SAMSUNG

Samsung Innovation Campus

Artificial Intelligence Course

Together for Tomorrow!
Enabling People
Education for Future Generations

Chapter 2.

Math for Data Science

Artificial Intelligence Course

Chapter Description

Chapter objectives

- ✓ Be able to install Anaconda to learn data science.
- ✓ Be able to know the types of numbers and math symbols that are required for data science.
- ✓ Be able to define and know the various equations in basic mathematics and algebra required for data science.
- ✓ Learn about sequences, absolute values, functions, graphs, linear algebra (vector and matrix), and derivatives in order to prepare for mathematical concepts and practices needed for AI.

Chapter contents

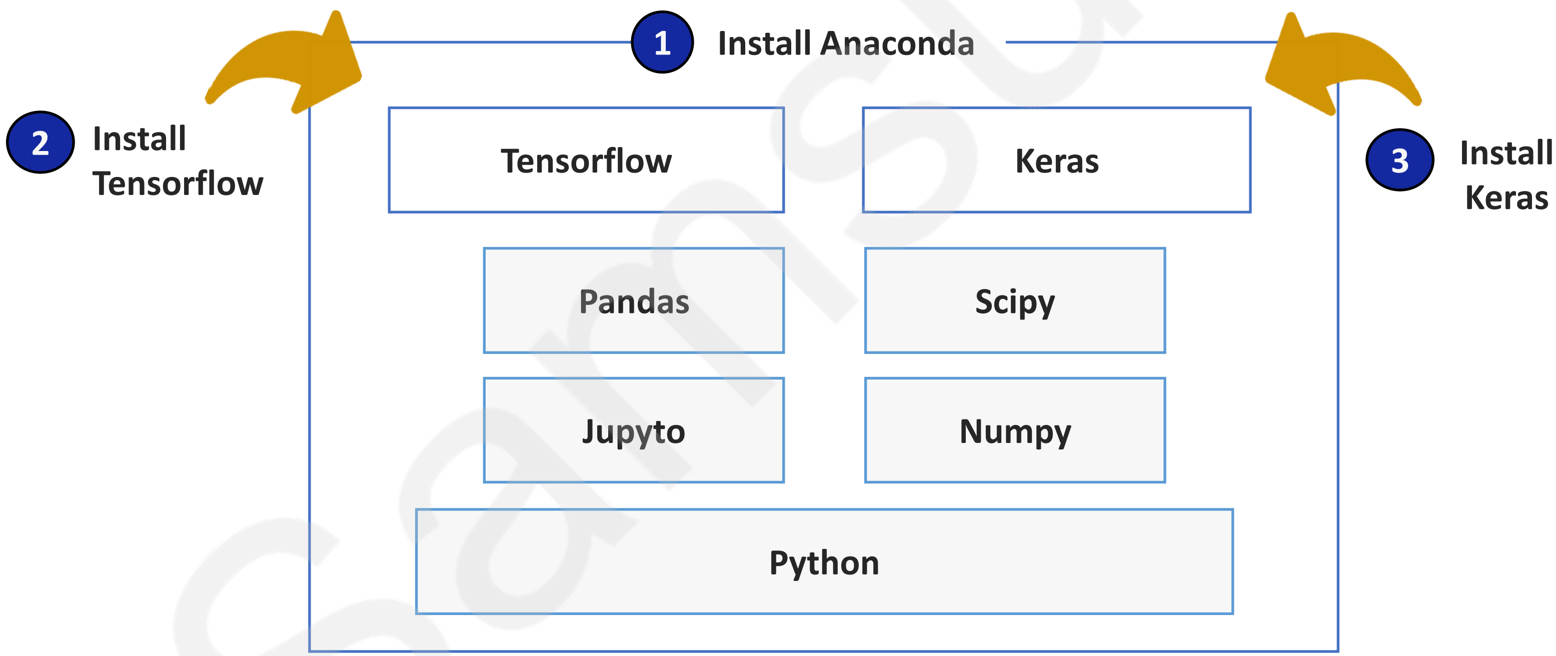
Unit 1.

Introduction

- 1.1. Installing Anaconda for Python
- 1.2. Intro to Mathematics
- 1.3. Mathematical Symbols

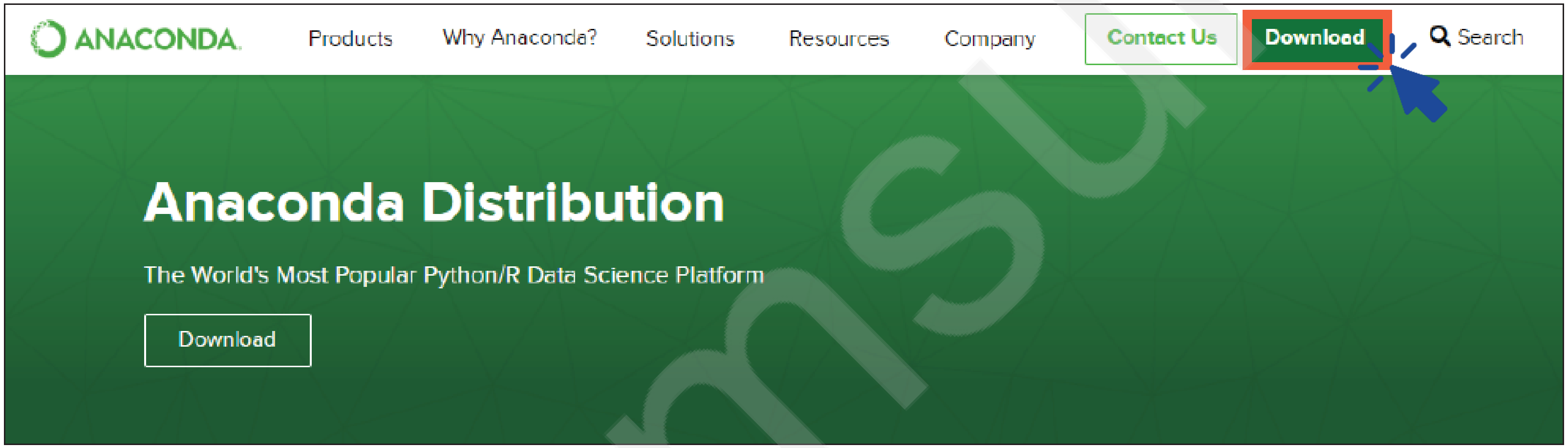
Installing Anaconda Jupyter Notebook

Environment Setting for Data Analysis



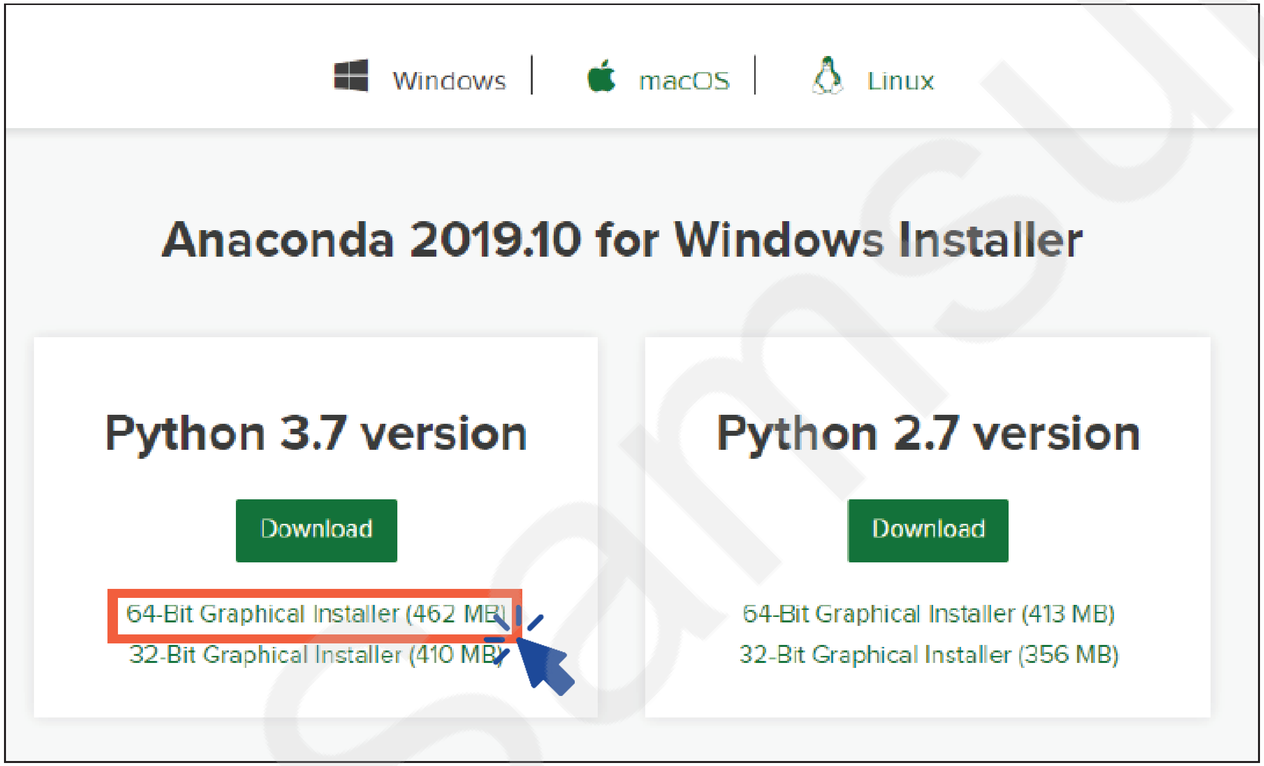
Installing Anaconda

- www.anaconda.com



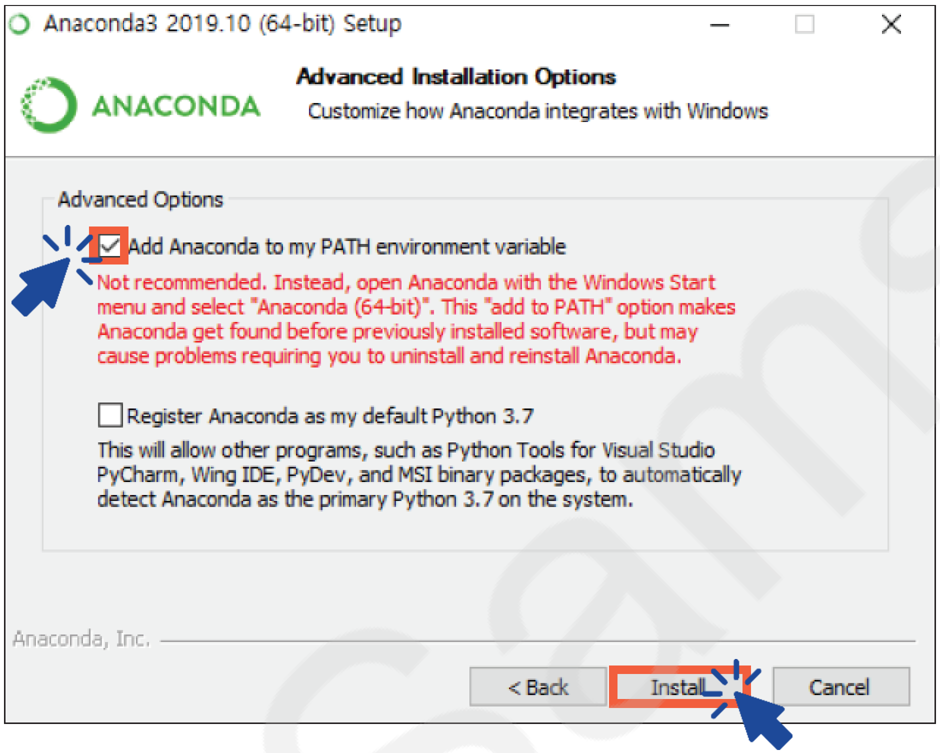
Installing Anaconda

- www.anaconda.com



Installing Anaconda

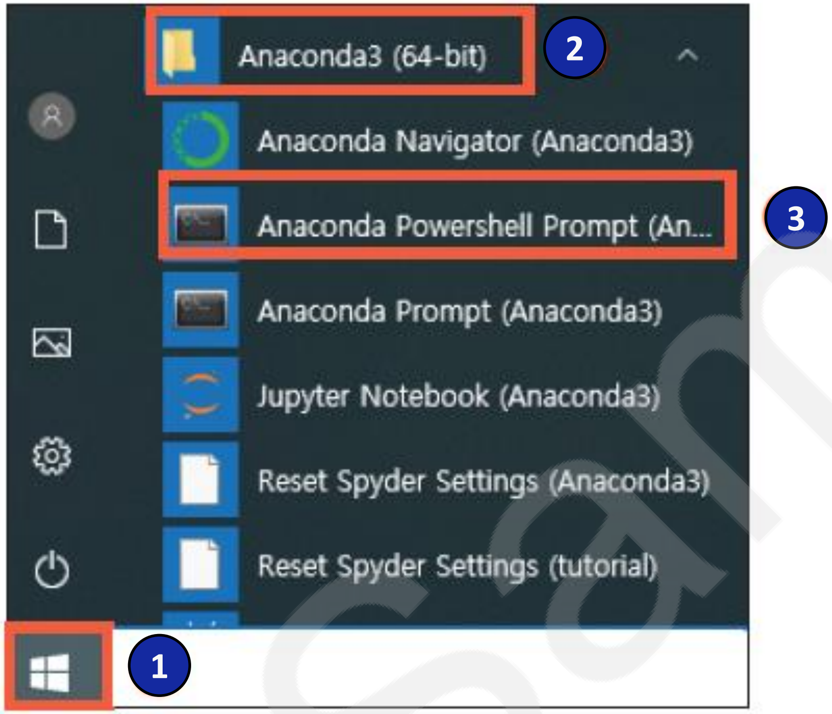
- www.anaconda.com



※ If you are not familiar with PATH environment, please click “Register Anaconda…”.

Installing Jupyter Notebook

- Once installation is complete, click the Windows Start menu, and then click on Anaconda 3(64-bit)/
- Select Anaconda Powershell Prompt



Installing Jupyter Notebook

- Enter “jupyter notebook” as instructed in the Anaconda prompt.

Anaconda Prompt (Anaconda3)

```
(base) C:\Users\it>jupyter notebook
```

Installing Jupyter Notebook

- The Juyper Notebook screen is shown as follows.

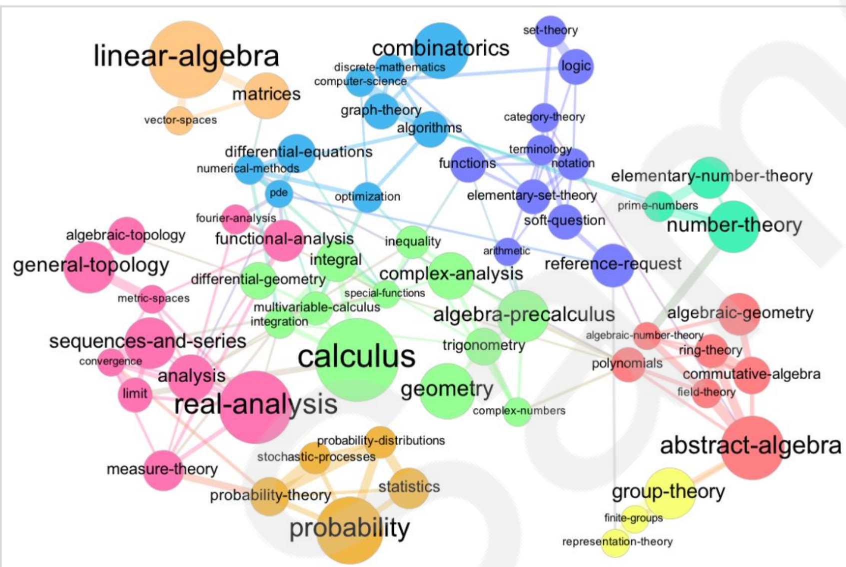
Unit 1.

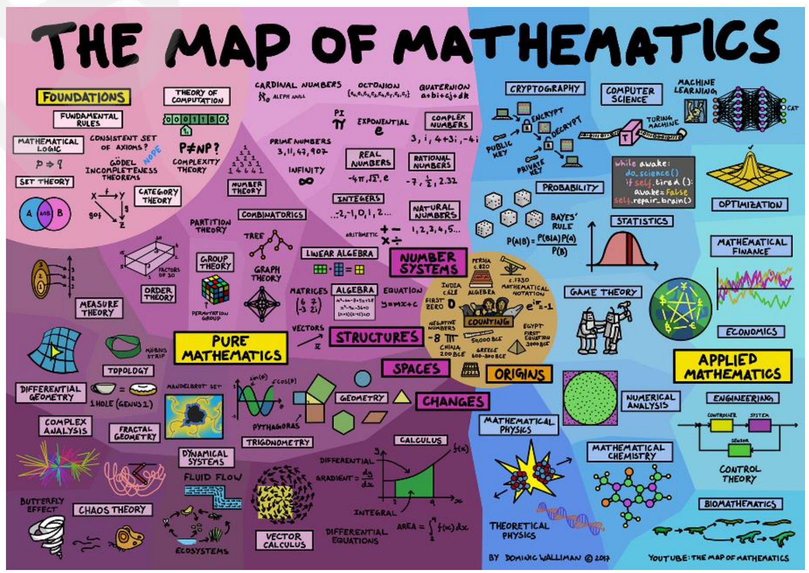
Introduction

Classification of Mathematics

Mathematics, in short, is the study of quantity. Traditionally, mathematics is largely divided into two areas: arithmetic and geometry.

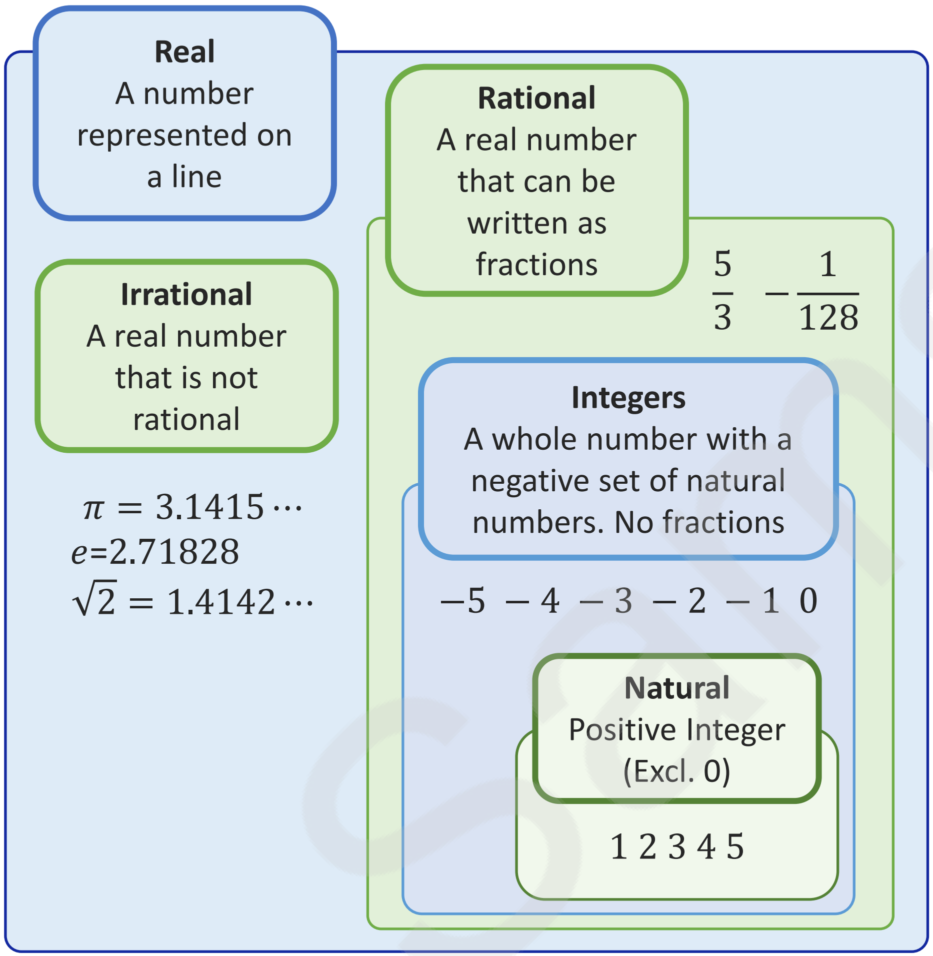
- Arithmetic is a branch of mathematics that deals with numerical calculations. It refers to a method of calculation using integers, rational numbers, real numbers, and complex numbers.
- Arithmetic covers all laws that combine two or more numbers.





https://www.quora.com/Is-there-any-diagram-or-tree-of-all-fields-of-mathematics-or-mathematics-evolution

The types of numbers are as shown below.



Type	Description
Natural Numbers	Common counting numbers (0 is not a natural number per se, but it is often included)
Integers	Whole numbers with a negative set of natural numbers. No fractions
Rational Numbers	All numbers which can be written as fractions.
Irrational Numbers	All numbers which cannot be written as fractions.
Real Numbers	A set of rational and irrational numbers, can be represented on a line.

Unit 1.

Introduction

Reading and Writing Math Symbols

In mathematics, it is important to learn how to read and write symbols because most of them are Greek letters.

Uppercase	Lowercase	Pronunciation	Uppercase	Lowercase	Pronunciation	Uppercase	Lowercase	Pronunciation
A	α	Alpha	I	ι	Iota	P	ρ	Rho
B	β	Beta	K	κ	Kappa	Σ	σ	Sigma
Γ	γ	Gamma	Λ	λ	Lambda	T	τ	Tau
Δ	δ	Delta	M	μ	Mu	Υ	υ	Upeilon
E	ε	Epsilon	N	ν	Nu	Φ	φ	Phi
Z	ζ	Zeta	Ξ	ξ	Xi	X	χ	Chi
H	η	Eta	O	o	omikron	Ψ	ψ	Psi
Θ	θ	theta	Π	π	mi	Ω	ω	omega

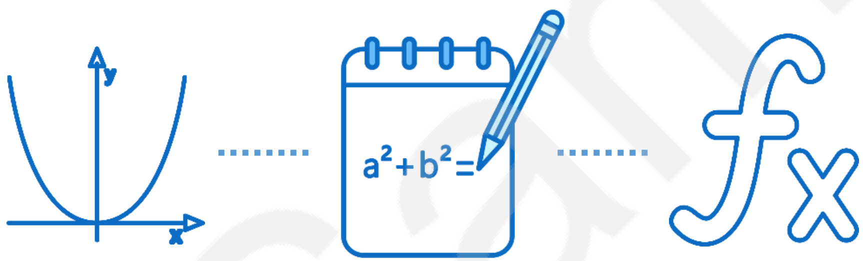
Unit 2.

Basic Math for Data Science

Algebra

Algebra is the study of numerical relationships, properties, and laws of calculation using general characters, which represent numbers instead of individual numbers.

- Currently, it encompasses the study of algebra, a set, where combinations between elements such as addition and multiplication are defined.
- In mathematics, an expression represents a rule, principle, fact, etc. by mathematical symbols.
- Equations usually represent a clear and unchanging relationship between certain quantities expressed in letters with algebraic symbols.



Algebra

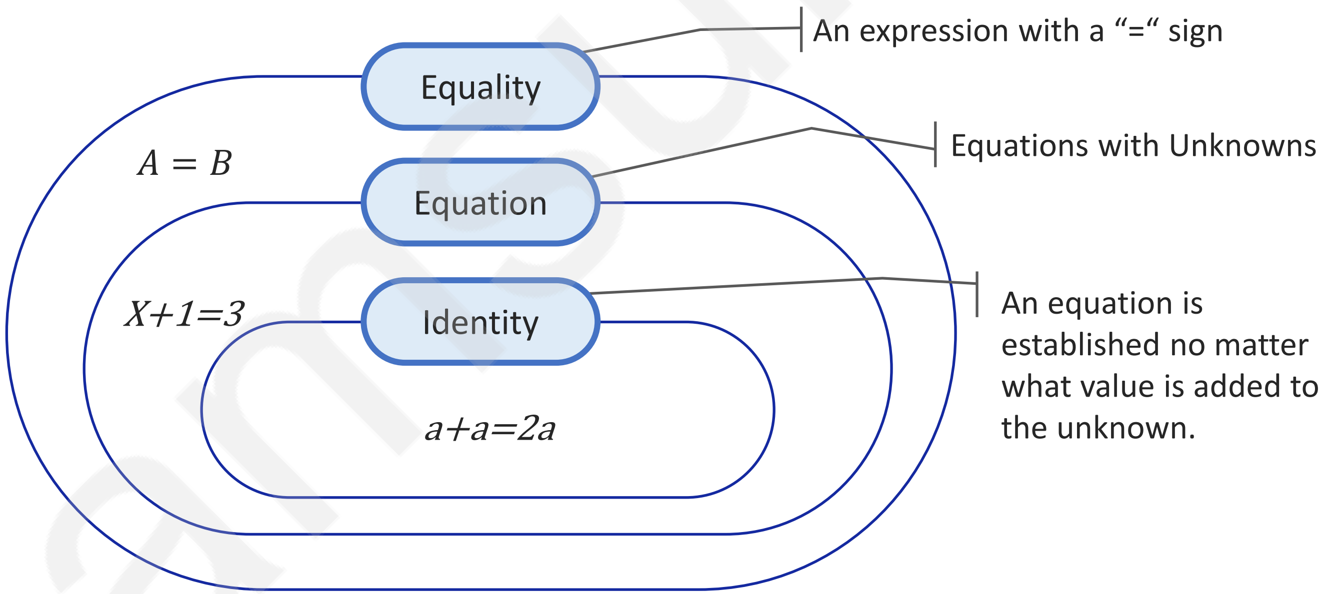
$$x = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a}$$

What is a Mathematical Expression?

- It is an expression using numbers, or variables, or both.

y
7
$9-5$
$5\times x - 9$
$4 + 8\times (5 - 4)$
$X + 5\times (7 - x)$

Equality, Equation, Identity



What is an Equation?

- It is an equation including unknowns.
- The equation is simply expressed by putting an equal sign between two equations.
- All of the following examples are equations.

$$7=7$$
$$x=9$$
$$y+9=14$$
$$x-4=15-x$$
$$5xy = 8xy^2 + 4$$

Algebraic Equation

- Variables or constants are combined with addition, subtraction, multiplication, and division (except when divided by 0).
- Variables are classified into independent and dependent variables. An independent variable is an amount that increases or decreases, or an amount that has countless values in the same equation. The dependent variable also changes, but its amount is generated according to the change in the independent variable.

Ex In y=f(x), x is an independent variable and y is a dependent variable.

- This type of algebraic equation is also called a "polynomial equation." In many algebraic equations, mathematical and scientific expressions have several variables that are conventionally used as follows.

  - n represents a natural number or integer.
  - x represents a real number.
  - z represents a complex number.

What is a Term?

It is an expression of a number or letter, or a product of the two. For example:

$$3, a, 3a, -4ab , \frac{x}{4}, a^2$$

The number of times each term is multiplied by a variable is called a degree.

In each term, the part excluding the characters corresponding to the variables is called a coefficient.

- The order of terms multiplied by two or more variables can be added to the indices of each variable.

Ex If a term is x2y3, the order for the variables x and y is $2 + 3 = 5$ which is 5.

What is a Polynomial Equation?

- It is an equation that includes the sum of powers with one or more unknowns.
- In this equation, the unknowns and numbers on both sides based on the equal sign are polynomials.=
- The equation below is a polynomial equation.

$$(x-2)^3 = x^3 - 6x^2 + 12x - 8$$

- The coefficient and the degree of the equation.

4 Terms

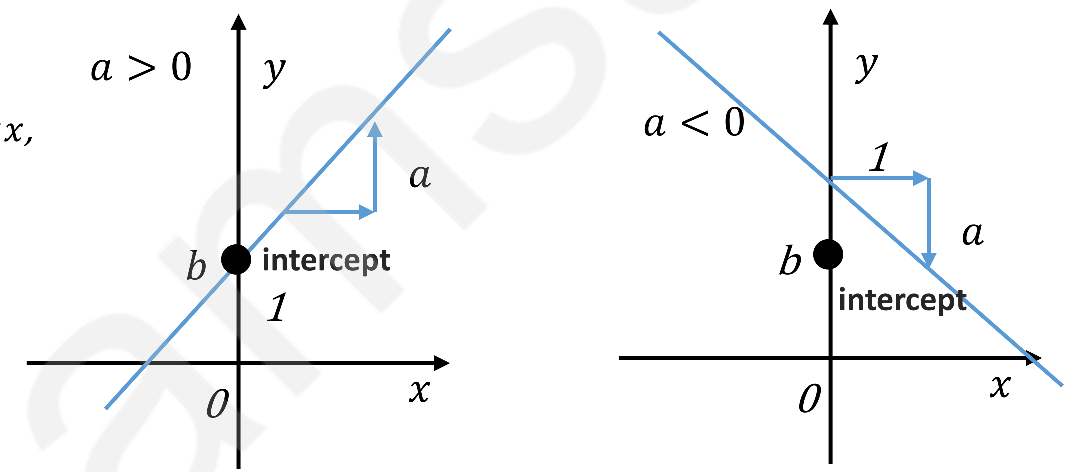
	$4a + (-3b)$		$+ \ 4a^2b$	$+ \ 6$
Coefficient	4	-3	4	6
Degree	1	1	3	0

First-Degree Equation

- The first-degree equation is represented by a straight line and the second equation is represented by a parabolic graph.
- The shape of the graph varies depending on whether the coefficient in front of the largest degree of n-degree is positive or negative.

Graph of a First-Degree Equation

First–Degree Equation for x,
$ax + b(but, a \neq 0)$



a slope
b intercept

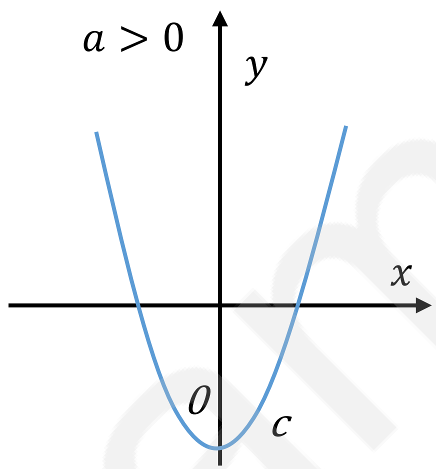
- a and b are treated as constants.
- The degree of term ax is 1, and the order of term b is zero because there is no letter x, and finally this equation becomes a first-degree equation.

Second-Degree Equation

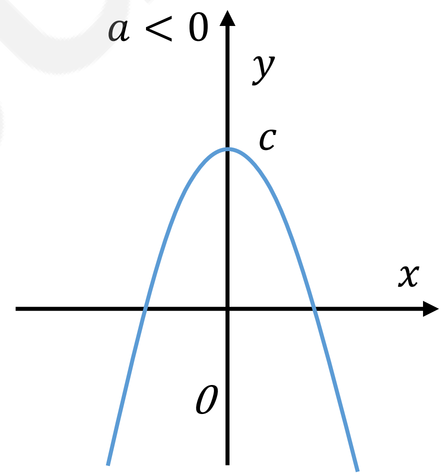
- If the coefficient a is positive, the parabola will convex downward, and if a is negative, the parabola will convex upward.

Graph of a Second-Degree Equation

Second-Degree Equation for x ,
$ax2 + bx + c$
$(but\ a \neq 0)$



Convex Downward



Convex Upward

- a and b are treated as constants. The degree of ax is 2, the degree of b is 1, and the degree of c is 0 because there is no letter x, therefore this equation is finally a second-degree equation.

Unit 2.

Basic Math for Data Science

Sequence

What is a Sequence?

- It is the listing of real numbers according to a rule in the order of natural numbers.
- Each number listed is referred to as the “term” of the sequence, and each term is distinguished by a comma.
- The sequence is usually placed in parentheses {} to represent the nth term.

Ex Example of a Sequence

$$1,2,3,4$$
$$x_1, x_2, x_3, x_4$$
$$x_1, x_2, \cdots\ x_n$$

The Sum and Product of Sequences

- Σ (sigma) : It is a symbol that briefly represents the sum of sequences and is read a sum.
- The Π symbol is an uppercase letter of the Greek letter pi, but it is not read as pi, but rather as a product.
- The starting index value is displayed below the sum and multiplication symbols, and the final index value is displayed above.
- Multiplication sign can be confused with the letter x, so it is marked as a dot, such as a·b, not a x b.

$$\sum_{i=1}^{N} x_i = x_1 + x_2 + \cdots + x_N$$

$$\prod_{i=1}^{N} x_i = x_1 \cdot x_2 \cdot \cdots \cdot x_N$$

Arithmetic Sequence

- The sequence created by adding a certain number d to the first term a is called a constant sequence. That is, the equivalent sequence $\{a_n\}$ satisfies the following for the natural number n.

$$a_{n+1} = a_n + d(n = 1,2,3,\cdots)$$

- In the definition of an arithmetic sequence, a constant number d is called a common difference of the sequence.

The Sum of Arithmetic Sequence

- The sum of S_n from the first term to the nth term of the arithmetic sequence $\{a_n\}$ with the first term a and the common difference d is as follows.

$$S_n = \frac{n(a_1 + a_n)}{2} = \frac{n(2a + (n-1)d)}{2}$$

Find the S_n for the arithmetic sequence $\{a_n\}$, and then solve for S_{10}.

- The first term is 5, and the common difference is 3.

In [1]:
```python
N = 10
a = 5
d = 3

sum1 = 0
for n in range(1,N+1):
    sum1 = sum1 + (a+(n-1)*d)
print("S10 = ", sum1, "by summation")
```

```
S10 =  185 by summation
```

Geometric Sequence

- The sequence produced by multiplying the first term a by a constant number r in turn is called an equal sequence. That is, the sequence $\{a_n\}$ satisfies the following for the natural number n.

$$a_{n+1} = a_n r \ (n = 1,2,3,\cdots)$$

- In the definition of an geometric sequence, the constant number r is called a common ratio.

The Sum of Geometric Sequences

- The sum S_n from the first term to the nth term of the geometric sequence $\{a_n\}$ with the first term a and the common ratio r is as follows.

$$S_n = \begin{cases} na & , r = 1 \\ \dfrac{a(1-r^n)}{1-r} = \dfrac{a(r^n-1)}{r-1} & , r \neq 1 \end{cases}$$

Properties of Σ

- The sign for the sum of sequences Σ has properties as listed below.

For sequences $\{a_n\}$, $\{b_n\}$ and constant c, the symbol Σ satisfies the following properties.

$$(1)\ \sum_{k=1}^{n}(a_k + bk) = \sum_{k=1}^{n} a_k + \sum_{k=1}^{n} b_k$$

$$(2)\ \sum_{k=1}^{n}(a_k - b_k) = \sum_{k=1}^{n} a_k - \sum_{k=1}^{n} b_k$$

$$(3)\ \sum_{k=1}^{n} ca_k = c\sum_{k=1}^{n} a_k$$

$$(4)\ \sum_{k=1}^{n} c = cn$$

The Sum of Powers of Natural Numbers

- The sum of the powers of the natural number is often used to find the sum S_n for a general sequence $\{a_n\}$. The following represents such formulas.

(1) $\sum_{k=1}^{n} k = \frac{n(n+1)}{2}$

(2) $\sum_{k=1}^{n} k^2 = \frac{n(n+1)(2n+1)}{6}$

(3) $\sum_{k=1}^{n} k^3 = \left(\frac{n(n+1)}{2}\right)^2$

Find the sum.

$$\sum_{k=1}^{10}(k^2-k)$$

In [2]:

```python
N = 10

sum1 = 0
for k in range(1,N+1):
    sum1 = sum1 + (k**2-k)
print("S10 = ", sum1, "by summation")
```

```
S10 =  330 by summation
```

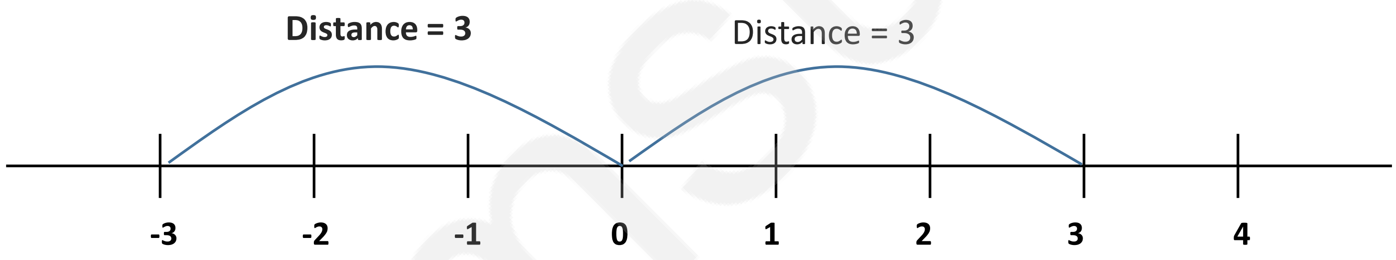
Unit 2.

Basic Math for Data Science

Definition of Absolute Value

Absolute Value

- The absolute value of a certain number means the distance on the vertical line between the number and zero.



Definition of the Euclidean Distance

Euclidean Distance

- It means the length of a line segment between two points as if measured with a ruler.
- The Euclidean distance in two dimensions is as follows.
- When there are points $p = (p_1, p_2, ..., p_n)$ and $q = (q_1, q_2, ..., q_n)$ are represented by orthogonal coordinate systems, the distances of two points p, q are calculated using the two Euclidean norms as follows.

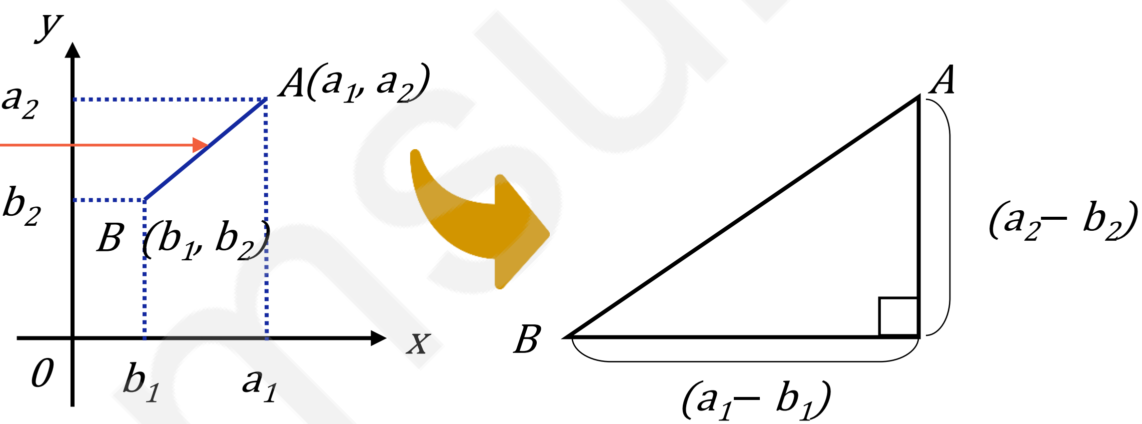
$$\| p - q \| = \sqrt{(p-q)\cdot(p-q)} = \sqrt{\| p \|^2 + \| q \|^2 - 2p \cdot q}$$

- The Euclidean distance between the point A and the origin can be expressed as $\| A \|$ using the symbol ‘ ‖ ’ ,and the distance between the two points of the point A and point B can be expressed as $\| A - B \|$.

Euclidean Distance

- As shown in the figure below, the length of the hypotenuse (segment AB) is calculated using the Pythagorean Theorem.

The distance of segment AB is the distance between point A and point B



Unit 2.

Basic Math for Data Science

The Concept of Sets

Sets and Elements

- Set: A group in which the target is clearly determined by a given condition
- Element: Each object that makes up the set.
- If A is the set of a measure 5, then 5 is an element of set A. In this case, the elements 1 and 5 belong to the set A, respectively, and are denoted by the symbols 1∈A, 5∈A. Measures that do not belong in set A, such as 2, 3, 4, are not elements and it is denoted by the ∈ symbol with a slash.

How to Represent a Set

- Enumeration: A method of representing a set by listing all elements in the set in { } (when representing a set in an enumeration, each element is written only once regardless of order, and can sometimes be omitted using the "…" if there are many elements and certain rules.

  Ex {1, 2, 3}

- Set-Builder Notation: A method of presenting and expressing the common properties of elements belonging to the set as conditions.

  Ex $\{x|x$ is a multiple of $3\}$

- Venn Diagram: A method of representing a set in a picture using a circle or rectangle, etc.

The Inclusion Relation Between Sets.

- For two sets A and B, if all elements of the set B belong to the set A, then the set B is called a subset of the set A. In this case, “Set B is included in Set A” or ”Set A contains Set B,” which is expressed as B⊂A.
- Every set is its own subset, and an empty set is a subset of every set.
- If the set B is not a subset of the set A, it is denoted by a single slash in the ⊂ symbol .
- When the elements of both sets A and B are the same, the sets A and B are said to be the same, and this is expressed as A=B. At this time, B is a subset of A (A⊂ B), and A is a subset of B (B⊂A).
- When the two sets A and B are not equal to each other, it is expressed with a single slash in the = symbol.
- The set A is called the true subset of the set B when the set A is a subset of the set B and is not equal, i.e, A⊂B and A=/=B (=/= symbolic replacement denoting ”is not equal”).

Calculation of Sets

Intersection and Union

- For two sets A and B, a set of all elements that belong to the set A and B is called the intersection of A and B, which is represented by the symbol A∩B. If this is expressed in the set-builder notation, it would be $\{x|x \in A$ and $x \in B\}$.
- For the two sets A and B, a set of all elements belonging to the set A or B is called the set of A and B, which is represented by the symbol A∪B. If this is expressed in the set-builder notation, it would be $\{x|x \in A$ or $x \in B\}$.
- In general, when two sets A and B are finite sets, n(A∪B)=n(A)+n(B)-n(A∩B).

Universal Set and Complement Set

- When considering a subset of a given set, the initial set is called the universal set, which is denoted by the symbol U.
- If the set A is a subset of the whole set U, then the set of all elements that belong to the set U and do not belong to the set A의 is called the complement set of A, and it is denoted as A^C. . If this is expressed in the set-builder notation, it would be A^C = $\{x|x \in U$ and $x \notin A$(does not belong)$\}$; and $n(A^C)=n(U)-n(A)$.
- For two sets A and B, a set of all elements that belong to the set A and do not belong to the set B is called the difference of set B for A, which is represented by the symbols A-B. If the difference of set B for set A is expressed in set-builder notation, it would be A-B= $\{x|x \in A$ and $x \notin B$(does not belong)$\}$; and $n(A-B)=n(A)-n(A\cap B)$. The A-B and B-A, here, are different.

Unit 2.

Basic Math for Data Science

Concept of Functions

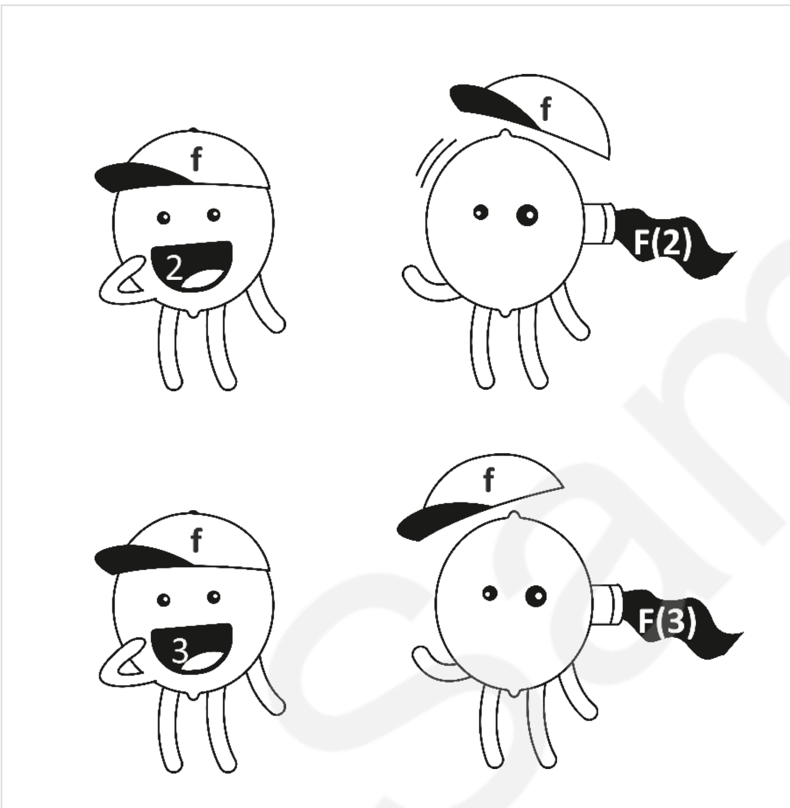
The rule, where each element x belonging to set A that corresponds with only one element y belonging to set B, is called a function.

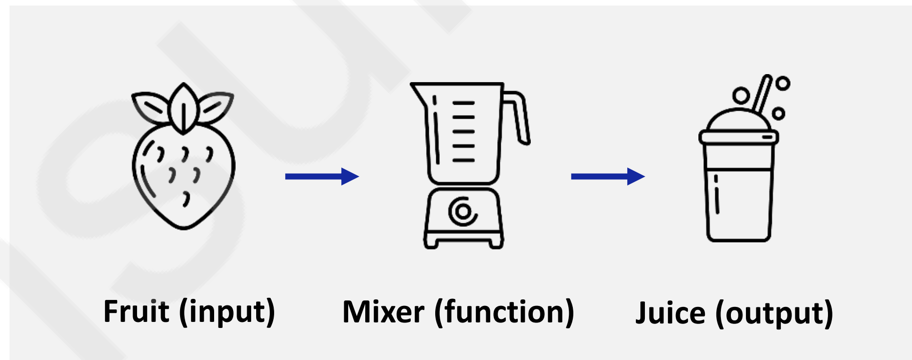
- Denoted by this symbol: f : A → B
- y corresponding to x is called y=f(x)
- The function is not always written as f(x), but also as g(x) depending on the equation.

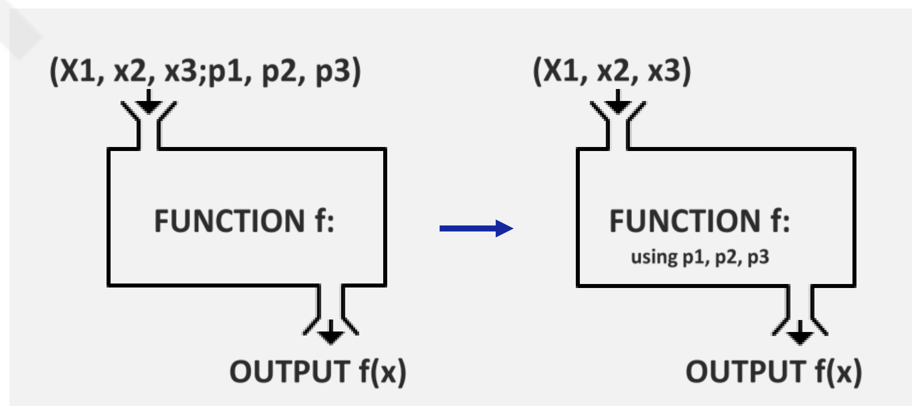
  However, the equation $x^2 + y^2 = 9$ is not a function because, in this case, both x and y are independent variables.

A Comparison Between a Function and a Computer

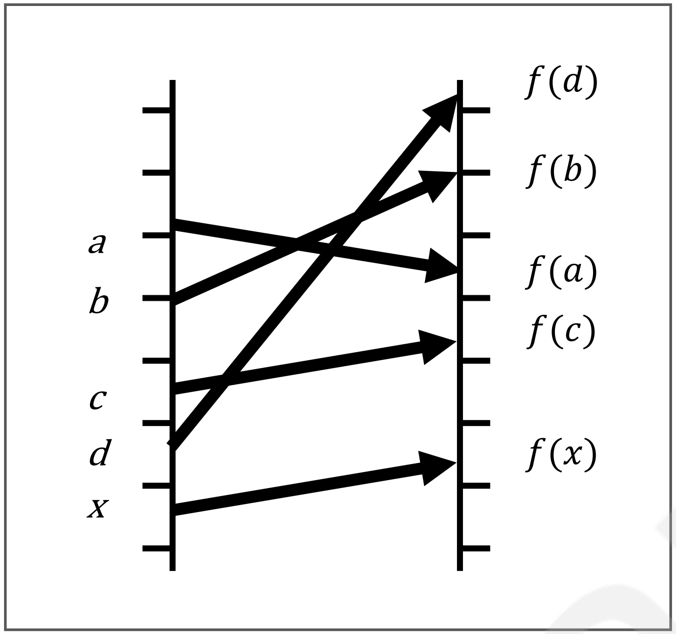
- When explaining functions and their correlation to computers, it is often compared to a mixer. That is, it is also described as an input-output device or a kind of number processor.







Function: A Set of Arrows.



- As shown in the figure, a function can be seen simply as a set of arrows pointing from one number to another.
- The arrow comes from each x in the domain of f and points to the value f(x).
- A function is an equation representing the relationship between variables and includes only algebraic operations.

Graph of a Function

Defining the Graph of a Function

- If f is a function with a set A as a domain, we define the graph of the function f as a set of ordered pairs as follows..

$$\{ (x, f(x)) \mid x \in A \}$$

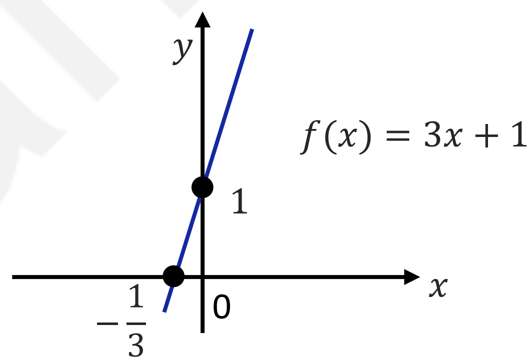
Find the domain and range of the following function and draw its graph.

$$f(x) = 3x + 1$$

- The given function f is defined for all real numbers x, and the phrase 3x+1 of x by f is also defined for all real numbers.
- Therefore, the domain and the range of the function f is the real set $\mathbb{R}$.

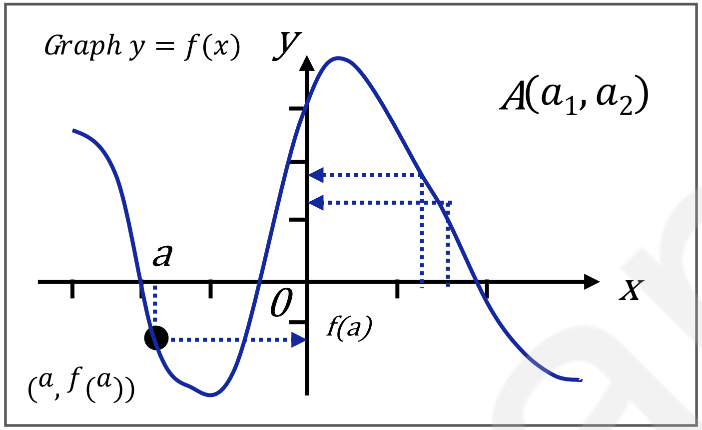
In [3]:

```python
import numpy as np
import matplotlib.pyplot as plt
#     f(x)=3x+1
x= np.linspace(-5,5,1001)
fx = 3*x+1
plt.plot(x,fx)
plt.show()
```



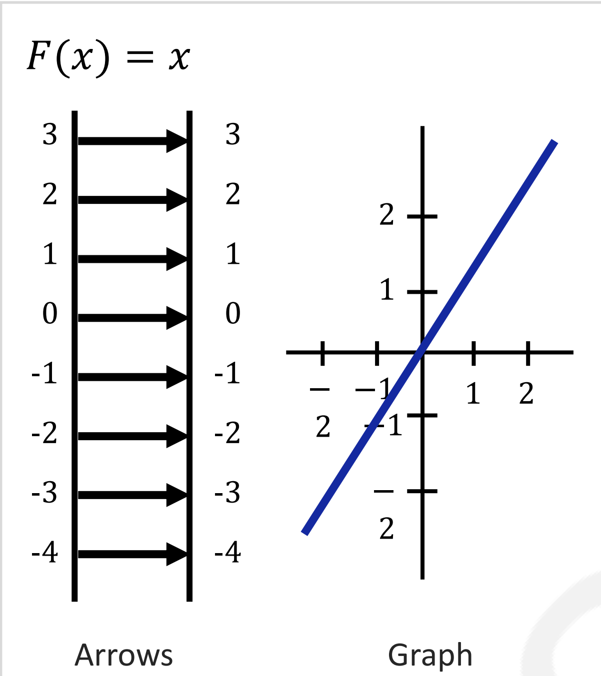
Graphic Representation of Functions

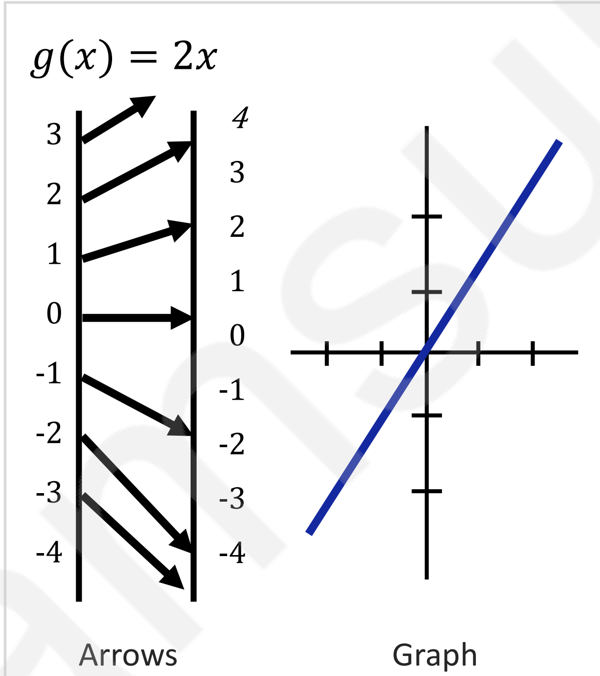
- When the first vertical line (or axis) is laid down, the function can be viewed in the form of a graph.
- The input value x is placed on the horizontal axis, and the output value y is placed on the vertical axis.
- And then position the point$(a, f(a))$ above (or below) the point with the coordinates of y corresponding to the value of the function f의 and point a and a.

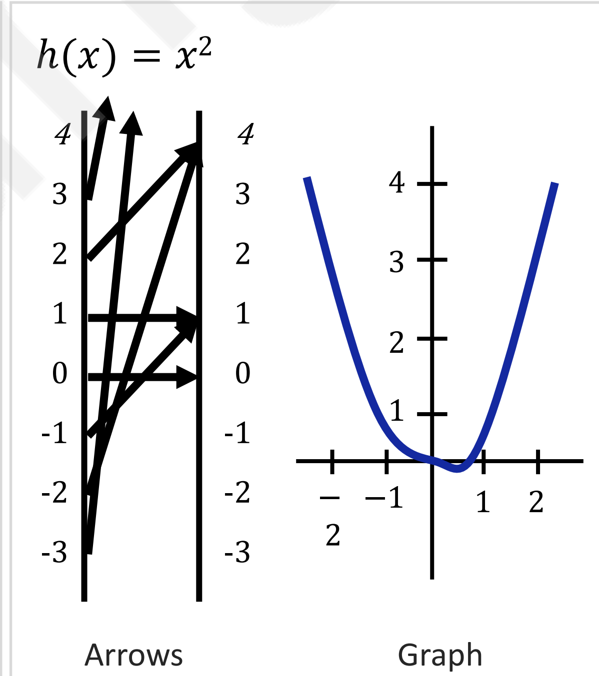


- The curve consists of all points (x, y) determined by $y = f(x)$ which is called the graph $y = f(x)$.

Few Examples of Graphic Representation of Functions





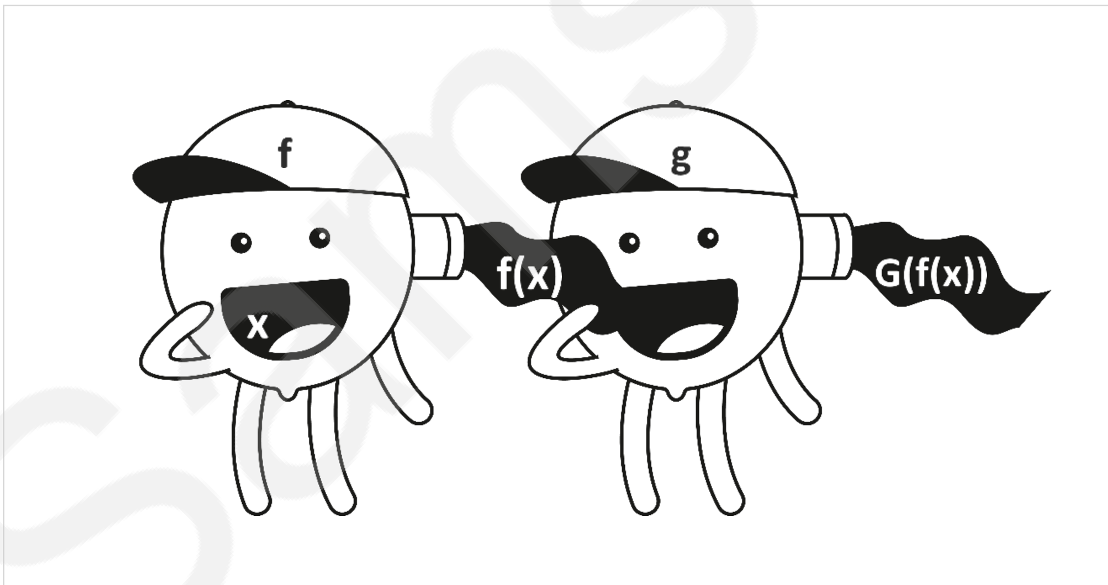


What is a Composite Function?

Composite Functions

- It is also possible to “insert” one function into another.

Ex $f(x) = 10 + x^2$ $g(x) = 3 + x$

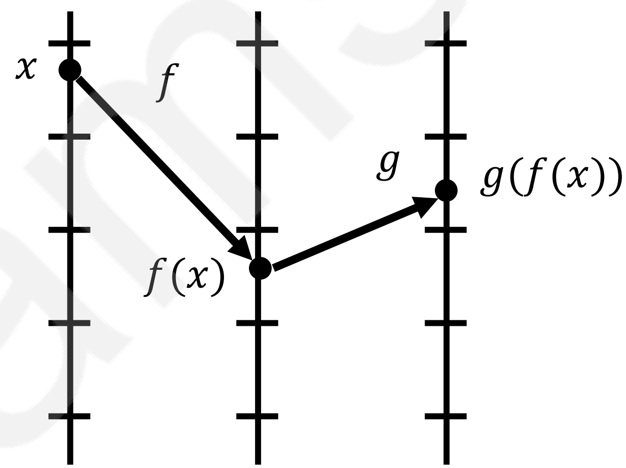


Composite Functions

- It is possible to substitute a value for x, or a function for x. In other words, the combination of several functions as follows is called a composite function.

$$f(g(x)) = 10 + g(x)^2 = 10 + (3 + x)^2$$

$$g(f(x)) = 3 + f(x) = 3 + (10 + x^2)$$



$$h : x \rightarrow f(x) \rightarrow g(f(x))$$

Find $(g \circ f)(2), (f \circ g)(2), (g \circ g)(2)$ for the functions f and g in $f(x) = x^3 + 1$, and $g(x) = \sqrt{x+2}$, respectively.

In [4]:
```
# f(x)=x^3+1
def f(x):
    return x**3 + 1

# g(x)=sqrt(x+2)
def g(x):
    return np.sqrt(x+2)

# (g o f)(2), (f o g)(2), (g o g)(2)
print("(g o f)(2) =", g(f(2)))
print("(f o g)(2) =", f(g(2)))
print("(g o g)(2) =", g(g(2)))
```

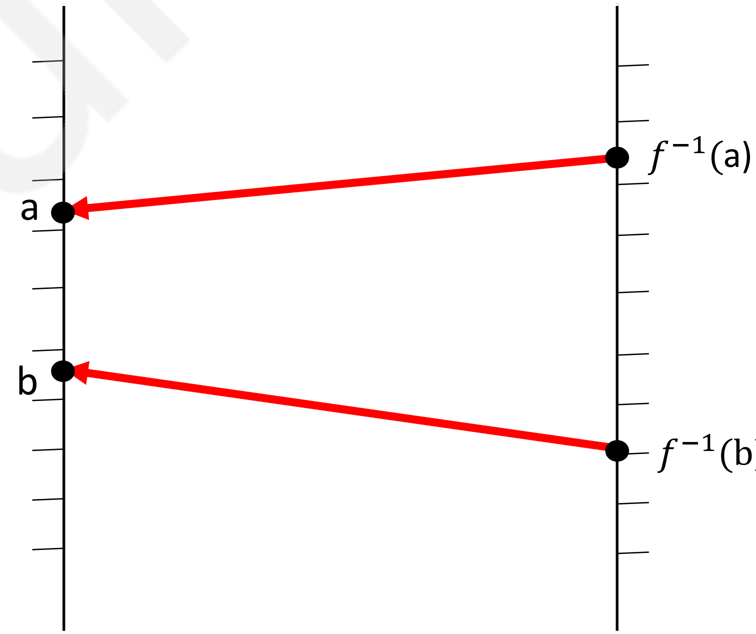
```
(g o f)(2) = 3.3166247903554
(f o g)(2) = 9.0
(g o g)(2) = 2.0
```

What is an Inverse Function?

Inverse Functions

- When combining two functions, sometimes, nothing happens.
- If $f(x) = x^{\frac{1}{3}}$ *and* $g(y) = y^3$, then $h(x) = g(f(x)) = (x^{\frac{1}{3}})^3 = x$
- If x is inserted into $g \circ f$, then x is returned. This is because h is the cube root of cube. Eventually, g returns the result of f to its original state.
- If f is a one-to-one function, we can create a new function, f^{-1}, that is, and inverse function of f.
- For every x in the domain, f^{-1} is defined as follows.

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



Unit 2.

Basic Math for Data Science

Exponents

An exponent is the number multiplied by a specific number called the base.

- The number expressed by the base and exponent is called power. It is usually represented in power of 10.

$$2^3 = 2 \cdot 2 \cdot 2 = 8$$

- As shown above, 2 is multiplied 3 times by the number of exponents. In this case, 3 is called the exponent and 2 is called the base.

Logarithms: The Opposite of Exponentials

$$2^3 = 2 \cdot 2 \cdot 2 = 8$$

- In the case above, the index 3 signifies how many times the base 2 must be multiplied, whereas the value that means 'how many times does 2 need to multiply in order to become 8?' is called a log, using the symbol $\log\_ 28 = 3$. At this time, 2 is called the base and 8 is called the antilogarithm.

Exponentials

$$8 = 2^3$$
$$100 = 10^2$$
$$81 = 3^4$$
$$0.01 = 10^{-2}$$

Logarithms

$$3 = \log_2 8$$
$$2 = \log_{10} 100$$
$$4 = \log_3 81$$
$$-2 = \log_{10} 0.01$$

Exponents **Log**

Base

Antilog

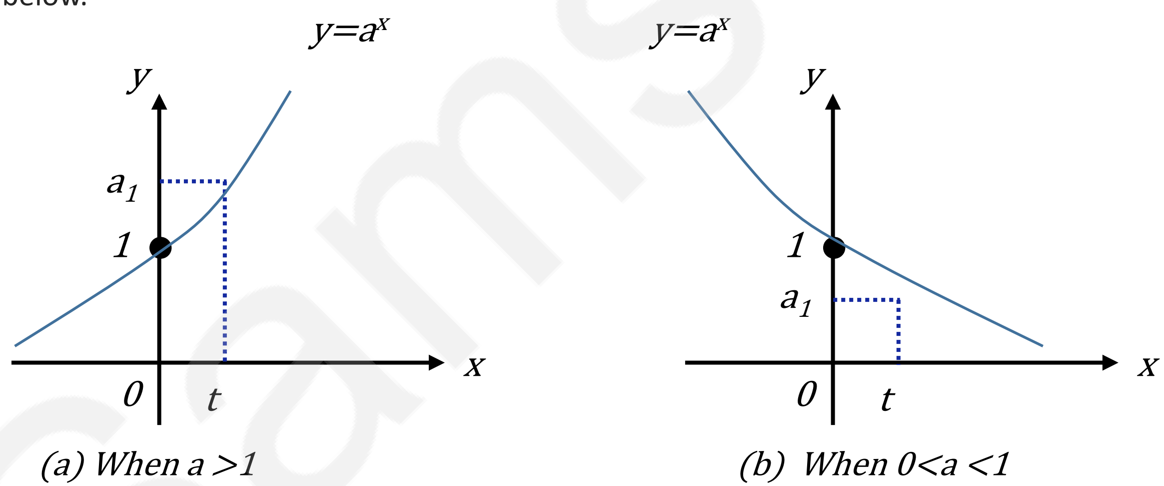
Exponential Functions

- A function in the form as shown below.

$$f(x) = a^x$$

- Here, the "base" a is a constant, and the exponent x is the variable.

For the real number a, satisfying $a > 0, a \neq 1$, the graph of the exponential function $y = ax$ with the range a can be drawn as shown below.



The graph of the exponential function $y = a^2$ with the range a.

- When the base is 1 in the exponent, it means nothing, so it is taken out altogether.

Draw the Graph of the Exponential Functions

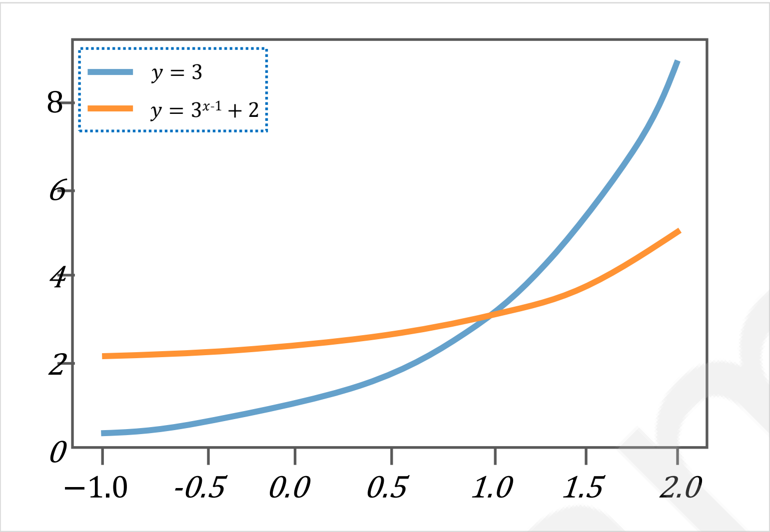
(a) $y = 3^{x-1} + 2$

b) $y = 2^{-x-2}$

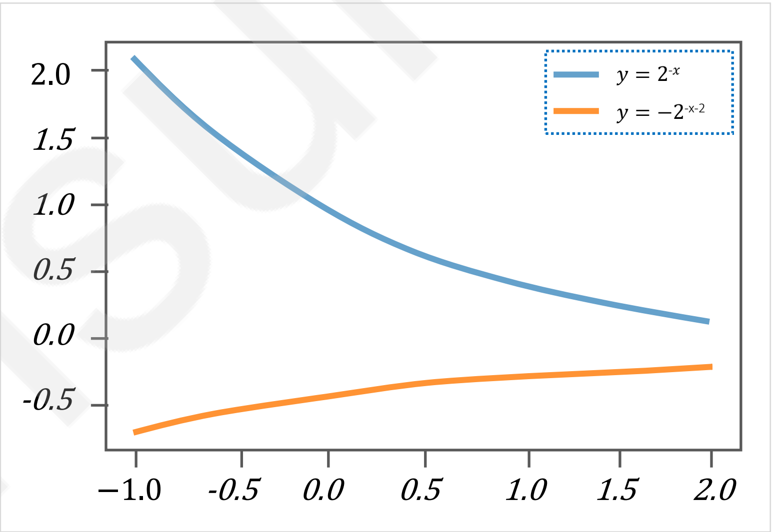
```
In [5]: import numpy as np
        import matplotlib.pyplot as plt

        # (a) y1=3^x, y2=3^(x-1)+2
        x = np.linspace(-1,2,301)
        y1 = 3**x
        y2 = 3**(x-1)+2

        # For visualization
        graph1, = plt.plot(x,y1)
        graph2, = plt.plot(x,y2)
        plt.legend(handles=(graph1,graph2), labels=(r'$y=3^x$',r'$y=3^{x-1}+2$'))
        plt.show()

        # (b) y1=2^(-x), y2=-2^(-x-2)}
        x = np.linspace(-1,2,301)
        y1 = 2**(-x)
        y2 = -2**(-x-2)

        # For visualization
        graph1, = plt.plot(x,y1)
        graph2, = plt.plot(x,y2)
        plt.legend(handles=(graph1,graph2), labels=(r'$y=2^{-x}$',r'$y=-2^{-x-2}$'))
        plt.show()
```

Draw the Graph of the Exponential Functions

(a) $y = 3^{x-1} + 2$



b) $y = 2^{-x-2}$



Logarithms (log)

Logarithmic Functions (log function)

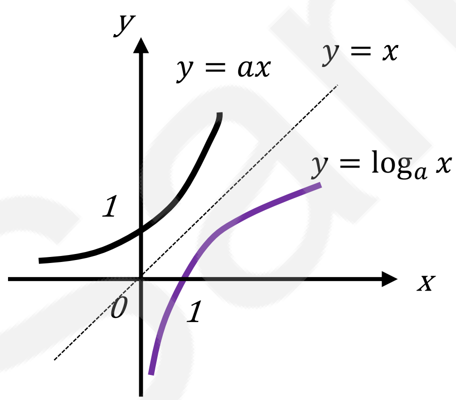
- The log function$f\ :\ \mathbb{R}^+\ \rightarrow \mathbb{R}$ for the real number a (where $a > 0, a \neq 1$) and the set of real numbers $\mathbb{R}$ and set $\mathbb{R}^+$ is as follows.

$$f(x) = log_a x$$

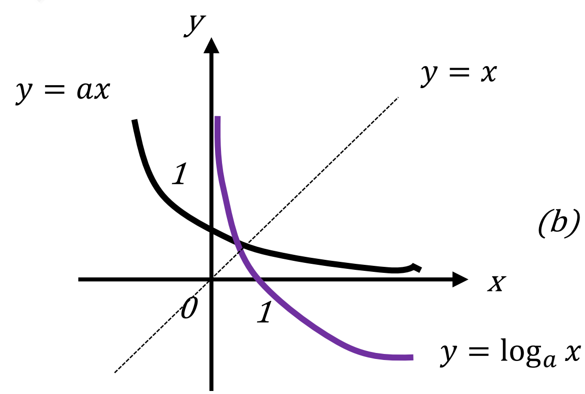
- The inverse function of $y = a^x$, according to the definition of logarithms, is the inverse function of the exponential function $x = \ log_a y$ and the inverse function of $y = a^x$: logarithmic function

For the real number a, satisfying $a > 0, a \neq 1$, the graph of the logarithmic function $\mathrm{y} = \log_{\mathrm{a}} \mathrm{x}$ with the range a can be drawn as shown below.

(a) When a >1



(b) When 0<a <1



The Graph of the Logarithmic Function $y = log_a x$ with the range a

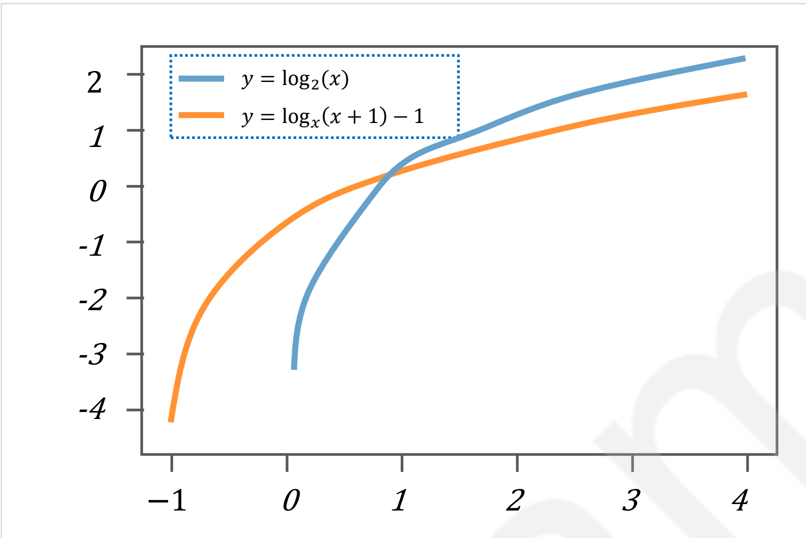
Draw the Graph of the Function Below.

- Draw the graph of the function $y = log_2(x+1) - 1$.

In [6]:
```
import numpy as np
import matplotlib.pyplot as plt

# y1=log_2(x), y2=log_2(x+1)-1
x1 = np.linspace(0.1,4,401)
y1 = np.log2(x1)
x2 = np.linspace(-0.9,4,501)
y2 = np.log2(x2+1)-1

# For visualization
graph1, = plt.plot(x1,y1)
graph2, = plt.plot(x2,y2)
plt.legend(handles=(graph1,graph2), labels=(r'$y=log_2(x)$',r'$y=log_2(x+1)-1$'))
plt.show()
```

Draw the Graph of the Function Below.

- Draw the graph of the function $y = log_2(x+1) - 1$.



Unit 2.

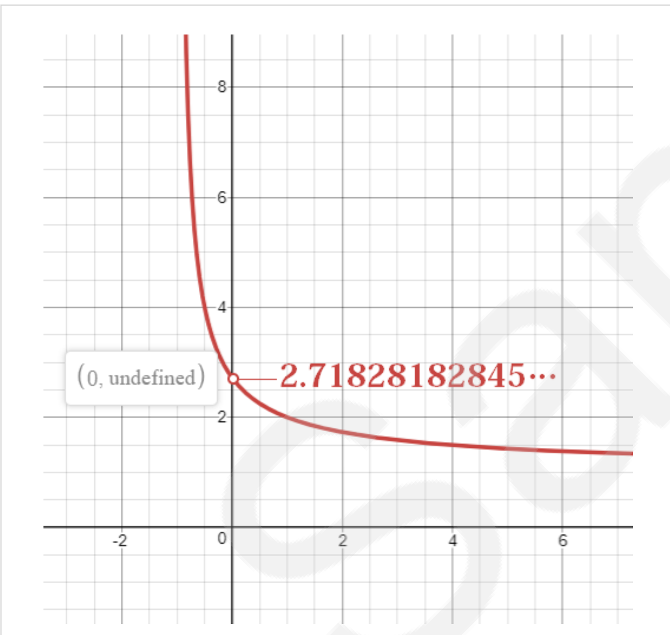
Basic Math for Data Science

- 2.1. Algebra
- 2.2. Sequence
- 2.3. Absolute Value and Euclidean Distance
- 2.4. Sets
- 2.5. Concept of Functions
- 2.6. Exponential and Logarithmic Functions
- 2.7. Natural Logarithms
- 2.8. Sigmoid Functions
- 2.9. Trigonometric Functions

What is Euler's Constant?

Euler's Constant

- e is also called the base of the natural logarithm and is an irrational constant expressed as 2.71828 18284…. For our sake, let's just know that it is an irrational number with an approximate value of 2.7182.



x	(1+x)^(1/x)
0.01	2.70481383
0.0099	2.7049473
0.0098	2.70508079
0.0097	2.70521431
0.0096	2.70534785
0.0095	2.70548142
0.0094	2.70561501
0.0093	2.70574863
0.0092	2.70588227
0.0091	2.70601593
0.009	2.70614962
0.0089	2.70628333
0.0088	2.70641707
0.0087	2.70655083
0.0086	2.70668461
0.0085	2.70681842
0.0084	2.70695225
0.0083	2.70708611
0.0082	2.70722
0.0081	2.70735390
0.008	2.70748783
0.0079	2.70762179
0.0078	2.70775577
0.0077	2.70788977
0.0076	2.70802380
0.0075	2.70815785
0.0074	2.70829193
0.0073	2.70842603
0.0072	2.70856016

x	(1+x)^(1/x)
0.0001	2.71814593
0.00009	2.7181595
0.00008	2.71817311
0.00007	2.71818669
0.00006	2.71820028
0.00005	2.71821387
0.00004	2.71822746
0.00003	2.71824106
0.00002	2.71825465
0.00001	2.71826824
0	#DIV/0!
-1.E-05	2.71829542
-2.E-05	2.71830901
-3.E-05	2.71832260
-4.E-05	2.71833620
-5.E-05	2.71834979
-6.E-05	2.71836338
-7.E-05	2.71837697
-8.E-05	2.71839057
-9.E-05	2.71840416
-0.0001	2.71841776
-0.0001	2.71843135
-0.0001	2.71844494
-0.0001	2.71845854
-0.0001	2.71847213
-0.0002	2.71848573
-0.0002	2.71849932
-0.0002	2.71851292
-0.0002	2.71852651

What are Natural Logarithms?

Natural Logarithms

- The log of the number represented by the symbol e is called the "natural logarithm."
- The natural logarithm of a certain number x is represented as $\ln x$.

The Euler's Constant is the Base of Natural Log

Natural log is a log that has e as its base.

- Natural logarithms, logs with e as its base, are used in many fields.
- Log functions can have multiple bases by definition, but in general, $\log x$, which does not have a separate base, signifies a natural logarithm. However, due to the confusion with common logarithms, it is now marked as ln x.

What are Common Logarithms?

Common Logarithms

- The logarithm is the number of powers that are multiplied by the base to obtain a certain positive number. For example, the logarithm of 100, whose base is 10, is 2 or $log_{10}100 = 2$. This is because $10^2 = 100$. The common logarithm is a positive value using 10 as its base and is expressed as log x.

Unit 2.

Basic Math for Data Science

Sigmoid Functions

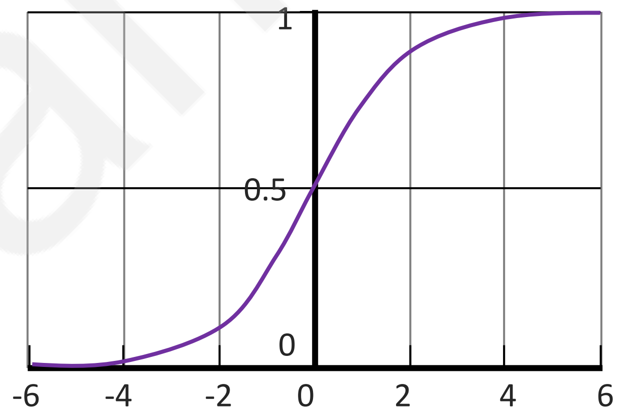
The sigmoid function has the entire real number as a domain, and the return value generally increases monotonically, but can also decrease monotonically.

- The return value (y-axis) of the sigmoid function often ranges from 0 to 1.
  For example, the logistic function below is defined by the following formula.

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

- Alternatively, it may range from -1 to 1.

Logistic Curve

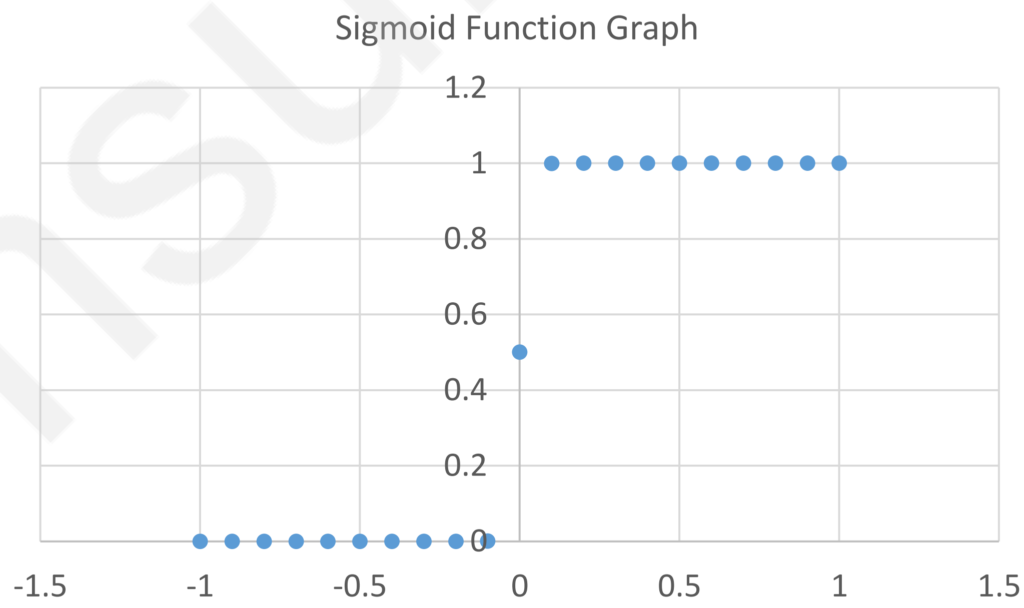


It can also be created through Excel.

- The sigmoid function is a function that is commonly seen in the field of artificial intelligence.
  When a is 1, it is called a standard sigmoid function.

$$\frac{1}{1+\exp(-ax)}$$

x	y
-1	3.72008E-44
-0.9	8.19401E-40
-0.8	1.80485E-35
-0.7	3.97545E-31
-0.6	8.75651E-27
-0.5	1.92875E-22
-0.4	4.24835E-18
-0.3	9.35762E-14
-0.2	2.06115E-09
-0.1	4.53979E-05
0	0.5
0.1	0.999954602
0.2	0.999999998
0.3	1
0.4	1
0.5	1
0.6	1
0.7	1
0.8	1
0.9	1
1	1



Unit 2.

Basic Math for Data Science

What are Trigonometric Functions?

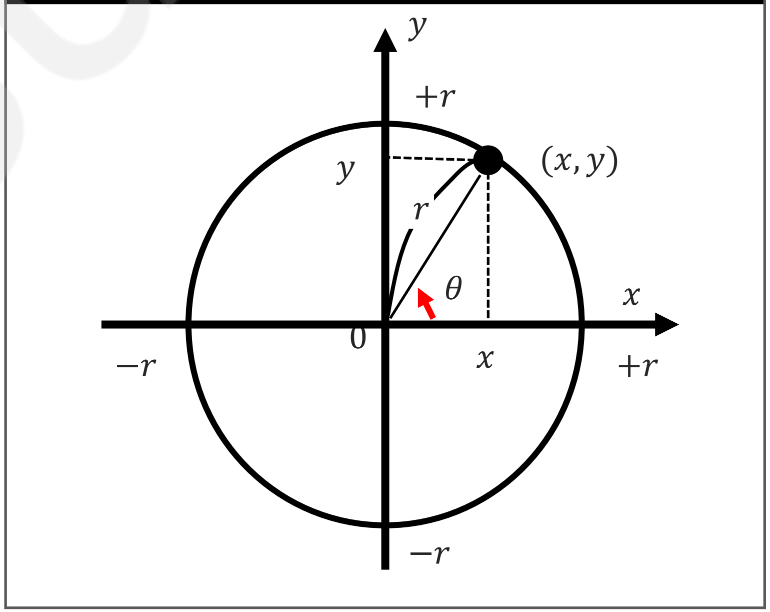
A trigonometric function refers to a function whose value varies depending on the size of the angle, or in other words, a function whose size is a variable.

$$\sin\theta = \frac{y}{r}, \qquad \cos\theta = \frac{x}{r}, \qquad \tan\theta = \frac{y}{x}$$

$$\csc\theta = \frac{r}{y}, \qquad \sec\theta = \frac{r}{x}, \qquad \cot\theta = \frac{x}{y}$$

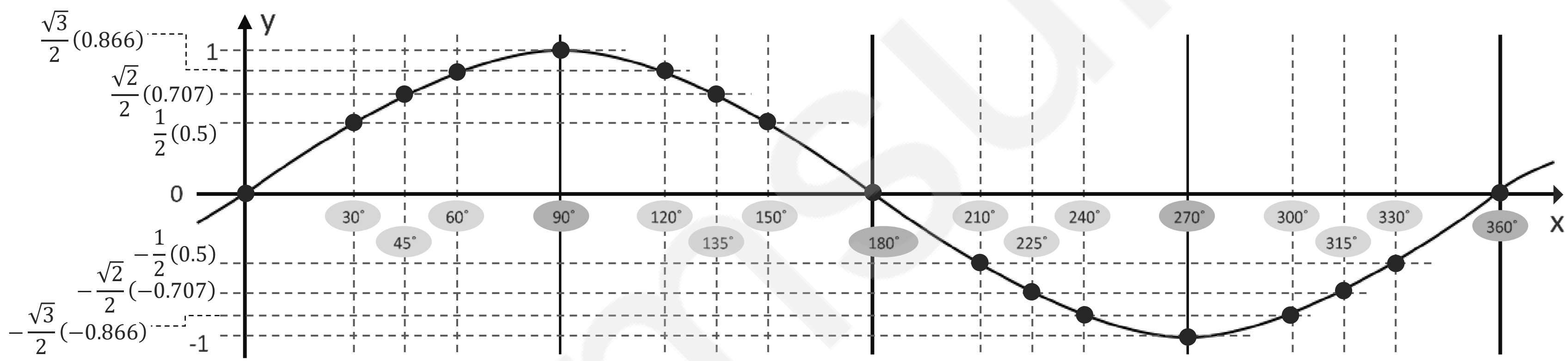
- The trigonometric function is defined using a point (x, y) on a circle with the radius r.

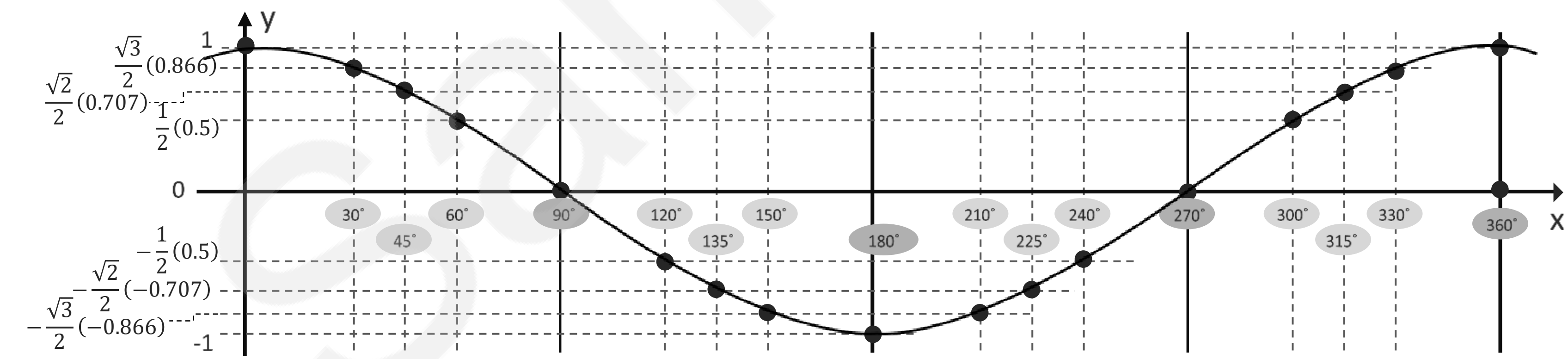
Definition of Trigonometric Funtions



Graph of Trigonometric Functions

Graphs of $y = \sin x$ and $y = \cos x$.





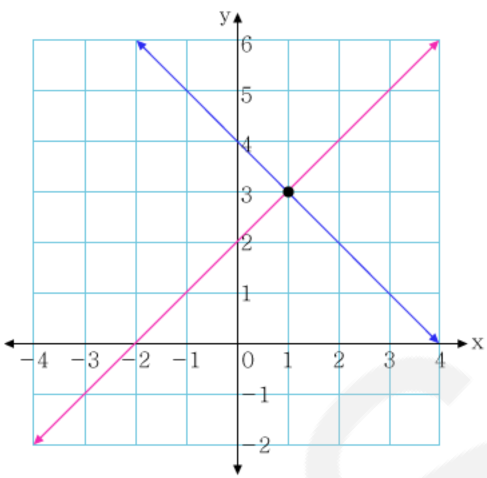
Unit 3.

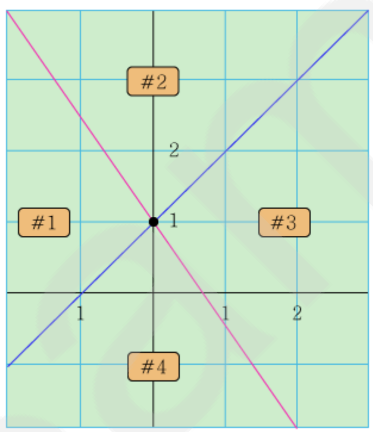
Understanding Data Science: Vector

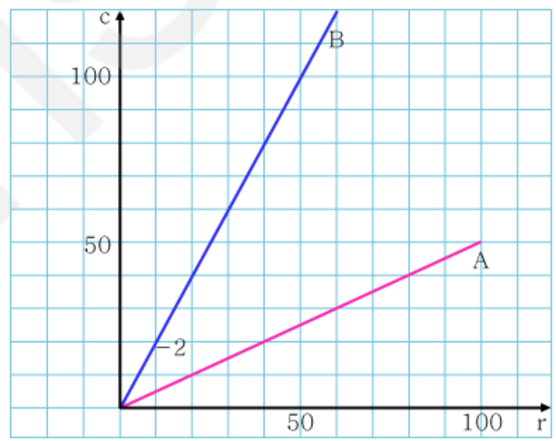
Concept of Vectors

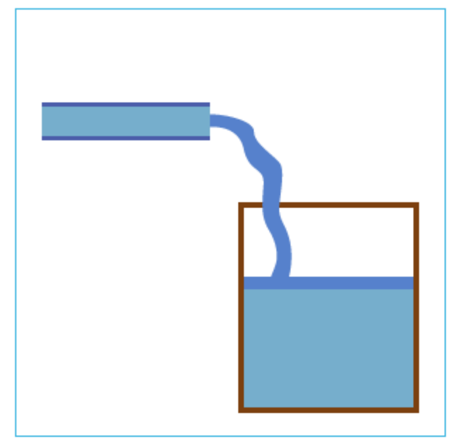
What is a vector?

- A vector refers to what can be expressed in the form of a linear combination regarding the elements of set A.
- It refers to the case where the elements of set A $x_1, x_2, x_3, \ldots, x_n$ is multiplied by each constant $a_1, a_2, a_3, \ldots, a_n$, then added into $a_1x_1 + a_2x_2 + a_3x_3 + \ldots + a_nx_n$ within set A. This type of equation is called the linear combination of $x_1, x_2, x_3, \ldots, x_n$.









Linear Relationship between Time and Height of Water

Various Definitions of Vector

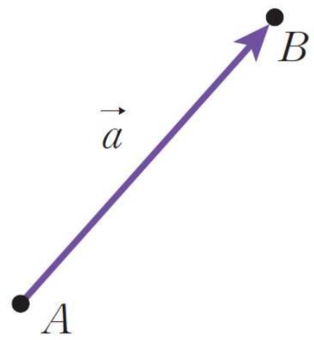
Definition of Vector (1)

- A one-dimensional array of numbers or symbols.

$$a = \begin{bmatrix} 2 \\ 6 \\ 3 \end{bmatrix}, \quad b = [2 \quad 6 \quad 3], \quad c = (2, 6, 3)$$

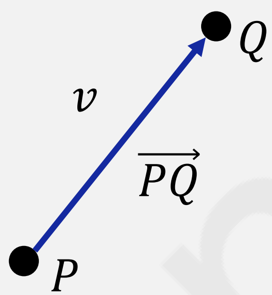
$$x = \begin{bmatrix} a \\ b \end{bmatrix}, \quad y = [c \quad d], \quad \vec{v} = (a, b, c)$$

- Vector : a quantity that has both sizeand direction.
- Starting Point of Arrow A : Tail, Ending Point of Arrow B : Head, Length of Arrow: Magnitude
- Its symbol is $\overrightarrow{AB}$ or $\vec{a}$.



Definition of Vector (2)

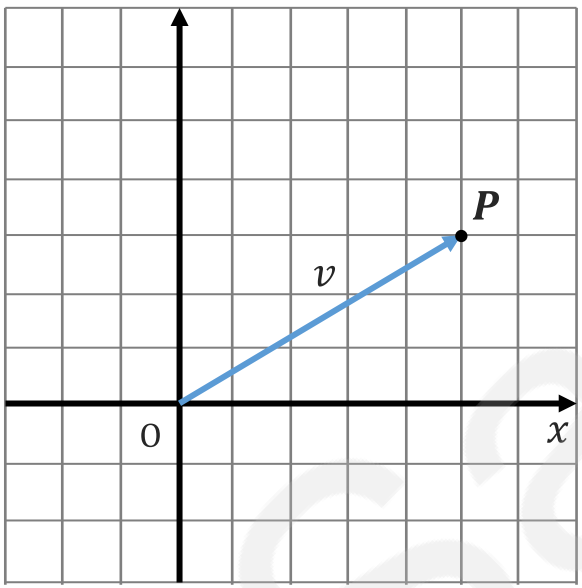
- The element of the vector space is called Vector, which intuitively represents an object whose ratio of direction and length is defined. As a point in space, the end of the arrow corresponds to the coordinates of the vector.
- A Vector that represents a line segment with a starting point and an ending point.



$$\overrightarrow{PQ} = \begin{bmatrix} 5 \\ 3 \end{bmatrix} \; or \; v = \begin{bmatrix} 5 \\ 3 \end{bmatrix}$$

Vector Space

- The vector space in linear algebra, mentioned here, refers to the space in which elements can be added to each other or increased or decreased to a given multiple. In other words, a space created by linear combinations is called a vector space.
- A vector representing the position in the coordinate space.



$$v = \begin{bmatrix} 5 \\ 3 \end{bmatrix}$$

Unit 3.

Understanding Data Science: Vector

Vector Norm

Vector Norm

- A function for calculating the magnitude of a vector is called Vector Norm**(norm).** The norm of the vector u is expressed as $||u||$ and satisfies the following properties. Here, u and v are vectors and α scalar.

(1) $||u|| \geq 0$

(2) $||\alpha u|| = |\alpha|\,||u||$

(3) $||u + v|| \leq ||u|| + ||v||$

(4) $||u|| = 0$ only when $u = 0$.

Unit 3.

Understanding Data Science: Vector

Inner Product of Vectors

Since the vector is defined as “a quantity with size and direction,” the multiplication of the vector should take both “size and direction” into account.

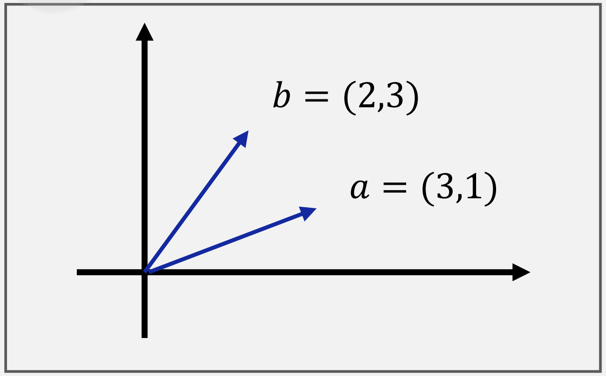
- The multiplication of vectors only concerned with the size (scalar), is called the Inner Product.
- The Inner Product of the two-dimensional vector can be calculated by the following equation.

$$\mathrm{a} = \begin{bmatrix} 3 \\ 1 \end{bmatrix} \quad \mathrm{b} = \begin{bmatrix} 2 \\ 3 \end{bmatrix}$$

$$\mathbf{a} \cdot \boldsymbol{b} = a_1 b_1 + a_2 b_2$$

If you transpose the vector and multiply it, it is the same as calculating the Inner Product.

$$\mathbf{a} \cdot \boldsymbol{b} = 3 \cdot 2 + 1 \cdot 3 = 9$$



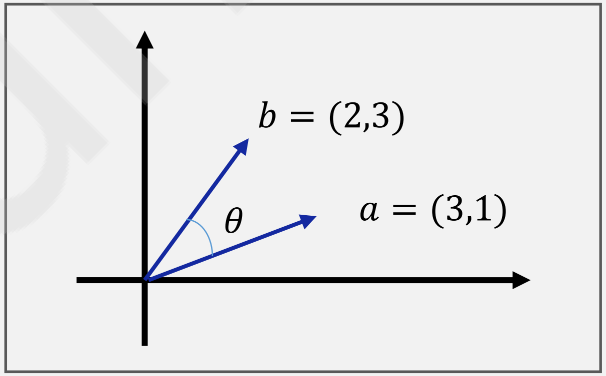
- In this way, if the vector’s Inner Product is taken, the result is in a form of a real number(size) rather than in a form of a vector.
- This real number is sometimes referred to as scalar, and so this process is often called the scalar multiplication.
- The operator symbol of the Inner Product is not x, but ·, and it is called a “dot.”

If the angle between the vectors a and b is θ, then the Inner Product may be expressed as the following.

$$\mathbf{a} \cdot \boldsymbol{b} = |\mathbf{a}| \cdot |\mathbf{b}| \cdot \cos\theta$$

The notation |a|, in the equation above, is the length of the vector.
For example, the length of the vector $a = (a_1, a_2)$is defined as the following.

$$|a| = \sqrt{a_1^2 + a_2^2}$$



Unit 3.

Understanding Data Science: Vector

Orthogonal Condition

Vectors with the Inner Product 0 are Orthogonal to Each Other

- If the angle between the vectors a and b is θ, the Inner Product can be expressed as follows.

$$<a, \boldsymbol{b}> = |a| \cdot |b| \cdot \cos\theta$$

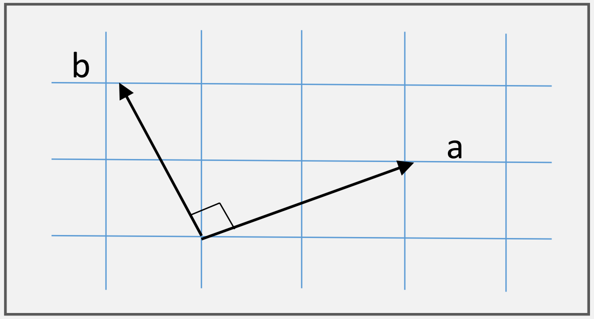
The fact that two vectors a and b are orthogonal (meeting vertically) means that the angles of the two vectors are at 90 °.

In other words, <a,b> is 0.

The combination of $\cos 90° = 0$ is $<\mathbf{a}, \boldsymbol{b}> = |\mathbf{a}| \cdot |\mathbf{b}| \cdot \cos\theta = 0$.

For example, if we find the inner product of a=(2,1) and b=(-1,2), as shown in the figure,

$$<a,b>=2 \times (-1) + 1 \times 2=0.$$

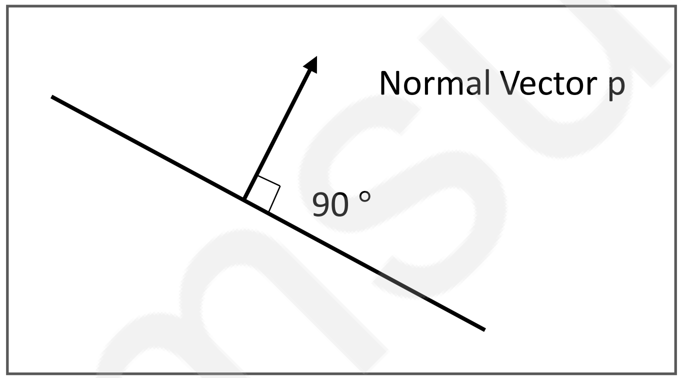


Unit 3.

Understanding Data Science: Vector

Normal Vector

A Normal Vector refers to a vector that is perpendicular to a certain straight line.



- If the equation of the straight line, shown above, is ax+by+c=0, the Normal Vector p becomes p=(a,b).

Unit 3.

Understanding Data Science: Vector

Cosine Similarity

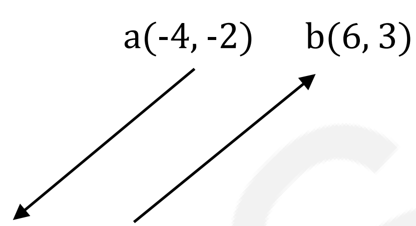
A high cosine similarity means that the vectors are more similar.

$$<a\,, \boldsymbol{b}> = \|\mathbf{a}\| \cdot \|\mathbf{b}\| \cdot \cos\theta$$

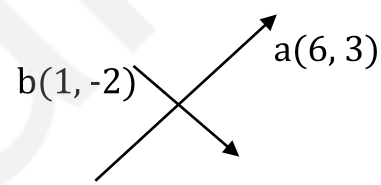
- The equation above solved around cos θ looks like the below.

$$\cos(a, b) = \frac{<a,b>}{\|\mathbf{a}\| \cdot \|\mathbf{b}\|}$$

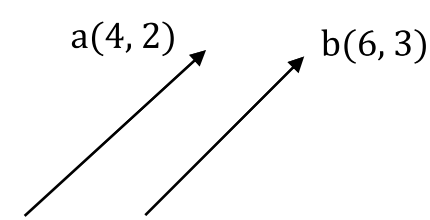
- Since the value of cosine similarity is within the interval of -1 ≤ cos ≤ 1, when the similarity is -1, the two vectors are parallel to each other in opposite directions; when it is 0, the vectors are orthogonal; and when it is 1, the two vectors are parallel in the same direction.



Cosine Similarity of -1



Cosine Similarity of 0



Cosine Similarity is 1

Unit 4.

Understanding Data Science: Matrix

- 4.1. Calculating Matrix
- 4.2. Reverse Matrix
- 4.3. Linear Transformation
- 4.4. Eigenvalues and Eigenvectors

Matrices

What is a Matrix?

- It contains several vectors.
- A matrix containing a training set is called a design matrix.

Ex The 150 samples in the Iris data are represented by the design matrix X.

$$X = \begin{pmatrix} 5.1 & 3.5 & 1.4 & 0.2 \\ 4.9 & 3.0 & 1.4 & 0.2 \\ 4.7 & 3.2 & 1.3 & 0.2 \\ 4.6 & 3.1 & 1.5 & 0.2 \\ \vdots & \vdots & \vdots & \vdots \\ 6.2 & 3.4 & 5.4 & 2.3 \\ 5.9 & 3.0 & 5.1 & 1.8 \end{pmatrix} = \begin{array}{ll} \begin{pmatrix} x_{1,1} & x_{1,2} & x_{1,3} & x_{1,4} \\ x_{2,1} & x_{2,2} & x_{2,3} & x_{2,4} \\ x_{3,1} & x_{3,2} & x_{3,3} & x_{3,4} \\ x_{4,1} & x_{4,2} & x_{4,3} & x_{4,4} \\ \vdots & \vdots & \vdots & \vdots \\ x_{149,1} & x_{149,2} & x_{149,3} & x_{149,4} \\ x_{150,1} & x_{150,2} & x_{150,3} & x_{150,4} \end{pmatrix} & \text{row} \\ \quad \text{column} & \end{array}$$

The Transposition Matrix $\mathbf{A}^{\mathrm{T}}$ of Matrix $\mathbf{A}$.

$$A = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1m} \\ a_{21} & a_{22} & \cdots & a_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nm} \end{pmatrix}, \qquad A^T = \begin{pmatrix} a_{11} & a_{21} & \cdots & a_{n1} \\ a_{12} & a_{22} & \cdots & a_{n2} \\ \vdots & \vdots & \ddots & \vdots \\ a_{1m} & a_{2m} & \cdots & a_{nm} \end{pmatrix}$$

- For example, if $A = \begin{pmatrix} 3 & 4 & 1 \\ 0 & 5 & 2 \end{pmatrix}$, then $A^T = \begin{pmatrix} 3 & 5 \\ 4 & 5 \\ 1 & 2 \end{pmatrix}$

Mathematical Expressions Using Matrices

A matrix allows for a concise mathematical expression.

Ex A Matrix Expression of Polynomials.

$$f(x) = f(x1, x2, x3)$$

$$= 2x_1x_1 - 4x_1x_2 + 3x_1x_3 + x_2x_1 + 2x_2x_2 + 6x_2x_3 - 2x_3x_1 + 3x_3x_2 + 2x_1 + 3x_2 - 4x_3 + 5$$

$$= (x_1\ x_2\ x_3)\begin{pmatrix} 2 & -4 & 3 \\ 1 & 2 & 6 \\ -2 & 3 & 2 \end{pmatrix}\begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} + (2 \quad 3 \quad -4)\begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} + 5$$

$$= X^T A_X + bTX + c$$

Special Matrices

Square Matrix, Diagonal Matrix, Unit Matrix, and Symmetric Matrix

$$\text{Square Matrix}\begin{pmatrix}2 & 0 & 1\\ 1 & 21 & 5\\ 4 & 5 & 12\end{pmatrix},\quad \text{Diagonal Matrix}\begin{pmatrix}50 & 0 & 0\\ 0 & 1 & 0\\ 0 & 0 & 8\end{pmatrix}$$

$$\text{Unit Matrix}\begin{pmatrix}1 & 0 & 0\\ 0 & 1 & 0\\ 0 & 0 & 1\end{pmatrix},\quad \text{Symmetric Matrix}\begin{pmatrix}1 & 2 & 11\\ 2 & 21 & 5\\ 11 & 5 & 1\end{pmatrix}$$

Multiplication of Matrices

How to multiply matrices

When $C = AB$, $c_{ij} = \sum_{k=1,s} a_{jk} b_{kj}$

If a $2 * 3$ Matrix $A = \begin{pmatrix} 3 & 4 & 1 \\ 0 & 5 & 2 \end{pmatrix}$ is multiplied by $3 * 3$ Matrix $B = \begin{pmatrix} 2 & 0 & 0 \\ 1 & 0 & 5 \\ 4 & 5 & 0 \end{pmatrix}$

$2 * 3$ Matrix $C = AB = \begin{pmatrix} 14 & 5 & 24 \\ 13 & 10 & 27 \end{pmatrix}$

- Commutative Law is not established
- Distributive Law and Associative Law are established. : $\boldsymbol{A}(\boldsymbol{B} + \boldsymbol{C}) = \boldsymbol{AB} + \boldsymbol{AC}$ *and* $\boldsymbol{A}(\boldsymbol{BC}) = (\boldsymbol{AB})\boldsymbol{C}$

Addition, Subtraction, and Scalar Multiplication of Matrices

Find $A + B, A - \frac{1}{2}B, 5A$ when the matrices A, B are as follows.

$$A = \begin{bmatrix} 1 & -4 \\ -4 & 1 \end{bmatrix}, B = \begin{bmatrix} 1 & -8 \\ -8 & 2 \end{bmatrix}$$

In [1]:
```python
import numpy as np
A = np.array([[1,-4],[-4,1]])
B = np.array([[2,-8],[-8,2]])
print("A+B=", A+B)
print("A-(1/2)B=", A-(1/2)*B)
print("5A=", 5*A)
```

```
A+B= [[  3 -12]
 [-12   3]]
A-(1/2)B= [[0. 0.]
 [0. 0.]]
5A= [[  5 -20]
 [-20   5]]
```

Find AB and BA when matrices A, B are as follows, respectively.

$$AB = \begin{bmatrix} 2 & 3 & 1 \\ 4 & 5 & 1 \end{bmatrix} \begin{bmatrix} -1 & 2 \\ 4 & -2 \\ 3 & 6 \end{bmatrix}$$

In [2]:
```
import numpy as np
A = np.array([[2,3,1],[4,5,1]])
B = np.array([[-1,2],[4,-2], [3,6]])
print("AB=", np.dot(A,B))
print("BA=", np.dot(B,A))
```

```
AB= [[13  4]
 [19  4]]
BA= [[ 6  7  1]
 [ 0  2  2]
 [30 39  9]]
```

Unit 4.

Understanding Data Science: Matrix

Reverse Matrix

Unit Matrix/Identity Matrix

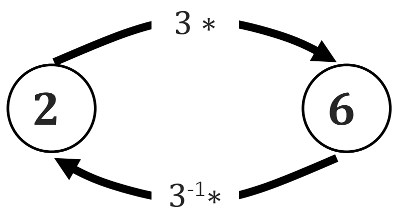
- All of the main diagonal components are square matrices with 1 and all of the remaining components are 0.
- $n \times n$ Unit Matrix ⇨ expressed as In

$$I = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \qquad I_3$$

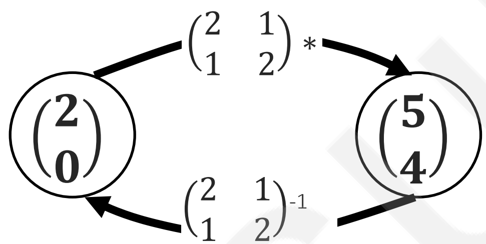
$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}$$

$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} a & b & c \\ d & e & f \\ g & h & i \end{bmatrix} = \begin{bmatrix} a & b & c \\ d & e & f \\ g & h & i \end{bmatrix}$$

Reverse Matrix is defined when the product of Matrix A and Reverse Matrix A^{-1} is Unit Matrix E.



$3^{-1} * 3 = 1$



$\begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix}^{-1} * \begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} = I$

(a) Principle of Reciprocity

(b) Principle of Reverse Matrix

Reverse Matrix A^{-1} of Square Matrix A

$$\boxed{A^{-1}A = AA^{-1} = I}$$

- For example,

  The Reverse Matrix of $\begin{pmatrix} 2 & 1 \\ 6 & 4 \end{pmatrix}$ is $\begin{pmatrix} 2 & -0.5 \\ -3 & 1 \end{pmatrix}$

The Reverse Matrix A^{-1} of the 2×2 Square Matrix $A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$ is the following.

$$A^{-1} = \frac{1}{ad - bc} \begin{bmatrix} d & -b \\ -c & a \end{bmatrix}$$

If $ad - bc = 0$, the Reverse Matrix of the given Matrix A does not exist.

Determinant

- The value defined only in a square matrix. Determines the reversibility of the matrix.
- Historically, the matrix was conceived with the question of "how to solve the system of linear equations." Arthur Cayley, in his research, saw that the solution of the system was different depending on the value of (ad-bc)(ad-bc), and thus coined the term 'determinant,' meaning "determining" the existence of a solution (the reversibility of the matrix). In addition to Cayley's research, William Rowan Hamilton thought, "Well, why don't we separate the coefficients and the variables of the system?" thus creating the matrix.

Determinant

The notation for the determinant is its abbreviation 'det' and the absolute value symbol (| · |).

For the Matrix $A = \begin{pmatrix} a & b \\ C & d \end{pmatrix}$

- $\det \begin{pmatrix} a & b \\ c & d \end{pmatrix} = ad - bc$
- $\begin{vmatrix} a & b \\ c & d \end{vmatrix} = ad - bc$
- $|A| = \left| \begin{pmatrix} a & b \\ C & d \end{pmatrix} \right| = ad - bc$ Etc.

- If the absolute value symbol is double=layered as seen in the latter two equations, it means to find the absolute value of the value after finding the determinant.
- The inner absolute value symbol means "calculate the matrix equation," and the outer absolute value symbol means "take the absolute value after the calculation."

$$\|A\| = \left| \begin{vmatrix} ab \\ C\ \mathrm{d} \end{vmatrix} \right| = |ad - bc| = \begin{cases} ad - bc & (ad - bc > 0) \\ -(ad - bc) & (ad - bc < 0) \end{cases}$$

Find the Reverse Matrix of A, B when the 2×2 Matrix A, B are as follows, respectively.

$$A = \begin{bmatrix} 1 & 3 \\ 5 & 7 \end{bmatrix}, B = \begin{bmatrix} 4 & -6 \\ -8 & 12 \end{bmatrix}$$

In [3]:
```
import numpy as np
A = np.array([[1,3], [5,7]])
B = np.array([[4,-6],[-8,12]])
detA = np.linalg.det(A)
Ainv = np.linalg.inv(A)
detB = np.linalg.det(B)
print("Determinant of A=", detA)
print("Inverse of A=", Ainv)
print("Determinant of B=", detB)
```

```
Determinant of A= -7.999999999999998
Inverse of A= [[-0.875  0.375]
 [ 0.625 -0.125]]
Determinant of B= 0.0
```

Unit 4.

Understanding Data Science: Matrix

Linear Transformation

Linear Transformation refers to a function that mathematically multiplies a vector by a matrix to create another vector. That is, it is a method of converting from one vector space to another vector space while maintaining the feature of the vector.

- Transformation : A function in two-dimensions that corresponds $P(x,y)$ to $P'(x', y')$.
- Linear Transformation : A transformation expressed by the linear expression for x, y without a constant term x', y'.

Transformation T: When x', y' in (x, y) → (x', y') is expressed as the following, it is called a Linear Transformation

$$\begin{cases} x' = ax + by \\ y' = cx + dy \end{cases} (but, \text{ a, b, c, d are constants})$$

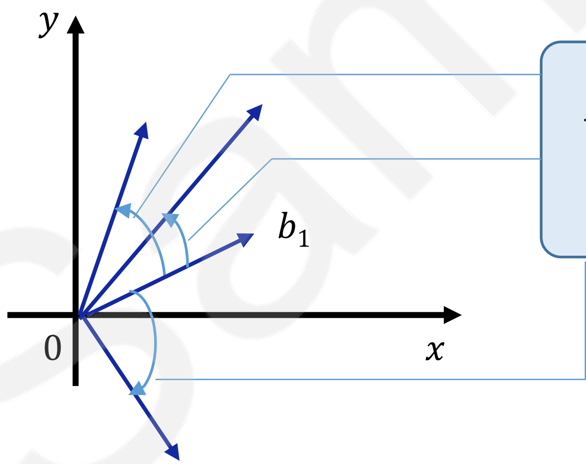
Since linear transformation is in the form of a linear system, it can be expressed in the form of a matrix.

- Transformation T: *When the constants $a, b, c,$ and d of x', y' in $(x, y) \rightarrow (x', y')$* is expressed as $\begin{cases} x' = ax + by \\ y' = cx + dy \end{cases}$

  Linear Transformation $\boldsymbol{T}$ can be expressed in the form of a matrix, as shown below.

$$\begin{bmatrix} x' \\ y' \end{bmatrix} = \begin{bmatrix} a & b \\ c & d \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix}$$

- Converting a vector in a vector space is called a Linear Transformation



When performing a Linear Transformation, the vector bi can be rotated, enlarged, or reduced using various matrices

Unit 4.

Understanding Data Science: Matrix

Eigenvalues and Eigenvectors

Characteristic Polynomials and Eigenvalues

- When A is an n × n matrix and I is an identity matrix,

$$Ax = \lambda x$$

In the linear system, as shown above, the necessary and sufficient condition for the existence of a solution for $x \neq 0$ is the determinant $|A - \lambda I| = 0$.

Since linear transformation is in the form of a linear system, it can be expressed in the form of a matrix.

- For the following $n \times n$ Matrix A = $\begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{11} \\ \vdots & \vdots & & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{bmatrix}$ $Det\ (A - \lambda E\ )$ = $\begin{bmatrix} a_{11} - \lambda & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} - \lambda & \cdots & a_{11} \\ \vdots & \vdots & & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} - \lambda \end{bmatrix}$ is called the characteristic polynomial of A.
- $p(\lambda) = |A - \lambda E| = 0$ is called the characteristic equation or determinantal equation.
- In this case, $p(\lambda)$ is the nth degree polynomial of λ. The root of the characteristic equation λ is called the eigenvalue,
- And a vector x is called the eigenvector of the eigenvalue λ.

Definition of Eigenvalues and Eigenvectors

- When a non-zero vector x goes through a nth order square matrix A for linear transformation, and the image of x is λx, λ is called the eigenvalue of A and x is called the eigenvector of λ. In other words, λ and x that satisfy the following relationship are called Eigenvalues and Eigenvectors, respectively.

$$Ax = \lambda x \ (but, x \neq 0)$$

In Matrix A = $\begin{bmatrix} 1 & 2 \\ 4 & 3 \end{bmatrix}$, if the Eigenvalues and Eigenvectors are

- x = $\begin{bmatrix} 1 \\ 2 \end{bmatrix}$:

$$Ax = \begin{bmatrix} 1 & 2 \\ 4 & 3 \end{bmatrix}\begin{bmatrix} 1 \\ 2 \end{bmatrix} = \begin{bmatrix} 5 \\ 10 \end{bmatrix} = 5\begin{bmatrix} 1 \\ 2 \end{bmatrix} = \lambda x$$

- Therefore, λ=5 becomes the eigenvalue of matrix A, and the vector $\begin{bmatrix} 1 \\ 2 \end{bmatrix}$ becomes the eigenvector of the matrix A for the eigenvalue 5.

Characteristic Equation (Determinantal Equation)

- The definition of the eigenvalue and characteristic equation is the following:

$$Ax = \lambda x \ (but, x \neq 0)$$

- If you move the right side of the equation to the left, you can make the following equation where E is a unit matrix.

$$(A - \lambda E)x = 0$$

- If $(A - \lambda E)$ of this equation has the Reverse Matrix $(A - \lambda E)^{-1}$, both sides can be multiplied by the Reverse Matrix as follows.

$$(A - \lambda E)^{-1}(A - \lambda E)x = (A - \lambda E)^{-1}0$$

- Therefore, $x = (A - \lambda E)^{-1}0 = 0$
- As a result, this equation must have a self-evident solution of $x = 0$. This contradicts the condition that the eigen vector x is initially assumed as $x \neq 0$.
- In order to eliminate this contradiction and to have an eigenvector, $(A - \lambda E)$ must not have a Reverse Matrix $(A - \lambda E)^{-1}$.
- After all, the conditional equation for the existence of an eigenvector is as follows, and this equation of λ is called the characteristic equation (or determinantal equation) of matrix A.

$$\det(A - \lambda E) = 0$$

Solving for Eigenvalues and Eigenvectors

- In order to obtain Eigenvalues and Eigenvectors, follow the procedure below.

1. Find det(A – λE).
2. The root of the obtained characteristic equation becomes the eigenvalue.
3. Obtain the eigenvectors by solving for each corresponding eigenvalue "(A – " λ"E)x=0."

Solve for the Characteristic Equation and the Eigenvalue of the given Matrix A.

$$\mathrm{A} = \begin{bmatrix} 2 & 3 \\ 3 & -6 \end{bmatrix}$$

- First, solve for the characteristic equation, then solve for the eigenvalue.

$$\mathrm{A} - \lambda\,\mathrm{E} = \begin{bmatrix} 2 & 3 \\ 3 & -6 \end{bmatrix} - \begin{bmatrix} \lambda & 0 \\ 0 & \lambda \end{bmatrix} = \begin{bmatrix} 2-\lambda & 3 \\ 3 & -6-\lambda \end{bmatrix}$$

$$\det(\mathrm{A} - \lambda\,\mathrm{E}) = (2-\lambda)(-6-\lambda) - (3)(3) = \lambda^2 + 4\lambda - 21$$

- The characteristic equation is $\lambda^2 + 4\lambda - 21$.
- If you solve $\lambda^2 + 4\lambda - 21$, it becomes ($\lambda$-3)($\lambda$+7)=0. Therefore, the eigenvalue of A is 3 and -7.

Solve for the Eigenvalue and the Corresponding Eigenvector of Matrix A.

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

- First, solve for the characteristic equation, then solve for the eigenvalue.

$$A - \lambda E = \begin{bmatrix} 3-\lambda & 2 \\ 3 & -2-\lambda \end{bmatrix} = 0$$

$$\det(A - \lambda E) = (3-\lambda)(-2-\lambda) - (3)(2) = \lambda^2 - \lambda - 12$$

- The characteristic equation is $\lambda^2 + \lambda - 12$.
- If you solve $\lambda^2 + \lambda - 12 = 0$, it becomes ($\lambda$-4)($\lambda$+3)=0. Therefore, the eigenvalue of matrix A is 4 and -3.

Solve for the Eigenvalue and the Corresponding Eigenvector of Matrix A.

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

- 1) In order to obtain the eigenvector corresponding to eigenvalue λ=4, solve for (A-4E)x=0. Then,

$$\begin{bmatrix} 3-4 & 2 \\ 3 & -2-4 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

$-x_1 + 2x_2 = 0$

$3x_1 - 6x_2 = 0$

Since $x_1 = 2x_2$, if x_2 is 1, it becomes $x_1 = 2$.

Therefore, the eigenvector corresponding to eigenvalue λ=4 is $\begin{bmatrix} 2 \\ 1 \end{bmatrix}$.

Solve for the Eigenvalue and the Corresponding Eigenvector of Matrix A.

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

- 2) In order to obtain the eigenvector corresponding to eigenvalue λ= -3, solve for (A-(-3)E)x=0. Then,

$$\begin{bmatrix} 3-(-3) & 2 \\ 3 & -2-(-3) \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

$$6x_1 + 2x_2 = 0$$

$$3x_1 - x_2 = 0$$

$$6x_1 = -2x_2$$

$$x_2 = -3x_1$$

If x_1 is 1, it becomes $x_2 = -3$.

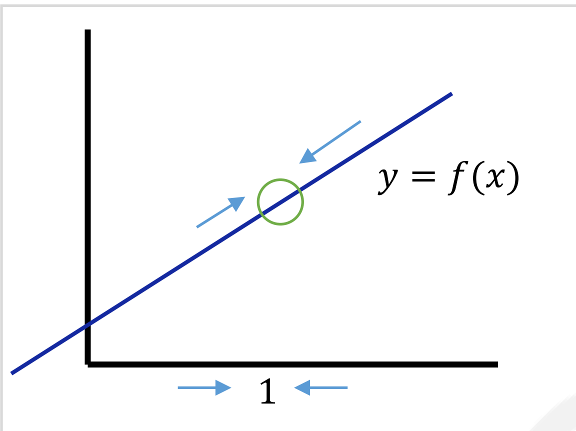
Therefore, the eigenvector corresponding to eigenvalue λ=-3 is $\begin{bmatrix} 1 \\ -3 \end{bmatrix}$.

Unit 5.

Understanding Deep Learning: Derivatives

Limits

- A Limit means that a sequence or function value is infinitely close to a certain value.
  - Take a look at the equation $f(x) = \frac{x^2-1}{x-1}$ and the figure below.



- $f(x)$ cannot define the value of y because the denominator becomes 0 when x=1.
- Instead, when $x \neq 1$, the value of the function f(x) may be determined.
- So, if we make the value of the variable x as close as possible to 1, but never let x=1, we can find the value of f(x).
- When we make the value of x as close as possible to any value a, we refer to the shape, in which the value of the function f(x) gets as close as possible to any value alpha (a), and express it as "converging."

$\lim_{x \to a} f(x) = \alpha$ **$f(x) \to \alpha (x \to a)$ α is the limit**

$$\lim_{x \to 1} \frac{x^2 - 1}{x - 1} = \lim_{x \to 1}(x + 1) = 2$$

Unit 5.

Understanding Deep Learning: Derivatives

Basics of Derivatives

Rate of Change

- When you look at the dashboard at the moment the car moves, it tell you the speed of the car at every moment.
- Since speed is the "rate of change in distance over time," the speed of a car means the "instantaneous rate of change" of the distance traveled by the car at the specific moment..
- As such, the rate of change at some point is called "derivative."
- For the function y=f(x), the amounts of change in x and y are denoted by Δx, Δy, respectively, and are called increments.
- The ration of the two increments is called the average rate of change and is defined as follows.
- For the function y=f(x), the average rate of change when the value of x changes from a to a+, Δx is expressed as $\frac{\Delta y}{\Delta x}$ and defined as follows.

$$\frac{\Delta y}{\Delta x} = \frac{f(a+\Delta x) - f(a)}{(a+\Delta x) - a} = \frac{f(a+\Delta x) - f(a)}{\Delta x}$$

Differential Coefficient and Derivatives

- The instantaneous rate of change can naturally represent the "moment" by making Δx close to 0 at the average rate of change $\frac{\Delta y}{\Delta x}$.
- In addition, with Limits, which we previously learned, the instantaneous rate of change can be mathematically defined as follows.
- For the function y=f(x), the instantaneous rate of change in x=a is expressed as $f'(a)$ and defined as follows.

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

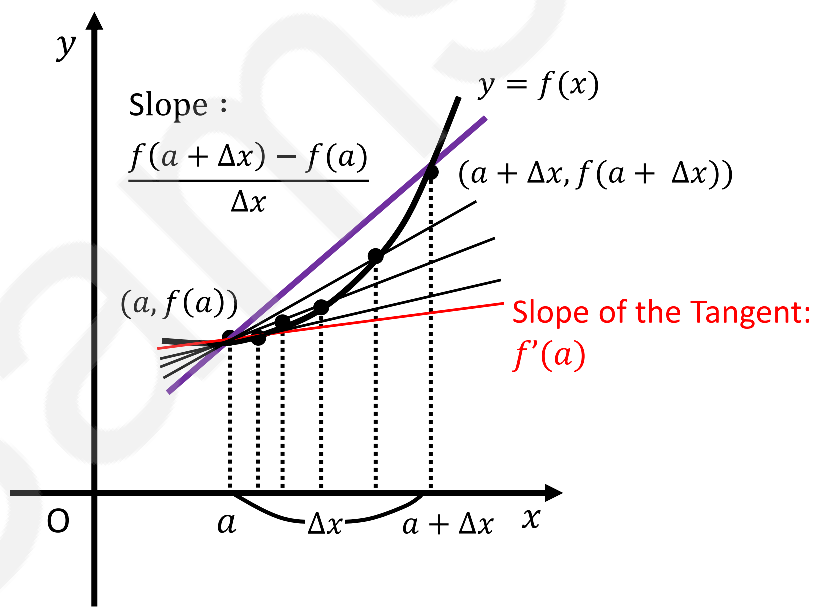
- The rate of instantaneous change for the function is called a differential coefficient.

Differential Coefficient and Derivatives

- Another definition of Differential Coefficients: $a+\Delta x=x$ (When $\Delta x \to 0$, $x \to a$)

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

- Geometric meaning of Differential Coefficients: The slope of the tangent of the function $y=f(x)$ in $x=a$



Differential Coefficient and Derivatives

- Derivatives: A function representing the "slope of the tangent" at all points above $y=f(x)$
- The derivative of the differentiable function y=f(x) is defined as follows.

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

Unit 5.

Understanding Deep Learning: Derivatives

Differential Method

We learned that derivatives represent the degree to which a function changes (slope) at a certain point in time, and that we also use Limits to find derivatives.

- However, since it is cumbersome to calculate in this way, the following formula is actually used to calculate differentials.
- For the two differentiable functions f(x), g(x) the following holds.

(1) $\{cf(x)\}' = cf'(x)$ (but c is a constant)

(2) $\{f(x) \pm g(x)\}' = f'(x) \pm g'(x)$

(3) $\{f(x)g(x)\}' = f'(x)g(x) + f(x)g'(x)$ (Product Rule)

(4) $\left\{\frac{f(x)}{g(x)}\right\}' = \frac{f'(x)g(x)-f(x)g'(x)}{\{g(x)\}^2}$ (but $g(x) \neq 0$)(Quotient Rule)

Differential Coefficients and Derivatives

- f(x) = x^n and the derivative of the constant function.

(1) $\frac{d}{dx}(x^n) = nx^{n-1}$ (but n is a natural number)

(2) $\frac{d}{dx}(c) = 0$ (but c is a constant)

Find the Derivative of the Constant Function

(a) $f(x) = x^{2020} + 3x^3 - 1$

In [1]:
```
import sympy as sp
# (a) f(x) = x**2020 + 3*x**3 - 1
x  = sp.Symbol('x') # Use the letter x as a symbol
fx = x**2020+3*x*x*x-1
```

In [2]:
```
fx
```

Out[2]: $x^{2020} + 3x^3 - 1$

In [3]:
```
sp.diff(fx,x)
```

Out[3]: $2020x^{2019} + 9x^2$

Find the Derivative of the Constant Function

(b) $g(x) = \frac{x}{x+1}$

$$g'(x) = \frac{\frac{d}{dx}(x)\cdot(x+1) - x\cdot\frac{d}{dx}(x+1)}{(x+1)^2} = \frac{1(x+1)-x}{(x+1^2)} = \frac{1}{(x+1)^2}$$

In [4]:
```
# (b) g(x) = x/(x+1)
x  = sp.Symbol('x')
gx = x/(x+1)
print("g'(x)=",sp.simplify(sp.diff(gx,x)))
```

```
g'(x)= (x + 1)**(-2)
```

Unit 5.

Understanding Deep Learning: Derivatives

Difference Between Logarithmic and Exponential

- Natural Log: Log with the irrational number e as its base. log_ex $\rightarrow$ log$x$ (base e omitted)

(1) $\frac{d}{dx}(e^x) = e^x$

(2) $\frac{d}{dx}(a^x) = a^x \log a$

(3) $\frac{d}{dx}(\log x) = \frac{1}{x}$

(4) $\frac{d}{dx}(\log_a x) = \frac{1}{x \log a}$

Find the Derivative

(a) $f(x) = \frac{e^x}{x}$ $(but\ \mathrm{x} \neq 0)$

$$f'(x) = \frac{\frac{d}{dx}(e^x)x - e^x\frac{d}{dx}(x)}{x^2} = \frac{(e^x)x-(e^x)}{x^2} = \frac{e^x(x-1)}{x^2}$$

In [5]:
```
# (a) f(x) = exp(x)/x
x  = sp.Symbol('x')
fx = sp.exp(x)/x
sp.diff(fx,x)
```

Out[5]: $\frac{e^x}{x} - \frac{e^x}{x^2}$

Find the Derivative

(b) $g(x) = 3\log_2 x - x\log x$

$$g'(x) = 3\frac{d}{dx}(log_2 x) - \left\{\frac{d}{dx}(x)\, log\, x + x\frac{d}{dx}(log\, x)\right\}$$

$$= 3\frac{1}{x\, log\, 2} - (1 \cdot log\, x + x \cdot \frac{1}{x})$$

$$= \frac{3}{x\, log\, 2} - log\, x \text{ -1}$$

```
In [6]: # (b) g(x) = 3log2(x)-xlogx
        x  = sp.Symbol('x')
        gx = 3*sp.log(x,2)-x*sp.log(x)
        print("g'(x)=",sp.simplify(sp.diff(gx,x)))

        g'(x)= -log(x) - 1 + 3/(x*log(2))
```

Unit 5.

Understanding Deep Learning: Derivatives

Derivatives of Composite Functions

Derivatives of Composite Functions consisting of quadratic functions f(x) and g(x).

$$f(x)=10+x^2$$

$$g(x)=3+x$$

$$f(g(x))=10+g(x)^2=10+(3+x)^2$$

- When the above composite function f(g(x)) is differentiated by x, it is replaced with a variable as follows.

$$y=f(u)$$

$$u=g(x)$$

- In this way, it can be differentiated step by step as follows.

$$\frac{dy}{dx}=\frac{dy}{d\mu}\cdot\frac{d\mu}{dx}$$

Derivatives of Composite Functions consisting of quadratic functions f(x) and g(x).

- In other words, you can differentiate y by μ and multiply μ by the differentiated x. In fact, the differential is as follows.

$$\frac{dy}{d\mu} = \frac{d}{d\mu}\mathrm{f}(\mu)$$
$$= \frac{d}{d\mu}(10 + \mu^2) = 2\mu$$

$$\frac{d\mu}{dx} = \frac{d}{dx}\, g(x)$$
$$= \frac{d}{dx}(3 + x) = 1$$

- Each result came out. Now multiply these two results.
- Then returning μ to g(x) yields the final differential result.

$$\frac{dy}{dx} = \frac{dy}{d\mu} \cdot \frac{d\mu}{dx}$$
$$=2\mu \cdot 1 = 2\mathrm{g}(\mathrm{x}) = 2(3+\mathrm{x})$$

Derivatives of Composite Functions

- In summary, it is as follows.
- For two differentiable functions $y = f(u), u = g(x)$, the derivative of the composite function $y = f\big(g(x)\big)$ is as follows.

$$\{f\big(g(x)\big)\}' = f'\big(g(x)\big)g'(x)$$

Find the Differential of the Function

(a) $\mathrm{f}(x) = (x^2 + 2)^3$

If μ=$(x^2 + 2)^2$, then f(μ)=μ^3,
$u' = 2x,\ and\ \ f'(\mu) = 3\,u^2.$

- Therefore, the Derivative of Composite Functions is as follows.

$$f'(x) = \{f\,(\mu)\}' = f'(\mu) \cdot u' = 3\,u^2 \cdot (2\,x) = 6(x^2 + 2)^2$$

In [1]:
```
import sympy as sp

# (a) f(x) = (x**2+2)**3
x  = sp.Symbol('x')
fx = (x**2+2)**3
sp.diff(fx,x)
```

Out[1]: $6x\left(x^2 + 2\right)^2$

Find the Differential of the Function

(b) $g(x) = \frac{1}{(x^4+1)^2}$

If $\mu = x^4 + 1$, then $g(\mu) = \frac{1}{\mu^2}$, $u' = 4x^3$, $and\ g'(\mu) = -\frac{2}{\mu^3}$

- Therefore, the Derivative of Composite Functions is as follows.

$$g'(x) = \{g(\mu)\}' = g'(\mu) \cdot u' = -\frac{2}{u^3} \cdot 4x^3 = -\frac{8x^3}{(x^4+1)^3}$$

In [2]:
```
# (b) g(x) = 1/(x**4+1)**2
x  = sp.Symbol('x')
gx = 1/(x**4+1)**2
sp.diff(gx,x)
```

Out[2]:

$$-\frac{8x^3}{(x^4+1)^3}$$

Unit 5.

Understanding Deep Learning: Derivatives

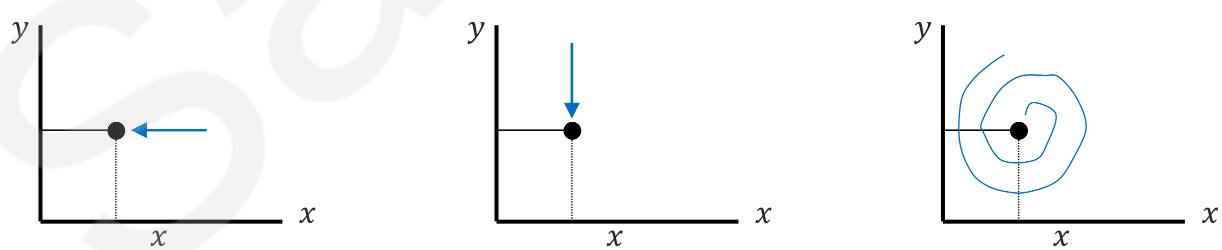
High Order Derivatives and Partial Derivatives

- So far, we have dealt with derivatives with only one variable, that is functions with only x, and these derivatives are called ordinary derivatives.

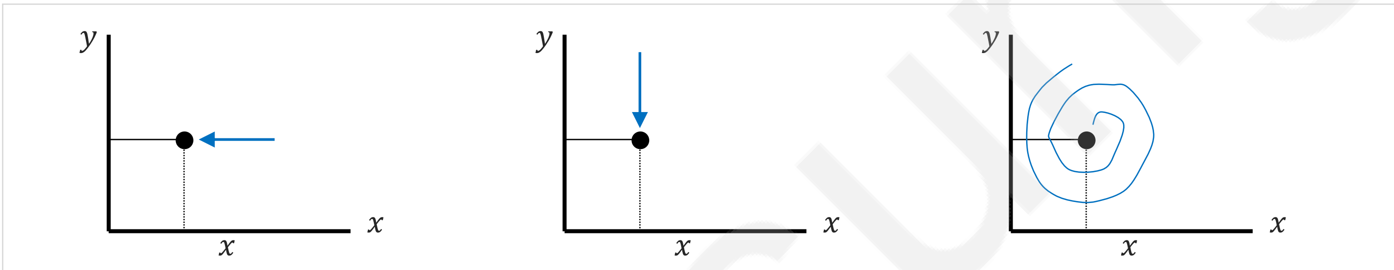
  - Let's learn how to differentiate a multi-variable function with two or more variables, such as $z=f'(x,y)=3x^2 + 2xy + 2y^2$.
  - Applying the method which we learned earlier, when the variables x, y change by Δx, Δy, solve for the rate of change of z, or Δz.

$$
\begin{aligned}
\Delta z &= f(x+\Delta x, y+\Delta y) - f(x,y) \\
&= 3(x+\Delta x)^2 + 2(x+\Delta x)(y+\Delta y) + 2(y+\Delta y)^2 - (3x^2 + 2xy + 2y^2) \\
&= (6x+2y)\,\Delta x + (2x+4y)\,\Delta y + 3\Delta x^2 + 2\,\Delta x\,\Delta y + 2\Delta y^2
\end{aligned}
$$

  - Then we can find the derivative of the function f(x,y) by calculating the limit value, such as Δx, Δy →0.
  - In order to obtain the limit value of this function, the point (x+ Δx, y+ Δy) must be approached as close as possible to point (x,y), and there are various approaches as shown below.



High Order Derivatives and Partial Derivatives



- In the case of the figure on the right, the limit value is difficult to calculate because x and y move together.
- Let's fix y as a constant as shown in the figure on the left.
- Changing x when $\Delta y=0$ can be considered as $\Delta x \to 0$, and this approach is called the partial derivative of x.

$$\frac{\partial f(x,y)}{\partial x} = \lim_{\Delta x \to 0} \frac{\Delta z}{\Delta x} = \lim_{\Delta x \to 0} 6x + 2y + 3\Delta x = 6x + 2y$$

- This time, fix x as the constant, then solve for the partial derivative of y. It will look as follows.

$$\frac{\partial f(x,y)}{\partial y} = 2x + 4y$$

Unit 5.

Understanding Deep Learning: Derivatives

Derivative of the Sigmoid Function

Differentiate the Sigmoid Function

- When function $\sigma(x)$ is defined as sigmoid function $\frac{1}{1+e^{-x}}$, the value of the differentiated $\frac{d\sigma(x)}{dx}$ is as follows.

$$\frac{d\sigma(x)}{dx} = \sigma(x) \cdot \left(1 - \sigma(x)\right)$$

How to Induce the Derivative of the Sigmoid Function

When $\sigma(x) = \dfrac{1}{1+e^{-x}}$

$$\frac{d}{dx}\sigma(x) = \frac{d}{dx}\left[\frac{1}{1+e^{-x}}\right]$$

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

$$= -(1+e^{-x})^{-2}(-e^{-x})$$

$$= \frac{e^{-x}}{(1+e^{-x})^2}$$

$$= \frac{1}{1+e^{-x}} \cdot \frac{e^{-x}}{1+e^{-x}}$$

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

$$= \frac{1}{1+e^{-x}} \cdot \left(1 - \frac{1}{1+e^{-x}}\right)$$

$$= \sigma(x) \cdot (1 - \sigma(x))$$

- Apply the Chain Rule

  1. When $f(x) = x^2(a = natural\ number)$, the differential value is ax^{a-1}

  2. The differential value of e^{-x} is $-e^{-x}$

- Proof : $\frac{d}{dx}[e^{-x}]$

$$= e^{-x} \cdot \frac{d}{dx}[-x]$$

$$= \left(-\frac{d}{dx}[x]\right)e^{-x}$$

$$= -1e^{-x}$$

$$= -e^{-x}$$

SAMSUNG

Together for Tomorrow!

Enabling People

Education for Future Generations