Artificial Intelligence

Basic Mathematics for Artificial Intelligence : Part 1

1. Variables & Constants
2. Linear Equations & Quadratic Equations
3. Concepts of Functions

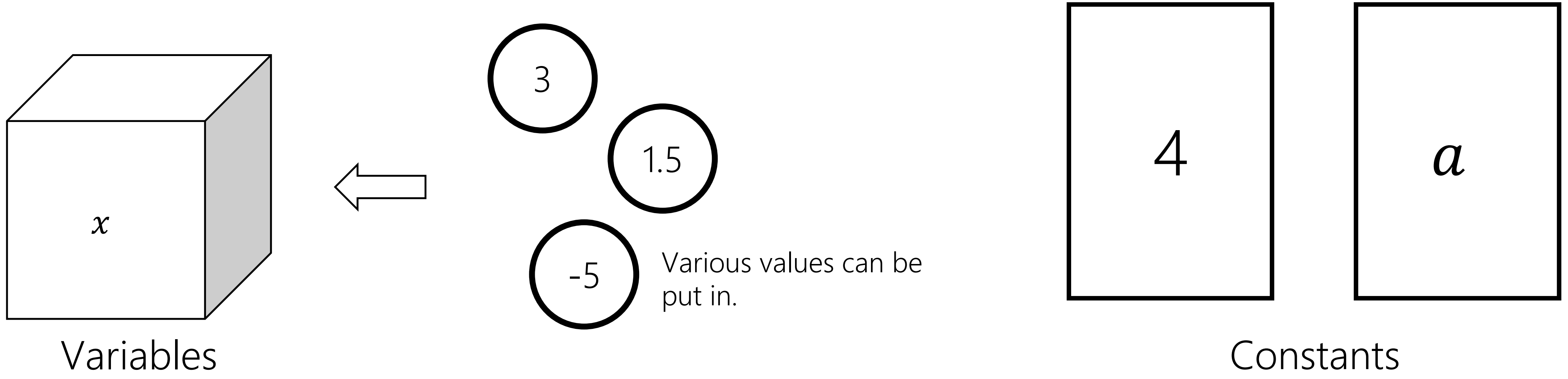
Artificial Intelligence

Basic Mathematics for Artificial Intelligence : Part 1

1. Variables & Constants
2. Linear Equations & Quadratic Equations
3. Concepts of Functions

Variables & Constants

<Figure 1.1.1>



Variables & Constants

$y = ax + b$	r	π	πr^2
$A\ formula\ for\ the\ relationship\ between\ x\ and\ y$	$Radius\ of\ a\ circle$	pi	$Circle\ area$

Variables & Constants

Exercise. Distinguish between Variable and Constant among a and b

$a\ cm$	$b\ cm$	$S\ cm^2$ $(S = ab)$
$Vertical\ Length$	$Horizontal\ Length$	$Rectangular\ Area$
Constant	Variable	

Artificial Intelligence

Basic Mathematics for Artificial Intelligence : Part 1

1. Variables & Constants
2. Linear Equations & Quadratic Equations
3. Concepts of Functions

Linear Equations & Quadratic Equations

$Term$

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

$Degree$

$Term$	$Degree$
3	0
A	1
$-4ab$	2
a^2	2

$Coefficient$

$Term$	$Coefficient$
3	3
$3a$	3
$\frac{x}{3}$	$\frac{1}{3}$ (※ $\frac{1}{3} \times x$)
$-4ab$	-4

Linear Equations & Quadratic Equations

$$Monomial$$

$$Ex)\ 3, a, -4ab, \frac{x}{3}\ , \ldots$$

$$Polynomial$$

$$Ex)\ 3a - 2b + 4a^2b + 6, \ldots$$

<Figure 1.2.1>

$$\underline{3a} + \underline{(-2b)} + \underline{4a^2b} + \underline{6}$$

$Coefficient$	3	-2	4	6
$Degree$	1	1	3	0

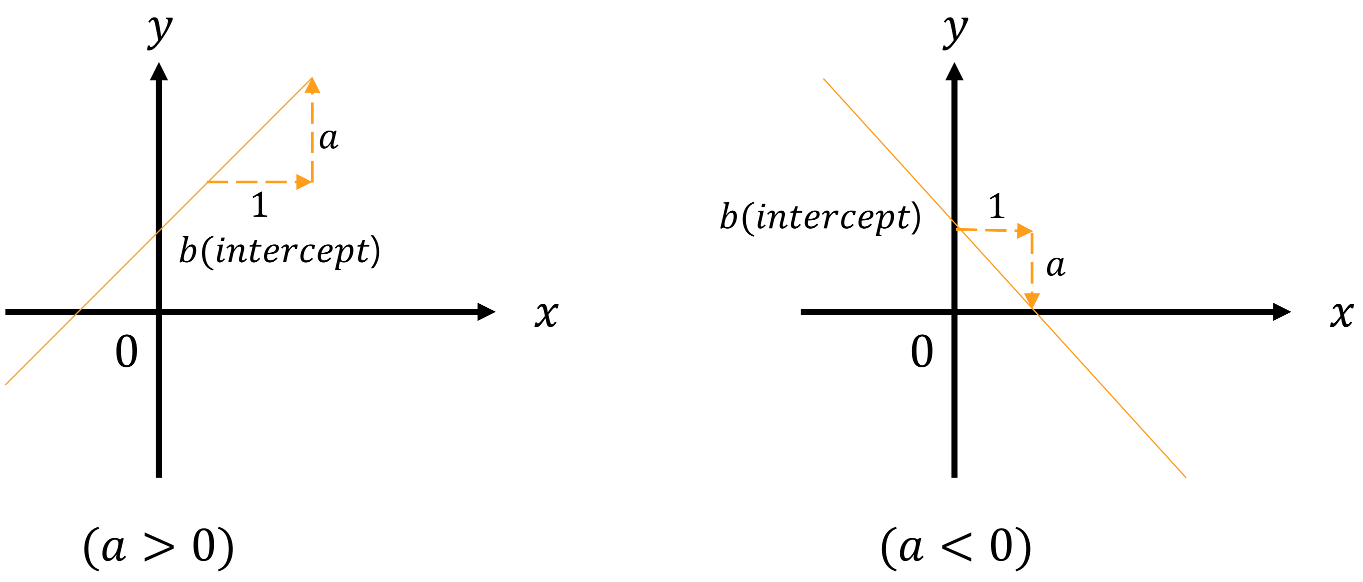
Linear Equations & Quadratic Equations

Definition : Linear Equations for x

$ax + b$ (※$a \neq 0$)

$$y = ax + b$$

<Figure 1.2.2>



Linear Equations & Quadratic Equations

Definition : Quadratic Equations for x

$ax^2 + bx + c$ (※$a \neq 0$)

$$y = ax^2 + bx + c$$

<Figure 1.2.3>

$(a > 0)$ $(a < 0)$

Linear Equations & Quadratic Equations

Definition : Nth Equations for x

$$a_0x^n + a_1x^{n-1} + a_2x^{n-2} + \cdots + a_{n-1}x + a_n (※\ a_0 \neq 0)$$

$$n = 4 \Rightarrow a_0x^4 + a_1x^3 + a_2x^2 + a_3x + a_4$$

$$n = 5 \Rightarrow a_0x^5 + a_1x^4 + \cdots + a_5$$

$$n = 6 \Rightarrow a_0x^6 + a_1x^5 + \cdots + a_6$$

…

Linear Equations & Quadratic Equations

Exercise 1

1-1) $-3ab$

$Monomial$	$Coefficient$	$Degree$
	-3	2

1-2) $2ab + b + 4$

$Polynomial$	$Coefficient$	$Degree$
	2,1,4	2

1-3) $3x^2 + 4$

$Polynomial$	$Coefficient$	$Degree$
	3,4	2

Exercise 2

$$3ax^2 + x + 2ab$$

$Coefficient$
$3a, \quad 1, \quad 2ab$

Artificial Intelligence

Basic Mathematics for Artificial Intelligence : Part 1

1. Variables & Constants
2. Linear Equations & Quadratic Equations
3. Concepts of Functions

Concepts of Functions

$$y = f(x)$$

<Figure 1.3.1>

x	$\rightarrow$	$f(x)$	$\rightarrow$	y
Input		Variation		Output

Input	Variation	Output
x	$f(x) = 2x$	y
0	$f(x) = 2 \times 0$	0
2	$f(x) = 2 \times 2$	4

John McCarthy
04.09.1927 ~ 24.10.2011

Artificial Intelligence

IBM
Deep Blue

Google DeepMind
Alpha Go

Artificial Intelligence

Cognitive Computing

Machine Learning

Experts System

Deep Learning

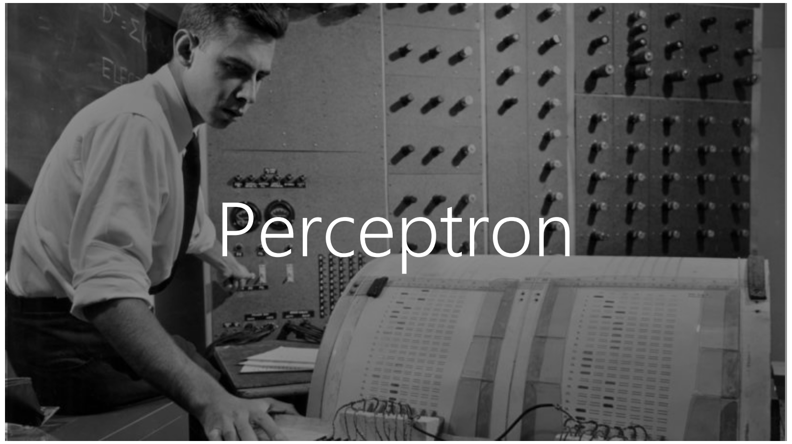
Artificial Neural Network

Deep Learning

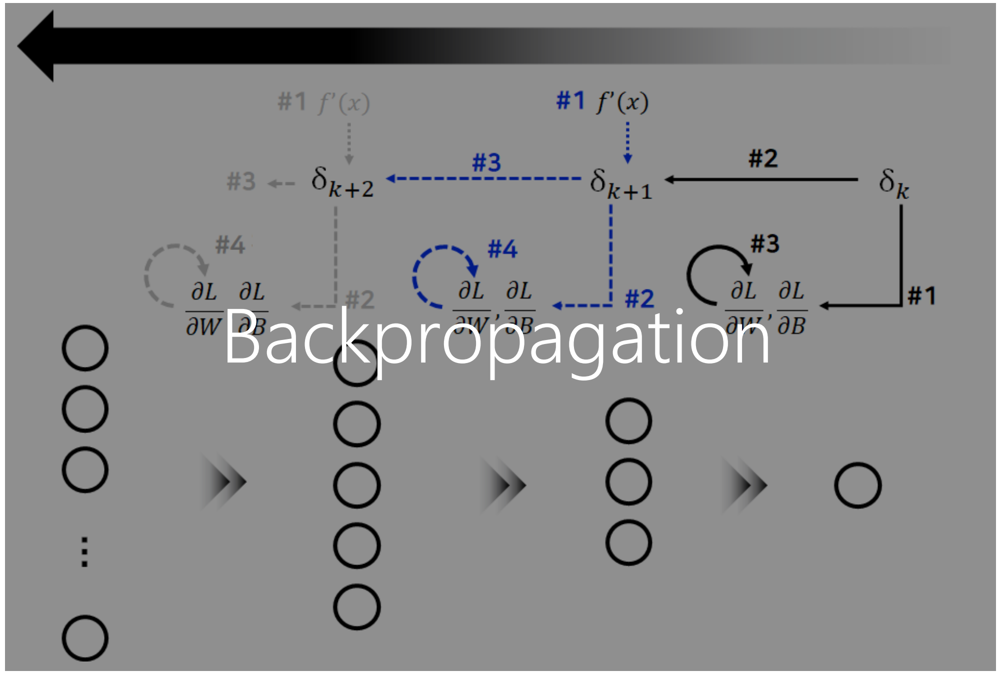
Unsupervised Learning

Supervised Learning

Reinforcement Learning



Frank Rosenblatt
11.07.1928 ~ 11.07.1971



Artificial Intelligence

Basic Mathematics for Artificial Intelligence : Part 1

4. Square root
5. Exponentiation
6. Exponential function & logarithmic function
7. Natural logarithm
8. Sigmoid function

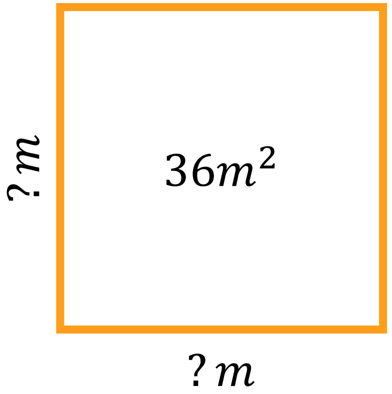
Artificial Intelligence

Basic Mathematics for Artificial Intelligence : Part 1

4. Square root
5. Exponentiation
6. Exponential function & logarithmic function
7. Natural logarithm
8. Sigmoid function

Square root

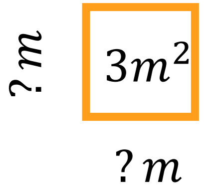
<Figure 1.4.1>



✓ $6 \times 6 = 36$

✓ $-6 \times -6 = 36$

<Figure 1.4.2>



✓ $\sqrt{3} \times \sqrt{3} = 3$

✓ $-\sqrt{3} \times -\sqrt{3} = 3$

Square root

Formula

$(※\ a > 0, b > 0, c > 0)$

1. $\sqrt{a^2} = a$
2. $a \times \sqrt{b} = a\sqrt{b}$
3. $b\sqrt{a} + c\sqrt{a} = (b + c)\sqrt{a}$
4. $\boxed{\sqrt{a} \times \sqrt{b} = \sqrt{ab}}$
5. $\boxed{\sqrt{a} \div \sqrt{c} = \frac{\sqrt{a}}{\sqrt{c}} = \sqrt{\frac{a}{c}}}$
6. $\sqrt{a^2 \times b} = a\sqrt{b}$

Square root of 5 is $\pm\sqrt{5}$

$2\sqrt{2} = 2 \times \sqrt{2}$

$\sqrt{2} + 2\sqrt{3}$

$\sqrt{12} = \sqrt{2^2 \times 3} = 2\sqrt{3}$

Square root

Exercise 1

1) Find the square root of 9

A : 3, -3

(※ $3^2 = 9, \quad (-3)^2 = 9$)

2-1) $\sqrt{18} + \sqrt{2}$

$= \sqrt{3^2 \times 2} + \sqrt{2}$

$= 3\sqrt{2} + \sqrt{2}$

$= 4\sqrt{2}$

A : $4\sqrt{2}$

2-2) $3\sqrt{6} \times 2\sqrt{2}$

$= 6\sqrt{12}$

$= 6 \times \sqrt{2^2 \times 3}$

$= 6 \times 2\sqrt{3}$

$= 12\sqrt{3}$

A : $12\sqrt{3}$

Artificial Intelligence

Basic Mathematics for Artificial Intelligence : Part 1

4. Square root
5. Exponentiation
6. Exponential function & logarithmic function
7. Natural logarithm
8. Sigmoid function

Exponentiation

2^2 $\quad$ 2^3 $\quad$ 2^N

$square$ $\quad$ $cube$ $\quad$ $to\ the\ power\ of\ N$

a^p ← exponent, index

↑

base

$\sqrt[p]{a}$

$4 \times 4 \times 4 = 64$

$\sqrt[3]{64} = 4$

$$\sqrt[2]{a} = \sqrt{a}$$

Exponentiation

Formula

(※ $a > 0, b > 0$)

1. $a^0 = 1$
2. $a^p a^q = a^{p+q}$
3. $(a^p)^q = a^{pq}$
4. $(ab)^p = a^p b^p$
5. $a^{-p} = \frac{1}{a^p}$
6. $\sqrt[p]{a}\sqrt[p]{b} = \sqrt[p]{ab}$
7. $\sqrt[p]{\sqrt[q]{a}} = \sqrt[pq]{a}$
8. $\sqrt[p]{a} = a^{\frac{1}{p}}$

- $2^{-1} \times 2^2 = \frac{1}{2} \times 4 = 2$
- $2^{-1} \times 2^2 = 2^{(-1+2)} = 2$
- $\sqrt{a} = a^{\frac{1}{2}}$
- $(\sqrt{a})^2 = a, \left(a^{\frac{1}{2}}\right)^2 = a$

Exponentiation

Exercise 1

1-1) $4^2 \times 2^{-1} \div 2^2$

$= (2^2)^4 \times 2^{-1} \times 2^{-2}$

$= 2^8 \times 2^{-1} \times 2^{-2}$

$= 2^{(8-1-8)} = 2^5 = 32$

A : 32

1-2) $\sqrt[3]{81} \times \sqrt[3]{9}$

$= 81^{\frac{1}{3}} \times 9^{\frac{1}{3}}$

$= (3^4)^{\frac{1}{3}} \times (3^2)^{\frac{1}{3}}$

$= 3^{\frac{4}{3}} \times 3^{\frac{2}{3}}$

$= 3^{\left(\frac{4}{3}+\frac{2}{3}\right)} = 3^2 = 9$

A : 9

1-3) $\sqrt[3]{\sqrt{64}}$

$= \sqrt[6]{64}$

$= 64^{\frac{1}{6}}$

$= (2^6)^{\frac{1}{6}} = 2^{\frac{6}{6}} = 2$

A : 2

Artificial Intelligence

Basic Mathematics for Artificial Intelligence : Part 1

4. Square root
5. Exponentiation
6. Exponential function & logarithmic function
7. Natural logarithm
8. Sigmoid function

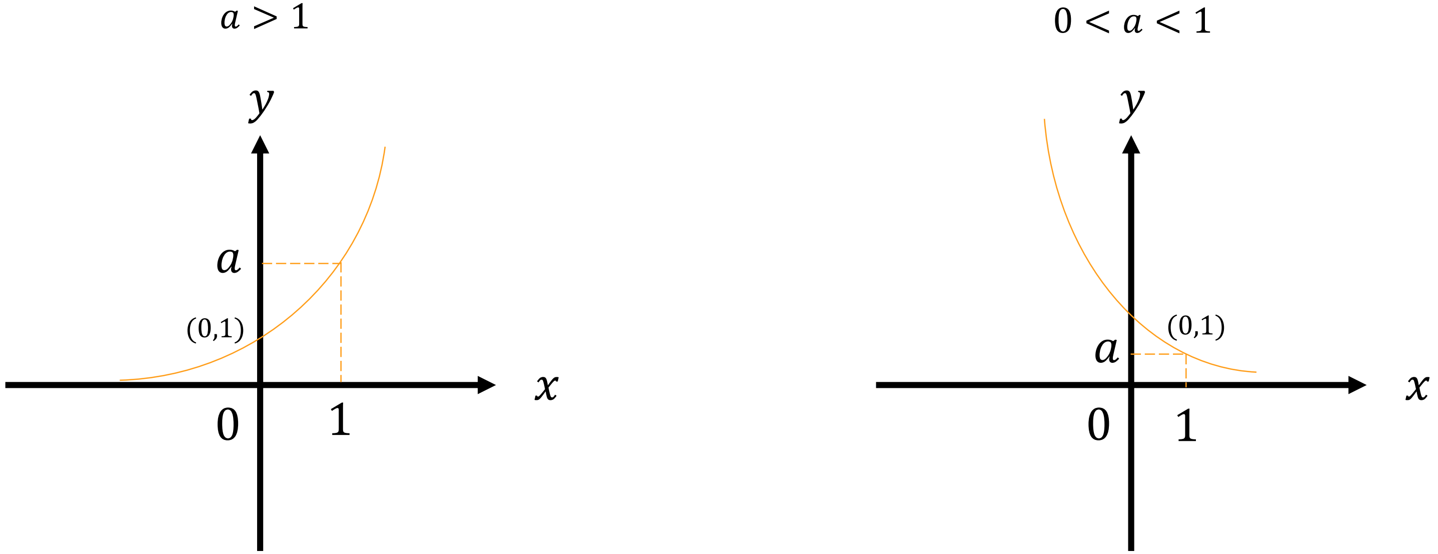
Exponential function & logarithmic function

Definition

$$y = a^x$$

$(※\ a > 0, a \neq 1)$

<Figure 1.6.1>



Exponential function & logarithmic function

Definition

$y = log_a x$

$\uparrow$

Antilogarithm

$(※ a > 0, a \neq 1, x > 0)$

$log_2 4 = ?$ $\quad 2^? = 4$ $\quad log_2 4 = 2$

$log_3 27 =?$ $\quad 3^? = 27$ $\quad log_3 27 = 3$

Formula

$(※ a > 0, \quad a \neq 1, \quad X, Y > 0)$

1. $log_a a = 1$
2. $log_a 1 = 0$
3. $log_a XY = log_a X + log_a Y$
4. $log_a \frac{X}{Y} = log_a X - log_a Y$
5. $log_a X^p = p \, log_a X$
6. $log_a X = \frac{log_c X}{log_c a}$ $(※ c > 0, c \neq 1)$

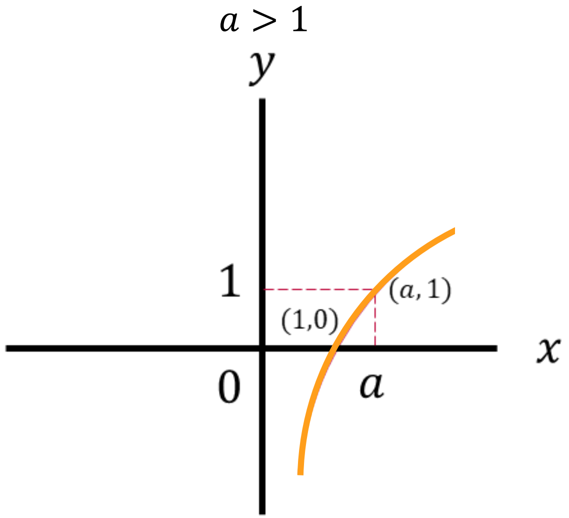
1. $log\_e(a)$

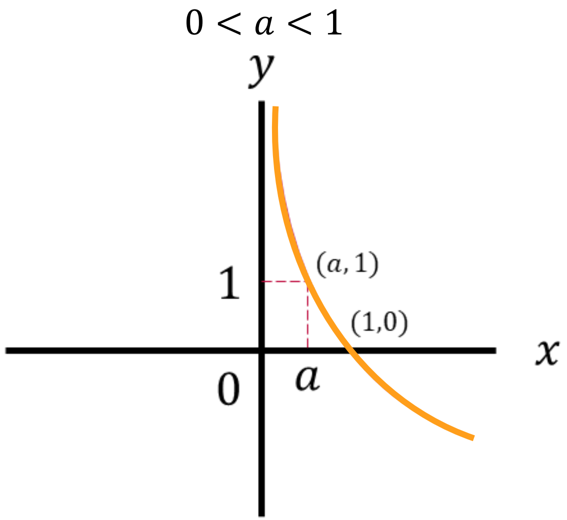
Exponential function & logarithmic function

Definition

$$y = log_a x$$

<Figure 1.6.2>





Exponential function & logarithmic function

Exercise 1

1-1) $log_3\sqrt{27}$

$$= log_3\, 27^{\frac{1}{2}}$$

$$= log_3(3^3)^{\frac{1}{2}} = log_3\, 3^{\frac{3}{2}} = \frac{3}{2} log_3\, 3 = \frac{3}{2}$$

A : $\frac{3}{2}$

1-2) $log_3\frac{3}{4} + 4\, log_3\sqrt{2}$

$$= log_3\frac{3}{4} + log_3\left(\sqrt{2}\right)^4$$

$$= log_3\frac{3}{4} + log_3\, 4$$

(※ $log_a XY = log_a X + log_a Y$)

$$= log_3\left(\frac{3}{4} \times 4\right) = log_3\, 3 = 1$$

A : 1

Artificial Intelligence

Basic Mathematics for Artificial Intelligence : Part 1

4. Square root
5. Exponentiation
6. Exponential function & logarithmic function
7. Natural logarithm
8. Sigmoid function

Natural logarithm

Definition : $Napier's\ number(euler's\ number)$

$$e = \lim_{n \to \infty} \left(1 + \frac{1}{n}\right)^n = 2.718281 \ldots$$

Natural logarithm

$$log_e = ln$$

✓ $\frac{d}{dx} e^x = e^x$

✓ $\frac{d}{dx} ln\, x = \frac{1}{x}$

✓ $e^x = exp\, x = exp(x)$

Artificial Intelligence

Basic Mathematics for Artificial Intelligence : Part 1

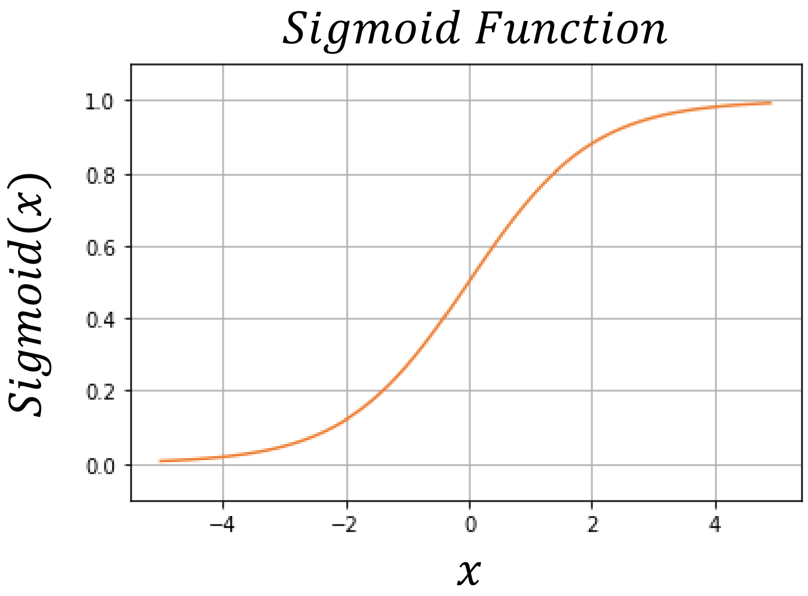
4. Square root
5. Exponentiation
6. Exponential function & logarithmic function
7. Natural logarithm
8. Sigmoid function

Sigmoid function

<Figure 1.8.1>

Definition

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



$Binary\ Classification$

- ✓ $ReLU$
- ✓ $Hyperbolic$
- ✓ Elu
- ✓ $Softmax$
- ✓ …

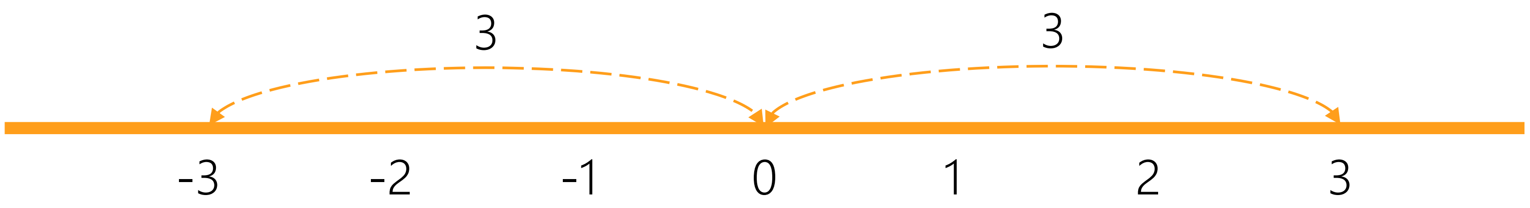
Artificial Intelligence

Basic Mathematics for Artificial Intelligence : Part 1

9. Absolute value and Euclidean distance
10. Sequence
11. Set and Elements

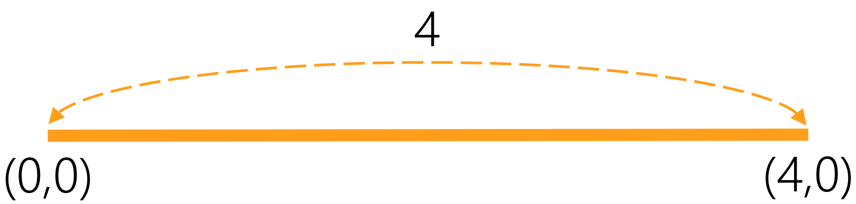
Absolute value and Euclidean distance

<Figure 1.9.1>



- ✓ $|3| = 3$
- ✓ $|3| = -3$

<Figure 1.9.2>



Absolute value and Euclidean distance

$1\ Dimension$

|A-B|

|4-(-1)| = |5| = 5

(A) (B)

(-1) (4)

$2\ Dimension$

$$\sqrt{(a_1 - b_1)^2 + (a_2 - b_2)^2}$$

$A(a_1, a_2)$ $B(b_1, b_2)$

$3\ Dimension$

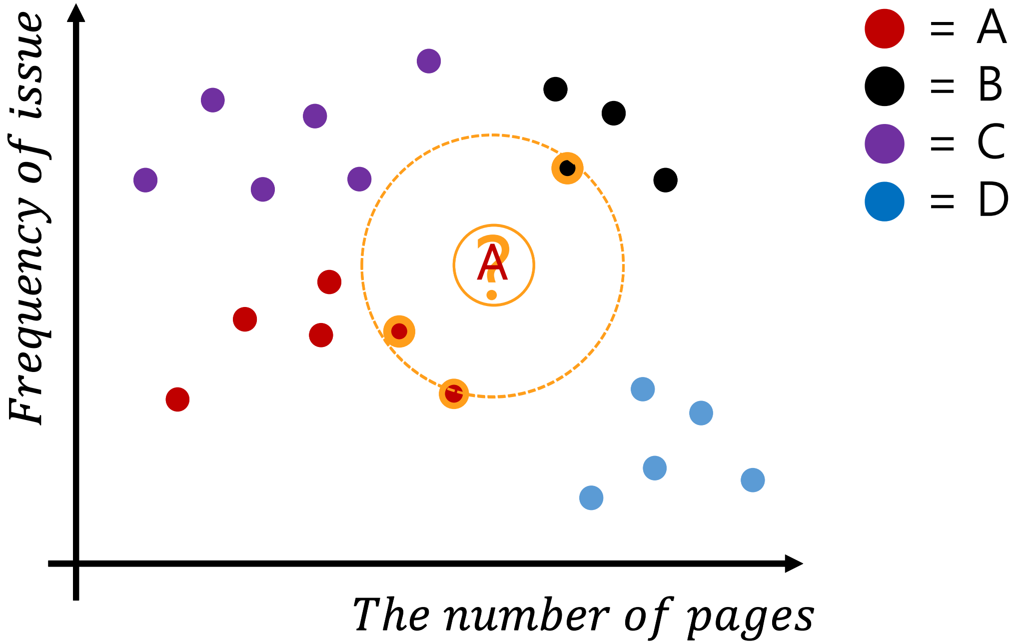
$$\sqrt{(a_1 - b_1)^2 + (a_2 - b_2)^2 + (a_3 - b_3)^2}$$

$A(a_1, a_2, a_3)$ $B(b_1, b_2, b_3)$

Absolute value and Euclidean distance

Ex) k-nearest neighbor

※ $k = 3$



Absolute value and Euclidean distance

Exercise 1. Find the absolute value of the following expression.

1-1) $|-2|$ A : 2

1-2) $|-\frac{3}{2}|$ A : $\frac{3}{2}$

1-3) $|x-3|$ (※ $x<3$)

A : $-x+3$

1-4) $|x-3|$ (※ $x>3$)

A : $x-3$

1-5) $|a-b|$ (※ $a<b$)

A : $-a+b$

Exercise 2. Find the Euclidean distance between the following two points.

2-1) A(3), B(-2)

2-2) A(2, -2), B(-3, 1)

2-3) A(1, 3, -1), B(-1, 0, 1)

Absolute value and Euclidean distance

Exercise 2. Find the Euclidean distance between the following two points.

2-1) A(3), B(-2) $= |3-(-2)| = |3+2| = |5| = 5$

A : 5

2-2) A(2, -2), B(-3, 1) $= \sqrt{\{2-(-3)\}^2+(-2-1)^2} = \sqrt{5^2+(-3)^2} = \sqrt{25+9} = \sqrt{34}$

A : $\sqrt{34}$

2-3) A(1, 3, -1), B(-1, 0, 1) $= \sqrt{\{1-(-1)\}^2+(3-0)^2+(-1-1)^2} = \sqrt{2^2+3^2+(-2)^2} = \sqrt{4+9+4} = \sqrt{17}$

A : $\sqrt{17}$

Sequence

$$\text{Term}$$

$$\underline{a_1}, a_2, a_3, a_4, \ldots, a_{n-1}, \underline{a_n}$$

$First\ term$ $\qquad$ $Last\ term$

✓ $Arithmetic\ sequence$

- $First\ term \rightarrow a$
- $Common\ difference \rightarrow d$

The nth term of an arithmetic sequence is given by. an = $a + (n - 1)d$.

2, 5, 8, 11, 14, 17, 20, 23,...

$$a_{n+1} = a_n + \underbrace{3}_{Common\ difference}$$

$$a_n = 2 + (n - 1) \times 3$$
$$= 3n - 1$$

✓ $Geometric\ sequence$

Sequence

✓ $Sum\ of\ Arithmetic\ sequence$

- $First\ term\ 2$
- $Common\ difference\ 3$
- $Last\ term\ 26$
- $Number\ of\ terms\ 9$

$$S = ?$$

- $First\ term\ a$
- $Last\ term\ l$
- $Number\ of\ terms\ n$
- $Sum\ of\ Arithmetic\ sequence\ S$

$$S = \frac{1}{2} n\,(a + l)$$

✓ 2+5+8+11+14+17+20+23+26

$$\begin{array}{rl} & 2+5+8+11+14+17+20+23+26 \\ +) & 26+23+20+17+14+11+8+5+2 \\ \hline \end{array}$$

2S = 28+28+28+28+28+28+28+28+28

⬇

(2+26) × 9 = 252

⬇

S = 252/2 = 126

Sequence

✓ *Geometric sequence*

3, 6, 12, 24, 48, 96, 196, …

✓ $a_n = 3 \times 2^{n-1}$

- *First term a*

$$a_{n+1} = 2a_n$$

Geometric ratio

$$a_n = ar^{n-1}$$

Sequence

✓ $Sum\ of\ Geometric\ sequence$

- $First\ term\ 3$
- $Last\ term\ 192$
- $Geometric\ ratio\ 2$
- $Number\ of\ terms\ 7$

S = (3 + 6 + 12 + 24 + 48 + 96 + 192)

$$S = 3 + 6 + 12 + 24 + 48 + 96 + 192$$

$$-)\ 2S = 3 \times 2 + 6 \times 2 + 12 \times 2 + 24 \times 2 + 48 \times 2 + 96 \times 2 + 192 \times 2$$

$$(1-2)S = 3 - 192 \times 2$$

$$= 381$$

Definition

- $First\ term\ a$
- $Geometric\ ratio\ r$

1) $S_n = \frac{a(1-r^n)}{1-r} = \frac{a(r^n-1)}{r-1}$ (※$r \neq 1$)

2) $S_n = na$ (※$r = 1$)

Sequence

✓ Σ, Π

$$a_1, a_2, a_3, \ldots, a_{n-1}, a_n$$

$$\Downarrow$$

$$a_1 + a_2 + a_3 + \cdots + a_{n-1} + a_n$$

$$\Downarrow$$

$$\sum_{k=1}^{n} a_k$$

$$\sum_{k=1}^{4} (3k+1)$$

$$= (3\times 1+1) + (3\times 2+1) + (3\times 3+1) + (3\times 4+1)$$

$$= 4 + 7 + 10 + 13 = 34$$

$$\sum_{k=1}^{n} k$$

$$= 1 + 2 + 3 + 4 + \cdots + (n-1) + n$$

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

Sequence

✓ Σ, Π

Definition

1) $\sum_{k=1}^{n} k = \frac{1}{2}n(n+1)$ 2) $\sum_{k=1}^{n} k^2 = \frac{1}{6}n(n+1)(2n+1)$ 3) $\sum_{k=1}^{n} k^3 = \left\{\frac{1}{2}n(n+1)\right\}^2$ 4) $\sum_{k=1}^{n} c = nc(※\ c = constant)$

Properties

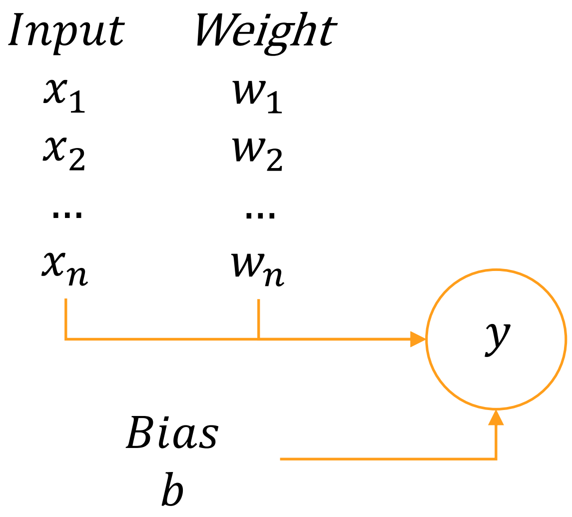
1) $\sum_{k=1}^{n}(a_k + b_k) = \sum_{k=1}^{n} a_k + \sum_{k=1}^{n} b_k$ 2) $\sum_{k=1}^{n} pa_k = p\sum_{k=1}^{n} a_k\ (※\ p = constant)$

Sequence

✓ Σ, Π

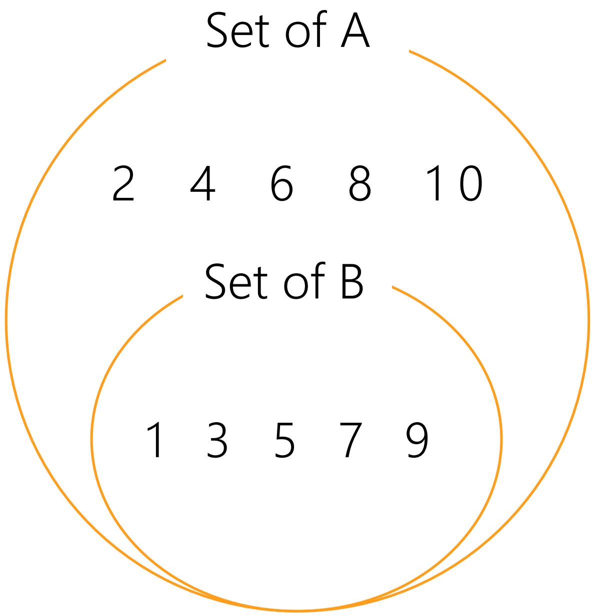
$$\prod_{k=1}^{4} a_k = a_1 \times a_2 \times a_3 \times a_4 = 1 \times 3 \times 5 \times 7 = 105$$

Sequence



$$y = b + x_1 \cdot w_1 + x_2 \cdot w_2 + \cdots + x_n w_n$$
$$= \sum_{k=1}^{n} x_k w_k + b$$

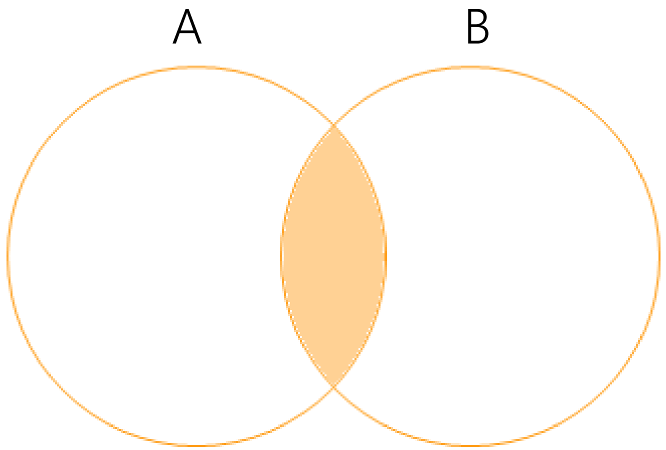
Set and elements



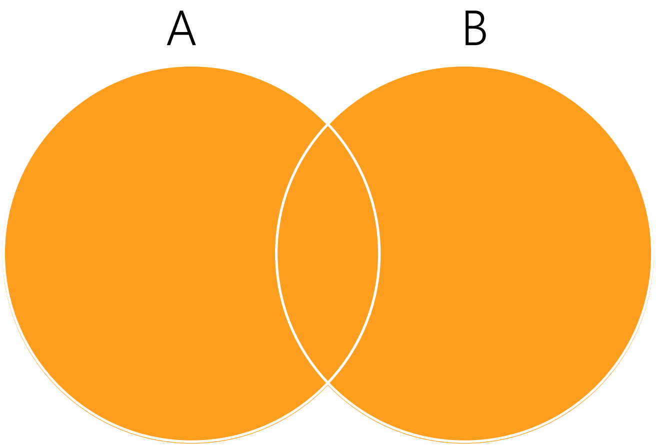
$B \subset A$

- $\{2, 4, 6, 8, 10\}$
- $A = \{1, 3, 5, 7, 9\}$
- $x \in A$
- $x \notin A$
- $A = B$
- ϕ
- $\phi \subset$ A

Set and elements



A ∩ B



A ∪ B

Set and elements

Exercise 1. Answer the following questions.

1-1) $\{x|x^2 = 9\}$

A : {-3, 3}

1-2) $\{x|x\ is\ positive\ factor\ of\ 12\}$

A : {1,2,3,4,6,12}

2-1) $A \cap B$ (※ $A = \{1,2,3,4,5,6\},\ B = \{4,5,6,7,8,9\}$)

A : {4,5,6}

2-2) $A \cup B$ (※ $A = \{1,2,3,4,5,6\},\ B = \{4,5,6,7,8,9\}$)

A : {1,2,3,4,5,6,7,8,9}

Limitations of Artificial Intelligence.

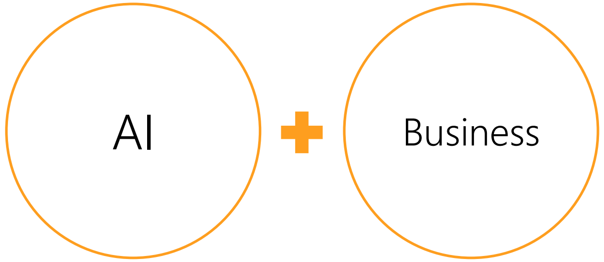
- Data Dependency
- Data Bias
- Human Labor
- AI fails to Explain
- Scalability of Algorithms

The Future of Artificial Intelligence

The Interaction of Human and Artificial Intelligence.

Emergence of Various Data

Overcoming Weakness



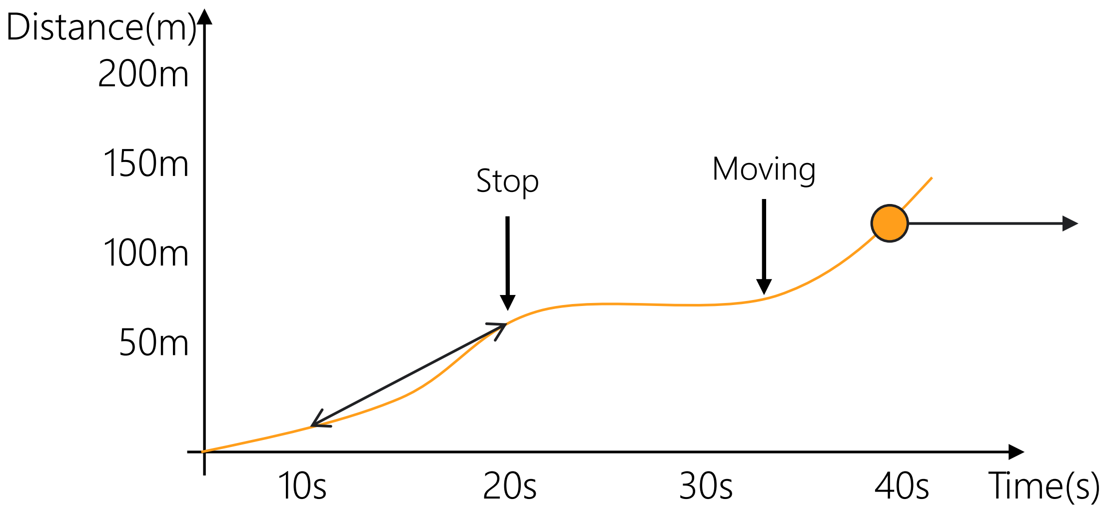
Artificial Intelligence

Basic Mathematics for Artificial Intelligence : Part 2

1. Derivative
2. Partial Derivative
3. Function Composite

Derivative

The Derivative as an Instantaneous Rate of Change.



$$\frac{120m}{40s} = 3m/s$$

$$\frac{60m}{10s} = 6m/s$$

Formula

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

Differential Operator

$$\frac{df(x)}{dx} \; or \; \frac{d}{dx}f(x) \; or \; f'(x)$$

$$Ex)10.0s \leftrightarrow 10.1s$$

$$(40m) \qquad (40.6m)$$

$$\frac{f(10+0.1)-f(10)}{0.1} = \frac{40.6-40}{0.1} = 6$$

Derivative

Properties

- $\frac{d}{dx}f(x) = nx^{n-1}$ ($※ f(x) = x^n$)
- $\frac{d}{dx}\big(f(x) + g(x)\big) = \frac{d}{dx}f(x) + \frac{d}{dx}g(x)$
- $\frac{d}{dx}\big(af(x)\big) = a\frac{d}{dx}f(x)$
- $\frac{d}{dx}a = 0$
- $\frac{d}{dx}\sum_{i=0}^{n} x^n = \sum_{i=0}^{n}\frac{d}{dx}x^n$

Exercise 1. Answer the following questions.

1. $\frac{d}{dx}5 = 0$

2. $\frac{d}{dx}x = \frac{d}{dx}x^1 = 1 \times x^0 = 1$

3. $\frac{d}{dx}x^3 = 3x^2$

4. $\frac{d}{dx}x^{-2} = -2x^{-3}$

5. $\frac{d}{dx}10x^4 = 10\frac{d}{dx}x^4 = 10 \times 4x^3$
$= 40x^3$

6. $\frac{d}{dx}(x^5 + x^6) = \frac{d}{dx}x^5 + \frac{d}{dx}x^6$
$= 5x^4 + 6x^5$

Derivative

Loss function

(cost function)

Loss

Label ⟷ Prediction

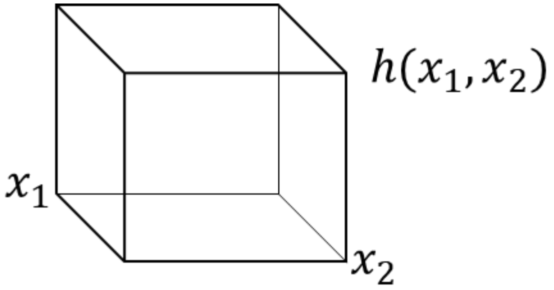
✓ Gradient Descent

Partial Derivative

multivariate function

$$g(x_1, x_2, \dots, x_n) = x_1 + x_2^2 + \dots + x_n^n$$

$$h(x_1, x_2) = x_1^2 + x_2^3$$



$$h(x_1, x_2) = x_1^2 + 1^3 \qquad \frac{\partial}{\partial x_1} h(x_1, x_2) = 2x_1$$

※ $\partial(del)$

$$h(x_1, x_2) = 1^2 + x_2^3 \qquad \frac{\partial}{\partial x_2} h(x_1, x_2) = 3x_2^2$$

Exercise 1. Find the derivative of the function.

1-1) $f(x) = ax^2 + bx + c$

A : $\frac{df(x)}{dx} = 2ax + b$

1-2) $f(x, y) = 3x^2 + 5xy + 3y^3$

A : $\frac{\partial f(x,y)}{\partial x} = 6x + 5y$

$\frac{\partial f(x,y)}{\partial y} = 5x + 9y^2$

Function Composite
(Chain rule)

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

$$f(1) = 10 + 1^2 = 11$$
$$f(2) = 10 + 2^2 = 14$$
$$f(3) = 10 + 3^2 = 19$$

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

$$g(1) = 3 + 1 = 4$$
$$g(2) = 3 + 2 = 5$$
$$g(3) = 3 + 3 = 6$$

Function Composite

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

Function Composite (Chain rule)

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

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

$Ex)\ \frac{d}{dx} f(g(x)) = ?$

$$y = f(u)$$

$$u = g(x)$$

$$\frac{dy}{dx} = \frac{dy}{du} \cdot \frac{du}{dx}$$

$$\frac{dy}{du} = \frac{d}{du} f(u)$$

$$=?$$

$$\frac{du}{dx} = \frac{d}{dx} g(x)$$

$$=?$$

$$\frac{dy}{du} = \frac{d}{du} f(u)$$

$$= \frac{d}{du}(10 + u^2)$$

$$= 2u$$

$$\frac{du}{dx} = \frac{d}{dx} g(x)$$

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

$$= 1$$

$$\frac{dy}{dx} = \frac{dy}{du} \cdot \frac{du}{dx}$$

$$= 2u \cdot 1$$

$$= 2g(x)$$

$$= 2(3 + x)$$

Function Composite
(Chain rule)

Exercise 1. Find the derivative of the function.

1-1) $f(x) = (3x-4)^{50}$

※ $u = 3x - 4$

$$\frac{df(x)}{dx} = \frac{df(x)}{du} \cdot \frac{du}{dx}$$

$$\frac{df(x)}{du} = \frac{du^{50}}{du} \cdot \frac{d(3x-4)}{dx}$$

$$= 50u^{49} \cdot 3$$

$$= 150(3x-4)^{49}$$

1-2) $f(x, y) = (3x-1)^2 + (x+y+1)^3$

※ $u = 3x - 1, \qquad v = x + y + 1$

$f(x, y) = u^2 + v^3$

$$\frac{\partial f(x,y)}{\partial x} = \frac{\partial f(x,y)}{\partial u} \cdot \frac{\partial u}{\partial x} + \frac{\partial f(x,y)}{\partial v} \cdot \frac{\partial v}{\partial x}$$

$$\frac{\partial f(x,y)}{\partial x} = \frac{\partial u^2}{\partial u} \cdot \frac{\partial u}{\partial x} + \frac{\partial v^3}{\partial v} \cdot \frac{\partial v}{\partial x}$$

$$= 2u \cdot 3 + 3v^2 \cdot 1$$

$$= 6(3x+1) + 3(x+y+1)^2$$

$$= 3x^2 + (6y+24)x + 3y^2 + 6y + 9$$

Artificial Intelligence

Basic Mathematics for Artificial Intelligence : Part 3

1. Vector
2. Inner Product
3. L1 norm and L2 norm
4. Cosine similarity

Vector

Ex)

- $\boldsymbol{a}, \boldsymbol{b}, \dots$
- $\vec{a}, \vec{b}, \vec{c}, \vec{d}, \dots$

Row vector

$$\boldsymbol{a} = (a_1, a_2, a_3, \dots, a_n)$$

Column vector

$$\boldsymbol{b} = \begin{pmatrix} b_1 \\ b_2 \\ b_3 \\ \dots \\ b_n \end{pmatrix}$$

Vector addition

$$\begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix} + \begin{pmatrix} 4 \\ 5 \\ 6 \end{pmatrix} = \begin{pmatrix} 1+4 \\ 2+5 \\ 3+6 \end{pmatrix} = \begin{pmatrix} 5 \\ 7 \\ 9 \end{pmatrix}$$

$$\begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix} + \begin{pmatrix} 4 \\ 5 \\ 6 \\ 7 \end{pmatrix} = \text{Unable to calculate}$$

Vector subtraction

$$\begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix} - \begin{pmatrix} 4 \\ 5 \\ 6 \end{pmatrix} = \begin{pmatrix} 1-4 \\ 2-5 \\ 3-6 \end{pmatrix} = \begin{pmatrix} -3 \\ -3 \\ -3 \end{pmatrix}$$

Scalar multiplication

$$2\begin{pmatrix} 1 \\ 2 \\ 3 \\ 4 \end{pmatrix} = \begin{pmatrix} 2\times 1 \\ 2\times 2 \\ 2\times 3 \\ 2\times 4 \end{pmatrix} = \begin{pmatrix} 2 \\ 4 \\ 6 \\ 8 \end{pmatrix}$$

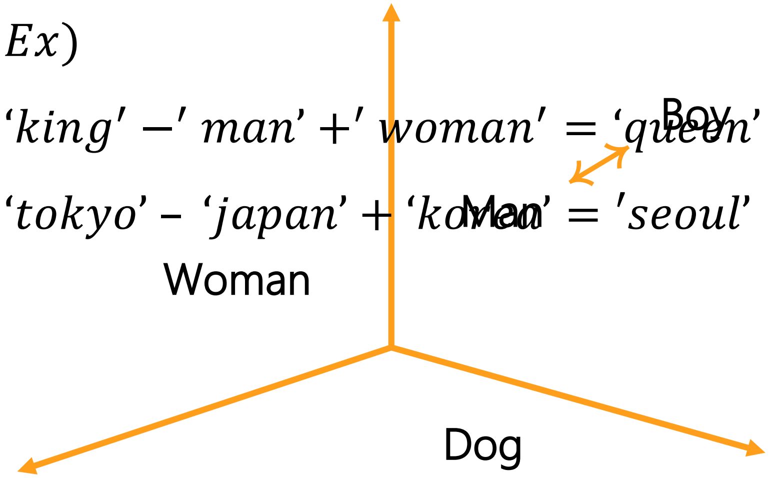
Scalar

Vector

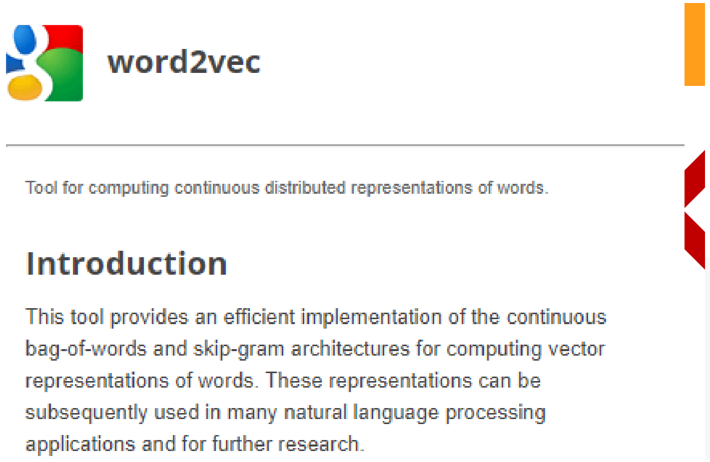
$Ex)$

$‘king' -' man' +' woman' = ‘queen'$

$‘tokyo' - ‘japan' + ‘korea' = 'seoul'$



Distributed Representation of Words in Vector Space

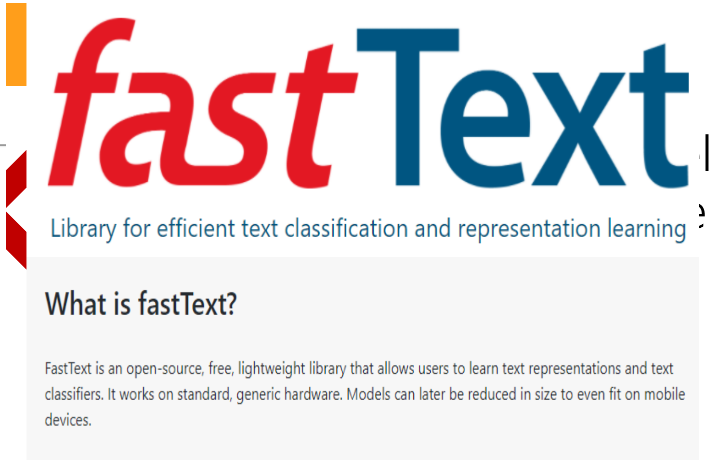


word2vec

Tool for computing continuous distributed representations of words.

Introduction

This tool provides an efficient implementation of the continuous bag-of-words and skip-gram architectures for computing vector representations of words. These representations can be subsequently used in many natural language processing applications and for further research.



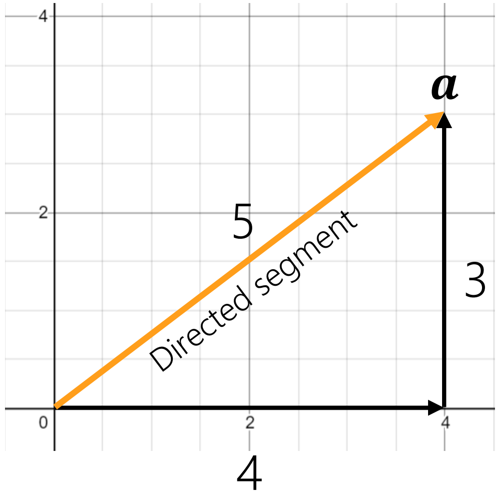
fastText

Library for efficient text classification and representation learning

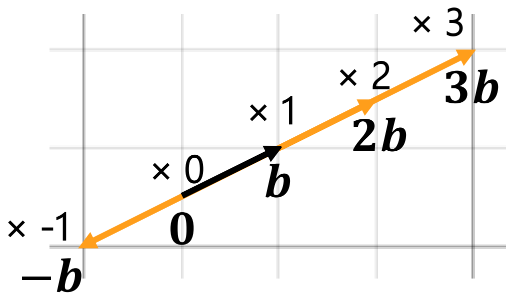
What is fastText?

FastText is an open-source, free, lightweight library that allows users to learn text representations and text classifiers. It works on standard, generic hardware. Models can later be reduced in size to even fit on mobile devices.

Vector

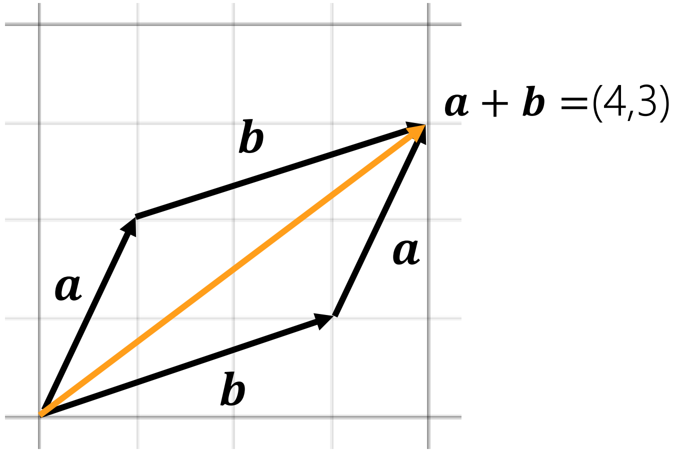


Directed line Segment of Vector a
$\boldsymbol{a} = (4,3)$

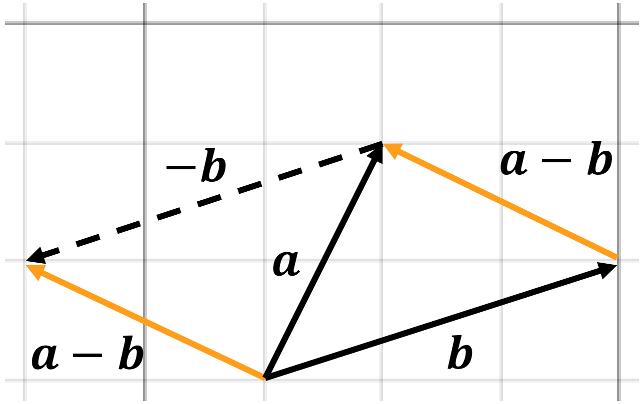


Directed line
Segment of Vector $3b, 2b, b, 0, -b$

Vector



Directed line Segment of Vector
$a, b, a + b$



Directed line Segment of Vector
$a, b, -b, a - b$

Inner Product

Definition 1.

$$\boldsymbol{a} = \begin{pmatrix} a_1 \\ a_2 \\ \cdots \\ a_n \end{pmatrix}, \ \boldsymbol{b} = \begin{pmatrix} b_1 \\ b_2 \\ \cdots \\ b_n \end{pmatrix}$$

$$\vec{a} \cdot \vec{b} = a_1 b_1 + a_2 b_2 + \cdots + a_n b_n = \sum_{i=1}^{n} a_i b_i$$

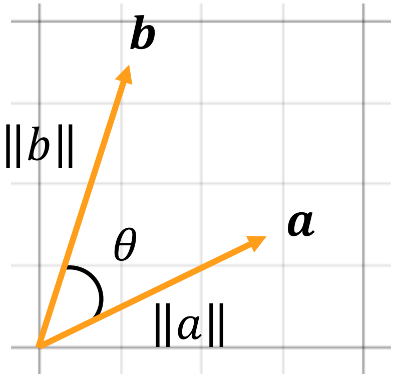
Definition 2.

$$\vec{a} \cdot \vec{b} = \|a\| \, \|b\| \cos\theta$$

$$\theta = \frac{\sqrt{2}}{2}$$

$$\|a\| = \sqrt{2^2 + 1^2} = \sqrt{5}$$

$$\|b\| = \sqrt{1^2 + 3^2} = \sqrt{10}$$



✓ $if \ \boldsymbol{a} = \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix}, \ \boldsymbol{b} = \begin{pmatrix} 4 \\ 5 \\ 6 \end{pmatrix} \ then \ find \ \vec{\boldsymbol{a}} \cdot \vec{\boldsymbol{b}}$

$$\begin{aligned} \vec{\boldsymbol{a}} \cdot \vec{\boldsymbol{b}} &= 1 \times 4 + 2 \times 5 + 3 \times 6 \\ &= 4 + 10 + 18 \\ &= 32 \end{aligned}$$

✓ $\boldsymbol{a} = (2,1), \ \boldsymbol{b} = (1,3), \ \theta = 45^{\circ}$

$$\vec{\boldsymbol{a}} \cdot \vec{\boldsymbol{b}} = 2 \times 1 + 1 \times 3 = 2 + 3 = 5$$

$$\vec{\boldsymbol{a}} \cdot \vec{\boldsymbol{b}} = \sqrt{5} \times \sqrt{10} \times \frac{\sqrt{2}}{2} = 5$$

Inner Product

Exercise 1.

$\boldsymbol{a} = (\sqrt{3}, 1), \boldsymbol{b} = (1, \sqrt{3}), \boldsymbol{c} = (-1, \sqrt{3})$

$Def\ 1.\ \sum_{i=1}^{n} a_i b_i$

$Def\ 2.\ \|\boldsymbol{a}\|\|\boldsymbol{b}\|\ cos\ \theta$

$cos\ 30°$	$\frac{\sqrt{3}}{2}$
$cos\ 60°$	$\frac{1}{2}$
$cos\ 90°$	0

- ✓ $angle\ between\ \boldsymbol{a}\ and\ \boldsymbol{b}\ is\ 30°$
- ✓ $angle\ between\ \boldsymbol{b}\ and\ \boldsymbol{c}\ is\ 60°$
- ✓ $angle\ between\ \boldsymbol{c}\ and\ \boldsymbol{a}\ is\ 90°$

✓ $\|\boldsymbol{a}\| = \|\boldsymbol{b}\| = \|\boldsymbol{c}\| = \sqrt{1^2 + (\sqrt{3})^2} = \sqrt{4} = 2$

1-1) $\vec{\boldsymbol{a}} \cdot \vec{\boldsymbol{b}}$ A1. $\sqrt{3} \times 1 + 1 \times \sqrt{3} = 2\sqrt{3}$

A2. $2 \times 2 \times cos30° = 2 \times 2 \times \frac{\sqrt{3}}{2} = 2\sqrt{3}$

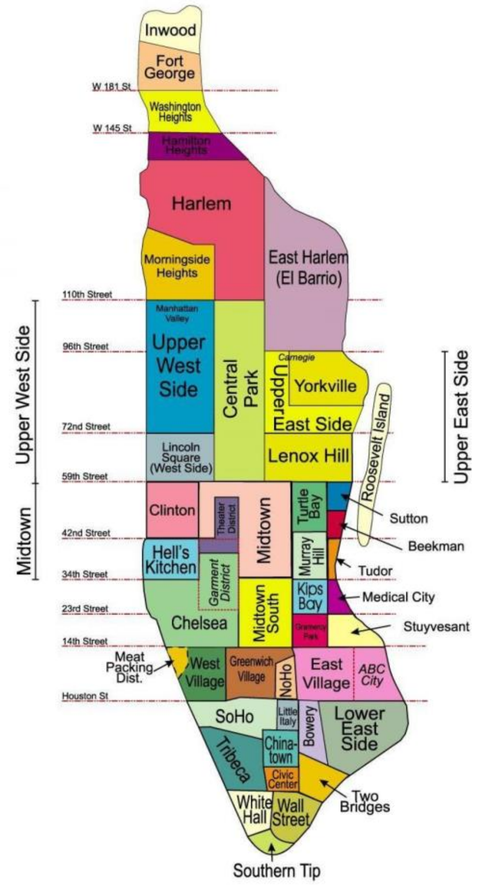
1-2) $\vec{\boldsymbol{b}} \cdot \vec{\boldsymbol{c}}$ A1. $1 \times (-1) + \sqrt{3} \times \sqrt{3} = 2$

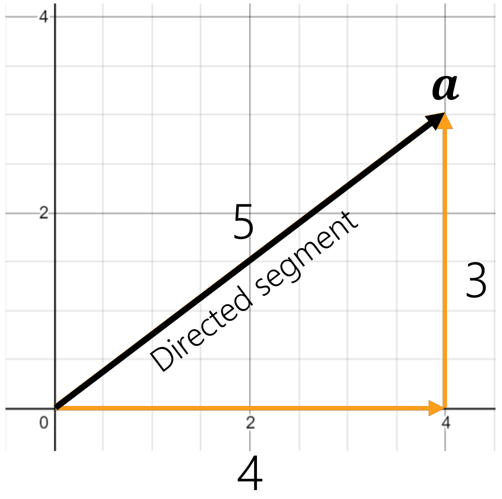
A2. $2 \times 2 \times cos60° = 2 \times 2 \times \frac{1}{2} = 2$

1-3) $\vec{\boldsymbol{c}} \cdot \vec{\boldsymbol{a}}$ A1. $-1 \times \sqrt{3} + \sqrt{3} \times 1 = 0$

A2. $2 \times 2 \times \cos 90° = 2 \times 2 \times 0 = 0$

L1 norm and L2 norm





Directed line Segment of Vector a

$\boldsymbol{a} = (4,3)$

✓ norm

- L1 norm

(※ $\|\boldsymbol{a}\|_1 = |a_1| + |a_2| + \cdots + |a_n| = \sum_{i=1}^{n} |a_i|$)

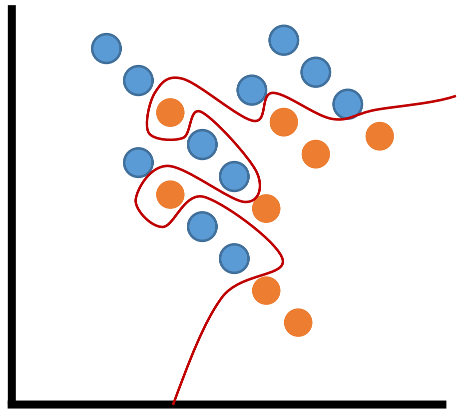
- L2 norm

(※ $\|\boldsymbol{a}\|_2 = \sqrt{\sum_{i=1}^{n} a_i^2} = \sqrt{a_1^2 + a_2^2 + \cdots + a_n^2}$)

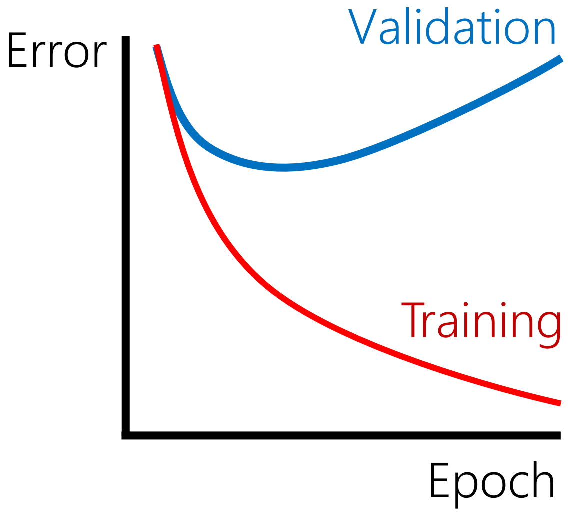
$\|\boldsymbol{a}\|_2 = \sqrt{< \boldsymbol{a}, \boldsymbol{a} >}$

L1 norm and L2 norm

Overfitting



Classification
(Training Dataset)



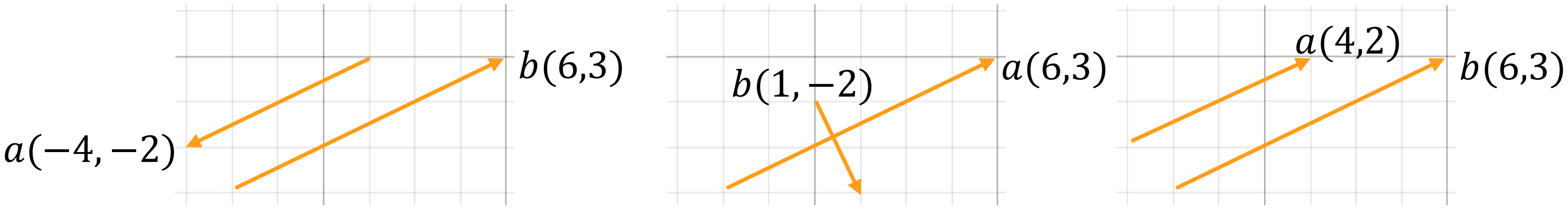
Cosine similarity

?

$\vec{a} \cdot \vec{b} = \|a\|\|b\| \ cos\ \theta$

Cosine similarity
$(-1 \leq \cos(\boldsymbol{a}, \boldsymbol{b}) \leq 1)$

$$\checkmark \cos\theta = \frac{\vec{a} \cdot \vec{b}}{\|a\|\|b\|}$$



$\cos(\boldsymbol{a}, \boldsymbol{b}) = -1$ $\cos(\boldsymbol{a}, \boldsymbol{b}) = 0$ $\cos(\boldsymbol{a}, \boldsymbol{b}) = 1$

- $\vec{a} \cdot \vec{b} = \sum_{i=1}^{n} a_i b_i$
- $\|\boldsymbol{a}\|_2 = \sqrt{\sum_{i=1}^{n} \boldsymbol{a}_i^2} = \sqrt{\boldsymbol{a}_1^2 + \boldsymbol{a}_2^2 + \cdots + \boldsymbol{a}_n^2}$)

$$\checkmark \cos(\boldsymbol{a}, \boldsymbol{b}) = \frac{\vec{a} \cdot \vec{b}}{\|\boldsymbol{a}\|\|\boldsymbol{b}\|} = \frac{\sum_{i=1}^{n} a_i b_i}{\sqrt{\sum_{i=1}^{n} \boldsymbol{a}_i^2}\sqrt{\sum_{i=1}^{n} \boldsymbol{b}_i^2}}$$

Cosine similarity

Exercise 1.

$cos\,(\boldsymbol{a}\,,\boldsymbol{b}) =$

※ $\boldsymbol{a} = (1,2,3) \quad \boldsymbol{b} = (6,5,4)$

- ✓ $\vec{a} \cdot \vec{b} = \mathbf{1 \times 6 + 2 \times 5 + 3 \times 4 = 28}$
- ✓ $||\boldsymbol{a}|| = \sqrt{1^2 + 2^2 + 3^2} = \sqrt{14}$
- ✓ $||\boldsymbol{b}|| = \sqrt{6^2 + 5^2 + 4^2} = \sqrt{77}$

$$cos(\boldsymbol{a},\boldsymbol{b}) = \frac{\sum_{i=1}^{n} a_i b_i}{\sqrt{\sum_{i=1}^{n} a_i^2}\sqrt{\sum_{i=1}^{n} b_i^2}}$$

$$cos(\boldsymbol{a},\boldsymbol{b}) = \frac{\mathbf{28}}{\sqrt{\mathbf{14}}\sqrt{\mathbf{77}}}$$

$$= \frac{\mathbf{4 \cdot 7}}{\sqrt{\mathbf{2 \cdot 7}}\sqrt{\mathbf{7 \cdot 11}}}$$

$$= \frac{\mathbf{2 \cdot \sqrt{2} \cdot \sqrt{2} \cdot 7}}{\sqrt{\mathbf{2}} \cdot \sqrt{\mathbf{7}} \cdot \sqrt{\mathbf{7}} \cdot \sqrt{\mathbf{11}}}$$

$$= \frac{\mathbf{2\sqrt{2}}}{\sqrt{\mathbf{11}}} = \frac{\mathbf{2\sqrt{22}}}{\mathbf{11}}$$

Cosine similarity

Natural Language Processing

A Sentence ← similarity → B Sentence

A : In June, Newton entered Trinity at the suggestion of a teacher.

B : In June, Newton entered the university at the suggestion of a teacher.

✓ $n-gram$

$Term - Document\ Matrix,\ TDM$

	June, Newton	Newton, Entered	Entered, Trinity	Trinity, Suggestion	Suggestion, Teacher	Teacher	
A	1	1	1	1	1	1	6
B	1	1	0	0	1	1	4

$$n-gram\ similarity = \frac{tf(A,B)}{tokens(A)}$$

$$n-gram\ similarity = \frac{4}{6} = 0.66\ldots$$

Cosine similarity

A : In June, Newton entered Trinity at the suggestion of a teacher.

B : In June, Newton entered the university at the suggestion of a teacher.

Word token in statements A,B

	June	Newton	Entered	Trinity	Suggestion	Teacher	University
A	1	1	1	1	1	1	0
B	1	1	1	0	1	1	1

$$\boldsymbol{a} = \{1,1,1,1,1,1,0\}$$

$$\boldsymbol{b} = \{1,1,1,0,1,1,1\}$$

$$\vec{\boldsymbol{a}} \cdot \vec{\boldsymbol{b}} = \sum_{i=1}^{n} \boldsymbol{a}_i \boldsymbol{b}_i$$

$$= (1 \times 1) + (1 \times 1) + (1 \times 1) + (1 \times 0) + (1 \times 1) + (1 \times 1) + (0 \times 1)$$

$$= 1 + 1 + 1 + 0 + 1 + 1 + 0 \quad = 5$$

$$||\boldsymbol{a}||\ ||\boldsymbol{b}|| = \sqrt{\sum_{i=1}^{n} (\boldsymbol{a}_i)^2} \times \sqrt{\sum_{i=1}^{n} (\boldsymbol{b}_i)^2}$$

$$= \sqrt{1^2 + 1^2 + 1^2 + 1^2 + 1^2 + 1^2 + 0^2} * \sqrt{1^2 + 1^2 + 1^2 + 0^2 + 1^2 + 1^2 + 1^2}$$

$$= \sqrt{6} \times \sqrt{6} = \sqrt{36} \quad = 6$$

Cosine similarity

A : In June, Newton entered Trinity at the suggestion of a teacher.

B : In June, Newton entered the university at the suggestion of a teacher.

Cosine similarity = 5/6 = 0.833..

n-gram similarity = 4/6 = 0.66...

Artificial Intelligence

Basic Mathematics for Artificial Intelligence : Part 3

5. Matrix Operations
6. Linear Regression Model

Matrix Operations

3 rows by 4 columns

$$A = \begin{array}{c} \\ \\ \text{1 row} \\ \text{2 row} \\ \text{3 row} \end{array} \begin{array}{cccc} 1 & 2 & 3 & 4 \\ \text{col} & \text{col} & \text{col} & \text{col} \\ \end{array}$$

$$A = \begin{array}{r} \text{1 row} \\ \text{2 row} \\ \text{3 row} \end{array} \begin{bmatrix} 3 & 4 & 0 & 10 \\ 0 & 1 & 0 & -3 \\ -1 & 1 & 9 & 0 \end{bmatrix}$$

(3, 2)

Exercise 1.

$$A = \begin{pmatrix} 0 & 7 & 2 & 2 \\ 1 & 2 & 6 & 1 \\ 5 & 3 & 3 & 4 \end{pmatrix}, \qquad B = \begin{pmatrix} 2 & 6 & 7 & -1 \\ 1 & 8 & 3 & 5 \\ 0 & -1 & 6 & 11 \end{pmatrix}$$

$$\mathrm{A} + \mathrm{B} = \begin{pmatrix} 0+2 & 7+6 & 2+7 & 2+(-1) \\ 1+1 & 2+8 & 6+3 & 1+5 \\ 5+0 & 3+(-1) & 3+6 & 4+11 \end{pmatrix}$$

$$= \begin{pmatrix} 2 & 13 & 9 & 1 \\ 2 & 10 & 9 & 6 \\ 5 & 2 & 9 & 15 \end{pmatrix}$$

$$\mathrm{A} - \mathrm{B} = \begin{pmatrix} 0-2 & 7-6 & 2-7 & 2-(-1) \\ 1-1 & 2-8 & 6-3 & 1-5 \\ 5-0 & 3-(-1) & 3-6 & 4-11 \end{pmatrix}$$

$$= \begin{pmatrix} -2 & 1 & -5 & 3 \\ 0 & -6 & 3 & -4 \\ 5 & 4 & -3 & -7 \end{pmatrix}$$

Matrix Operations

Definition 1.

$$\boldsymbol{a} = (a_1, a_2, \ldots, a_n),\ \ \boldsymbol{b} = \begin{pmatrix} b_1 \\ b_2 \\ \cdots \\ b_n \end{pmatrix}$$

$$\boldsymbol{ab} = \vec{a} \cdot \vec{b} = a_1 b_1 + a_2 b_2 + \cdots + a_n b_n = \ \sum_{i=1}^{n} a_i b_i$$

Exercise 1.

$$\boldsymbol{a} = (-1,2) \qquad \boldsymbol{b} = \begin{pmatrix} 3 \\ 2 \end{pmatrix}$$

$1 \times 2\ matrix \qquad 2 \times 1\ matrix$

$$\boldsymbol{ab} = -1 \times 3 + 2 \times 2 = 1$$

Matrix Operations

$$A = \begin{pmatrix} -1 & 2 \\ 1 & 1 \end{pmatrix}, \quad \boldsymbol{b_1} = \begin{pmatrix} 3 \\ 2 \end{pmatrix}$$

Definition 1.

$$\boldsymbol{a} = (a_1, a_2, \dots, a_n), \quad \boldsymbol{b} = \begin{pmatrix} b_1 \\ b_2 \\ b_3 \end{pmatrix}$$

$$\boldsymbol{ab} = \vec{a} \cdot \vec{b} = a_1 b_1 + a_2 b_2 + \cdots + a_n b_n = \sum_{i=1}^{n} a_i b_i$$

$$\boldsymbol{Ab_1} = \begin{pmatrix} -1 & 2 \\ 1 & 1 \end{pmatrix} \begin{pmatrix} 3 \\ 2 \end{pmatrix} = \begin{pmatrix} -1 \times 3 + 2 \times 2 \\ 1 \times 3 + 1 \times 2 \end{pmatrix} = \begin{pmatrix} 1 \\ 5 \end{pmatrix}$$

$$\boldsymbol{A'b_1} = \begin{pmatrix} -1 & 2 \\ 1 & 1 \\ 3 & 0 \end{pmatrix} \begin{pmatrix} 3 \\ 2 \end{pmatrix} = \begin{pmatrix} -1 \times 3 + 2 \times 2 \\ 1 \times 3 + 1 \times 2 \\ 3 \times 3 + 0 \times 2 \end{pmatrix} = \begin{pmatrix} 1 \\ 5 \\ 9 \end{pmatrix}$$

Matrix Operations

$$\underset{(m \times n)}{A} \qquad \underset{(n \times l)}{B} \qquad \underset{\substack{(? \times ?) \\ (m \times l)}}{AB}$$

$$\begin{array}{cc} & \begin{array}{cccc} 1d & 2d & \dots & nd \end{array} \\ \begin{array}{c} m_1 \\ m_2 \\ \cdots \\ m_n \end{array} & \begin{pmatrix} 1 & 5 & \cdots & 1 \\ 3 & 7 & \cdots & 4 \\ \vdots & \vdots & \vdots & \vdots \\ 4 & 8 & \dots & 8 \end{pmatrix} \end{array} \qquad \begin{array}{cc} & \begin{array}{cccc} 1l & 2l & \dots & nl \end{array} \\ \begin{array}{c} n_1 \\ n_2 \\ \cdots \\ n_n \end{array} & \begin{pmatrix} 2 & 1 & \cdots & 4 \\ 3 & 4 & \cdots & 2 \\ \vdots & \vdots & \vdots & \vdots \\ 8 & 1 & \dots & 5 \end{pmatrix} \end{array}$$

Ex) $\cancel{\underset{(2 \times 2)}{\begin{pmatrix} 6 & 3 \\ 8 & 10 \end{pmatrix}} \underset{(3 \times 1)}{\begin{pmatrix} 2 \\ 5 \\ 2 \end{pmatrix}}}$

Matrix Operations

$$A = \begin{pmatrix} \boldsymbol{a_1} \\ \boldsymbol{a_2} \\ \cdots \\ \boldsymbol{a_m} \end{pmatrix} = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \vdots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{pmatrix}, \quad B = (\boldsymbol{b_1}, \boldsymbol{b_2}, \ldots, \boldsymbol{b_l}) = \begin{pmatrix} b_{11} & b_{12} & \cdots & b_{1l} \\ b_{21} & b_{22} & \cdots & b_{2l} \\ \vdots & \vdots & \vdots & \vdots \\ b_{n1} & b_{n2} & \cdots & b_{nl} \end{pmatrix}$$

$$AB = \begin{pmatrix} \boldsymbol{a_1} \\ \boldsymbol{a_2} \\ \cdots \\ \boldsymbol{a_m} \end{pmatrix} (\boldsymbol{b_1}, \boldsymbol{b_2}, \ldots, \boldsymbol{b_l}) = \begin{pmatrix} \boldsymbol{a_1 b_1} & \boldsymbol{a_1 b_2} & \cdots & \boldsymbol{a_1 b_l} \\ \boldsymbol{a_2 b_1} & \boldsymbol{a_2 b_2} & \cdots & \boldsymbol{a_2 b_l} \\ \vdots & \vdots & \vdots & \vdots \\ \boldsymbol{a_m b_1} & \boldsymbol{a_m b_2} & \cdots & \boldsymbol{a_m b_l} \end{pmatrix}$$

$$\boldsymbol{a_p b_q} = \sum_{i=1}^{n} a_{pi} b_{qi} = a_{p1} b_{1q} + a_{p2} b_{2q} + \cdots + a_{pn} b_{nq}$$

Matrix Operations

Exercise 1.

$$AB = \begin{pmatrix} 1 & 4 \\ 0 & 0 \\ 8 & 0 \end{pmatrix}\begin{pmatrix} 0 & 1 & 0 \\ 0 & -3 & 11 \end{pmatrix}$$

(3×2) (2×3)

A. $$AB = \begin{pmatrix} 1\times0+4\times0 & 1\times1+4\times(-3) & 1\times0+4\times11 \\ =0 & =-11 & =44 \\ 0\times0+0\times0 & 0\times1+0\times(-3) & 0\times0+0\times11 \\ =0 & =0 & =0 \\ 8\times0+0\times0 & 8\times1+0\times(-3) & 8\times0+0\times11 \\ =0 & =8 & =0 \end{pmatrix}$$

(3×3)

Properties

- ✓ $(m\times n)(n\times l) = (m\times l)$
- ✓ $AB \neq BA$

Ex) $$\begin{pmatrix} 1 & 4 \\ 0 & 0 \\ 8 & 0 \end{pmatrix}\begin{pmatrix} 0 & 1 & 0 \\ 0 & -3 & 11 \end{pmatrix} \neq \begin{pmatrix} 0 & 1 & 0 \\ 0 & -3 & 11 \end{pmatrix}\begin{pmatrix} 1 & 4 \\ 0 & 0 \\ 8 & 0 \end{pmatrix}$$

(3×2) (2×3) (2×3) (3×2)

Matrix Operations

Transpose of A Matrix

$$\boldsymbol{a} = \begin{pmatrix} 2 \\ 5 \\ 2 \end{pmatrix}, \qquad \boldsymbol{a}^T = (2,5,2)$$

$$A = \begin{pmatrix} 2 & 1 \\ 5 & 3 \\ 2 & 8 \end{pmatrix}, \qquad A^T = \begin{pmatrix} 2 & 5 & 2 \\ 1 & 3 & 8 \end{pmatrix}$$

$$\boldsymbol{a} = \begin{pmatrix} 2 \\ 5 \\ 2 \end{pmatrix}, \boldsymbol{b} = \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix}$$

(3×1) (3×1)

$$\boldsymbol{a}^T\boldsymbol{b} = (2,5,2)\begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix}$$

(1×3) (3×1)

$$= (2 \times 1 + 5 \times 2 + 2 \times 3) = (18)$$

Linear Regression Model

Linear Regression Model

Housing Dataset

AI model
(Housing Price Prediction Model)

Dependent variable

- ✓ The Subway Station
- ✓ Number of rooms in a house
- ✓ Size of rooms in a house
- ✓ …

Independent variable

Linear Regression Model

Size 40
Housing Price 40,000$

Size 1 × Housing Price 1,000$ →

Size 80
Housing Price 80,000$

- ✓ The Subway Station
- ✓ Number of rooms in a house
- ✓ Size of rooms in a house
- ✓ …

Data bias

⇨ Housing Price Prediction

Linear Regression Model

Boston Housing Dataset

Dataset

Training Data	Test Data

- ✓ (506 × 14)
- ✓ Dependent variable : MEDV

CRIM	per capita crime rate by town.	Count
ZN	proportion of residential land zoned for lots over 25,000 sq.ft.	%
INDUS	proportion of non-retail business acres per town.	%
CHAS	Charles River dummy variable (= 1 if tract bounds river; 0 otherwise).	1, 0
NOX	nitrogen oxides concentration (parts per 10 million).	%
RM	average number of rooms per dwelling.	Count
AGE	proportion of owner-occupied units built prior to 1940.	%
DIS	weighted mean of distances to five Boston employment centres.	-
RAD	index of accessibility to radial highways.	1 ~ 24
TAX	full-value property-tax rate per ₩$10,000.	$
PTRATIO	pupil-teacher ratio by town.	Count
B	1000(Bk - 0.63)^2 where Bk is the proportion of blacks by town.	-
LSTAT	lower status of the population (percent).	%
MEDV	median value of owner-occupied homes in ₩$1000s.	1,000 %

Linear Regression Model

$$y = \underline{w_0} + \underline{w_1}\overline{x_1} + \underline{w_2}\overline{x_2} + \cdots + \underline{w_l}\overline{x_l}$$

Independent variable

Weight

$$y = w_0 + \sum_{k=1}^{l} w_k x_k$$

Optimal Weight

Linear Regression

Definition.

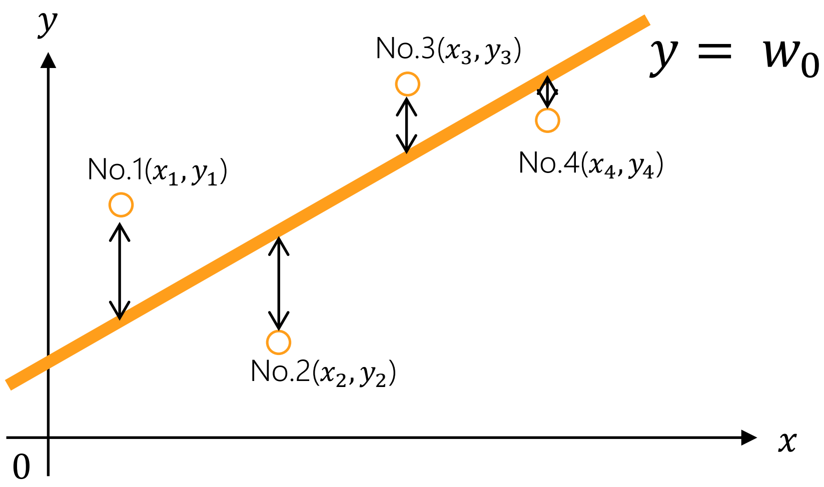
$$\begin{bmatrix} y_1 \\ y_2 \\ \cdots \\ y_n \end{bmatrix} = \begin{bmatrix} 1 & x_{11} & x_{12} & & x_{1l} \\ 1 & x_{21} & x_{22} & \cdots & x_{2l} \\ \cdots & \cdots & \cdots & & \cdots \\ 1 & x_{n1} & x_{n2} & & x_{nl} \end{bmatrix} \begin{bmatrix} w_0 \\ w_1 \\ \cdots \\ w_l \end{bmatrix}$$

Definition.

$$Y = XW$$

Linear Regression Model

Least Square Method



$$y = w_0 + w_1 x$$

Label ← Error → Prediction

Ex)

Number	x(Distance)	y(Price)
1	0.5	8.7
2	0.8	7.5
3	1.1	7.1
4	1.5	6.8

$$D = \sum_{l=1}^{4} |y_l - (w_0 + w_1 x_l)| \Rightarrow D = \sum_{l=1}^{4} \{y_l - (w_0 + w_1 x_l)\}^2$$

Linear Regression Model

Least Square Method

Number	x(Distance)	y(Price)
1	0.5	8.7
2	0.8	7.5
3	1.1	7.1
4	1.5	6.8

$$D = \sum_{l=1}^{4} \{y_l - (w_0 + w_1 x_l)\}^2$$

$$D = \{8.7 - (w_0 + 0.5w_1)\}^2 + \{7.5 - (w_0 + 0.8w_1)\}^2 + \{7.1 - (w_0 + 1.1w_1)\}^2 + \{6.8 - (w_0 + 1.5w_1)\}^2$$

$$= 4w_0^2 + 4.35w_1^2 + 7.8w_0w_1 - 60.2w_0 - 56.72w_1 + 228.59$$

$$\frac{\partial D}{\partial w_0} = 8w_0 + 7.8w_1 - 60.2 = 0$$

$$\frac{\partial D}{\partial w_1} = 8.7w_1 + 7.8w_0 - 56.72 = 0$$

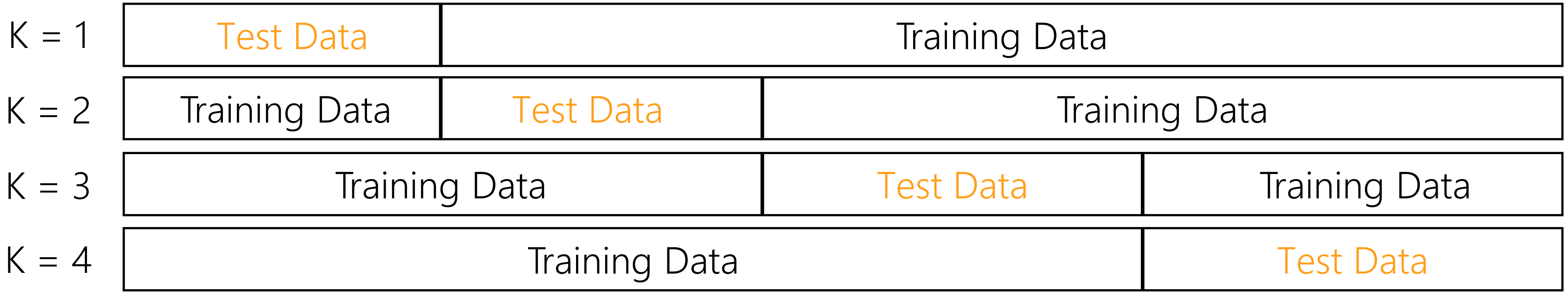
$$w_0 \fallingdotseq 9.2836 \quad w_1 \fallingdotseq -1.8037$$

$$y = -1.8037x + 9.2836$$

Linear Regression Model

Holdout Cross Validation

K − Fold Cross Validation



Linear Regression Model

Training, Validation and Test Sets

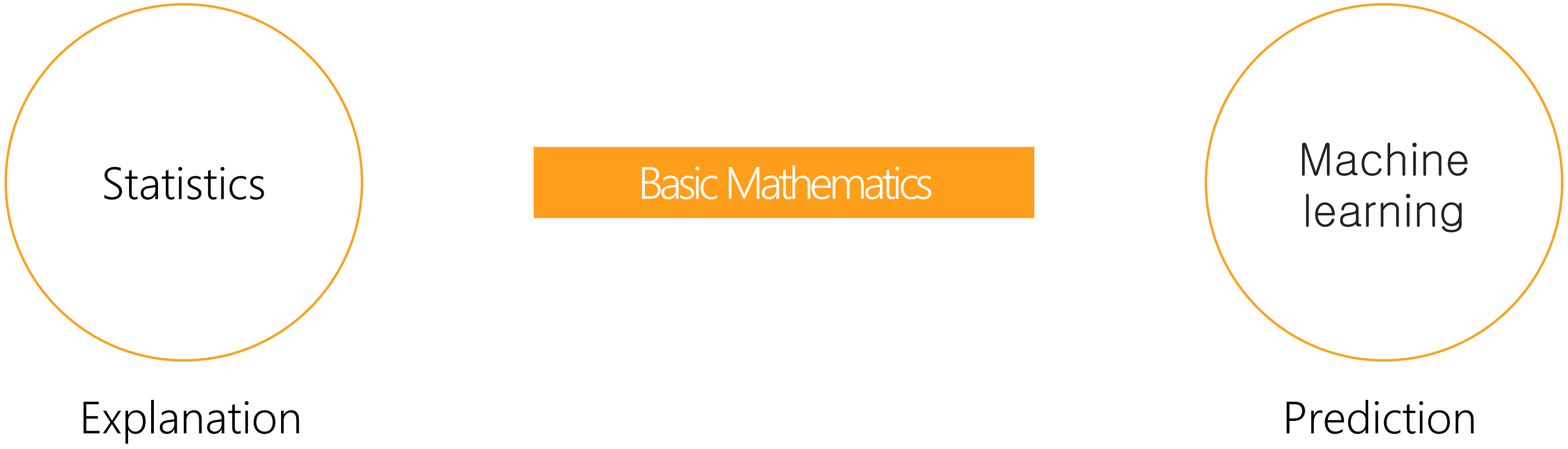
Training Data (75%)	Test Data (25%)

Training Data (60%)	Validation data (20%)	Test Data (20%)

Way to evaluate a machine learning model's performance

$$MSE = \frac{1}{n} \sum_{i=1}^{n} (y_i - \hat{y}_i)^2$$

Linear Regression Model



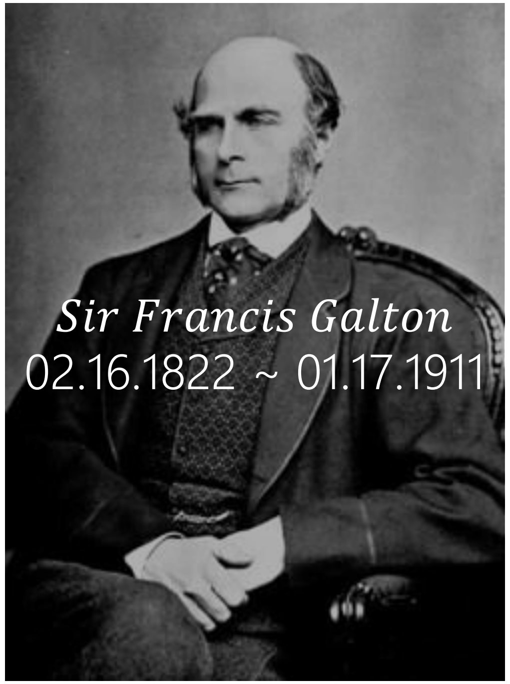
Linear Regression Model

Tensorflow

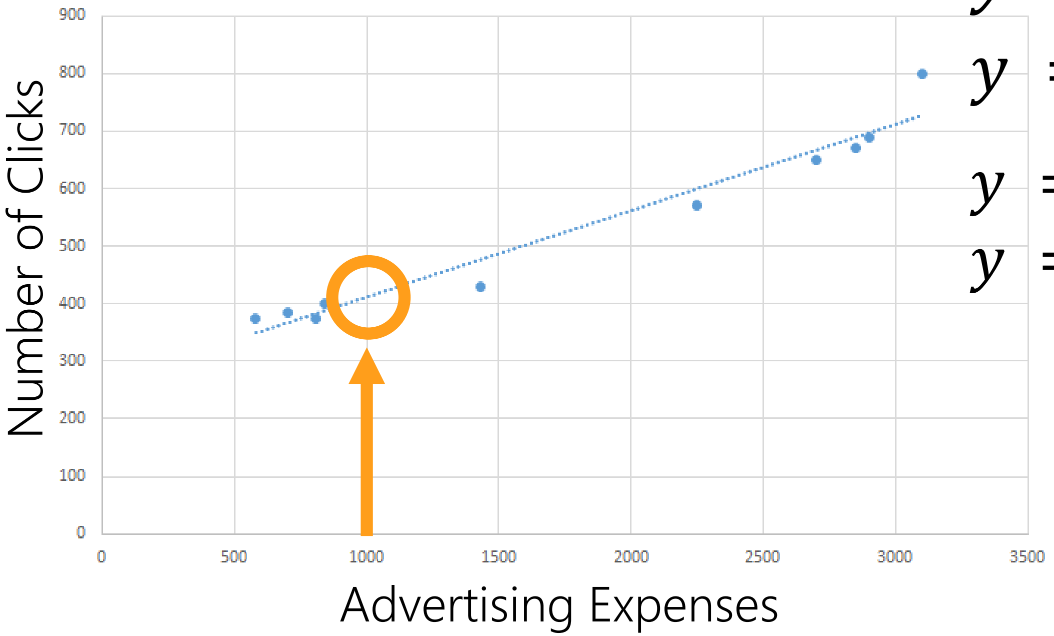
Basic Mathematics

Regression

Parent Child Height Correlation



Linear Regression Model



$$y = ax + b$$

$$y = \theta_0 + \theta_1 x$$

$if\ \theta_0 = 1,\ \theta_1 = 2$

$$y = 1 + 2x$$

$$y = 1 + 2 \times 1000(Advertising\ Expenses)$$

$$y = 2001(Number\ of\ Clicks)$$

※ $\theta(theta) : Unknown\ Value$

Q. Number of Clicks when Advertising Expenses 2,000

A. 500 ~

Linear Regression Model

Optimization Problem

$$y = f_\theta(x) \Rightarrow y - f_\theta(x) = 0$$

$$E(\theta) = \frac{1}{2}\sum_{i=1}^{n}\left(y^{(i)} - f_\theta(x^{(i)})\right)^2$$

$Advertising\ Expenses\ (x)$	$Number\ of\ Clicks\ (y)$	$if\ \theta_0 = 1, \theta_0 = 2\ then\ \hat{y}$
580	374	1161
700	385	1401
810	375	1621
840	401	1681

minimize

$x^{(1)} = 580 \quad y^{(1)} = 374$

$x^{(2)} = 700 \quad y^{(2)} = 385$

Linear Regression Model

$$E(\theta) = \frac{1}{2}\sum_{i=1}^{n}\left(y^{(i)} - f_\theta\left(x^{(i)}\right)\right)^2$$

$$y^{(i)} - f_\theta\left(x^{(i)}\right) = 1$$

$$y^{(i)} - f_\theta\left(x^{(i)}\right) = 10$$

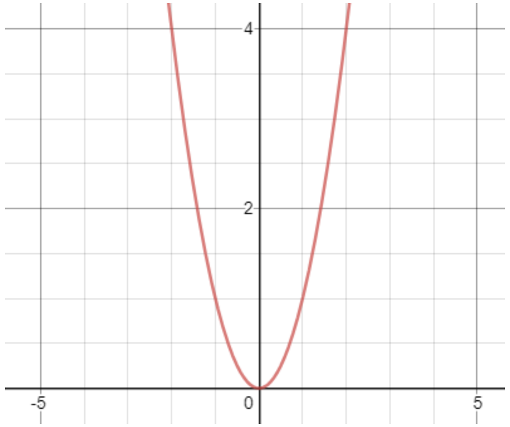
×10

$$\left(y^{(i)} - f_\theta\left(x^{(i)}\right)\right)^2 = 1^2$$

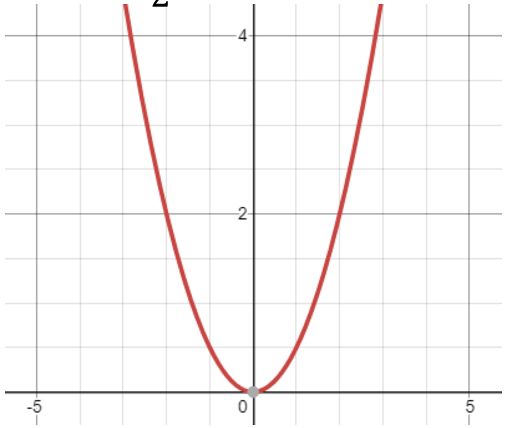
$$\left(y^{(i)} - f_\theta\left(x^{(i)}\right)\right)^2 = 10^2$$

×100

$f_\theta(x) = x^2$



$f_\theta(x) = \frac{1}{2} \times x^2$



Linear Regression Model

$$E(\theta) = \frac{1}{2}\sum_{i=1}^{4}\left(y^{(i)} - f_\theta\left(x^{(i)}\right)\right)^2$$

$= \frac{1}{2}$ (374−1161)^2 +(385−1401)^2 + (375−1621)^2+(401−1684)^2

$Advertising\ Expenses$ (x)	$Number\ of\ Clicks$ (y)	$if\ \theta_0 = 1, \theta_0 = 2$ $then\ \hat{y}$
580	374	1161
700	385	1401
810	375	1621
840	401	1681

$$E(\theta) = \frac{1}{2}\sum_{i=4}^{n}\left(y^i - f_\theta\left(x^i\right)\right)^2$$

$$= \frac{1}{2}\left((374 \ - 1161)^2 + (385 \ - 1401)^2 + (375 \ - 1621)^2 + (401 \ - 1681)^2\right)$$

$$= \frac{1}{2}(619369 + \ 1032256 + 1552516 + 1638400)$$

$$= 2421270.5$$

Artificial Intelligence

Basic Mathematics for Artificial Intelligence : Part 3

6. Linear Regression Model

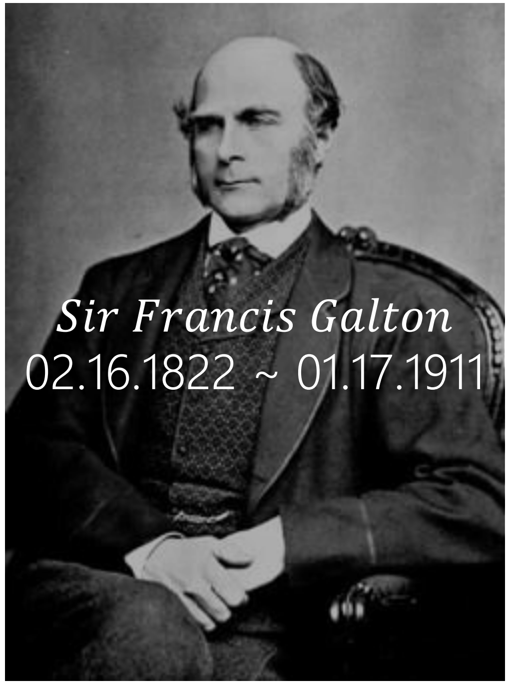
Linear Regression Model

Tensorflow

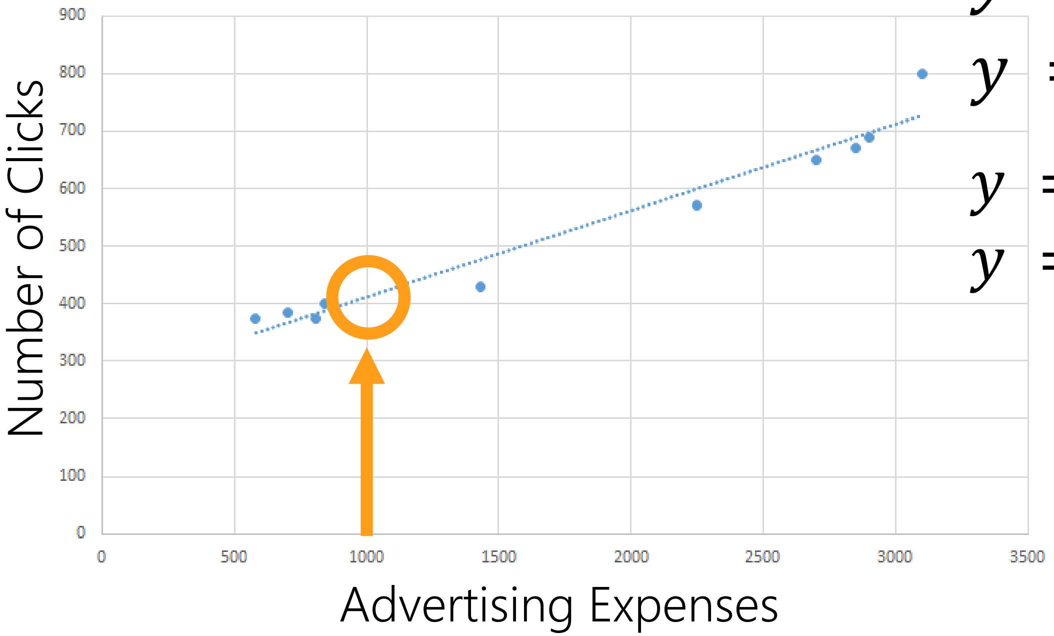
Basic Mathematics

Regression

Parent Child Height Correlation



Linear Regression Model



$$y = ax + b$$

$$y = \theta_0 + \theta_1 x$$

$$y = 1 + 2 \times 1000(Advertising\ Expenses)$$

$$y = 2001(Number\ of\ Clicks)$$

$if\ \theta_0 = 1,\ \theta_1 = 2$

$$y = 1 + 2x$$

※ $\theta(theta) : Unknown\ Value$

Q. Number of Clicks when Advertising Expenses 2,000

A. 500 ~

Linear Regression Model

$$y = f_\theta(x) \Rightarrow y - f_\theta(x) = 0$$

Optimization Problem

$$E(\theta) = \frac{1}{2}\sum_{i=1}^{n}\left(y^{(i)} - f_\theta(x^{(i)})\right)^2$$

Advertising Expenses (x)	*Number of Clicks* (y)	$if\ \theta_0 = 1, \theta_0 = 2$ $then\ \hat{y}$
580	374	1161
700	385	1401
810	375	1621
840	401	1681

minimize

$x^{(1)} = 580 \quad y^{(1)} = 374$

$x^{(2)} = 700 \quad y^{(2)} = 385$

Linear Regression Model

$$E(\theta) = \frac{1}{2}\sum_{i=1}^{n}\left(y^{(i)} - f_\theta\left(x^{(i)}\right)\right)^2$$

$$y^{(i)} - f_\theta\left(x^{(i)}\right) = 1$$

$$y^{(i)} - f_\theta\left(x^{(i)}\right) = 10$$

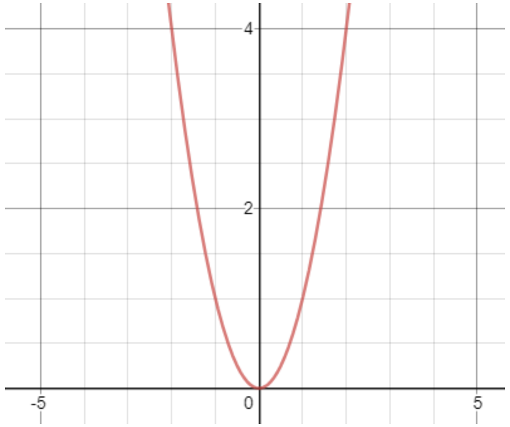
×10

$$\left(y^{(i)} - f_\theta\left(x^{(i)}\right)\right)^2 = 1^2$$

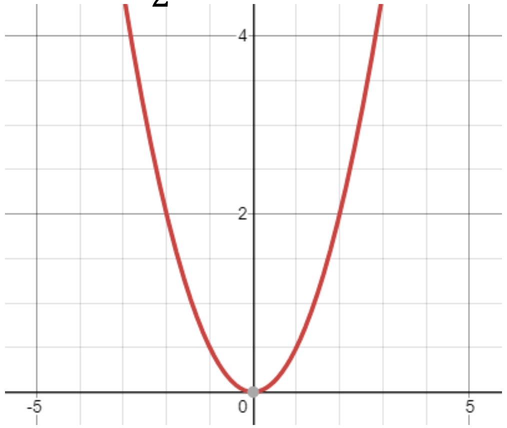
$$\left(y^{(i)} - f_\theta\left(x^{(i)}\right)\right)^2 = 10^2$$

×100

$$f_\theta(x) = x^2$$



$$f_\theta(x) = \frac{1}{2} \times x^2$$



Linear Regression Model

$$E(\theta) = \frac{1}{2}\sum_{i=1}^{4}\left(y^{(i)} - f_{\theta}\left(x^{(i)}\right)\right)^{2}$$

$$= \frac{1}{2}((374 - 1161)^2 + (385 - 1401)^2 + (375 - 1621)^2 + (401 - 1681)^2)$$

$$= \frac{1}{2}(619369 + 1032256 + 1552516 + 1638400)$$

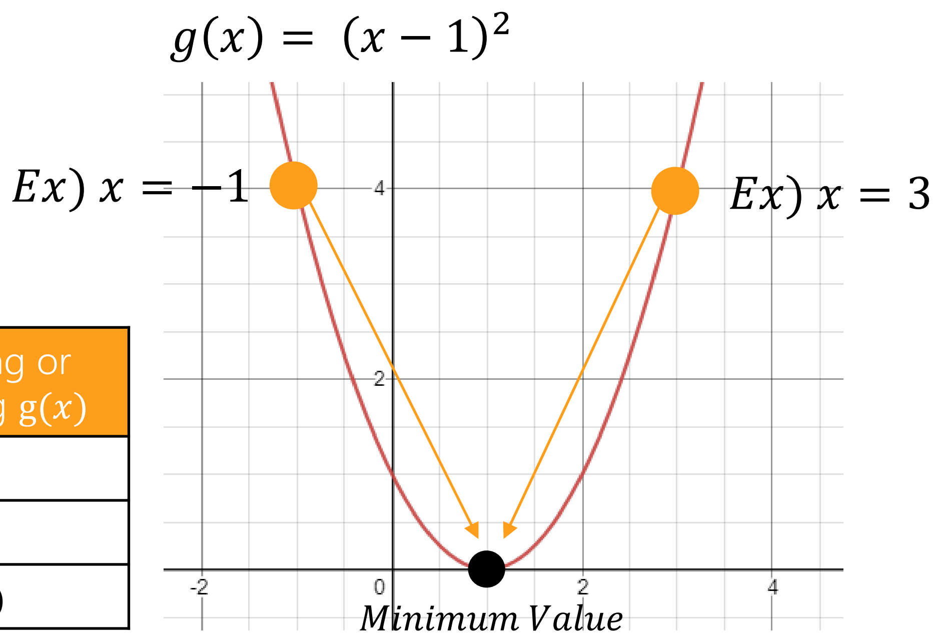
$$= 2421270.5$$

$Advertising\ Expenses$ (x)	$Number\ of\ Clicks$ (y)	$if\ \theta_0 = 1, \theta_0 = 2$ $then\ \hat{y}$
580	374	1161
700	385	1401
810	375	1621
840	401	1681

Linear Regression Model

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

Range of x	The sign of $\frac{d}{dx}(g)$	Increasing or Decreasing g(x)
$x < 1$	$-$	↘
$x = 1$	0	$-$
$x > 1$	$+$	↗ (↙)



Linear Regression Model

$$x := x - \eta \frac{d}{dx} g(x)$$

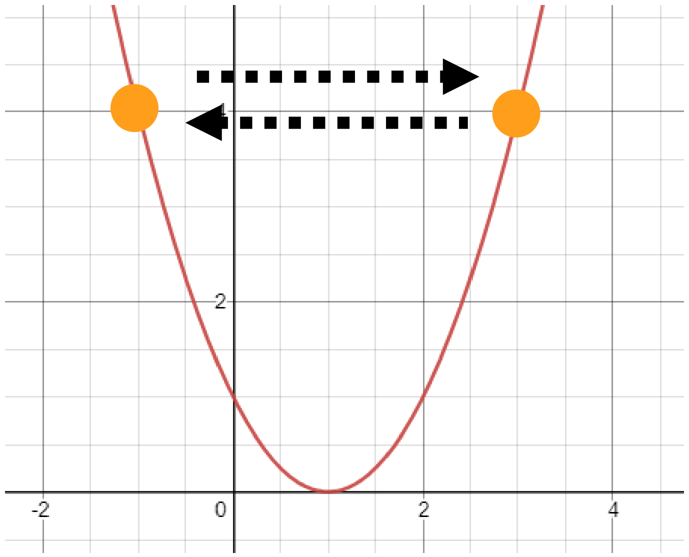
※ $\eta(ETA) : Learning\ Rate$

Ex) $if\ \eta = 1, x = 3\ then\ motion\ of\ x$

$x := 3 - 1(2x - 2)$

$x := 3 - 1(2 \times 3 - 2) \quad = 3 - 4 \quad = -1$
$x := -1 - 1(2 \times -1 - 2) = -1 + 4 = 3$
$x := 3 - 1(2 \times 3 - 2) \quad = 3 - 4 \quad = -1$

※ $:= \quad is\ defined\ as$



$g(x) = (x - 1)^2$

Linear Regression Model

$$x := x - \eta \frac{d}{dx} g(x)$$

※ $\eta(ETA)$: $Learning\ Rate$

Ex) $if\ \eta = 0.1, x = 3\ then\ motion\ of\ x$

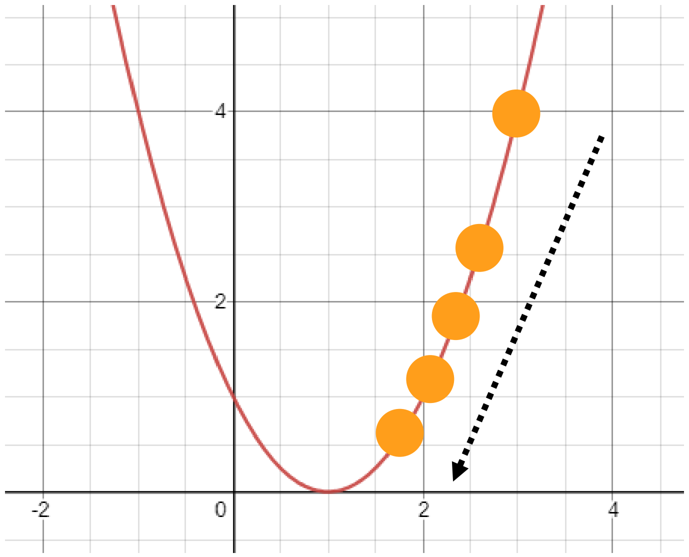
$x \coloneqq 3 - 0.1(2x - 2)$

$x \coloneqq 3 - 0.1(2 \times 3 - 2) \quad = 3 - 0.4 \quad = 2.6$

$x \coloneqq 2.6 - 0.1(2 \times 2.6 - 2) = 2.6 - 0.3 = 2.3$

$x \coloneqq 2.3 - 0.1(2 \times 2.3 - 2) = 2.3 - 0.2 = 2.1$

$x \coloneqq 2.1 - 0.1(2 \times 2.1 - 2) = 2.1 - 0.2 = 1.9$



$g(x) = (x-1)^2$

Linear Regression Model

$$E(\theta) = \frac{1}{2}\sum_{i=1}^{n}\left(y^{(i)} - f_\theta\left(x^{(i)}\right)\right)^2$$

$$x := x - \eta\frac{d}{dx}g(x)$$

$$\theta_0 := \theta_0 - \eta\frac{\partial E}{\partial \theta_0}g(x)$$

$$\theta_1 := \theta_1 - \eta\frac{\partial E}{\partial \theta_1}g(x)$$

※ $f_\theta(x) = \theta_0 + \theta_1 x$

$u = E(\theta)$

$v = f_\theta(x)$

$$\frac{\partial u}{\partial \theta_0} = \boxed{\frac{\partial u}{\partial v}} \cdot \frac{\partial v}{\partial \theta_0}$$

$$\frac{\partial u}{\partial v} = \frac{\partial}{\partial v}E(\theta)$$
$$= \frac{\partial}{\partial v}\left(\frac{1}{2}\sum_{i=1}^{n}\left(y^{(i)} - v\right)^2\right) = \frac{1}{2}\sum_{i=1}^{n}\left(\frac{\partial}{\partial v}\left(y^{(i)2} - 2y^{(i)}v + v^2\right)\right)$$
$$= \frac{1}{2}\sum_{i=1}^{n}\left(-2y^{(i)} + 2v\right) = \sum_{i=1}^{n}\left(v - y^{(i)}\right)$$

Linear Regression Model

$$E(\theta) = \frac{1}{2}\sum_{i=1}^{n}\left(y^{(i)} - f_\theta\left(x^{(i)}\right)\right)^2 \qquad x := x - \eta\frac{d}{dx}g(x)$$

$$\theta_0 := \theta_0 - \eta\frac{\partial E}{\partial \theta_0}g(x)$$

$$\theta_1 := \theta_1 - \eta\frac{\partial E}{\partial \theta_1}g(x)$$

※ $f_\theta(x) = \theta_0 + \theta_1 x$

$$u = E(\theta)$$

$$v = f_\theta(x)$$

$$\frac{\partial u}{\partial \theta_0} = \frac{\partial u}{\partial v}\cdot\boxed{\frac{\partial v}{\partial \theta_0}}$$

$$\frac{\partial v}{\partial \theta_0} = \frac{\partial}{\partial \theta_0}f_\theta(x)$$

$$= \frac{\partial}{\partial \theta_0}(\theta_0 + \theta_1 x) = 1$$

Linear Regression Model

$$E(\theta) = \frac{1}{2}\sum_{i=1}^{n}\left(y^{(i)} - f_\theta\left(x^{(i)}\right)\right)^2 \qquad x := x - \eta\frac{d}{dx}g(x)$$

$$\theta_0 := \theta_0 - \eta\frac{\partial E}{\partial \theta_0}g(x)$$

$$\theta_1 := \theta_1 - \eta\frac{\partial E}{\partial \theta_1}g(x)$$

※ $f_\theta(x) = \theta_0 + \theta_1 x$

$$u = E(\theta)$$

$$v = f_\theta(x)$$

$$\frac{\partial u}{\partial \theta_0} = \frac{\partial u}{\partial v}\cdot\frac{\partial v}{\partial \theta_0}$$

$$\frac{\partial u}{\partial \theta_0} = \sum_{i=1}^{n}(v - y^{(i)}) \times 1$$

$$= \sum_{i=1}^{n}(f_\theta(x^{(i)}) - y^{(i)})$$

Linear Regression Model

$$E(\theta) = \frac{1}{2}\sum_{i=1}^{n}\left(y^{(i)} - f_\theta\left(x^{(i)}\right)\right)^2 \qquad x := x - \eta \frac{d}{dx} g(x)$$

$$\theta_0 := \theta_0 - \eta \frac{\partial E}{\partial \theta_0} g(x)$$

$$\theta_1 := \theta_1 - \eta \frac{\partial E}{\partial \theta_1} g(x)$$

※ $f_\theta(x) = \theta_0 + \theta_1 x$

$u = E(\theta)$

$v = f_\theta(x)$

$$\Rightarrow \frac{\partial u}{\partial \theta_1} = \frac{\partial u}{\partial v} \cdot \boxed{\frac{\partial v}{\partial \theta_1}} \qquad \frac{\partial v}{\partial \theta_1} = \frac{\partial}{\partial \theta_1} f_\theta(x) = \frac{\partial}{\partial \theta_1}(\theta_0 + \theta_1 x) = x$$

$$\Rightarrow \frac{\partial u}{\partial \theta_1} = \frac{\partial u}{\partial v} \cdot \frac{\partial v}{\partial \theta_1} \Rightarrow \sum_{i=1}^{n}\left(f_\theta(x^{(i)}) - y^{(i)}\right)x^{(i)}$$

Linear Regression Model

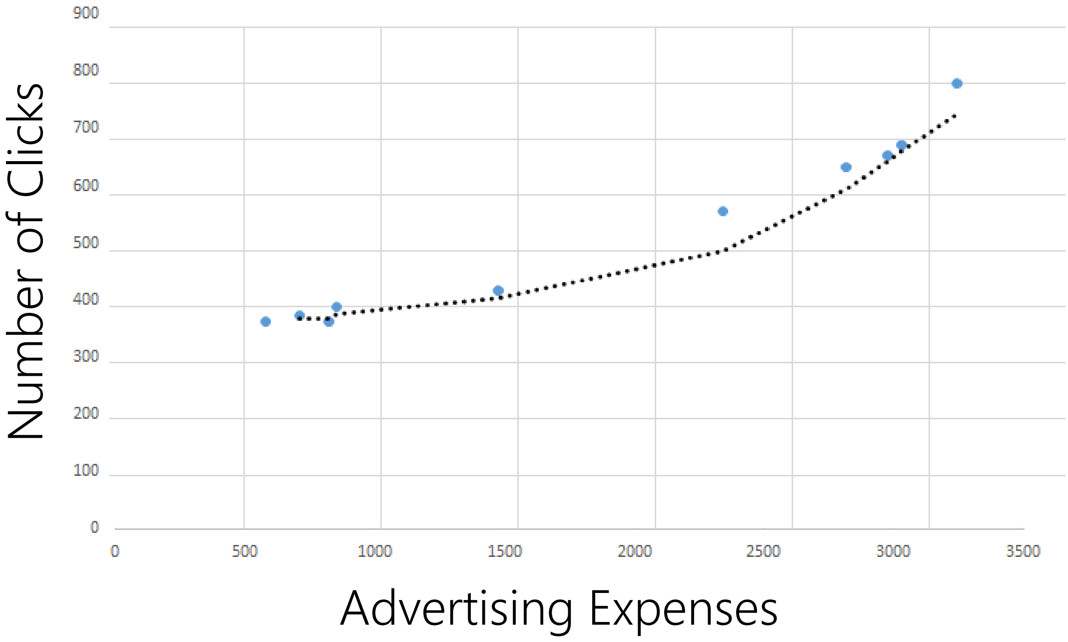
$$x := x - \eta \frac{d}{dx} g(x)$$

$$\theta_0 := \theta_0 - \eta \frac{\partial E}{\partial \theta_0} g(x) \rightarrow \theta_0 := \theta_0 - \eta \sum_{i=1}^{n} (f_\theta(x^{(i)}) - y^{(i)})$$

$$\theta_1 := \theta_1 - \eta \frac{\partial E}{\partial \theta_1} g(x) \rightarrow \theta_1 := \theta_1 - \eta \sum_{i=1}^{n} (f_\theta(x^{(i)}) - y^{(i)}) x^{(i)}$$

Linear Regression Model

$$f_\theta(x) = \theta_0 + \theta_1 x + \theta_2 x^2$$



$$\frac{\partial u}{\partial \theta_2} = \frac{\partial u}{\partial v} \cdot \frac{\partial v}{\partial \theta_2}$$

$$\frac{\partial u}{\partial v} = \sum_{i=1}^{n} \left(f_\theta(x^{(i)}) - y^{(i)}\right)$$

$$\frac{\partial v}{\partial \theta_2} = \frac{\partial}{\partial \theta_2} f_\theta(x) = \frac{\partial}{\partial \theta_2}(\theta_0 + \theta_1 x + \theta_2 x^2) = x^2$$

$$\frac{\partial u}{\partial \theta_2} = \sum_{i=1}^{n} \left(f_\theta(x^{(i)}) - y^{(i)}\right) {x^{(i)}}^2$$

Linear Regression Model

$$x := x - \eta \frac{d}{dx} g(x)$$

$$\theta_0 := \theta_0 - \eta \frac{\partial E}{\partial \theta_0} g(x) \rightarrow \theta_0 := \theta_0 - \eta \sum_{i=1}^{n} (f_\theta(x^{(i)}) - y^{(i)})$$

$$\theta_1 := \theta_1 - \eta \frac{\partial E}{\partial \theta_1} g(x) \rightarrow \theta_1 := \theta_1 - \eta \sum_{i=1}^{n} (f_\theta(x^{(i)}) - y^{(i)}) x^{(i)}$$

$$\theta_2 := \theta_2 - \eta \frac{\partial E}{\partial \theta_2} g(x) \rightarrow \theta_2 := \theta_2 - \eta \sum_{i=1}^{n} (f_\theta(x^{(i)}) - y^{(i)}) x^{(i)^2}$$

Artificial Intelligence

Thank You