ a_{1n} & a_{2n} & \cdots & a_{mn} \end{bmatrix} = \begin{bmatrix} \mathbf{a}_1^\top \\ \mathbf{a}_2^\top \\ \vdots \\ \mathbf{a}_n^\top \end{bmatrix}.$$

Linear Algebra – Matrix

- **Scalar-Matrix Operation:**

$$2\begin{bmatrix} 1 & -1 \\ 3 & 5 \end{bmatrix} =$$

- **Matrix-Vector Multiplication:**

$$\begin{bmatrix} 1 & -1 \\ -1 & 2 \\ 3 & 5 \end{bmatrix}\begin{bmatrix} x \\ y \end{bmatrix} = \qquad \begin{bmatrix} 1 & -1 & 0 \\ 2 & 3 & 4 \end{bmatrix}\begin{bmatrix} x \\ y \\ z \end{bmatrix} =$$

Daegu Campus of KOREA POLYTECHNIC

Linear Algebra – Matrix

- **Matrix-Matrix Addition:**

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix} + \begin{bmatrix} w & x \\ y & z \end{bmatrix} =$$

- **Matrix-Matrix Multiplication:**

$$\begin{bmatrix} 1 & -1 & 0 \\ 2 & 3 & 4 \end{bmatrix} \begin{bmatrix} 1 & -1 \\ -1 & 2 \\ 3 & 5 \end{bmatrix} =$$

 © KOPO-Dept.
AI Engineering Limited

Linear Algebra – Matrix

- **Inverse(A^{-1}) and Transpose (A^T)**

Definition

Consider a square matrix $\mathbf{A} \in \mathbb{R}^{n \times n}$. The matrix $\mathbf{B} \in \mathbb{R}^{n \times n}$, which satisfies $\mathbf{AB} = \mathbf{BA} = \mathbf{I}$, is called the **inverse** and denoted by $\mathbf{A}^{-1}$.

Definition

For $\mathbf{A} \in \mathbb{R}^{m \times n}$, the matrix $\mathbf{B} \in \mathbb{R}^{n \times m}$ with $b_{ij} = a_{ji}$ is called the **transpose** of $\mathbf{A}$ and denoted by $\mathbf{A}^{\top}$.

© KOPO-Dept.
AI Engineering Limited

Linear Algebra – Matrix

- **Properties of Inverse(A^{-1}) and Transpose (A^T)**

$$
\begin{aligned}
\mathbf{A}\mathbf{A}^{-1} &= \mathbf{A}^{-1}\mathbf{A} = \mathbf{I}, \\
(\mathbf{A}\mathbf{B})^{-1} &= \mathbf{B}^{-1}\mathbf{A}^{-1}, \\
(\mathbf{A}+\mathbf{B})^{-1} &\neq \mathbf{A}^{-1}+\mathbf{B}^{-1}, \\
(\mathbf{A}^{\top})^{\top} &= \mathbf{A}, \\
(\mathbf{A}+\mathbf{B})^{\top} &= \mathbf{A}^{\top}+\mathbf{B}^{\top}, \\
(\mathbf{A}\mathbf{B})^{\top} &= \mathbf{B}^{\top}\mathbf{A}^{\top}.
\end{aligned}
$$

© KOPO-Dept.
AI Engineering Limited

Linear Algebra – Matrix

- **Symmetric Matrix**

Definition

A square matrix $\mathbf{A} \in \mathbb{R}^{n \times n}$ is **symmetric** if $\mathbf{A} = \mathbf{A}^\top$

Example:

$$\mathbf{A} = \begin{bmatrix} 1 & -1 \\ -1 & 5 \end{bmatrix}$$

$$\mathbf{A} = \mathbf{X}\mathbf{X}^\top$$

$$(\mathbf{AB})^\top = \mathbf{B}^\top \mathbf{A}^\top.$$

© KOPO-Dept.
AI Engineering Limited

- **Positive Definite Matrix**
  - 특정한 성질을 가지는 행렬에 대해 양수/음수와 같이 부호를 정의하는 것

Definition
A symmetric matrix $\mathbf{A} \in \mathbb{R}^{n \times n}$ is said to be **positive definite** if

$$\mathbf{x}^\top \mathbf{A} \mathbf{x} > 0, \quad \forall \mathbf{x} \in \mathcal{V} \backslash \{\mathbf{0}\}.$$

Definition
A symmetric matrix $\mathbf{A} \in \mathbb{R}^{n \times n}$ is said to be **positive semidefinite** if

$$\mathbf{x}^\top \mathbf{A} \mathbf{x} \geq 0, \quad \forall \mathbf{x} \in \mathcal{V} \backslash \{\mathbf{0}\}.$$

Example:

$$\mathbf{A} = \begin{bmatrix} 9 & 6 \\ 6 & 5 \end{bmatrix}, \quad \mathbf{x}^\top \mathbf{A} \mathbf{x} = 9x_1^2 + 12x_1x_2 + 5x_2^2 = (3x_1 + 2x_2)^2 + x_2^2 > 0.$$

 © KOPO-Dept. AI Engineering Limited

Linear Independence

© ERAE AMS Limited

Linear Algebra – Linear Independence (선형독립)

Definition

Let us consider a vector space $\mathcal{V}$ with $k \in \mathbb{N}$ and $\mathbf{x}_1, \ldots, \mathbf{x}_k \in \mathcal{V}$. If only trivial solution to $\lambda_1 \mathbf{x}_1 + \cdots \lambda_k \mathbf{x}_k = \mathbf{0}$ exists, i.e., $\lambda_1 = \cdots = \lambda_k = 0$, then the vectors $\mathbf{x}_1, \ldots, \mathbf{x}_k$ are **linearly independent**.

Examples:

1. Two vectors $[1, 0]^\top$ and $[0, 1]^\top$ are independent?
2. Two vectors $[1, 0]^\top$ and $[-1, 0]^\top$ are independent?

$$\lambda_1 \begin{bmatrix} 1 \\ 0 \end{bmatrix} + \lambda_2 \begin{bmatrix} 0 \\ 1 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}.$$

This equation holds only when $\lambda_1 = \lambda_2 = 0$, implying that those two vectors are **linearly independent**.

Daegu Campus of KOREA POLYTECHNIC

Span and Basis

The **span** of $\mathcal{A} = \{\mathbf{a}_1, \ldots, \mathbf{a}_n\}$ is the set of all linear combinations of vectors in $\mathcal{A}$:

$$\text{span}(\mathcal{A}) = \{\alpha_1 \mathbf{a}_1 + \cdots + \alpha_n \mathbf{a}_n \mid \alpha_1, \ldots, \alpha_n \in \mathbb{R}\}.$$

If $\mathcal{A}$ spans the vector space $\mathcal{V}$, then we write

$$\mathcal{V} = \text{span}(\mathcal{A}),$$

and $\mathcal{A}$ is called a **generating set** of $\mathcal{V}$.

Every linear independent generating set of $\mathcal{V}$ is minimal and is called a **basis** of $\mathcal{V}$.

© KOPO-Dept.
AI Engineering Limited

Span and Basis

The canonical basis in $\mathbb{R}^3$ is

$$\left\{ \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} \right\}.$$

Or alternative basis in $\mathbb{R}^3$ is

$$\left\{ \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} \right\}.$$

© KOPO-Dept.
AI Engineering Limited

Thank you

Q&A

www.kopo.ac.kr
jsshin7@kopo.ac.kr

 © KOPO-Dept.
AI Engineering Limited