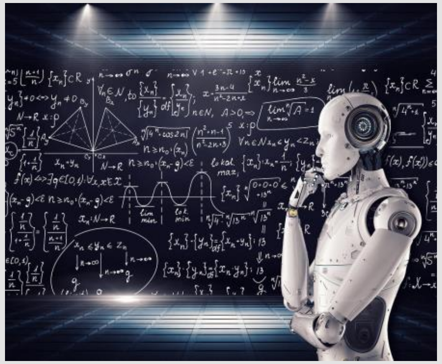
인공지능을 위한 수학 (Mathematics for AI)

세종대학교 수학통계학부 수학전공
이지은 교수

세종대학교
SEJONG UNIVERSITY



출처:https://www.flickr.com/photos/mikemacmarketing/30212411048

출처:www. Unsplash. com

[학습목표]

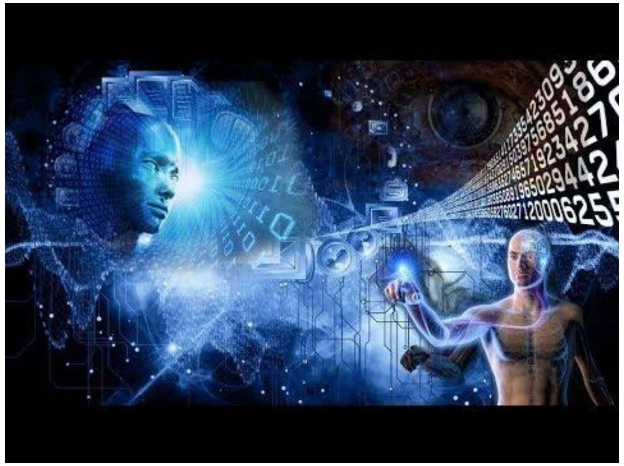
인공지능(AI)에
사용되는
수학에 대하여
공부한다.

세종대학교
SEJONG UNIVERSITY

(Question) 인공지능과 수학의 관계는 무엇인가?

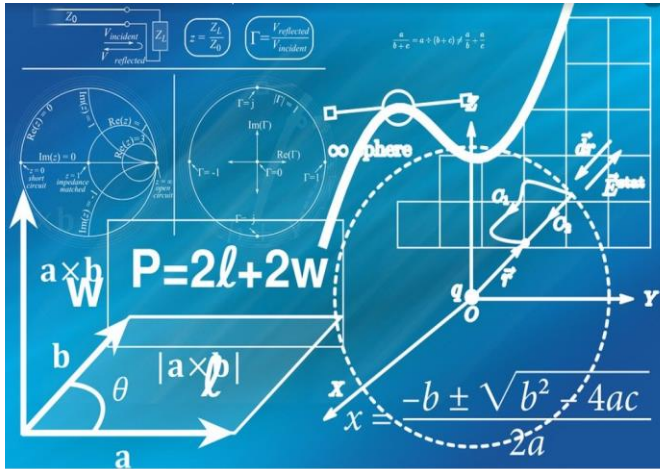
[학습목표]

인공지능 (AI)에 사용되는 수학에 대하여 공부한다.



(인공지능)AI

[출처] https://sagaciousnewsnetwork.com/the-beast-system-5g-the-ai-control-grid/



수학(Mathematics)

[출처] https://pixabay.com/es/illustrations/geometría-matemáticas-volume-1044090/

세종대학교
SEJONG UNIVERSITY

[학습목표]

인공지능 (AI)에 사용되는 수학에 대하여 공부한다.

- [인공지능과 수학 관련 기사들]

- [칼럼] 인공지능 과 수학은 무슨 상관이 있을까?

(https://m.post.naver.com/viewer/postView.nhn?volumeNo=11808498&memberNo=36498508)

- AI 시대, 수학 실력이 최고의 경쟁력이다

  (http://news.chosun.com/site/data/html_dir/2019/11/11/2019111100009.html)

- 수학 없으면 AI도 없다..."국내 산업 수학 생태계 마련 시급"

(https://biz.chosun.com/site/data/html_dir/2019/09/16/2019091601712.html)

[학습목표]

인공지능(AI)에 사용되는 수학에 대하여 공부한다.

- [동기부여(Motivation)]

(2011) 한국 최초 KBS 시각장애인 앵커 **이창훈**

[출처]http://www.seoul.co.kr/news/newsView.php?id=20111108029021

AI로 날개 단 시각장애인, 아테네 국제마라톤 완주

(2019) 시각장애인 마라토너 선수 **한동훈**

[출처]www.sisaon.co.kr/news/articleView.html?idxno=105396

세종대학교
SEJONG UNIVERSITY

[학습목표]

인공지능 (AI)에
사용되는
수학에 대하여
공부한다.

[목차(Contents)]

1. 미분적분학 (Calculus)
2. 선형대수 (Linear algebra)
3. 확률과 통계
(Probability and Statistics)

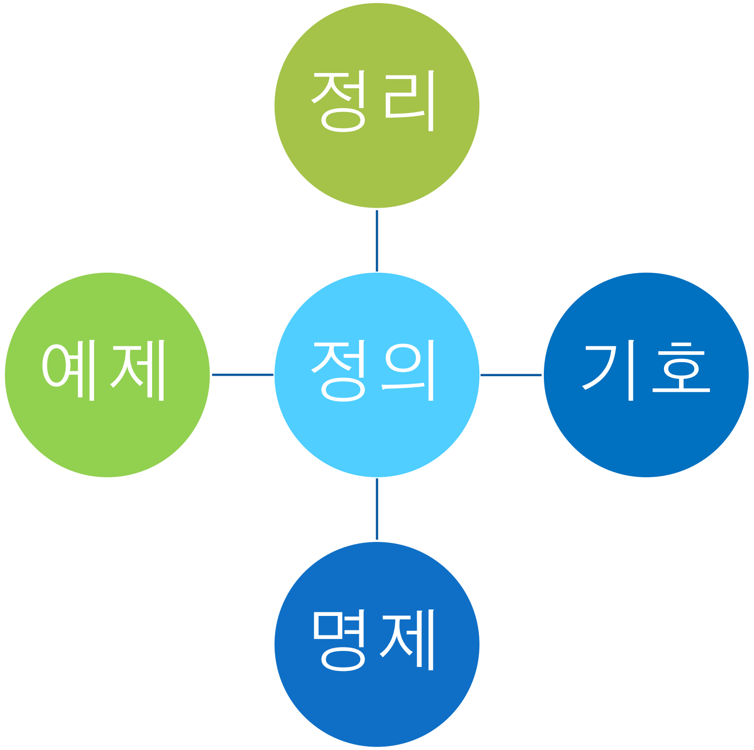
[학습목표]

인공지능(AI)에
사용되는
수학에 대하여
공부한다.

❶정의(Definition): 어떤 용어의 뜻을 명확하게 정한 문장

❷명제(Proposition): 그 내용이 참인지 거짓인지를 명확하게 판별할 수 있는 문장이나 식

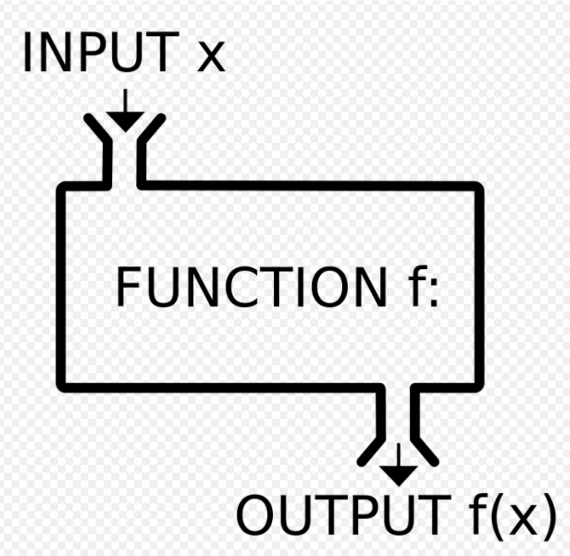
❸정리(Theorem): 증명된 명제 중에서 기본이 되는 것이나 앞으로 여러 가지 성질을 증명할 때, 활용되는 것



[Section 1]

미분적분학 (Calculus)

- (Question) 아래의 이미지로 부터 컴퓨터 프로그래밍과 인공지능에서 필요한 수학 개념은 무엇일까요 ?



[출처: https://ko.wikipedia.org/wiki/함수]

세종대학교
SEJONG UNIVERSITY

[Section 1]

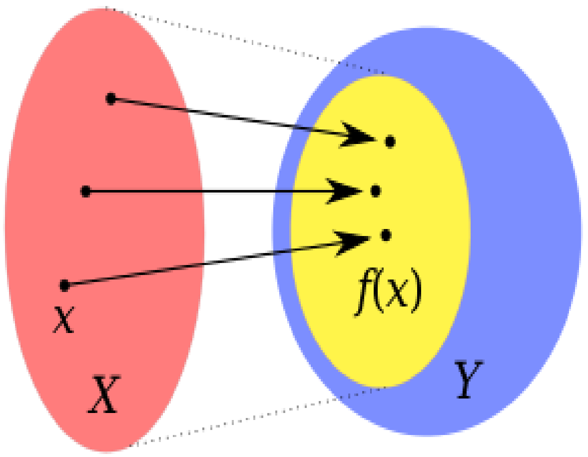
미분적분학
(Calculus)

- **함수(Function)**

[정의] 함수 (function)

:공집합이 아닌 두 집합 X, Y에 대하여 $\forall$ (임의의) $x \in X$에 대하여 $y \in Y$가 존재하여 $y = f(x)$를 만족하는 대응 규칙을 함수 $f : X \to Y$라고 한다.

(단, X = 정의역, Y = 치역)



$f : X \to Y$

즉, $f(X) \subseteq Y$

[출처] https://ko.wikipedia.org/wiki/함수

- [Example] 대수 함수, 삼각함수, 로그함수, 지수함수, 쌍곡선 함수, 시그모이드 함수, 역삼각함수, 역쌍곡선 함수등이 있다.

[Section 1]

미분적분학
(Calculus)

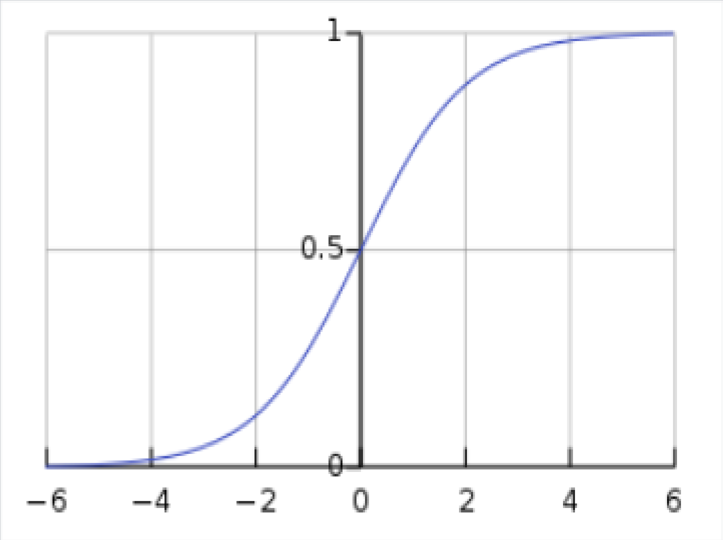
- **시그모이드 함수(Function)**

[정의] 시그모이드(Sigmoid) 함수

: S자형 곡선 또는 **시그모이드 곡선**을 갖는 수학 함수 이다.

즉, $S_a(x)=\dfrac{1}{1+e^{-ax}}$, $a\in R$(실수).

특히, $a=1$일 때, $S_1(x)=\dfrac{1}{1+e^{-x}}$를 표준 시그모이드 함수라 한다.



로지스틱 곡선

[출처] https://ko.wikipedia.org/wiki/시그모이드_함수

세종대학교 SEJONG UNIVERSITY

[Section 1]

미분적분학 (Calculus)

- 시그모이드 함수(Function) 의 예

[Ex] (1) 오차 함수 (Error function)

$$f(x) = erf(x) = \frac{2}{\sqrt{\pi}} \int_0^x e^{-t^2} dt$$

(2) 로지스틱 함수(Logistic function)

$$f(x) = \frac{1}{1+e^{-x}}$$

(3) 역탄젠트함수(Inverse tangent function)

$$f(x) = \tan^{-1} x$$

(4) 쌍곡탄젠트 함수(Hyperbolic tangent function)

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

[참고] https://ko.wikipedia.org/wiki/시그모이드_함수

세종대학교
SEJONG UNIVERSITY

[Section 1]

미분적분학
(Calculus)

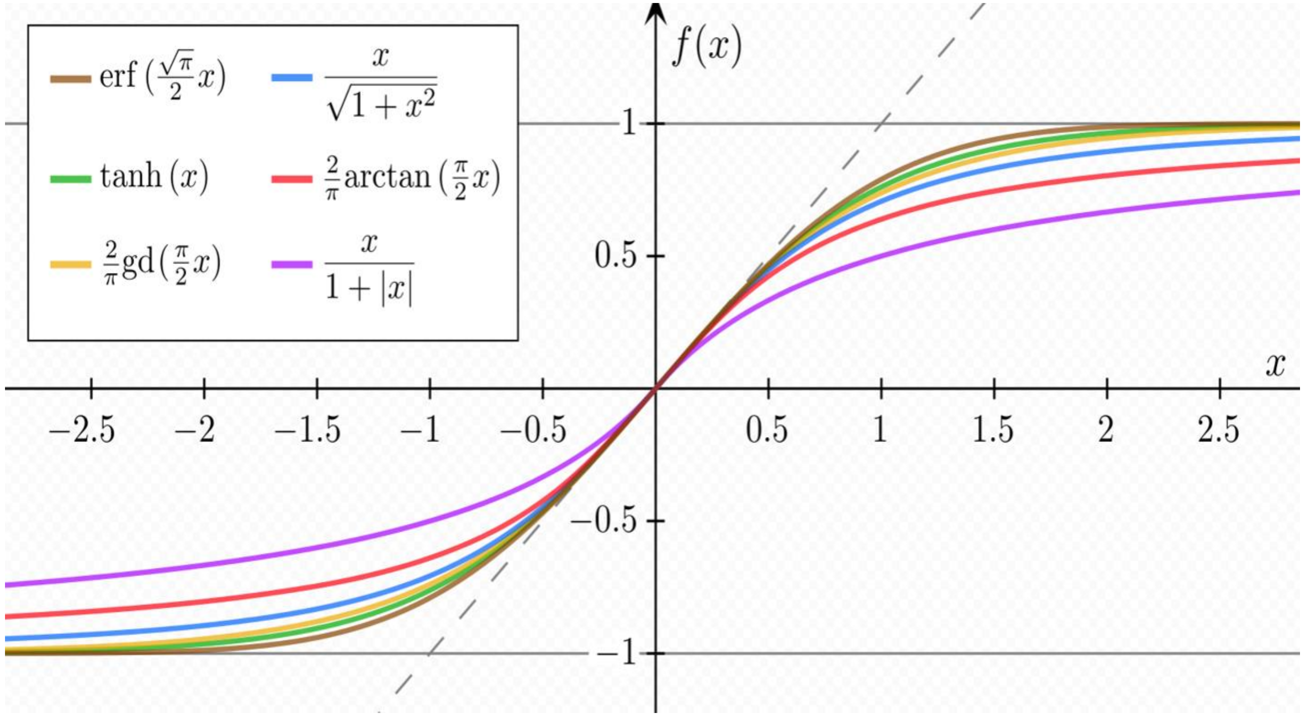
- 시그모이드 함수(Function) 의 그래프(직접 그리기)

세종대학교 SEJONG UNIVERSITY

[Section 1]

미분적분학 (Calculus)

- 시그모이드 함수(Function)의 그래프



[출처] https://ko.wikipedia.org/wiki/시그모이드_함수

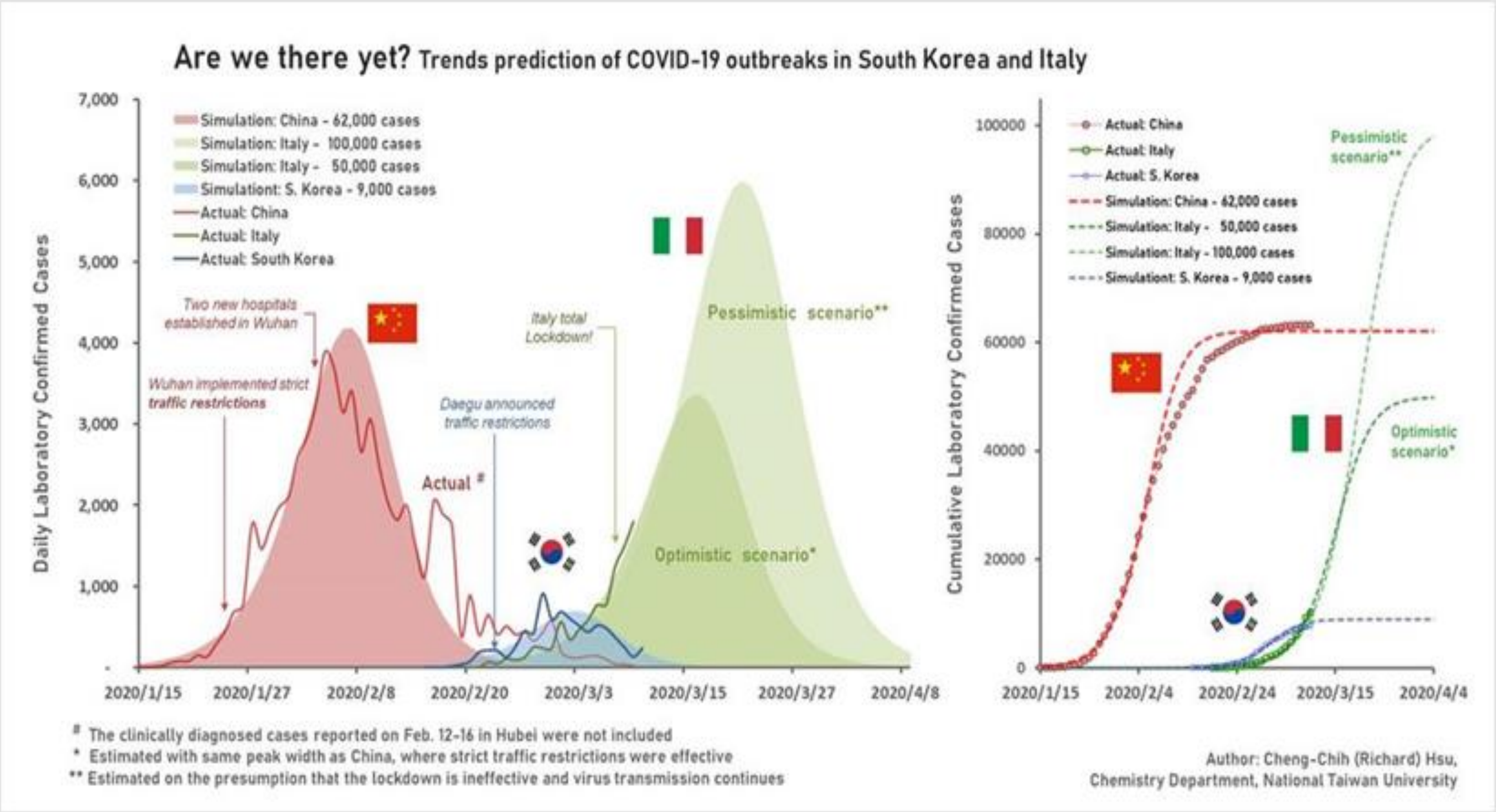
[참고] 그래프 그리기 https://www.geogebra.org/graphing?lang=ko

[Section 1]

미분적분학
(Calculus)

- 시그모이드 함수(Logistic Function)의 응용

: 코로나 -19 그래프(대만, 화학과, Cheng-Chin 교수)



[Section 1]

미분적분학 (Calculus)

- [인공지능에서 활용] (참고문헌 2번)]
- 활성화 함수(activation function)란 인공지능 모델(신경망:neural network)의 표현력을 높이기 위해 사용하는 함수이다.

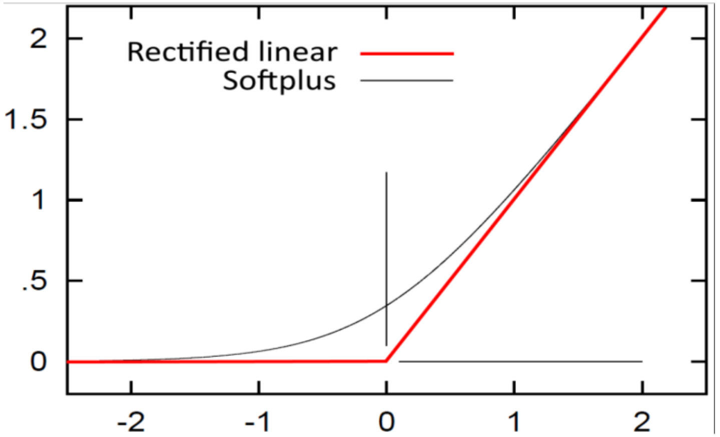
[출처] https://www.shutterstock.com/image-illustration/digital-xray-human-brain-on-blue-577251916?irgwc=1&utm_medium=Affiliate&utm_campaign=Pixabay+GmbH&utm_source=44814&utm_term=https%3A%2F%2Fpixabay.com%2Fes%2Fimages%2Fsearch%2Fneural%2520network%2F

세종대학교
SEJONG UNIVERSITY

[Section 1]

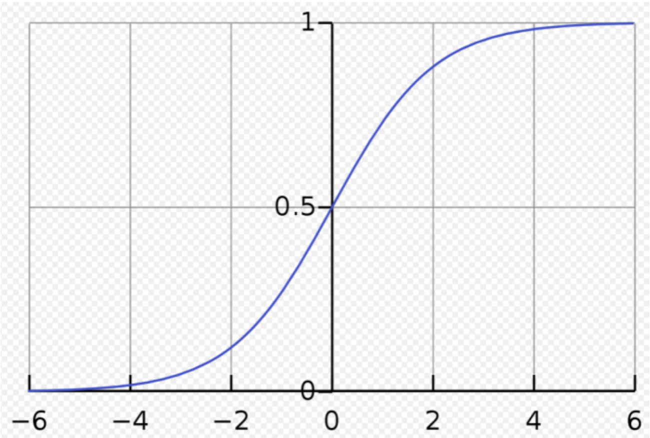
미분적분학 (Calculus)

- **[인공지능에서 활용] (참고문헌 2번)]**
- 시그모이드 함수(Function)는 주로 **활성화 함수**로 빈번하게 사용된다.



ReLU 함수: DNN(Deep neural network)과 CNN(Convolutional neural network)에 사용

[ReLU 함수] $f(x) = \max(0, x) = \begin{cases} x, x > 0 \\ 0, x \leq 0 \end{cases}$



tanh 함수: RNN(Recurrent neural network)과 LSTM(Long Short Term Memory network)에 사용

[출처] https://en.wikipedia.org/wiki/Activation_function

[Section 1]

미분적분학
(Calculus)

- [퀴즈1]

역탄젠트함수 $y = \arctan x$의 그래프를 직접 그려보세요.
(힌트: 역함수가 존재하려면 함수가 일대일함수여야 한다.)

[Section 1]

미분적분학
(Calculus)

[미분]

- x가 정해진 값 x_1에서 다른 값 x_2로 변하였을 때,

  Δx : $x_2 - x_1$ 을 x의 증분(increment) (델타 x 라고 읽음)

- 함수 $y = f(x)$에서 x가 Δx만큼 변할 때,

  Δy : y의 증분 $f(x+\Delta x) - f(x)$,

  $\dfrac{\Delta y}{\Delta x}$: x와 $x+\Delta x$사이에서 x에 대한 $y = f(x)$의 **평균변화율** (average rate of change) 로 표시함.

[Section 1]

미분적분학 (Calculus)

[미분]

[정의] (미분가능성, 순간 변화율, 미분계수, 미분가능한 함수)

주어진 함수 f 에 대하여 극한값 $\lim_{x \to a} \frac{f(x)-f(a)}{x-a} = L$가 존재하면

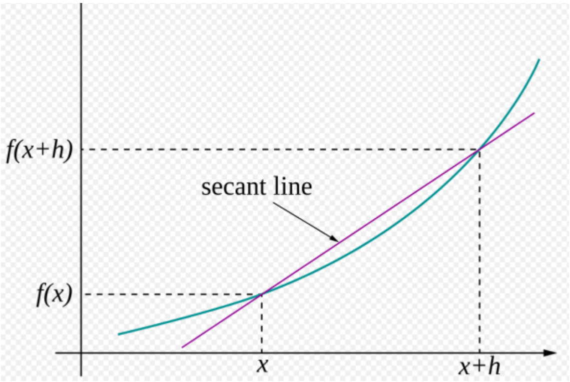
"$f(x)$는 $x=a$에서 미분가능하다" 하고

극한값 L: $x=a$에서 함수 f의 **순간변화율** 또는 **미분계수**(differential coefficient).

∘$f'(a)$: $x=a$에서의 f 의 **미분계수**

$$f'(a) = \lim_{x \to a} \frac{f(x)-f(a)}{x-a} \quad \text{또는}$$

$$f'(a) = \lim_{\Delta x \to 0} \frac{\Delta y}{\Delta x} = \lim_{\Delta x \to 0} \frac{f(a+\Delta x)-f(a)}{\Delta x} \ .$$



[출처] https://ko.wikipedia.org/wiki/미분

세종대학교
SEJONG UNIVERSITY

[Section 1]

미분적분학
(Calculus)

- [Example] 한 커플이 제주공항에 도착해서 렌터카를 빌려서 제주도 해안도로 270km를 3시간 동안에 운전을 했다. 이때, 이 커플은 벌금을 내야하는가?
- (단, 제주도에서는 자동차의 속도가 시속 60km이상인 경우 벌금을 내야한다.)

[출처] https://pixabay.com/es/images/search/jeju/

[Section 1]

미분적분학 (Calculus)

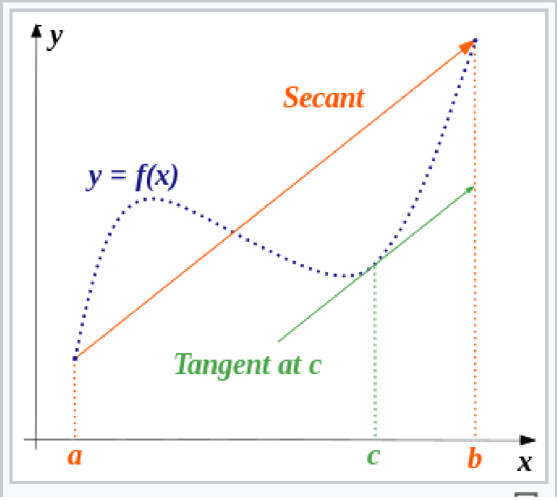
[정리] (평균값의 정리, mean value theorem)

함수 f가 닫힌구간 $[a, b]$에서 연속이고, 열린구간 (a, b)에서 미분가능하면

$a < c < b$ 인 점 c가 존재하여 $f'(c) = \dfrac{f(b) - f(a)}{b - a}$ 이다.

- [Solution]
- 평균속도 $f'(c) = \dfrac{270-0}{3-0} = 90km/\text{h}$

따라서 이 커플은 벌금을 내야한다.



[출처] https://ko.wikipedia.org/wiki/평균값_정리

세종대학교
SEJONG UNIVERSITY

[Section 1]

미분적분학
(Calculus)

[미분]

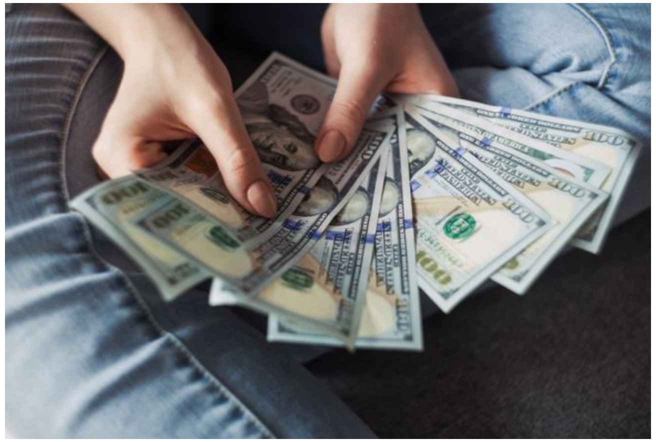
[Example] $f(x) = \frac{1}{1+\mathrm{e}^{-ax}}$ 를 미분하시오.

[Section 1]

미분적분학
(Calculus)

[인공지능에서 활용] (참고문헌 2번)]

- 함수의 값이 최소가 되는 지점을 알아내는 것은 인공지능에서 중요하다.
- 손실함수(Loss function)는 정답과 측정값 사이의 오차를 표현하는 함수이다. 인공지능 분야에서 이 함수의 값을 최소로 만들기 위해 다양한 기법을 활용한다.



[출처] https://unsplash.com/photos/lCPhGxs7pww

세종대학교
SEJONG UNIVERSITY

[Section 1]

미분적분학
(Calculus)

[편도함수]

[정의] (편도함수: partial derivative)

이변수 함수 $z=f(x,y)$에서

1) x에 관한 f의 도함수(y: 고정) $\exists \lim_{h\to 0}\frac{f(x+h,y)-f(x,y)}{h}$가 존재할 때,

이 도함수를 <u>x에 관한 f의 **편도함수**</u>

[기호] $\frac{\partial z}{\partial x}, \frac{\partial f}{\partial x}, \frac{\partial}{\partial x}f(x,y), f_x(x,y), z_x$

2) y에 관한 f의 도함수(x: 고정) $\exists \lim_{k\to 0}\frac{f(x,y+k)-f(x,y)}{k}$가 존재할 때,

이 도함수를 <u>y에 관한 f의 **편도함수**</u>

[기호] $\frac{\partial z}{\partial y}, \frac{\partial f}{\partial y}, \frac{\partial}{\partial y}f(x,y), f_y(x,y), z_y$

[Section 1]

미분적분학 (Calculus)

[Example] (1) $f(x,y) = x^3 + 3xy + y^2$ 일때 편도함수를 구하여라.
(2) $f(x,y) = e^{xy}$ 일때 편도함수를 구하여라.

[Sol]

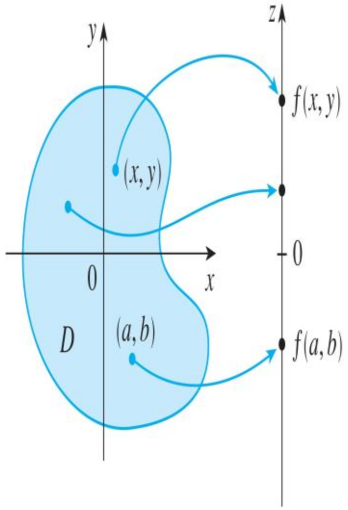


FIGURE 1

[Section 1]

미분적분학 (Calculus)

[인공지능에서 활용(참고문헌 2번)]

- 인공지능관련 책 및 논문에서 미분 $\frac{dy}{dx}$와 편미분 $\frac{\partial f(x \cdot y)}{\partial x}$이 많이 나타난다.
(두 기호의 차이에 대하여 명확하게 아는 것이 중요하다.)

- 신경망에서 가중치(weight)의 조정량은 오차 값을 가중치로 편미분한 것을 이용한다.

- 표준 시그모이드 함수를 미분하면 최대값이 $\frac{1}{4}$인데 신경망의 계층이 많아질수록 **오차역전파법(backpropagation)**에서 오차 전파되기 어려워지는 상황이 발생할 수도 있다. 이러한 현상을 **기울기 소실 문제**(vanishing gradient)라고 한다.

연쇄법칙(Chain Rule)

$$\frac{\partial z}{\partial s} = \frac{\partial z}{\partial x}\frac{\partial x}{\partial s} + \frac{\partial z}{\partial y}\frac{\partial y}{\partial s} \qquad \frac{\partial z}{\partial t} = \frac{\partial z}{\partial x}\frac{\partial x}{\partial t} + \frac{\partial z}{\partial y}\frac{\partial y}{\partial t}$$

세종대학교
SEJONG UNIVERSITY

[Section 2]

선형대수
(Linear algebra)

선형대수학(linear algebra)은 벡터 공간, 벡터, 선형 변환, 행렬, 연립 선형 방정식 등을 연구하는 대수학의 한 분야이다.

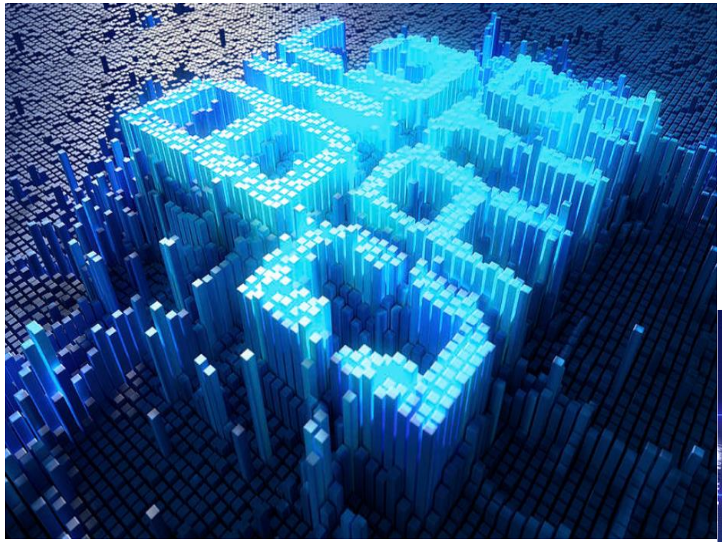
현대 선형대수학은 그 중에서도 벡터 공간이 주 연구 대상이다.
추상대수학, 함수해석학에 널리 쓰이고 있다.
[출처: https://ko.wikipedia.org/wiki/선형대수학]

머신러닝 [출처] https://pixabay.com/es/vectors/a-me-ai-anatomía-2729781/

[Section 2]

선형대수
(Linear algebra)

(Question) 실 생활에서 나타나는 많은 데이터를 어떻게 다룰 것인가?



[Section 2]

선형대수 (Linear algebra)

[벡터(vector)]

[벡터(Vector)의 정의]: 크기와 방향을 가지는 양

(기호) $\vec{v}$, $\overrightarrow{PQ}$

[벡터의 성질]

[정리] 두 벡터 $\vec{v}, \vec{w}$에 대하여,

1) 벡터의 합; $\vec{v}+\vec{w}=\overrightarrow{v+w}$

2) 벡터의 차: $\vec{v}-\vec{w}=\overrightarrow{v-w}$

3) 스칼라 곱; $c\vec{v}=\overrightarrow{cv}$($c$: 상수)

4) 교환 법칙; $\vec{v}+\vec{w}=\vec{w}+\vec{v}$

5) 결합 법칙; $(\vec{v}+\vec{w})+\vec{u}=\vec{v}+(\vec{w}+\vec{u})$ 이 성립한다.

[성분벡터의 정의]: $\vec{\alpha}=a_1\vec{e_1}+a_2\vec{e_2}+a_2\vec{e_3}$ 일 때, $\vec{\alpha}:=<a_1,a_2,a_3>$

(단, $\vec{i}=\vec{e_1}=\overrightarrow{(1,0,0)}$

$\vec{j}=\vec{e_2}=\overrightarrow{(0,1,0)}$

$\vec{k}=\vec{e_3}=\overrightarrow{(0,0,1)}$)

벡터의 크기: $|\vec{\alpha}|:=\sqrt{a_1^2+a_2^2+a_3^2}$

세종대학교
SEJONG UNIVERSITY

[Section 2]

선형대수 (Linear algebra)

[내적(Inner product)]

[내적의 정의]: 두 벡터 $\vec{\alpha} = < a_1, a_2 >$와 $\vec{\beta} = < b_1, b_2 >$의 내적(Inner product)를

$$\vec{\alpha} \circ \vec{\beta} = < a_1, a_2 > \circ < b_1, b_2 > \ = a_1 b_1 + a_2 b_2$$ 로 정의

※ $\vec{\alpha} \circ \vec{\alpha} = < a_1, a_2 > \circ < a_1, a_2 > \ = |a_1|^2 + |a_2|^2 = |\vec{\alpha}|^2$

[내적의 성질]

[정리] 세 벡터 $\vec{u}, \vec{v}, \vec{w}$에 대하여,

1) $\vec{v} \circ \vec{w} = \vec{w} \circ \vec{v}$

2) $(\vec{u} + \vec{v}) \circ \vec{w} = \vec{u} \circ \vec{w} + \vec{v} \circ \vec{w}$

3) $\vec{u} \circ (\vec{v} + \vec{w}) = \vec{u} \circ \vec{v} + \vec{u} \circ \vec{w}$

4) $a\vec{v} \circ b\vec{w} = ab(\vec{u} \circ \vec{v})$, ($a, b$ 는 실수)이 성립한다.

[정리] 두 벡터 $\vec{\alpha}, \vec{\beta}$가 이루는 각이 θ 일 때, $\cos\theta = \dfrac{\vec{\alpha} \circ \vec{\beta}}{|\vec{\alpha}||\vec{\beta}|}$ 가 성립한다.

특히, $\vec{\alpha}, \vec{\beta} \neq \vec{0}$인 경우, $\vec{\alpha} \perp \vec{\beta} \Leftrightarrow \vec{\alpha} \circ \vec{\beta} = 0$

위의 정리는 **코사인 유사도**(Cosine similarity)이다.

세종대학교
SEJONG UNIVERSITY

[Section 2]

선형대수
(Linear algebra)

[벡터(vector)]

[Example] 아래의 두 벡터가 이루는 각을 구하시오.

(1) $a = < 4,2 >, b = < -3\sqrt{2}, \sqrt{2} >$

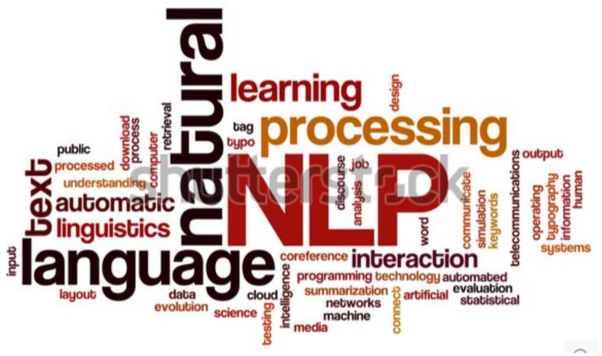
(2) $a = < 1,1,2 >, b = < 1,1,-1 >$

[Section 2]

선형대수
(Linear algebra)

[인공지능에서 활용(참고문헌 2번)]

- [벡터공간 분석] 인공지능에서 문장의 의미를 분석할 때 벡터의 개념을 활용한다.
- 벡터 공간 분석은 어떤 문장에 포함된 단어 개수와 같은 정보들을 벡터로 사용함으로써 해당 문장의 특징을 수학적으로 표현하려는 접근방법이다.



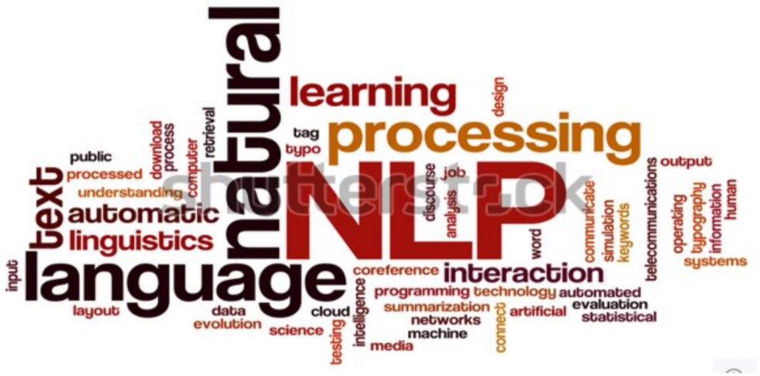
[출처] https://www.shutterstock.com/image-illustration/natural-language-processing-concept-word-cloud-201226490?irgwc=1&utm_medium=Affiliate&utm_campaign=Pixabay+GmbH&utm_source=44814&utm_term=https%3A%2F%2Fpixabay.com%2Fes%2Fimages%2Fsearch%2Fnatural%2520language%2520processing%2F

[Section 2]

선형대수 (Linear algebra)

[인공지능에서 활용(참고문헌 2번)]

- 문장을 벡터로 표현하면 내적(Inner product)과 같은 수학적 처리를 할 수 있고, 코사인 유사도 (Cosine similarity)를 통해 서로 다른 문장을 비교할 수 있다.
- 코사인 유사도가 높을 수록 해당하는 단어나 문장들은 더 가까운 관계라 는 것이라고 분석한다.
- 이는 **자연어 처리**(natural language processing)와 관련이 있다.



[출처] https://www.shutterstock.com/image-illustration/natural-language-processing-concept-word-cloud-201226490?irgwc=1&utm_medium=Affiliate&utm_campaign=Pixabay+GmbH&utm_source=44814&utm_term=https%3A%2F%2Fpixabay.com%2Fes%2Fimages%2Fsearch%2Fnatural%2520language%2520processing%2F

[Section 2]

선형대수
(Linear algebra)

[행렬(matrix)]

- [정의] 행렬(matrix) 는 수 또는 문자를 괄호 안에 직사각형 형태로 배열한 것이다. 즉,

$$\mathbf{A} = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix} = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{pmatrix}$$

$$= [a_{ij}] = (a_{ij}) = [a_{ij}]_{1\le i\le m,\ 1\le j\le n} = [a_{ij}]_{m\times n}$$

https://ko.wikipedia.org/wiki/행렬

세종대학교
SEJONG UNIVERSITY

[Section 2]

선형대수 (Linear algebra)

[행렬(matrix)의 성질]i

[정리] $A = (a_{ij})$, $B = (b_{ij})$ 가 $m \times n$ 행렬 일 때,

(1) $\mathrm{A} + B = (a_{\mathrm{ij}}) + (b_{ij})$ (=B+A)

(2) k $\mathrm{A} = (ka_{\mathrm{ij}})$, k는 실수

(3) $AB \neq BA$

[Section 2]

선형대수
(Linear algebra)

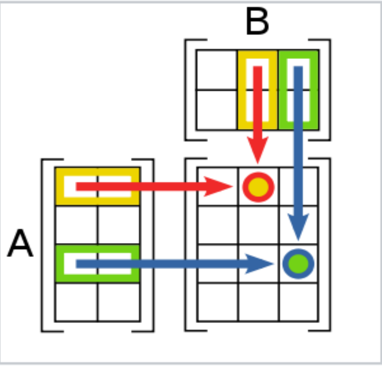
[행렬(matrix) 의 곱(Product)]

주어진 $m \times n$ 행렬 $\mathbf{A}$와 $n \times p$ 행렬 $\mathbf{B}$의 곱은 $m \times p$ 행렬이며, 각 (i, j) 성분은 $\mathbf{A}$의 i행벡터와 $\mathbf{B}$의 j열벡터의 점곱으로 정의된다.

$$(\mathbf{AB})_{ij} = \sum_{k=1}^{n} \mathbf{A}_{ik}\mathbf{B}_{kj} = A_{i1}B_{1j} + A_{i2}B_{2j} + \cdots + A_{in}B_{nj}$$

예를 들어,

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



행렬 **A**와 **B**의 곱 **AB**의 계산법

https://ko.wikipedia.org/wiki/행렬

[Section 2]

선형대수 (Linear algebra)

[역행렬(Inverse matrix)]

- [정의] A가 n $\times n$ 행렬일 때, B $A = AB = I_n$ 를 만족하는 행렬 B를 A의 **역행렬**이라고 한다. 이때, $B = A^{-1}$ 로 표시한다.

(단, det(A)가 0이 아닐 때 존재한다. 여기서 I_n은 단위행렬)

- 특히, A=$\begin{pmatrix} a & b \\ c & d \end{pmatrix}$일 때, $\det(A) = ad - bc$이다.
- A=$\begin{pmatrix} a & b \\ c & d \end{pmatrix}$ 이고 $\det(A) \neq 0$ 이면 $A^{-1} = \frac{1}{ad-bc}\begin{pmatrix} a & -b \\ -c & d \end{pmatrix}$

세종대학교
SEJONG UNIVERSITY

[Section 2]

선형대수
(Linear algebra)

[역행렬(Inverse matrix)]

[Example]

$A=\begin{pmatrix} 1 & 3 \\ 2 & 4 \end{pmatrix}$ 과 $B=\begin{pmatrix} 1 & 2 \\ 2 & 4 \end{pmatrix}$ 일 때, 행렬의 역행렬 존재하면 역행렬을 구하여라.

[Sol]

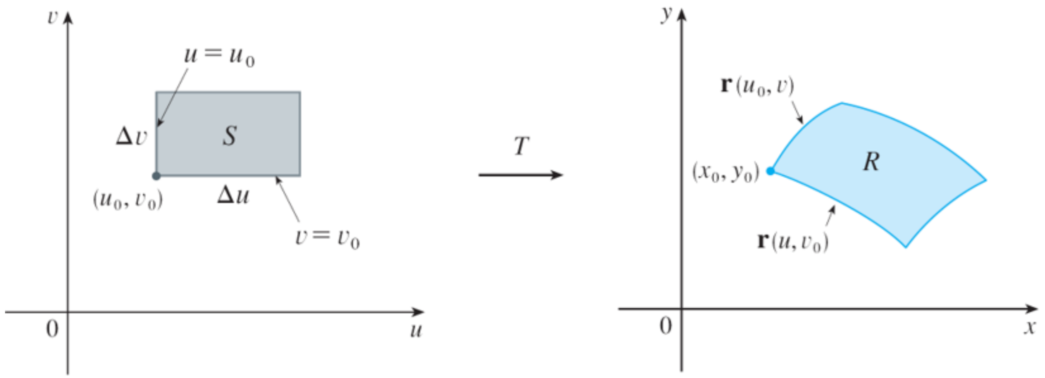
세종대학교
SEJONG UNIVERSITY

[Section 2]

선형대수
(Linear algebra)

[선형변환(linear transformation)]

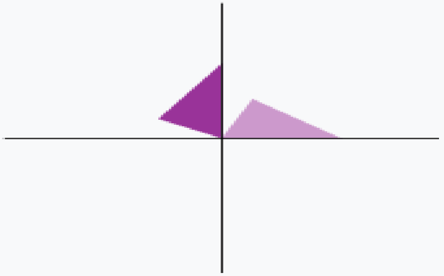
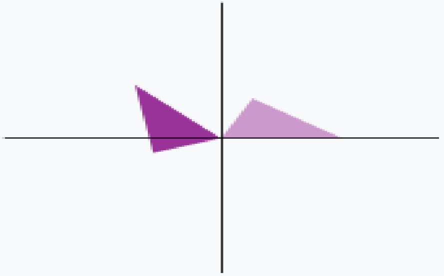
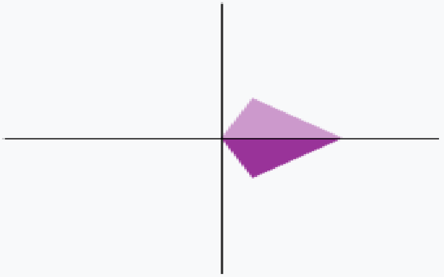
[정의] 실수 위에서 정의된 벡터공간 V, W에 대하여,
사상 $T: V \rightarrow$ W 가 성질 (i) $T(u+v) = \mathrm{T}(\mathrm{u}) + T(v)$
(ii) $\mathrm{T}(\alpha\ \ u) = \alpha T(u)$ 를 만족할 때,
사상 $T: V \rightarrow$ W를 선형변환(linear transformation)이라고 한다.



[Section 2]

선형대수 (Linear algebra)

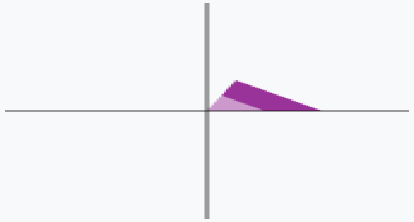
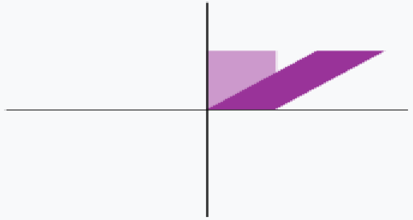
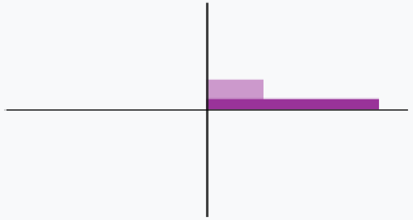
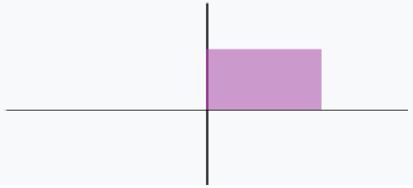
[선형변환의 Example] (출처: https://ko.wikipedia.org/wiki/선형_변환)

선형 변환	행렬 표현	도해
시계 반대 방향 90도 회전	$\begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}$	
시계 반대 방향 135도 회전	$\begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix} \quad (\theta = 135^\circ)$	
x축에 대한 반사	$\begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$	

[Section 2]

선형대수 (Linear algebra)

[선형변환의 Example] (출처: https://ko.wikipedia.org/wiki/선형_변환)

모든 방향에서 2배 확대	$\begin{pmatrix} 2 & 0 \\ 0 & 2 \end{pmatrix}$	
x축에 대한 전단	$\begin{pmatrix} 1 & \lambda \\ 0 & 1 \end{pmatrix} \quad (\lambda = 1)$	
쌍곡 회전(영어: hyperbolic rotation)	$\begin{pmatrix} c & 0 \\ 0 & \frac{1}{c} \end{pmatrix} \quad (c = 3)$	
y축에 사영	$\begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}$	

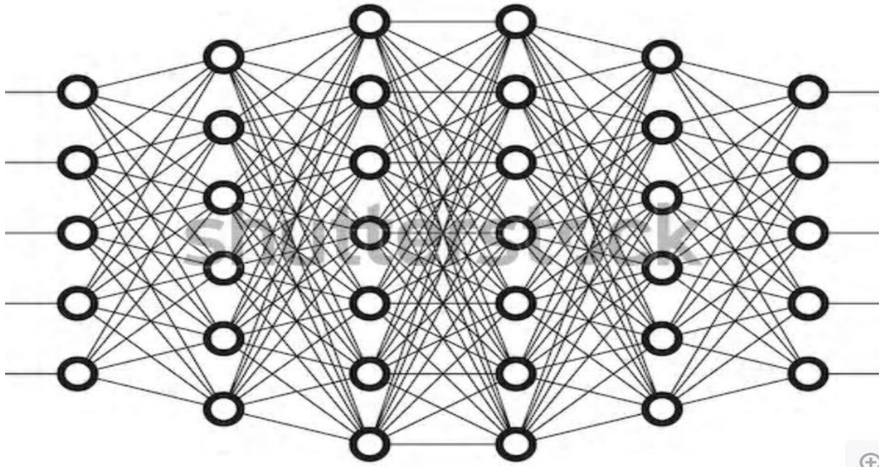
세종대학교
SEJONG UNIVERSITY

[Section 2]

선형대수 (Linear algebra)

[인공지능에서 활용(참고문헌 2번)]

- 신경망은 인공지능 분야에서 많이 사용되는 알고리즘 중 하나이다.
- 신경망의계산에서는 파라미터와 가중치를 곱하는 과정이 있는데 이를 선형변환으로 생각할 수 있다.

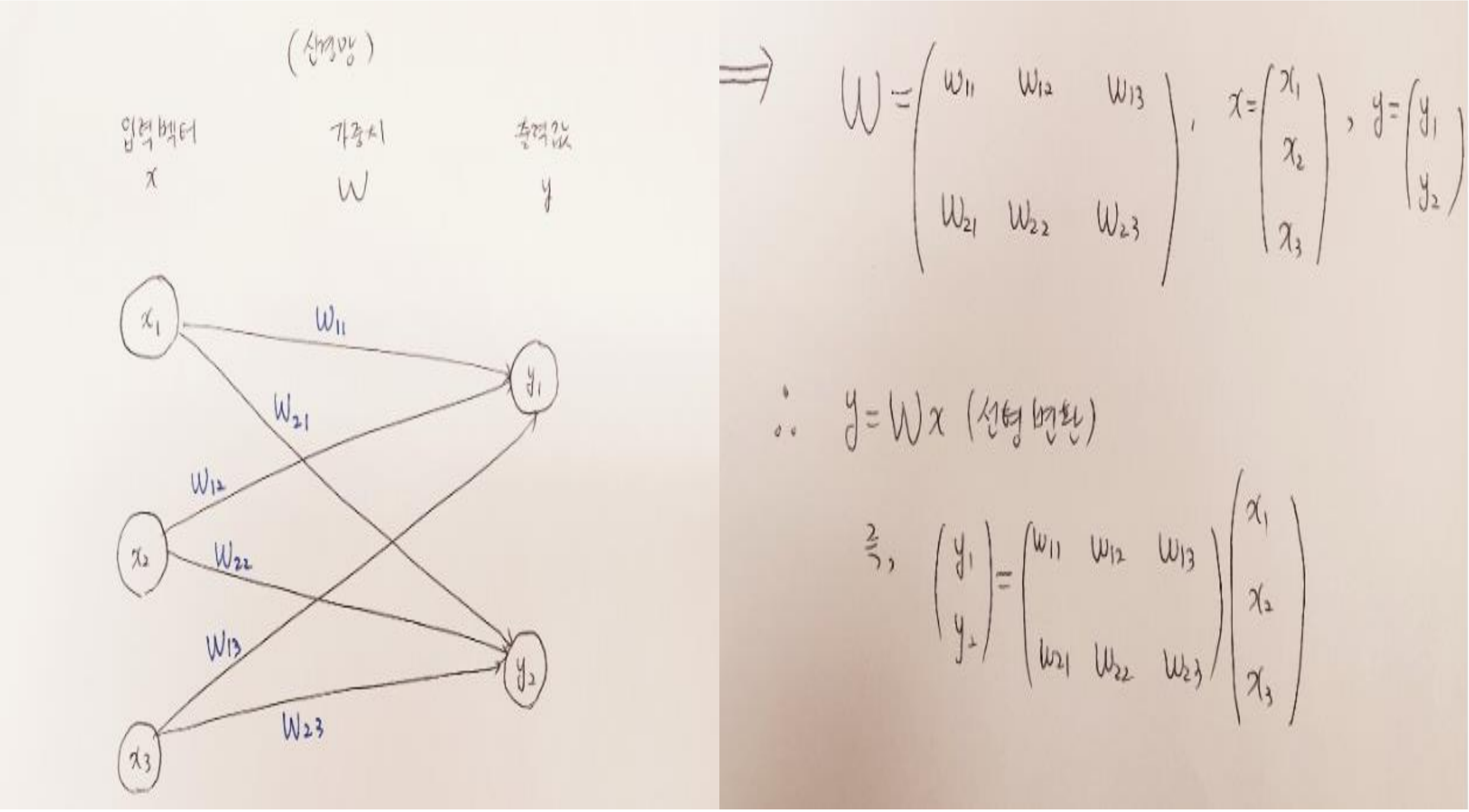


[출처] https://www.shutterstock.com/image-vector/neural-net-neuron-network-deep-learning-484275199?irgwc=1&utm_medium=Affiliate&utm_campaign=Pixabay+GmbH&utm_source=44814&utm_term=https%3A%2F%2Fpixabay.com%2Fes%2Fimages%2Fsearch%2Fneral%252onetwork%2F

[Section 2]

선형대수
(Linear algebra)

- (Ex) 신경망과 선형변환



[Section 2]

선형대수
(Linear algebra)

[고윳값과 고유벡터(eigenvalue and eigenvector)]

- [정의] 정방행렬 A 와 영이 아닌 열벡터 x에 대하여,

  Ax= $\lambda I x$ 를 만족 할 때,

  λ 를 A의 고윳값(eigenvalue)이라 하고

  x 를 A의 고유벡터(eigenvector)이라 한다.

(단, I 는 단위행렬)

- 만약 $(A - \lambda I)$ 가 역행렬을 가지면
- $(A - \lambda I)^{-1}(A - \lambda I)x = (A - \lambda I)^{-1}x = 0$ 이다.
- $x = 0$이 라는 자명한 해(trivial solution)를 갖는다. 따라서 가정에 모순이 된다.

- [정리] 고유방정식 $\det(A - \lambda I) = 0$ 일 때, 고유벡터 x 가 존재한다.

[Example] 아래 행렬 의 모든 고유값과 고유벡터를 구하여라.

[Sol]

$$A=\begin{pmatrix} 5 & -6 \\ 2 & -2 \end{pmatrix}$$

[Section 2]

선형대수
(Linear algebra)

세종대학교
SEJONG UNIVERSITY

[Section 2]

선형대수
(Linear algebra)

[퀴즈 2] 아래 행렬 의 모든 고유값과 고유벡터를 구하여라.

$$A=\begin{pmatrix} 2 & 4 \\ -1 & -3 \end{pmatrix}$$

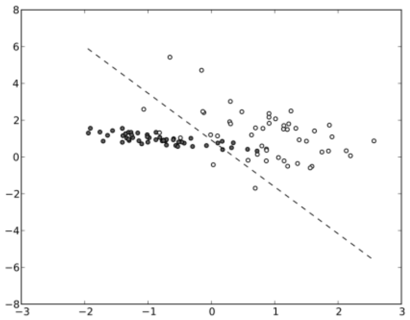
세종대학교
SEJONG UNIVERSITY

[Section 2]

선형대수 (Linear algebra)

[인공지능에서 활용(참고문헌 2번)]

- **비지도 학습**(unsupervised Learning: 선생님이 없는 학습방법) 에서 **주성분 분석**(PCA: principal component analysis) 기법을 사용한다.
- **주성분 분석**(PCA)은 다차원 데이터를 다루기 쉽게 만들기 위해 2차원이나 3차원으로 압축한 방법이다.
- 고유값과 고유벡터의 개념을 이용하여 데이터가 많이 흩어져 있는 분포상황의 문제를 해결한다.
- **기여율**(coefficient of determination) 는 고유벡터에 대응하는 고유값을 전체고유값의 총합으로 나눈 것이다.
- 기여율은 주성분이 우리가 가진 데이터를 얼마나 잘 설명할 수 있는 지를 평가하는 척도로 사용된다.



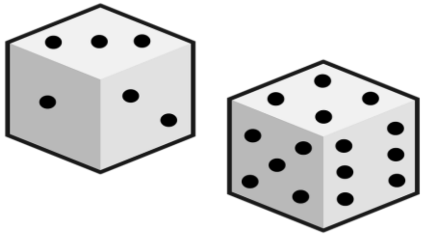
[Section 3]

확률과 통계 (Probability and Statistics)

[확률(Probability)의 정의]

A$\subset U$ (전체)는 사건이고 n(A)는 사건 A가 일어날 횟수 일 때,

확률 P = $\frac{n(A)}{n(U)}$ 이다.



- 확률의 최대값은 1이고 최솟값은 0이다.

[인공지능에서 활용(참고문헌 2번)]

인공지능 분야에서 상황을 판단하는 방법으로 확률이 높은 쪽을 정답으로 택하는 방법을 자주 사용한다.

세종대학교
SEJONG UNIVERSITY

[Section 3]

확률과 통계 (Probability and Statistics)

- **[확률변수와 이산확률분포의 정의]**
- 이산확률변수 X 의 확률함수 $f(x_i) = P(X = x_i)$는 아래의 두 조건을 만족한다.
- 모든 x_i값에 대하여 $0 \leq f(x_i) \leq 1$
- $\sum_{\text{모든 } x_i} f(x_i) = 1$

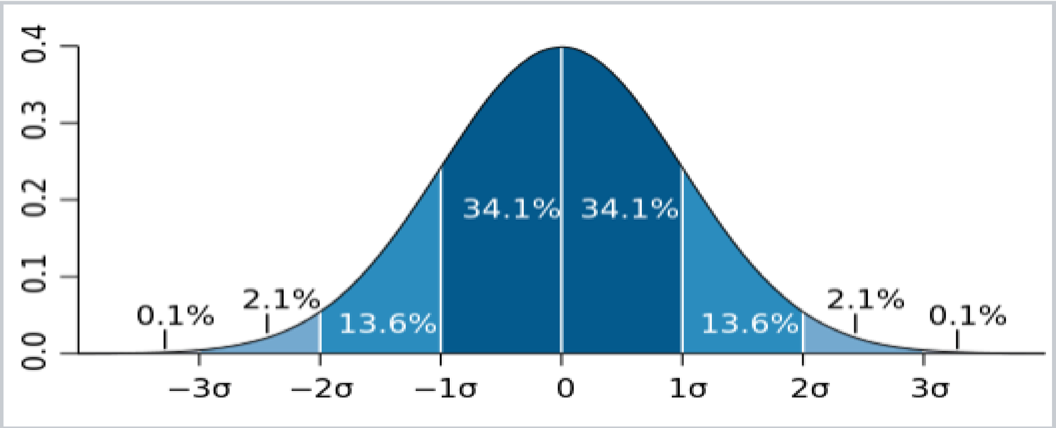
[출처] https://ko.wikipedia.org/wiki/확률_분포

세종대학교
SEJONG UNIVERSITY

[Section 3]

확률과 통계
(Probability and Statistics)

- **[확률변수와 연속확률분포의 정의]**
- 연속확률변수X의 확률밀도함수 $f(x)$ 에 대하여 아래의 조건을 만족한다.
- *모든 x 값에 대하여* $f(x) \geq 0$
- $P(a \leq X \leq b) = \int_a^b f(x)dx$
- $P(-\infty \leq X \leq \infty) = \int_{-\infty}^{\infty} f(x)dx = 1$



정규 분포

[출처] https://ko.wikipedia.org/wiki/확률_분포

세종대학교
SEJONG UNIVERSITY

[Section 3]

확률과 통계 (Probability and Statistics)

- **[인공지능에서 활용(참고문헌 2번)]**
- 어떤 현상에 대한 관측 결과들을 이산확률변수로 택하고, 이에 대한 이산확률분포를 구할 수 있으면 다음에 일어날 사건에 대한 확률을 과거의 데이터로부터 예측할 수가 있다.
- 적당한 연속확률분포를 선택한다면 적은 수의 시행만으로도 앞으로 일어날 사건의 확률을 상당히 높은 정확도로 추측할 수 있다.

[Section 3]

확률과 통계 (Probability and Statistics)

- **[평균, 분산, 표준편차]**

[정의][평균(mean)]

$x_1, x_2, x_3, \ldots, x_n$ 가 n의 확률변수일 때, 평균값 $\overline{x}$ 는 다음과 같다.

$$\overline{x} = \frac{1}{n}\sum_{k=1}^{n} x_k$$

[정의][분산(variance)과 표준편차(standard deviation)]

$x_1, x_2, x_3, \ldots, x_n$ 가 n의 확률변수이고 평균값이 $\overline{x}$ 일 때, 분산 σ^2 과 표준 편차 σ 는 다음과 같다.

$$\sigma^2 = \frac{1}{n}\sum_{k=1}^{n}(x_k - \overline{x})^2$$

$$\sigma = \sqrt{\sigma^2} = \sqrt{\frac{1}{n}\sum_{k=1}^{n}(x_k - \overline{x})^2}$$

세종대학교
SEJONG UNIVERSITY

[Section 3]

확률과 통계 (Probability and Statistics)

- **[인공지능에서 활용(참고문헌 2번)]**
- 평균과 분산, 그리고 표준편차는 과거의 데이터로부터 어떤 특징이나 경향을 밝혀낼 수 있는 가장 기본적인 방법으로, 인공지능 모델을 만들기 전에 데이터의 특징을 파악할 때 사용된다

세종대학교
SEJONG UNIVERSITY

[Section 3]

확률과 통계
(Probability and Statistics)

- [공분산(covariance)과 상관계수(correlation coefficient)]

[정의][공분산(covariance)]

두 종류 데이터에 대한 n의 확률변수$(X, Y)=\{(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)\}$가 있다고 하자. X의 평균이 λ_x이고 Y의 평균이 λ_y일 때, 공분산 $Cov(X, Y)$는 다음과 같다.

$$Cov(X, Y) = \frac{1}{n}\sum_{k=1}^{n}(x_k - \lambda_x)(y_k - \lambda_y)$$

[정의][상관관계(correlation coefficient)]

확률변수 X와 Y의 분산이 양수이고 각각의 표준편차가 σ_X, σ_Y, 공분산 $Cov(X, Y)$라 할 때의 상관계수는 다음과 같다. (이때, $-1 \le \rho \le 1$)

$$\rho = \frac{Cov(X, Y)}{\sigma_X \sigma_Y}$$

[Section 3]

확률과 통계 (Probability and Statistics)

- [EX] X= 작업시간, Y=월급여 인 결합 확률분포가 다음과 같이 주어져 있다고 하자. (단위: X=1 시간, Y=만원)

X (1시간) / Y (만원)	100	200
2	0.3	0.1
4	0.1	0.5

- (1) E(X)=3.2 이고 E(Y)=160일 때, 공분산은?
- (2) (1)의 가정이 성립하고 X의 분산은 64 이고 Y의 분산은 49일 때 상관관계를 구하시오.

[Section 3]

확률과 통계
(Probability and Statistics)

[Sol] (1) $Cov(X,Y)=E\left[\left(X-E(X)\right)\left(Y-E(Y)\right)\right]$

=(2-3.2)(100-160)(0.3)+(2-3.2)(200-160)(0.1)

+(4-3.2)(100-160)(0.1)+(4-3.2)(200-160)(0.5)

=28(1시간 ×만원)

- (2) $\rho=\frac{Cov(X,Y)}{\sigma(X)\sigma(Y)}=\frac{28}{42}=0.5$

[Section 3]

확률과 통계
(Probability and Statistics)

- **[인공지능에서 활용(참고문헌 2번)]**
- 사람이 직관적으로 분석하기 어려울 만큼의 대량의 데이터가 있다면 컴퓨터로 하여금 무수히 많은 파라미터(parameter)를 조합하고 그들의 상관계수를 계산하면서 상관관계가 강한 조합을 찾아내게 만들 수 있습니다. 이런 과정을 거치면 사람이 미처 발견하지 못한 숨은 관계나 데이터의 특징을 찾을 수 있어 데이터를 보다 유용하게 활용할 수 있게 한다.

[Section 3]

확률과 통계 (Probability and Statistics)

- [최대가능도 추정(maximum likelihood estimation)]

[정리][최대가능도추정(maximum likehood estimation)]

(1) 최대가능도 추정이란 어떤 파라미터 θ의 값을 추정하는 방법이며, θ에 대한 가능도 함수 $L(\theta)$를 최대로 만드는 θ를 찾으면 된다. 즉,

$$\frac{dL(\theta)}{d\theta}=0$$

(2) 가능도함수 $L(\theta)$를 최대로 만드는 θ는 로그가능도 함수 $\ln L(\theta)$에 대하여 아래의 방정식을 만족한다. (단, $\ln x = \log_e x$)

$$\frac{d}{d\theta} ln\, L(\theta)=0$$

[Section 3]

확률과 통계 (Probability and Statistics)

[Example] 주사위를 100번 던졌을 때, 20번이 1이 나왔다. 어떤 주사위를 던졌을 때 숫자 1이 나올 확률은?

[Solution] 주사위를 던졌을 때 1이 나올 확률을 θ라고 하자.

이때, 가능도함수는 $L(\theta) = {}_{100}C_{20}\theta^{20}(1-\theta)^{80}$ 이다.

차수를 낮추기 위해 위에 식에 양변에 로그를 적용한다.

$$\begin{aligned} \ln L(\theta) &= \ln[{}_{100}C_{20}\theta^{20}(1-\theta)^{80}] \\ &= \ln {}_{100}C_{20} + 20\ln\theta + 80\ln(1-\theta) \end{aligned}$$

위의 정리로부터 $\Rightarrow \dfrac{d}{d\theta} ln\, L(\theta) = 0 + \dfrac{20}{\theta} - \dfrac{80}{1-\theta} = 0.$

따라서 $\theta = \dfrac{1}{5}.$

세종대학교
SEJONG UNIVERSITY

[Section 3]

확률과 통계 (Probability and Statistics)

- **[인공지능에서 활용(참고문헌 2번)]**
- 최대가능도추정은 이미 확보한 데이터를 사용해서 미처 발견한 확률 모델의 파라미터를 추정할 때 사용하는 통계 기법입니다. 실제로 과거의 데이터로 부터 미래를 예측할 때 이러한 방법을 많이 사용합니다.

[Section 3]

확률과 통계 (Probability and Statistics)

[퀴즈 3] 사격에서 400발을 쏘았습니다. 한 발 쏠 때마다 탄착점이 표적이 중심에 가까운 순을 10점에서 0점까지의 점수가 주어집니다.

(1) 400발 중에 10점은 30번이 나왔습니다. 이때, 10점을 얻을 확률 Θ 의 최대가능도를 추정하시오.

(2) 추가로 400발을 더 쏘았습니다. 10점이 나온 횟수는 800발 중에 50번이었습니다. 이 때, 10점을 얻을 확률 Θ 의 최대가능도를 추정하시오.

[참고문헌]
References

1. M. P. Deisenroth, A. A. Faisal, and C. S. Ong, Mathematics for Machine Learning, Cambridge University Press(2020).
2. 이시카와 아카히코, 인공지능을 위한 수학, 프리텍(2018).
3. 하길찬, 해설이 있는 선형대수학, 교우사(2019).
4. Howard Anton, Elementary linear algebra (2015)
5. 이원우, 알기쉽게 풀어 쓴 통계학, 박영사 (2014).
6. James Stewart, 다변수함수 벡터해석 (2014)