

Inclass 02: Linear Algebra and Probability

[SCS4049] Machine Learning and Data Science

Seongsik Park (s.park@dgu.ac.kr)

School of AI Convergence & Department of Artificial Intelligence, Dongguk University

Linear Algebra

Notation: 약속, 편리,

가인자 인자	Scalar 소문자 regular	vector 소문자 <u>bold</u>	matrix 대문자 bold
판서	x, y, z a, b, c	X, Y, Z a, b, c, d	X, Y, Z A, B, C, D

Vector

- Notation for a column vector

$$\mathbf{x} = (x_1, x_2, \dots, x_n) = \begin{bmatrix} \underline{x_1} \\ \underline{x_2} \\ \dots \\ \underline{x_n} \end{bmatrix} \in \mathcal{R}^n \quad (1)$$

Handwritten notes:
 $\mathcal{R}$: 실수 집합
 $\mathcal{R}^n$: n차원 벡터 집합
 $n, n \times 1$

- Transpose and a row vector

$$\underline{\mathbf{x}}^T = (x_1, x_2, \dots, x_n)^T = \begin{bmatrix} x_1 & x_2 & \dots & x_n \end{bmatrix} \quad (2)$$

- Magnitude of a vector (= 2-norm)

$$\|\mathbf{x}\| = \|\mathbf{x}\|_2 = \sqrt{x_1^2 + x_2^2 + \dots + x_n^2} \quad (3)$$

$$\|\mathbf{x}\| = \|\mathbf{x}\|_2 = 0 \iff \mathbf{x} = \mathbf{0} \quad (4)$$

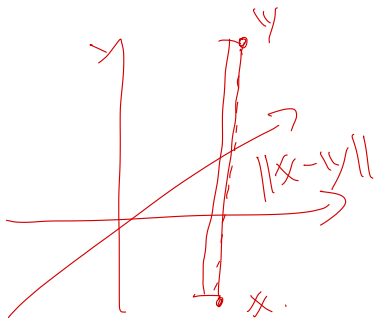
~~$\|\mathbf{x}\|$~~

$$\underline{x} = \begin{bmatrix} 1 \\ 0 \\ -1 \end{bmatrix}$$

$$\|\underline{x}\| = \sqrt{1^2 + 0^2 + (-1)^2}$$

$$\underline{y} = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}$$

$$\|\underline{x} - \underline{y}\| = \sqrt{0^2 + 2^2 + 4^2}$$



$$x, y, z \in \mathbb{R}^3$$

$$A, B \in \mathbb{R}^{10 \times 100}$$

Vector operation

- Addition : 두개의 size가 동일.

shape

$$\mathbf{x} + \mathbf{y} = \begin{bmatrix} x_1 \\ x_2 \\ \dots \\ x_n \end{bmatrix} + \begin{bmatrix} y_1 \\ y_2 \\ \dots \\ y_n \end{bmatrix} = \begin{bmatrix} x_1 + y_1 \\ x_2 + y_2 \\ \dots \\ x_n + y_n \end{bmatrix} \quad (5)$$

where $\mathbf{x}, \mathbf{y} \in \mathcal{R}^n$

- Scalar product

$$\lambda \mathbf{x} = \lambda \begin{bmatrix} x_1 \\ x_2 \\ \dots \\ x_n \end{bmatrix} = \begin{bmatrix} \lambda x_1 \\ \lambda x_2 \\ \dots \\ \lambda x_n \end{bmatrix} \quad (6)$$

Vector operation

- Inner product : ^{내적} size가 동일한 두 벡터 $\longrightarrow$ scalar
projection 의미.

$$\mathbf{x} = (x_1, x_2, \dots, x_n) \quad \mathbf{y} = (y_1, y_2, \dots, y_n) \quad (7)$$

$$\mathbf{x} \cdot \mathbf{y} = \mathbf{x}^T \mathbf{y} = \sum_{i=1}^n x_i y_i = \|\mathbf{x}\| \|\mathbf{y}\| \cos \theta \quad \theta \rightarrow 0. \quad (8)$$

- Property
 - Distributiveness: $(\mathbf{x} + \mathbf{y}) \cdot \mathbf{z} = \mathbf{x} \cdot \mathbf{z} + \mathbf{y} \cdot \mathbf{z}$ and $\mathbf{x} \cdot (\mathbf{y} + \mathbf{z}) = \mathbf{x} \cdot \mathbf{y} + \mathbf{x} \cdot \mathbf{z}$
 - Linearity: $(\lambda \mathbf{x}) \cdot \mathbf{y} = \mathbf{x} \cdot (\lambda \mathbf{y}) = \lambda(\mathbf{x} \cdot \mathbf{y})$
 - Symmetry: $\mathbf{x} \cdot \mathbf{y} = \mathbf{y} \cdot \mathbf{x}$
 - Non-negativity: $\forall \mathbf{x} \neq \mathbf{0}, \mathbf{x} \cdot \mathbf{x} > 0$ and $\mathbf{x} = \mathbf{0} \iff \mathbf{x} \cdot \mathbf{x} = 0$

Vector Norm

- A norm on a vector space Ω is a function $f: \Omega \rightarrow \mathcal{R}$ that satisfies following properties:

- Positive scalability: $f(\lambda \mathbf{x}) = |\lambda|f(\mathbf{x})$
- Triangle inequality: $f(\mathbf{x} + \mathbf{y}) \leq f(\mathbf{x}) + f(\mathbf{y})$
- If $f(\mathbf{x}) = 0$, then $\mathbf{x} = 0$

바탕이 3기.

- Examples of norm

- 1-norm: $\|\mathbf{x}\|_1 = \sum_{i=1}^n |x_i|$
- 2-norm: $\|\mathbf{x}\|_2 = (\sum_{i=1}^n |x_i|^2)^{\frac{1}{2}}$
- p-norm: $\|\mathbf{x}\|_p = (\sum_{i=1}^n |x_i|^p)^{\frac{1}{p}}$
- Infinity norm: $\|\mathbf{x}\|_\infty = \max_j |x_j|$

$$\mathbf{x} = \begin{bmatrix} 1 \\ 0 \\ -2 \end{bmatrix}$$

$$\|\mathbf{x}\|_1 = 1 + 0 + |-2| = 3$$

$$\|\mathbf{x}\|_2 = \|\mathbf{x}\| = \sqrt{1^2 + 0^2 + (-2)^2} = \sqrt{5}$$

$$\|\mathbf{x}\|_\infty = \max(1, 0, 2) = 2$$

Vector Norm

- Inner product revisited

$$\mathbf{x} = (x_1, x_2, \dots, x_n) \quad \mathbf{y} = (y_1, y_2, \dots, y_n) \quad (9)$$

$$\mathbf{x} \cdot \mathbf{y} = \mathbf{x}^T \mathbf{y} = \sum_{i=1}^n x_i y_i = \|\mathbf{x}\| \|\mathbf{y}\| \cos \theta \quad (10)$$

$$\|\mathbf{x}\|_2^2 = \sum_{i=1}^n |x_i|^2 = \mathbf{x} \cdot \mathbf{x} = \mathbf{x}^T \mathbf{x} \quad (11)$$

$\downarrow$

$$\Rightarrow \left(\sqrt{x_1^2 + \dots + x_n^2} \right)^2 = \underline{x_1^2 + x_2^2 + \dots + x_n^2}$$

$$\Rightarrow \underline{x_1^2 + x_2^2 + \dots + x_n^2}$$

Matrix

- Notation for a 2D array of scalars

$$\begin{array}{c}
 \text{row} \rightarrow \\
 \text{row} \rightarrow \\
 \text{row} \rightarrow
 \end{array}
 \begin{array}{c}
 \downarrow \text{col} \\
 \downarrow \text{col} \\
 \downarrow \text{col}
 \end{array}
 \mathbf{A} \in \mathcal{R}^{m \times n} \quad \mathbf{A} = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix}, \quad (\mathbf{A})_{ij} = a_{ij} \quad (12)$$

- Identity matrix $\mathbf{I}_n$ is the $n \times n$ square matrix with ones on the main diagonal and zeros elsewhere *off-diagonal*

$$\mathbf{I}_m \mathbf{A} = \mathbf{A} \mathbf{I}_n = \mathbf{A} \quad (13)$$

$$\mathbf{I}_2 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \quad \mathbf{I}_3 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

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

$$a_1^T = [1 \ 2 \ 3]$$

$$a_2^T = [4 \ 5 \ 6]$$

row by row $\Rightarrow$ A의 row vector $\begin{matrix} \xrightarrow{2} \\ \xrightarrow{3} \end{matrix}$

col by col $\Rightarrow$ A의 column vector $\begin{matrix} \xrightarrow{2} \\ \xrightarrow{3} \end{matrix}$

$$a_1 = \begin{bmatrix} 1 \\ 4 \end{bmatrix} \quad a_2 = \begin{bmatrix} 2 \\ 5 \end{bmatrix} \quad a_3 = \begin{bmatrix} 3 \\ 6 \end{bmatrix}$$

Matrix operation

- Addition : Size가 동일.

$$\mathbf{A} + \mathbf{B} = \begin{bmatrix} a_{11} + b_{11} & a_{12} + b_{12} & \cdots & a_{1n} + b_{1n} \\ a_{21} + b_{21} & a_{22} + b_{22} & \cdots & a_{2n} + b_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} + b_{m1} & a_{m2} + b_{m2} & \cdots & a_{mn} + b_{mn} \end{bmatrix} \quad (14)$$

$$(\mathbf{A} + \mathbf{B})_{ij} = a_{ij} + b_{ij} \quad (15)$$

- Commutative: $\mathbf{A} + \mathbf{B} = \mathbf{B} + \mathbf{A}$
- Associative: $(\mathbf{A} + \mathbf{B}) + \mathbf{C} = \mathbf{A} + (\mathbf{B} + \mathbf{C})$

○ Transpose: $(\mathbf{A}^T)_{ij} = (\mathbf{A})_{ji}$

- $(\mathbf{A}^T)^T = \mathbf{A}$
- $(\mathbf{AB})^T = \mathbf{B}^T \mathbf{A}^T$

행 ↔ 열

$$\mathbf{A} \in \mathbb{R}^{m \times n}$$

$$\mathbf{A}^T \in \mathbb{R}^{n \times m}$$

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

2x3

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

3x2

$$X = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}$$

3x1

$$X^T = [1 \ 2 \ 3]$$

1x3

Matrix multiplication

- Multiplication

$$(\mathbf{AB})_{ij} = \sum_k a_{ik} b_{kj} \quad (16)$$

where $\mathbf{A} \in \mathcal{R}^{m \times k}$, $\mathbf{B} \in \mathcal{R}^{k \times n}$, then $\mathbf{AB} \in \mathcal{R}^{m \times n}$

- Associative: $(\mathbf{AB})\mathbf{C} = \mathbf{A}(\mathbf{BC})$
- Distributive: $\mathbf{A}(\mathbf{B} + \mathbf{C}) = \mathbf{AB} + \mathbf{AC}$
- Not commutative: $\mathbf{AB} \neq \mathbf{BA}$

① 한가지 순서로 가능? $\Rightarrow$ 그럴수록 size

② 그다음 계산분?

①

A

B

A|B

3x2

2x3

T

3x3

3x3

6x4

F

3x5

5x1

T

3x1

3x2

3x2

F

아, 여
1, 2남, 여
1, 2

②

 A 3×2

$$\begin{bmatrix} -a_1^T - \\ -a_2^T - \\ -a_3^T - \end{bmatrix}$$

 B 2×3

$$\begin{bmatrix} | & | & | \\ b_1 & b_2 & b_3 \\ | & | & | \end{bmatrix}$$

 $A|B$ 3×3

$$\begin{bmatrix} a_1^T b_1 & a_1^T b_2 & a_1^T b_3 \\ a_2^T b_1 & a_2^T b_2 & a_2^T b_3 \\ a_3^T b_1 & a_3^T b_2 & a_3^T b_3 \end{bmatrix}$$

$$(A|B)_{ij} = a_i^T b_j = \text{scalar.}$$

i 번째 행 벡터 j 번째 열 벡터
 행 벡터 열 벡터

Matrix vector multiplication

For $\mathbf{x} \in \mathcal{R}^n, \mathbf{y} \in \mathcal{R}^m, \mathbf{A} \in \mathcal{R}^{m \times n}$,

$$\mathbf{y} = \mathbf{A} \mathbf{x}$$

$m \times 1 \quad m \times n \quad n \times 1$

$$\mathbf{y} = \mathbf{A} \mathbf{x}$$

(17)

$$\textcircled{1} \quad = \begin{bmatrix} | & | & \cdots & | \\ \mathbf{a}_1 & \mathbf{a}_2 & \cdots & \mathbf{a}_n \\ | & | & \cdots & | \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} = \mathbf{a}_1 x_1 + \mathbf{a}_2 x_2 + \cdots + \mathbf{a}_n x_n \quad (18)$$

$$\textcircled{2} \quad = \begin{bmatrix} - & \tilde{\mathbf{a}}_1^T & - \\ - & \tilde{\mathbf{a}}_2^T & - \\ & \vdots & \\ - & \tilde{\mathbf{a}}_m^T & - \end{bmatrix} \mathbf{x} = \begin{bmatrix} \tilde{\mathbf{a}}_1^T \mathbf{x} \\ \tilde{\mathbf{a}}_2^T \mathbf{x} \\ \vdots \\ \tilde{\mathbf{a}}_m^T \mathbf{x} \end{bmatrix} \quad (19)$$

$$A = \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{bmatrix} \quad x = \begin{bmatrix} 1 \\ 0 \\ -1 \end{bmatrix} \quad y = Ax$$

① span $y = Ax$

$$= \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}$$

$$= \begin{bmatrix} 1 \\ 4 \end{bmatrix} x_1 + \begin{bmatrix} 2 \\ 5 \end{bmatrix} x_2 + \begin{bmatrix} 3 \\ 6 \end{bmatrix} x_3$$

$$= \begin{bmatrix} 1 \\ 4 \end{bmatrix} 1 + \begin{bmatrix} 2 \\ 5 \end{bmatrix} 0 + \begin{bmatrix} 3 \\ 6 \end{bmatrix} (-1) \quad \swarrow$$

$$A = \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{bmatrix} \quad x = \begin{bmatrix} 0 \\ 1 \\ -1 \end{bmatrix}$$

$$\textcircled{2} \quad y = Ax$$

$$= \begin{bmatrix} a_1^T \\ a_2^T \end{bmatrix} x$$

$$= \begin{bmatrix} a_1^T x \\ a_2^T x \end{bmatrix} = \begin{bmatrix} [1 \ 2 \ 3] \begin{bmatrix} 0 \\ 1 \\ -1 \end{bmatrix} \\ [4 \ 5 \ 6] \begin{bmatrix} 0 \\ 1 \\ -1 \end{bmatrix} \end{bmatrix}$$

Probability

Random variables

- A random variable X takes on a defined set of values with different probabilities
 - For example, if you roll a die, the outcome is random (not fixed) and there are 6 possible outcomes, each of which occur with probability $1/6$
 - For example, if you poll people about their voting preferences, the percentage of the sample that responds “Yes on Proposition A” is a random variable (the percentage will be slightly different every time you poll)
- Roughly, **probability** is how frequently we expect different outcomes to occur if we repeat the experiment over and over

변수 $\Rightarrow$ 가변함
~~~~~

$$\underbrace{\sum_{i=1}^n \sum_{j=1}^n X}_{\text{sum of all } X} \leftarrow \sum_{i=1}^n \sum_{j=1}^n \underbrace{\frac{1}{6}}_{\text{Pr}(X)}$$

| $X$            | 1             | 2             | 3             | 4             | 5             | 6             |
|----------------|---------------|---------------|---------------|---------------|---------------|---------------|
| $\text{Pr}(X)$ | $\frac{1}{6}$ | $\frac{1}{6}$ | $\frac{1}{6}$ | $\frac{1}{6}$ | $\frac{1}{6}$ | $\frac{1}{6}$ |

# Discrete and continuous random variable

- **Discrete** random variables have a countable number of outcomes
  - Dead/live, dice, counts, etc.
  - Probability mass function (pmf, p.m.f.)
- **Continuous** random variables have an infinite continuum of possible values
  - Blood pressure, weight, real number, etc.
  - Probability density function (pdf, p.d.f.)
- **Probability distribution**  $F_X(x) = \Pr(X \leq x)$
- **Probability mass function**  $p_X(x) = \Pr(X = x)$
- **Probability density function**  $f_X(x) = \frac{d}{dx}F_X(x)$



# Conditional probability

**Definition** For any two events  $A$  and  $B$  (with  $\Pr\{B\} > 0$ ), the conditional probability of  $A$ , conditional on  $B$ , is defined by

$$\Pr\{A|B\} = \Pr\{A \cap B\} / \Pr\{B\}$$

|          | COVID<br>확진<br>확률 | 아니<br>확률 |         |
|----------|-------------------|----------|---------|
| 바깥<br>확률 | 100               | 99900    | 100,000 |
| 아니<br>확률 | 1                 | 999      | 100     |

$\Pr(\text{COVID} | \text{확진}) = \frac{100}{100,000} = \frac{1}{1000}$

# Independent

**Definition** Two events, A and B, are statistically independent if

$$\Pr\{A \cap B\} = \Pr\{A\}\Pr\{B\}$$

Two rv's, say X and Y, are *statistically independent* if

$$F_{XY}(x, y) = F_X(x)F_Y(y) \text{ for each value } x_i \text{ of } X \text{ and } y_j \text{ of } Y$$

Discrete rv's

$$p_{XY}(x_i, y_j) = p_X(x_i)p_Y(y_j)$$

$$p_{X|Y}(x_i | y_j) = p_X(x_i)$$

Continuous rv's

$$f_{XY}(x, y) = f_X(x)f_Y(y)$$

$$f_{X|Y}(x | y) = f_X(x)$$

# Expectation

The expected value  $E[X]$  of a random variable  $X$  is also called the expectation or the mean and is frequently denoted as  $\bar{X}$ . Considering nonnegative discrete rv's, the expected value is then given by

$X \sim p(x)$        $E[X] = \sum_x xp_X(x)$ .      :  $X$ 가  $E[X]$ 만큼 평균  $X$ .

- Discrete case:  $E(X) = \sum_x xp_X(x)$
- Continuous case:  $E(X) = \int_x xf_X(x)dx$

|                   |                                                                                                 |
|-------------------|-------------------------------------------------------------------------------------------------|
| $\textcircled{X}$ | $\downarrow \downarrow \downarrow$                                                              |
|                   | <u>1</u> <u>2</u> <u>3</u>                                                                      |
| $p(x)$            | <u><math>\frac{1}{4}</math></u> <u><math>\frac{1}{2}</math></u> <u><math>\frac{1}{4}</math></u> |

$$\begin{aligned} E(X) &= \textcircled{1} \cdot \frac{1}{4} + \textcircled{2} \cdot \frac{1}{2} + \textcircled{3} \cdot \frac{1}{4} \\ &= \textcircled{2} \end{aligned}$$

$\nwarrow$   
 ch 평균 : 평균, 중앙값  
 mean median

$\searrow$   
 $\hookrightarrow$

무게 중심.

$\nwarrow$   
 $1, 1, 1, 1, (100) \Rightarrow 20 \sim$  평균  
 $1$  중앙값

## Variance and standard deviation

The *variance* is denoted by  $\sigma_X^2$  or  $\text{Var}[X]$ . It is given by

$$\sigma_X^2 = \text{E}[(X - \bar{X})^2] = \text{E}[X^2] - \bar{X}^2$$

The *standard deviation*  $\sigma_X$  of  $X$  is the square root of the variance and provides a measure of dispersion of the rv around the mean. Thus the mean is a rough measure of typical values for the outcome of the rv, and  $\sigma_X$  is a measure of the typical difference between  $X$  and  $\bar{X}$ .

## Example: expectation and variance

**Table 1:** Example: discrete random variable

|          |     |     |     |     |     |
|----------|-----|-----|-----|-----|-----|
| $X$      | 10  | 11  | 12  | 13  | 14  |
| $p_X(x)$ | 0.4 | 0.2 | 0.2 | 0.1 | 0.1 |

- Expectation

$$E(X) = \sum_x x p_X(x) \quad (20)$$

$$= 10 \cdot 0.4 + 11 \cdot 0.2 + 12 \cdot 0.2 + 13 \cdot 0.1 + 14 \cdot 0.1 \quad (21)$$

$$= 11.3 \quad (22)$$

## Example: expectation and variance

**Table 2:** Example: discrete random variable

|          |     |     |     |     |     |
|----------|-----|-----|-----|-----|-----|
| $X$      | 10  | 11  | 12  | 13  | 14  |
| $p_X(x)$ | 0.4 | 0.2 | 0.2 | 0.1 | 0.1 |

- Variance

$$E[(X - E(X))^2] = \sum_x (x - 11.3)^2 p_X(x) \quad (23)$$

$$= (10 - 11.3)^2 \cdot 0.4 + (11 - 11.3)^2 \cdot 0.2 \quad (24)$$

$$+ (12 - 11.3)^2 \cdot 0.2 + (13 - 11.3)^2 \cdot 0.1 \quad (25)$$

$$+ (14 - 11.3)^2 \cdot 0.1 \quad (26)$$

$$= 1.81 \quad (27)$$

## Example: expectation and variance

**Table 3:** Example: discrete random variable

|          |     |     |     |     |     |
|----------|-----|-----|-----|-----|-----|
| $X$      | 10  | 11  | 12  | 13  | 14  |
| $p_X(x)$ | 0.4 | 0.2 | 0.2 | 0.1 | 0.1 |

- Variance

$$E[X^2] - \{E(X)\}^2 = \sum_x x^2 p_X(x) - 11.3^2 \quad (28)$$

$$= 10^2 \cdot 0.4 + 11^2 \cdot 0.2 + 12^2 \cdot 0.2 + 13^2 \cdot 0.1 \quad (29)$$

$$+ 14^2 \cdot 0.1 - 11.3^2 \quad (30)$$

$$= 40 + 24.2 + 28.8 + 16.9 + 19.6 - 127.69 \quad (31)$$

$$= 1.81 \quad (32)$$



# Distribution

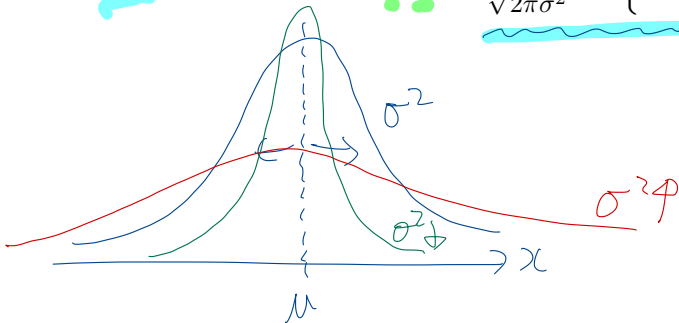
→ Uniform distribution (discrete)

모든 행을 동일

$$U(x) = \frac{1}{n} \quad (33)$$

→ Gaussian distribution (continuous)  $\Leftarrow$  정규분포.

$$p(x) = p(x; \mu, \sigma^2) = \mathcal{N}(\mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp \left\{ -\frac{(x - \mu)^2}{2\sigma^2} \right\} \quad (34)$$



# Bayes' theorem

Bayes' theorem

$$\Pr(A | B) = \frac{\Pr(B | A)\Pr(A)}{\Pr(B)}$$

$\Rightarrow$  기밀한  
확률값.  
(35)

where

- $\Pr(A)$ : prior probability of  $A$
- $\Pr(A | B)$ : posterior probability of  $A$  given  $B$
- $\Pr(B | A)$ : likelihood of  $A$  given  $B$