Math for AI
(Linear Algebra and Differential Calculus)

成均館大學校
SUNG KYUN KWAN UNIVERSITY

Vector and Matrix :

to process various input and output at once

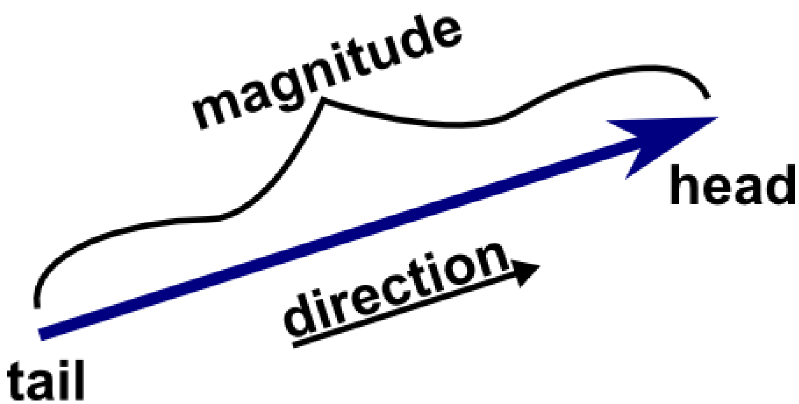
3.1 Vector

❖ 1. Scalar, Vector, Matrix, Tensor

- **Scalar :** Any real number, or any quantity that can be measured using a single real number.

  ex) length, width, temperature

- **Vector :** A vector is an object that has both a magnitude and a direction.



ex) 2D coordinate plane : $\vec{a}$ = [2,1], $\vec{b}$ = [1,2]

https://mathinsight.org/vector_introduction

3.1 Vector

❖ 1. Scala, Vector, Matrix, Tensor

- **Vector :** column vector, row vector
  - **column vector :** n by 1 matrix consisting of a single column of n elements

$$\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} \in \mathbb{R}^n = \mathbb{R}^{n \times 1}$$

  - **row vector :** 1 by n matrix consisting of a single row of n elements

$$\mathbf{x}^T = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}^T = [x_1 \quad x_2 \quad \cdots \quad x_n] \in \mathbb{R}^{1 \times n}$$

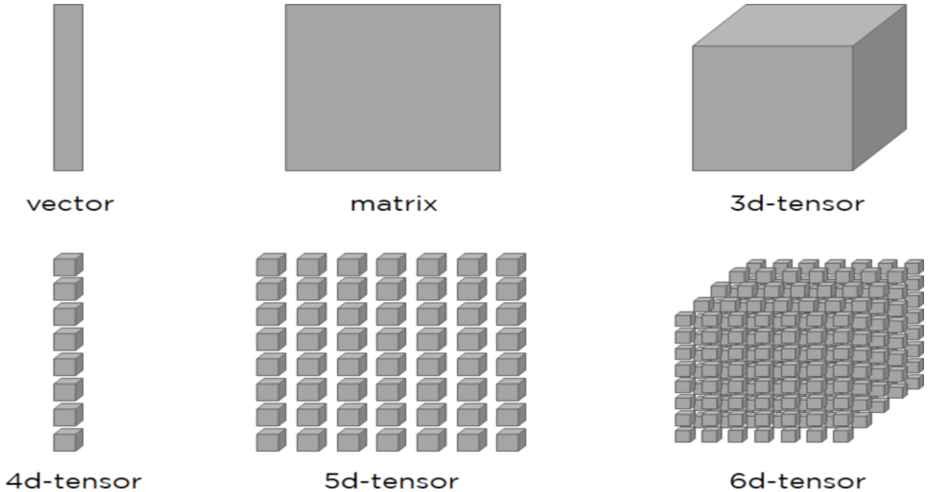
3.1 Vector

❖ 1. Scala, Vector, Matrix, Tensor

- **Matrix :** a rectangular array of numbers, symbols, or expression, arranged in rows and columns.

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

- **Tensor** : a type of data structure used in linear algebra. you can calculate arithmetic operations with tensors. → **multidimensional data array**



https://rekt77.tistory.com/102
https://machinelearningmastery.com/introduction-to-tensors-for-machine-learning/

3.1 Vector

❖ 1. Scala, Vector, Matrix, Tensor

Rank	Math entity	Python example
0	Scalar (magnitude only)	`s = 483`
1	Vector (magnitude and direction)	`v = [1.1, 2.2, 3.3]`
2	Matrix (table of numbers)	`m = [[1, 2, 3], [4, 5, 6], [7, 8, 9]]`
3	3-Tensor (cube of numbers)	`t = [[[2], [4], [6]], [[8], [10], [12]], [[14], [16], [18]]]`
n	n-Tensor (you get the idea)	`....`

https://tensorflowkorea.gitbooks.io/tensorflow-kr/content/g3doc/resources/dims_types.html

3.1 Vector

❖ 2. Addition, Subtraction, Magnitude of Vector

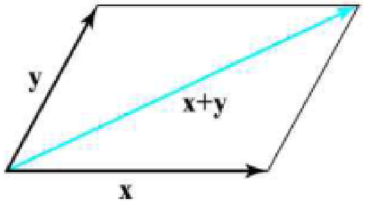
- **Addition and Subtraction of Vector:** element-wise addition of vector and element-wise subtraction of vector

$C = A + B$: Element-wise addition, i.e., $C_{ij} = A_{ij} + B_{ij}$

- A, B, C should have the same size, i.e., $A, B, C \in \mathbb{R}^{m \times n}$

ex) $\vec{a}$ **= [3,2],** $\vec{b}$ **= [1,4]**

$\vec{a} + \vec{b}$ **= [3+1, 2+4] = [4,6],** $\vec{a} - \vec{b}$ **= [3-1, 2-4] = [2,-2]**

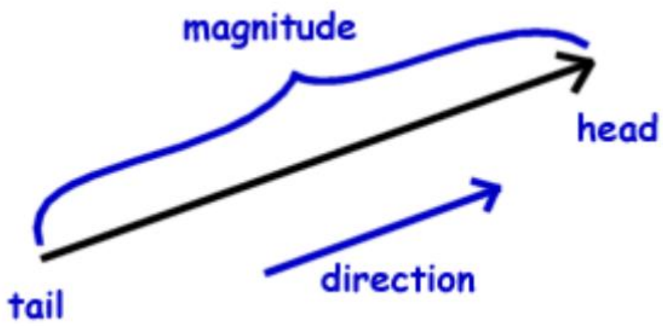


addition of vector ($\vec{x} + \vec{y}$) is called **parallelogram rule** ← $\vec{x}, \vec{y}$ form the sides of a parallelogram and $\vec{x} + \vec{y}$ is one of the diagonals

https://en.wikipedia.org/wiki/Euclidean_vector

❖ 2. Addition, Subtraction, Magnitude of Vector

- **Magnitude (or Length or Norm) of Vector :** magnitude of a vector is represented by the length of the arrow



- **Norm :** compute the length of the vector **a**

$$\|\mathbf{a}\| = \sqrt{a_1^2 + a_2^2 + a_3^2}$$

- ex) when $\vec{a}$ is [3,2], $\|\vec{a}\|$ =$\sqrt{3^2 + \mathbf{2^2}}$ = $\sqrt{13}$

https://www.ducksters.com/science/physics/scalars_and_vectors.php
https://www.physicsclassroom.com/class/vectors/Lesson-1/Vectors-and-Direction

❖ 3. Dot Product of Vectors

- **Dot product of two vector $\vec{a}$ and $\vec{b}$:** The dot product can be defined **algebraically or geometrically**. These two definitions is equal.
  Dot product is denoted by '·'

- **Algebraic definition :**
  The dot product of two vectors a = $[a_1, a_2, \ldots, a_n]$ and b = $[b_1, b_2, \ldots, b_n]$ is defined as $\vec{a} \cdot \vec{b} = a_1b_1 + a_2b_2 + \cdots + a_1b_n$.

  ex) when $\vec{a}$ is (1,3,-2) and $\vec{b}$ is (4,2,1),
  (1,3,-2) · (4,2,1) = 1*4 + 3*2 + (-2)*1 = 8

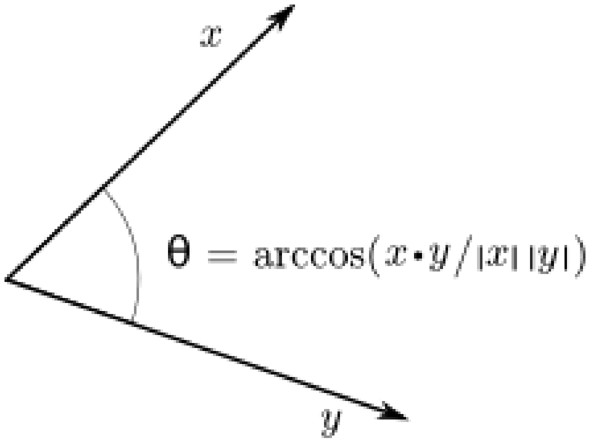
https://en.wikipedia.org/wiki/Dot_product

❖ 3. Dot Product of Vectors

- **Geometric definition :**
  based on the notions of magnitude ($||\vec{a}||$) and angle($\cos\theta$)

$$\mathbf{a}\cdot\mathbf{b} = \begin{cases} \|\mathbf{a}\|\,\|\mathbf{b}\|\cos\measuredangle(\mathbf{a},\mathbf{b}) & \mathbf{a}\neq\mathbf{0}\wedge\mathbf{b}\neq\mathbf{0} \\ 0 & \mathbf{a}=\mathbf{0}\vee\mathbf{b}=\mathbf{0} \end{cases}$$

- $||\vec{a}||$ **:** the magnitude of a vector $\vec{a}$.
- $||\vec{b}||$ **:** the magnitude of a vector $\vec{b}$.
- $\cos\theta$ **:** the angle of $\vec{a}$ and $\vec{b}$

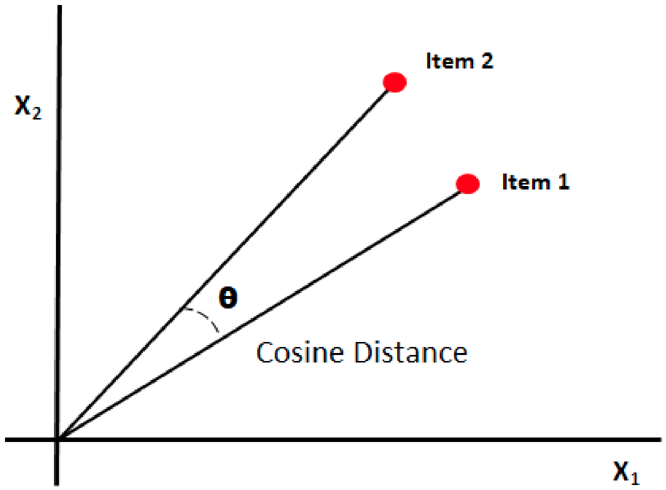


https://en.wikipedia.org/wiki/Dot_product

❖ 4. Cosine Similarity

- Cosine Similarity : measure of similarity between two non-zero vectors by measuring the cosine of the angle between them.

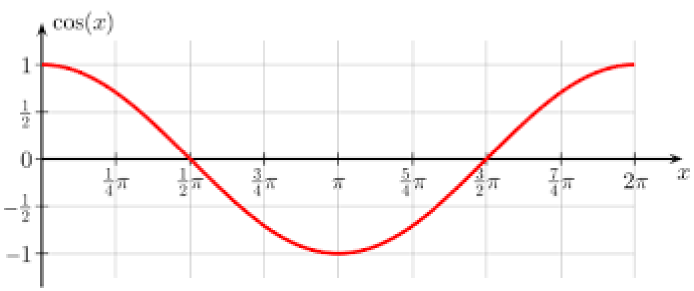
$$\mathbf{A} \cdot \mathbf{B} = \|\mathbf{A}\| \, \|\mathbf{B}\| \cos\theta \quad \rightarrow \quad \cos(\theta) = \frac{\mathbf{A} \cdot \mathbf{B}}{\|\mathbf{A}\| \|\mathbf{B}\|} = \frac{\sum_{i=1}^{n} A_i B_i}{\sqrt{\sum_{i=1}^{n} A_i^2} \sqrt{\sum_{i=1}^{n} B_i^2}}$$

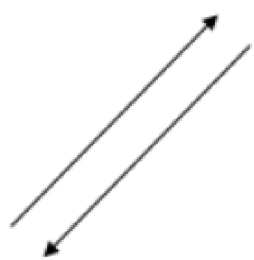


https://en.wikipedia.org/wiki/Cosine_similarity

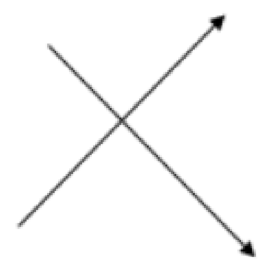
❖ 4. Cosine Similarity

- **Cosine similarity's properties:**
  - **measure of similarity between two vectors that measures the cosine of the angle between them.**
  - cosine similarity is neatly bounded in [0, 1]
  - **cosine similarity of -1 :** two vectors diametrically opposed
    cosine similarity of 0 : two vectors oriented at 90° relative to each other
    cosine similarity of 1 : two vectors with the same orientation

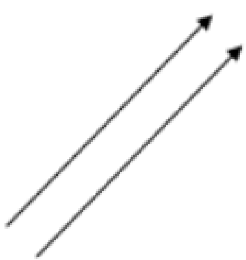




Cosine similarity : -1



Cosine similarity : 0



Cosine similarity : 1

https://wikidocs.net/24603

❖ 1. Matrix Addition and Product

- **Matrix addition:**

$C = A + B$: Element-wise addition, i.e., $C_{ij} = A_{ij} + B_{ij}$

- A, B, C should have the same size, i.e., $A, B, C \in \mathbb{R}^{m \times n}$

예) If we have

$$\mathbf{A} = \begin{bmatrix} 1 & 2 \\ 3 & 4 \\ 5 & 6 \end{bmatrix} \quad \text{and} \quad \mathbf{B} = \begin{bmatrix} 7 & 10 \\ 8 & 11 \\ 9 & 12 \end{bmatrix}$$

then we can calculate **C** = **A** + **B** by

$$\mathbf{C} = \mathbf{A} + \mathbf{B} = \begin{bmatrix} 1 & 2 \\ 3 & 4 \\ 5 & 6 \end{bmatrix} + \begin{bmatrix} 7 & 10 \\ 8 & 11 \\ 9 & 12 \end{bmatrix} = \begin{bmatrix} 8 & 12 \\ 11 & 15 \\ 14 & 18 \end{bmatrix}$$

❖ 1. Matrix Addition and Product

- **Matrix product :** If A is an m × n matrix and B is an n × p matrix, the matrix product C = AB is defined to be the m × p matrix

$$c_{ij} = a_{i1}b_{1j} + a_{i2}b_{2j} + \cdots + a_{in}b_{nj} = \sum_{k=1}^{n} a_{ik}b_{kj},$$

$$\mathbf{C} = \begin{pmatrix} a_{11}b_{11} + \cdots + a_{1n}b_{n1} & a_{11}b_{12} + \cdots + a_{1n}b_{n2} & \cdots & a_{11}b_{1p} + \cdots + a_{1n}b_{np} \\ a_{21}b_{11} + \cdots + a_{2n}b_{n1} & a_{21}b_{12} + \cdots + a_{2n}b_{n2} & \cdots & a_{21}b_{1p} + \cdots + a_{2n}b_{np} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1}b_{11} + \cdots + a_{mn}b_{n1} & a_{m1}b_{12} + \cdots + a_{mn}b_{n2} & \cdots & a_{m1}b_{1p} + \cdots + a_{mn}b_{np} \end{pmatrix}$$

ex)

If we have $\mathbf{A} = \begin{bmatrix} 1 & 4 & 7 \\ 2 & 5 & 8 \\ 3 & 6 & 9 \end{bmatrix}$ and $\mathbf{B} = \begin{bmatrix} 1 & 4 \\ 2 & 5 \\ 3 & 6 \end{bmatrix}$

then $\mathbf{AB} = \begin{bmatrix} 1 & 4 & 7 \\ 2 & 5 & 8 \\ 3 & 6 & 9 \end{bmatrix} \times \begin{bmatrix} 1 & 4 \\ 2 & 5 \\ 3 & 6 \end{bmatrix} = \begin{bmatrix} c_{11} & c_{12} \\ c_{21} & c_{22} \\ c_{31} & c_{32} \end{bmatrix} = \begin{bmatrix} 30 & 66 \\ 36 & 81 \\ 42 & 96 \end{bmatrix}$

https://ko.wikipedia.org/wiki/행렬_곱셈

❖ 2. Transpose of Matrix and Invertible Matrix

- **Transpose of matrix :** operator which flips a matrix over its diagonal → switches the row and column indices of the matrix

ex) $A = \begin{bmatrix} 1 & 6 \\ 3 & 4 \\ 5 & 2 \end{bmatrix}$ $A^T = \begin{bmatrix} 1 & 3 & 5 \\ 6 & 4 & 2 \end{bmatrix}$

- If A is equal to A^T, A is called a **symmetric matrix**

https://en.wikipedia.org/wiki/Transpose
https://en.wikipedia.org/wiki/Orthogonal_matrix

❖ 2. Transpose of Matrix and Invertible Matrix

- **Invertible matrix :** n-by-n square matrix A, if there exists an n-by-n square matrix B such that $AB = BA = I_n$.
  The matrix B is called the inverse of A

Example :

$M = \begin{bmatrix} 4 & 3 \\ 3 & 2 \end{bmatrix}$ and it's inverse is $M^{-1} = \begin{bmatrix} -2 & 3 \\ 3 & -4 \end{bmatrix}$ since

$MM^{-1} = \begin{bmatrix} 4 & 3 \\ 3 & 2 \end{bmatrix} \begin{bmatrix} -2 & 3 \\ 3 & -4 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$,

$M^{-1}M = \begin{bmatrix} -2 & 3 \\ 3 & -4 \end{bmatrix} \begin{bmatrix} 4 & 3 \\ 3 & 2 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$.

Therefore, $\begin{bmatrix} 4 & 3 \\ 3 & 2 \end{bmatrix}$ and $\begin{bmatrix} -2 & 3 \\ 3 & -4 \end{bmatrix}$ are inverses of each other

https://en.wikipedia.org/wiki/Invertible_matrix
https://www.slideshare.net/zakilivebuzz/presentation-on-inverse-matrix

❖ 3. Transformation Matrix

- **linear transformation :** mapping $V \rightarrow W$ between two modules (for example, two vector spaces)

  linear transformation maps linear subspaces onto linear subspaces

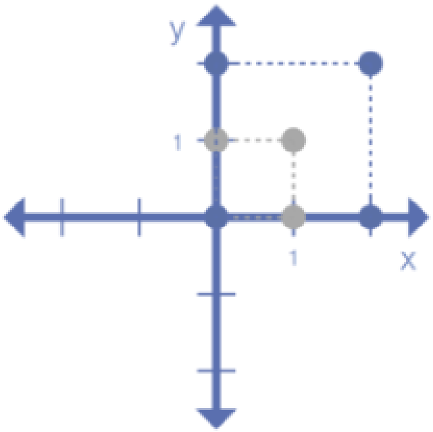
- **Invertible matrix :** linear transformations can be represented by matrices.

  If T is a linear transformation mapping $\mathbb{R}^n$ to $\mathbb{R}^m$ and $\vec{x}$ is a column vector with n entries, then
  $T(\vec{x}) = A\vec{x}$
  for some $m \times n$ matrix $\boldsymbol{A}$, called **the transformation of T**

❖ 3. Transformation Matrix

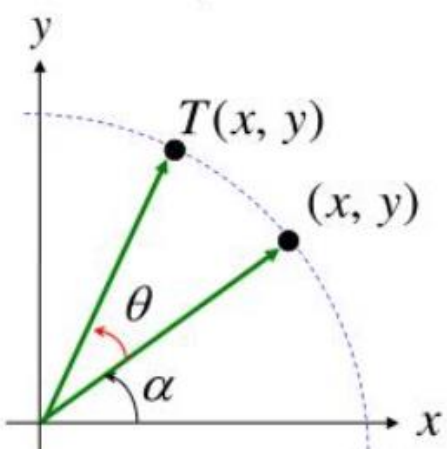
- Example 1: Transformation matrix for scale

$$\begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} 2x \\ 2y \end{bmatrix}$$



- Example 2: Transformation matrix for rotation by an angle θ clockwise about the origin the functional form

$$\begin{bmatrix} x' \\ y' \end{bmatrix} = \begin{bmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix}$$

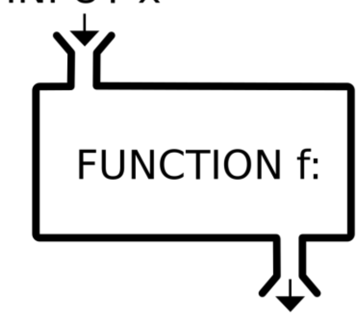


https://neutrium.net/mathematics/basics-of-affine-transformation/
https://en.wikipedia.org/wiki/Transformation_matrix

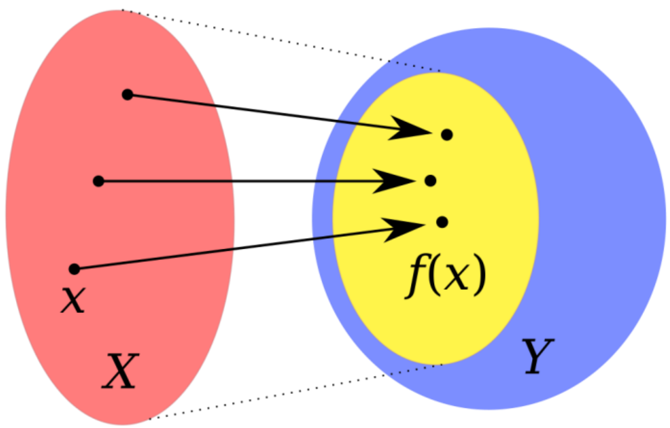
Function and Differential Calculus:

How to converge for the optimization

❖ 1. Function



- **Function's definition** : a relation between sets that associates to every element of the first set exactly one element of the second set.

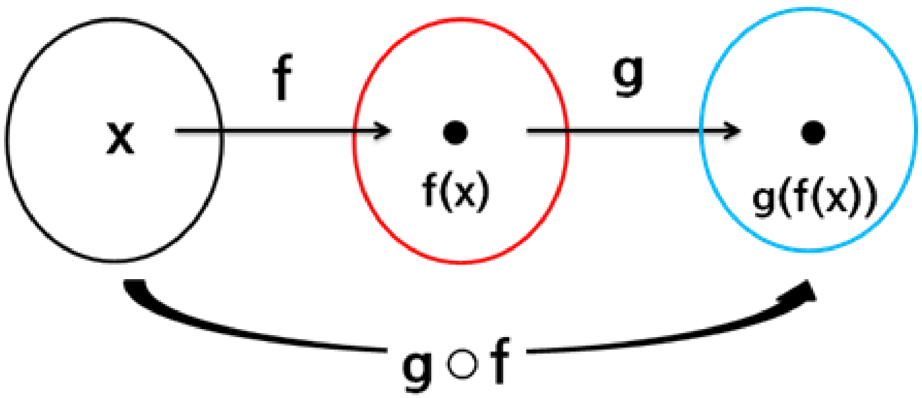


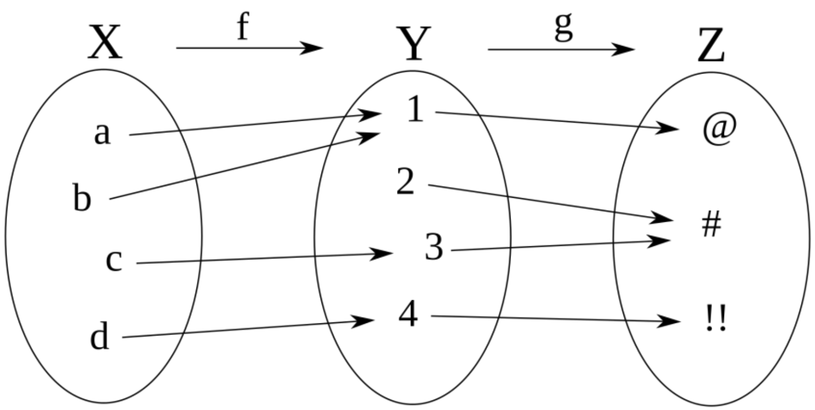
$f : X \to Y$

https://en.wikipedia.org/wiki/Function_(mathematics)

❖ 2. Composition of Function

- **Definition of Composite Function :** operation that takes two functions f and g and produces a new function h such that h(x) = g(f(x))





$g \circ f$, the **composition** of f and g.
For example, $(g \circ f)(c) = \#$.

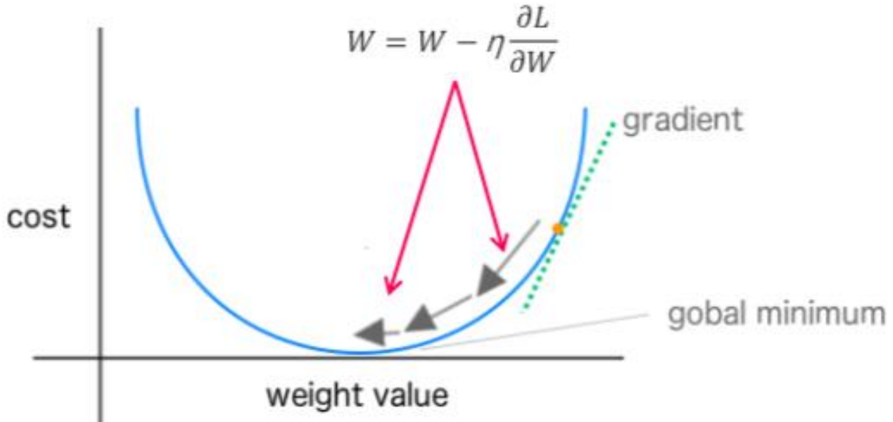
- **Properties of Function composition :**
  - $f \circ (g \circ h) = (f \circ g) \circ h$
  - $g \circ f \neq f \circ g$

https://en.wikipedia.org/wiki/Function_composition

❖ 3. Loss Function

- **loss function** : a function that maps an event or values of one or more variables onto a real number intuitively representing some "cost" associated with the event.
- **optimization problem** → seeking to minimize a loss function.

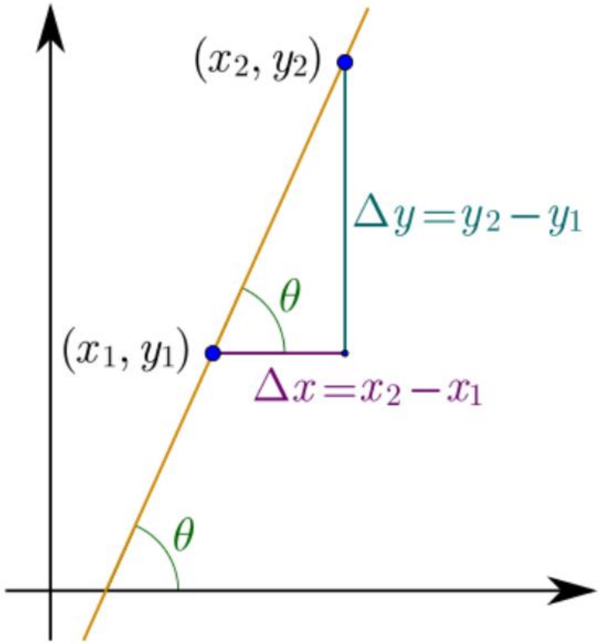
- **loss functions and optimization**
  - step 1: loss function's setting.
  - step 2: loss function's partial differential calculation
  - step 3: find global minimum of the loss function
    using gradient descent



https://en.wikipedia.org/wiki/Loss_function

❖ 1. Derivative

- The **derivative** of a function of a real variable measures **the sensitivity to change (gradient)** of the function value (output value) with respect to a change in its argument (input value)



y = f(x)

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

https://en.wikipedia.org/wiki/Derivative

➢ Derivatives

1. $\frac{d}{dx}(c) = 0$
2. $\frac{d}{dx}[cf(x)] = cf'(x)$
3. $\frac{d}{dx}[f(x)+g(x)] = f'(x)+g'(x)$
4. $\frac{d}{dx}[f(x)-g(x)] = f'(x)-g'(x)$
5. $\frac{d}{dx}[f(x)g(x)] = f(x)g'(x)+g(x)f'(x)$
6. $\frac{d}{dx}[\frac{f(x)}{g(x)}] = \frac{g(x)f'(x)-f(x)g'(x)}{[g(x)]^2}$
7. $\frac{d}{dx}f(g(x)) = f'(g(x))g'(x)$
8. $\frac{d}{dx}(x^n) = nx^{n-1}$

2. Derivative, Chain Rule

❖ 2. Chain Rule

- **Definition :** formula to compute the derivative of a composite function.

Suppose F(x) = f(g(x)),

$F'(x) = f'(g(x))g'(x)$ **<-Lagrange's notation**

$$\frac{dy}{dx} = \frac{dy}{du} \cdot \frac{du}{dx}$$ **<- Leibniz's notation**

https://en.wikipedia.org/wiki/Chain_rule

2. Derivative, Chain Rule

❖ 2. Chain Rule

- example : $f = \left(g(x)\right)^2, \;\; g = (2x^2 + 3x + 1)$

  solution :

$$\frac{df}{dg} = \frac{d}{dg}\left(g(x)\right)^2 = 2g(x)$$

$$\frac{dg}{dx} = \frac{d}{dx}(2x^2 + 3x + 1) = 4x + 3$$

$$\frac{df}{dx} = \frac{df}{dg} \cdot \frac{dg}{dx} = 2(2x^2 + 3x + 1\ )(4x + 3)$$

❖ 2. Chain Rule

- example : $f = (g(h))^2,\ g = (h(x))^2,\ h = (2x+1)$

solution :

$$\frac{df}{dg} = \frac{d}{dg}\big(g(h)\big)^2 = 2g(h)$$

$$\frac{dg}{dh} = \frac{d}{dh}\big(h(x)\big)^2 = 2h(x)$$

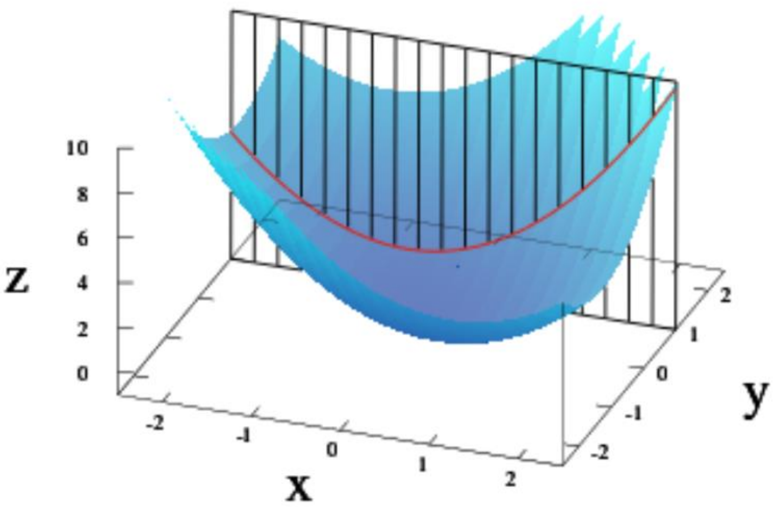
$$\frac{dh}{\alpha x} = \frac{d}{\alpha x}(2x+1) = 2$$

$$\frac{df}{dx} = \frac{df}{dg}\cdot\frac{dg}{dh}\cdot\frac{dh}{\alpha x} = 2(2x+1)^2 * 2(2x+1) * 2$$

$$= 8(2x+1)^3$$

❖ 3. Partial Derivative

- Functions used in machine learning → have multi-variable
- derivative of function having multi-variable → Partial Derivative

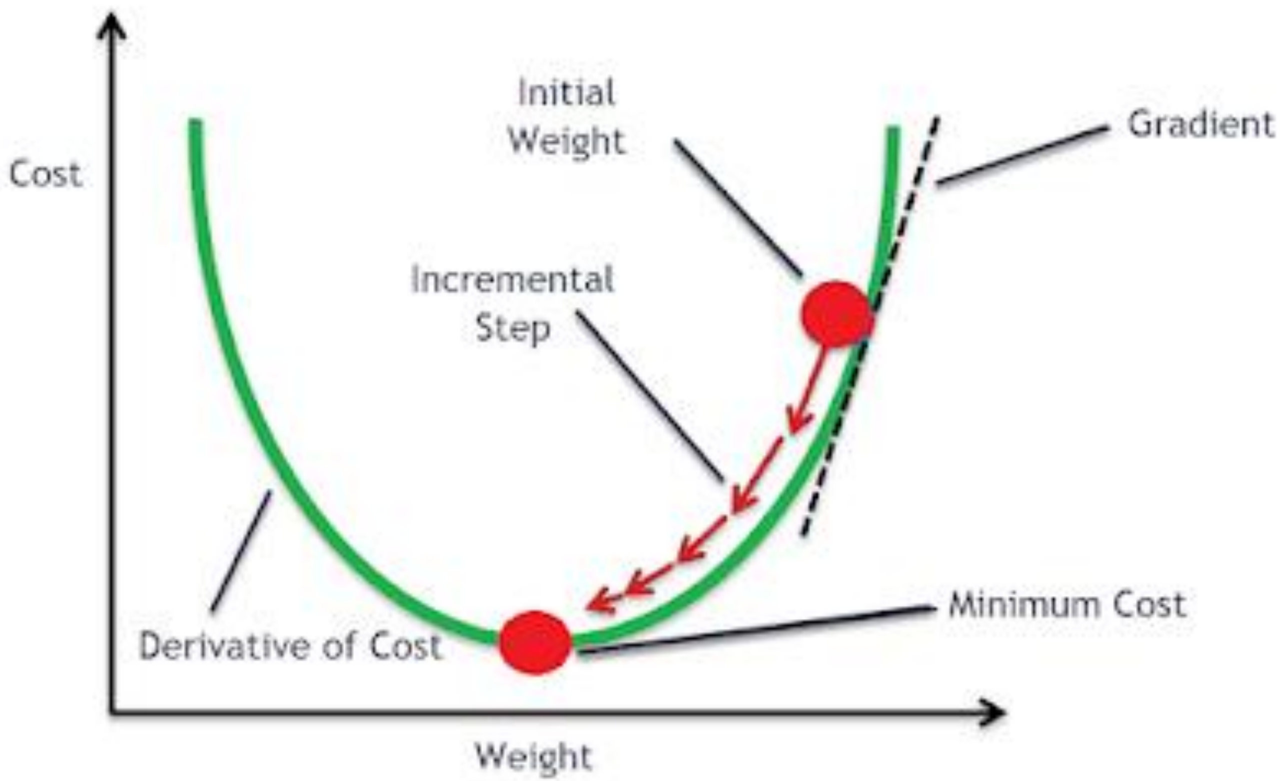


예) $f(x, y) = x^2 + 3xy + 2y^2$

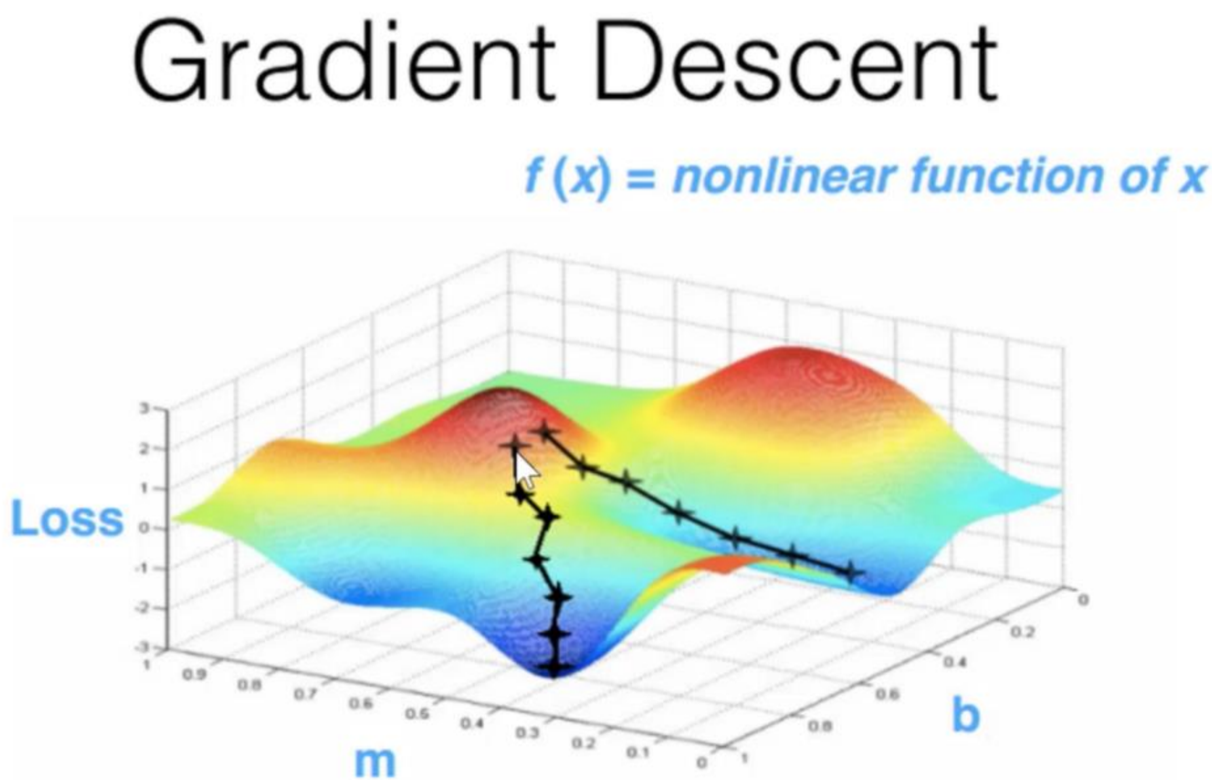
$$\frac{\partial f}{\partial x} = 2x + 3y, \qquad \frac{\partial f}{\partial y} = 3x + 4y$$

https://en.wikipedia.org/wiki/Partial_derivative

❖ 1. Gradient Descent of single variable



❖ 2. Gradient Descent of multi-variable



https://mc.ai/stochastic-gradient-descent-in-plain-english/