[SWCON253] Machine Learning – Lec.07a

Probability Review

김 휘 용

경희대학교
KYUNG HEE UNIVERSITY

Contents

- ✓ Probability
- ✓ Conditional Probability & Bayes Theorem
- ✓ Random Variables
- ✓ Independence
- ✓ Expectations
- ✓ Correlation & Covariance
- ✓ Gaussian Distribution

References

- *"Schaum's Outline of Probability, Random Variables, and Random Processes,"* by Hwei P. Hsu

Probability

- Random Experiment
  - *experiment*: any process of observation
  - *outcomes*: the results of an observation
  - *random experiment*: if outcome cannot be predicted with certainty
- Sample Space (S) and Event Space (E)
  - *sample space S*: the set of all possible outcomes
  - *event*: any subset of the sample space S
    - ★ Note that ∅ and S are also events.
  - *event space E*: the set of all possible events
- *Probability Space* (S, E, P)
  - *probability measure P*: a function defined over the event space E
  - *probability space*: the triplet (S, E, P)

Example: *Rolling a Dice*

- *sample space S*:
- *event space E*:
- *probability measure P*:

Probability (cont'd)

- **Axiomatic Definition of *Probability***
  - Consider a probability space (S, E, P).
  - The probability $P(A)$ of an event $A \in E$ is defined as a real number assigned to A which satisfies the following three axioms:
    1. $P(A) \geq 0$
    2. $P(S) = 1$
    3. $P(A \cup B) = P(A) + P(B)$ *if* $P(A \cap B) = \emptyset$ (disjoint)

- **Properties of Probability**
  - ★ $P(A^c) = 1 - P(A)$
  - ★ $P(\emptyset) = 0$
  - ★ $P(A) \leq P(B)$ *if* $A \subset B$
  - ★ $P(A) \leq 1$
  - ★ $P(A \cup B) = P(A) + P(B) - P(A \cap B)$

Conditional Probability & Bayes' Theorem

Conditional Probability

- The ***conditional probability*** of an event A given event B, $P(A|B)$, is defined as:

$$\boxed{P(A|B) = \frac{P(A \cap B)}{P(B)}} \qquad P(B) > 0$$

★ $P(A \cap B)$ is the ***joint probability*** of A and B.

★ Note that $A|B$ is not a set (i.e., not an event).
'$|B$' is just a notation saying that event B has occurred already.

Bayes' Rule

- Note that $P(A \cap B) = P(A|B)P(B) = P(B|A)P(A)$.
- Thus, we can obtain the following ***Bayes' Rule***:

$$\boxed{P(A|B) = \frac{P(B|A)P(A)}{P(B)}}$$

<u>**Example**</u>: *Rolling a Dice*

Assume all outcomes are equally likely.
And let A={1, 2, 3, 4} and B={4, 5, 6}.

- P(A) =
- P(B) =
- P(A∩B) =
- P(A|B) =

Conditional Probability & Bayes' Theorem (cont'd)

Bayes' Theorem

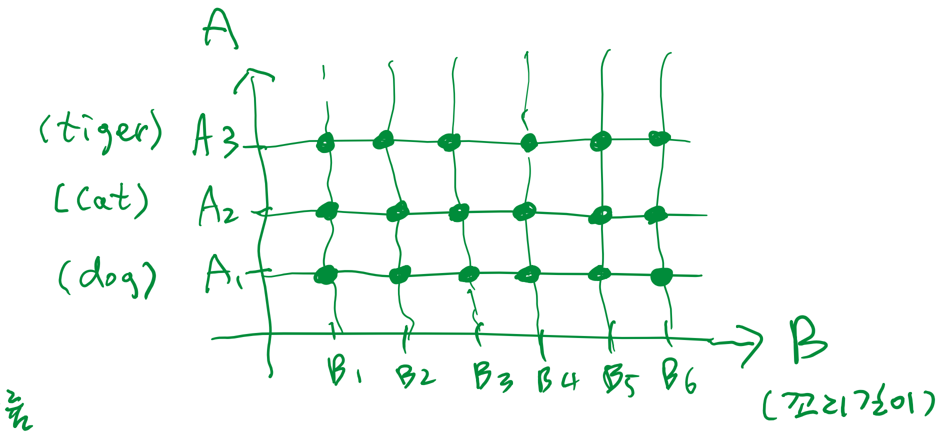
- Suppose the events $A_1, A_2, \ldots, A_n$ are a ***partition*** of S, i.e.,
  - ★ $A_i \cap A_j = \emptyset$ for $\forall i \neq j$: *mutually exclusive (disjoint)*
  - ★ $\bigcup_{i=1}^{n} A_i = S$

- Let B be any event in S. Then we can obtain $P(B)$ by:

  $P(B) = \sum_{i=1}^{n} P(B \cap A_i) = \sum_{i=1}^{n} P(B|A_i)P(A_i)$: *the total probability*

- Using Bayes' Rule, we obtain ***Bayes' Theorem***:

$$\boxed{P(A_i|B) = \frac{P(B|A_i)P(A_i)}{\sum_{i=1}^{n} P(B|A_i)P(A_i)}}$$



★ Sometimes, we call each component:

- $P(A_i|B)$: a ***posteriori*** probability → 꼬리길이를 알때 각 동물의 확률
- $P(B|A_i)$: a ***conditional*** probability → 각 동물의 꼬리길이 분포
- $P(A_i)$: a ***priori*** probability → dog, cat, tiger의 개체수 비율

Random Variables & PDFs

- Random Variable (확률 변수)
  - **(정의)** S에 속한 각 원소 ζ에 실수값을 대응시키는 함수
  - **CDF** (Cumulative Distribution Function)

$$\boxed{F_X(x) = P(X \le x) \qquad -\infty < x < \infty}$$

    - Note that F(x) is a non-decreasing function of x.

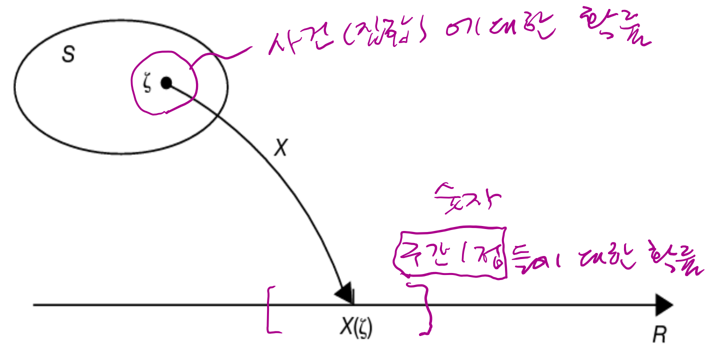


Fig. 2-1 Random variable X as a function.

  - **PMF** (Probability Mass Function)

★ 확률 변수 X가 discrete할 때, X의 PMF $p_X(x_i)$는

$$p_X(x_i) = F_X(x_i) - F_X(x_{i-1}) = P(X \le x_i) - P(X \le x_{i-1}) = P(X = x_i)$$

$$F_X(x) = P(X \le x) = \sum_{x_k \le x} p_X(x_k)$$

  - **PDF** (Probability Density Function)

★ 확률 변수 X가 continuous할 때, X의 PDF $f_X(x)$는

$$f_X(x) = \frac{dF_X(x)}{dx}$$

$$F_X(x) = P(X \le x) = \int_{-\infty}^{x} f_X(\xi)\, d\xi$$

Independence

- **Independent Events**
  - $P(A \cap B) = P(A)P(B)$
  - $P(\bigcap_{i=1}^{n} A_i) = \prod_{i=1}^{n} P(A_i)$

- **Independent Random Variables**
  - Concept: $P(X = x, Y = y) = P(X = x)P(Y = y)$ for any x and y
  - Discrete: $p_{XY}(x_i, y_j) = p_X(x_i)p_Y(y_j)$ for any x_i and y_j
  - Continuous: $f_{XY}(x, y) = f_X(x)f_Y(y)$ for any x and y

Expectations

Mean (*Expectation*) of a Random Variable

The *mean* (or *expected value*) of a r.v. X, denoted by μ_X or $E(X)$, is defined by

$$\mu_X = E(X) = \begin{cases} \sum_k x_k p_X(x_k) & X: \text{ discrete} \\ \int_{-\infty}^{\infty} x f_X(x)\, dx & X: \text{ continuous} \end{cases}$$

Variance of a Random Variable

The *variance* of a r.v. X, denoted by σ_X^2 or $\mathrm{Var}(X)$, is defined by

$$\sigma_X^2 = \mathrm{Var}(X) = E\{[X - E(X)]^2\}$$

$$= \begin{cases} \sum_k (x_k - \mu_X)^2 p_X(x_k) & X: \text{ discrete} \\ \int_{-\infty}^{\infty} (x - \mu_X)^2 f_X(x)\, dx & X: \text{ continuous} \end{cases}$$

$$\mathrm{Var}(X) = E(X^2) - [E(X)]^2$$

[Note] Mean & Variance from Samples

- Sample Mean (empirical mean)

- Sample Variance (empirical variance)

Expectations (cont'd)

- ***Conditional*** **Expectation**
  - Two random variables: X and Y

$$E(Y|X) = \begin{cases} \sum_k y_k \, p(y_k|X) & Y: \text{discrete} \\ \int_{-\infty}^{\infty} y f(y|x)\, dy & Y: \text{continuous} \end{cases}$$

- **Expectation of a *Function* of a Random Variable**
  - $Y = g(X)$

$$E(g(X)) = \begin{cases} \sum_k g(x_k)\, p(x_k) & X: \text{discrete} \\ \int_{-\infty}^{\infty} g(x) f(x)\, dx & X: \text{continuous} \end{cases}$$

$$\text{cf. } E(Y) = \begin{cases} \sum_k y_k \, p(y_k) & Y: \text{discrete} \\ \int_{-\infty}^{\infty} y f(y)\, dy & Y: \text{continuous} \end{cases}$$

PDF for a function of a random variable

Monotonic Function of a Random Variable

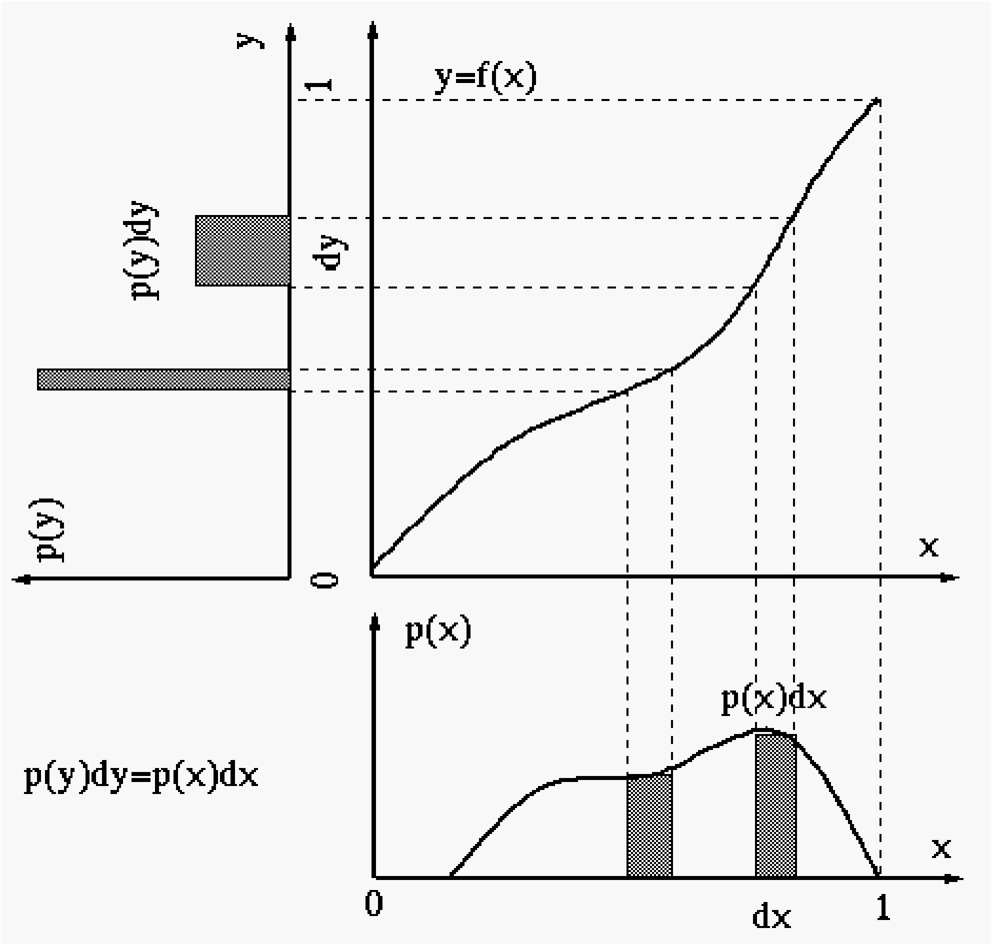
- For a given random variable X,
  let us define a new random variable Y s.t.

  $$Y = g(X)$$

  where $g(X)$ is a *monotonic* function of x.

- Then

$$\boxed{p_y(y) = p_x(x)\left|\frac{dx}{dy}\right|}$$



Correlation & Covariance

- **Two Random Variables: *X* and *Y***
  - *Correlation*: $E(XY)$
    - ★ *orthogonal*: $E(XY) = 0$
    - ★ *uncorrelated*: $E(XY) = E(X)E(Y)$
  - *Covariance*: $Cov(X,Y) = \sigma_{XY} = E[(X - E(X))(Y - E(Y))]$

    ★음수도 될수 있음에 유의!

    $$= E(XY) - E(X)E(Y)$$
    - ★ *uncorrelated*: $\sigma_{XY} = 0$
  - *Correlation Coefficient*: a normalized covariance

    $$\rho_{XY} = \frac{\sigma_{XY}}{\sigma_X \sigma_Y} \qquad |\rho_{XY}| \le 1$$
  - Note
    - ★ Independence implies uncorrelatedness - (1)
    - ★ Uncorrelatedness does NOT imply independence - (2)

(1) independent

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

$$= \left[\sum_{x_i} x_i p_X(x_i)\right]\left[\sum_{y_j} y_j p_Y(y_j)\right] = E(X)E(Y)$$

(2)

$$p_{XY}(x_i, y_j) = \begin{cases} \frac{1}{3} & (0,1), (1,0), (2,1) \\ 0 & \text{otherwise} \end{cases}$$

$$E(X) = \sum_{x_i} x_i p_X(x_i) = (0)\left(\frac{1}{3}\right) + (1)\left(\frac{1}{3}\right) + (2)\left(\frac{1}{3}\right) = 1$$

$$E(Y) = \sum_{y_j} y_j p_y(y_j) = (0)\left(\frac{1}{3}\right) + (1)\left(\frac{2}{3}\right) = \frac{2}{3}$$

$$E(XY) = \sum_{y_j}\sum_{x_i} x_i y_j p_{XY}(x_i, y_j)$$

$$= (0)(1)\left(\frac{1}{3}\right) + (1)(0)\left(\frac{1}{3}\right) + (2)(1)\left(\frac{1}{3}\right) = \frac{2}{3}$$

$E(XY) = E(X)E(Y)$

uncorrelated

$$p_{XY}(0,1) = \frac{1}{3} \neq p_X(0)p_Y(1) = \frac{2}{9} \quad \Rightarrow \text{NOT independent}$$

Correlation & Covariance (cont'd)

Correlation Coefficient & Linear Dependence

Let $Y = aX + b$.

(a) Find the covariance of X and Y.

(b) Find the correlation coefficient of X and Y.

(a) By Eq. (4.131), we have

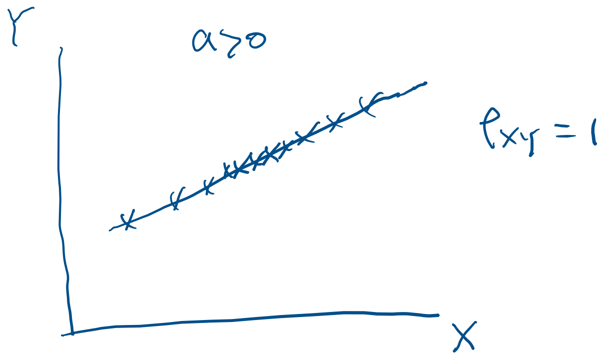
$$E(XY) = E[X(aX+b)] = aE(X^2) + bE(X)$$
$$E(Y) = E(aX+b) = aE(X) + b$$

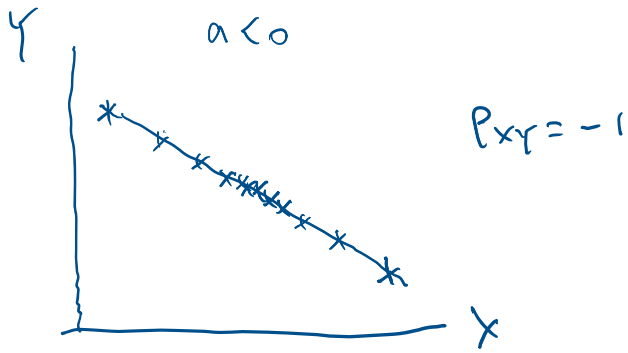
Thus, the covariance of X and Y is [Eq. (3.51)]

$$\begin{aligned}\text{Cov}(X,Y) = \sigma_{XY} &= E(XY) - E(X)E(Y) \\ &= aE(X^2) + bE(X) - E(X)[aE(X)+b] \\ &= a\{E(X^2) - [E(X)]^2\} = a\sigma_X^2\end{aligned}$$

(b) By Eq. (4.130), we have $\sigma_Y = |a|\,\sigma_X$. Thus, the correlation coefficient of X and Y is

$$\rho_{XY} = \frac{\sigma_{XY}}{\sigma_X\sigma_Y} = \frac{a\sigma_X^2}{\sigma_X|a|\sigma_X} = \frac{a}{|a|} = \begin{cases} 1 & a > 0 \\ -1 & a < 0 \end{cases}$$





Correlation & Covariance (cont'd)

Covariance Matrix of a Random Vector

- *random vector*: an array of random variables

$$\mathbf{X} = [X_1 \quad \dots \quad X_n]^T$$

- *covariance matrix* of $\mathbf{X}$:

$$K_{\mathbf{X}} = \begin{bmatrix} \sigma_{11} & \cdots & \sigma_{1n} \\ \vdots & \ddots & \vdots \\ \sigma_{n1} & \cdots & \sigma_{nn} \end{bmatrix} \quad \text{where } \sigma_{ij} = Cov(X_i, X_j)$$

★ If X_i's are *uncorrelated*, then K becomes a diagonal matrix since $\sigma_{ij} = 0 \;\; for \; \forall i \neq j$.

$$K_{\mathbf{X}} = \begin{bmatrix} \sigma_{11} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \sigma_{nn} \end{bmatrix}$$

Correlation & Covariance (cont'd)

Estimating Mean & Covariance from a Dataset

- Consider *n training samples* of *d-dimensional data*:

$$\mathbb{X} = \{\mathbf{x}^{(1)}, \mathbf{x}^{(2)}, \cdots, \mathbf{x}^{(n)}\}, \quad \mathbf{x}^{(k)} = \left[x_1^{(k)} \;\; \dots \;\; x_d^{(k)}\right]^T$$

- The *mean* of each component can be estimated from the given dataset:

$$\mu_i \equiv E[x_i] \approx \frac{1}{n}\sum_{k=1}^{n} x_i^{(k)} \qquad (1 \le i \le d)$$

or we can collectively estimate the ***mean vector*** by:

$$\boldsymbol{\mu} = E[\mathbf{x}] = [\mu_1 \;\; \dots \;\; \mu_d]^T \approx \frac{1}{n}\sum_{k=1}^{n} \mathbf{x}^{(k)}$$

- The *covariance* of each pair of data components (i.e., feature components) is:

$$\sigma_{ij} \equiv E\left[(x_i - \mu_i)(x_j - \mu_j)\right] \approx \frac{1}{n}\sum_{k=1}^{n}\left(x_i^{(k)} - \mu_i\right)\left(x_j^{(k)} - \mu_j\right) \qquad (1 \le i, j \le d)$$

or we can collectively estimate the ***covariance matrix*** by:

$$K \equiv [\sigma_{ij}] \approx \frac{1}{n}\sum_{k=1}^{n}\left(\mathbf{x}^{(k)} - \boldsymbol{\mu}\right)\left(\mathbf{x}^{(k)} - \boldsymbol{\mu}\right)^T$$

$$\because \; (\mathbf{x} - \boldsymbol{\mu})(\mathbf{x} - \boldsymbol{\mu})^T = \begin{bmatrix} (x_1 - \mu_1) \\ \vdots \\ (x_d - \mu_d) \end{bmatrix} [(x_1 - \mu_1) \;\dots\; (x_d - \mu_d)] = \begin{bmatrix} (x_1 - \mu_1)(x_1 - \mu_1) & \cdots & (x_1 - \mu_1)(x_d - \mu_d) \\ \vdots & \ddots & \vdots \\ (x_d - \mu_d)(x_1 - \mu_1) & \cdots & (x_d - \mu_d)(x_d - \mu_d) \end{bmatrix} \Rightarrow \begin{bmatrix} \sigma_{11} & \cdots & \sigma_{1d} \\ \vdots & \ddots & \vdots \\ \sigma_{d1} & \cdots & \sigma_{dd} \end{bmatrix}$$

Correlation & Covariance (cont'd)

◆ 평균 벡터와 공분산 행렬 예제

Iris 데이터베이스의 샘플 중 8개만 가지고 공분산 행렬을 계산하자.

$$\mathbb{X} = \{\mathbf{x}_1 = \begin{pmatrix}5.1\\3.5\\1.4\\0.2\end{pmatrix}, \mathbf{x}_2 = \begin{pmatrix}4.9\\3.0\\1.4\\0.2\end{pmatrix}, \mathbf{x}_3 = \begin{pmatrix}4.7\\3.2\\1.3\\0.2\end{pmatrix}, \mathbf{x}_4 = \begin{pmatrix}4.6\\3.1\\1.5\\0.2\end{pmatrix}, \mathbf{x}_5 = \begin{pmatrix}5.0\\3.6\\1.4\\0.2\end{pmatrix}, \mathbf{x}_6 = \begin{pmatrix}5.4\\3.9\\1.7\\0.4\end{pmatrix}, \mathbf{x}_7 = \begin{pmatrix}4.6\\3.4\\1.4\\0.3\end{pmatrix}, \mathbf{x}_8 = \begin{pmatrix}5.0\\3.4\\1.5\\0.2\end{pmatrix}\}$$

먼저 평균벡터를 구하면 $\boldsymbol{\mu} = (4.9125,\ 3.3875,\ 1.45,\ 0.2375)^{\mathrm{T}}$이다. 첫 번째 샘플 $\mathbf{x}_1$을 식 (2.39)에 적용하면 다음과 같다.

$$(\mathbf{x}_1 - \boldsymbol{\mu})(\mathbf{x}_1 - \boldsymbol{\mu})^{\mathrm{T}} = \begin{pmatrix}0.1875\\0.1125\\-0.05\\-0.0375\end{pmatrix}(0.1875 \quad 0.1125 \quad -0.05 \quad -0.0375)$$

$$= \begin{pmatrix}0.0325 & 0.0211 & -0.0094 & -0.0070\\0.0211 & 0.0127 & -0.0056 & -0.0042\\-0.0094 & -0.0056 & 0.0025 & 0.0019\\-0.0070 & -0.0042 & 0.0019 & 0.0014\end{pmatrix}$$

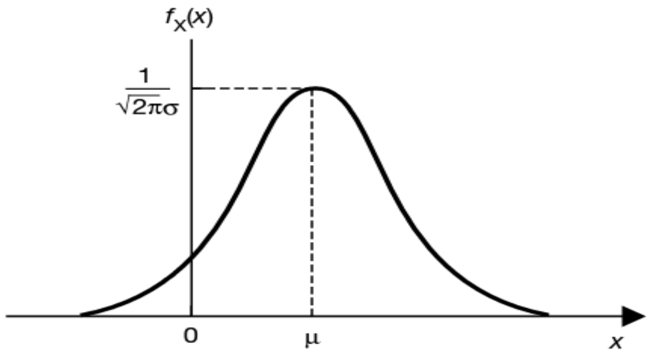
나머지 7개 샘플도 같은 계산을 한 다음, 결과를 모두 더하고 8로 나누면 다음과 같은 공분산 행렬을 얻는다.

$$\boldsymbol{\Sigma} = \begin{pmatrix}0.0661 & 0.0527 & 0.0181 & 0.0083\\0.0527 & 0.0736 & 0.0181 & 0.0130\\0.0181 & 0.0181 & 0.0125 & 0.0056\\0.0083 & 0.0130 & 0.0056 & 0.0048\end{pmatrix}$$

Gaussian Distribution

- **Univariate**
  - A r.v. X is called a ***normal*** (or ***Gaussian***) r.v. if its pdf is given by

$$\boxed{f_X(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-(x-\mu)^2/(2\sigma^2)}} \Rightarrow \begin{aligned} \mu_X &= E(X) = \mu \\ \sigma_X^2 &= \mathrm{Var}(X) = \sigma^2 \end{aligned}$$



- **Bivariate**

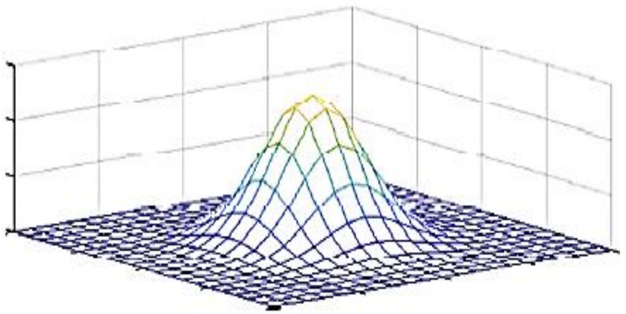
$$f_{XY}(x,y) = \frac{1}{2\pi\sigma_x\sigma_y(1-\rho^2)^{1/2}} \exp\left[-\frac{1}{2}q(x,y)\right]$$

$$q(x,y) = \frac{1}{1-\rho^2}\left[\left(\frac{x-\mu_x}{\sigma_x}\right)^2 - 2\rho\left(\frac{x-\mu_x}{\sigma_x}\right)\left(\frac{y-\mu_y}{\sigma_y}\right) + \left(\frac{y-\mu_y}{\sigma_y}\right)^2\right]$$

  - If the correlation coefficient $\rho = 0$ (i.e., *uncorrelated*), *then* X and Y are *independent*.

$$f_{XY}(x,y) = \frac{1}{2\pi\sigma_X\sigma_Y} \exp\left\{-\frac{1}{2}\left[\left(\frac{x-\mu_X}{\sigma_X}\right)^2 + \left(\frac{y-\mu_Y}{\sigma_Y}\right)^2\right]\right\}$$

$$= \frac{1}{\sqrt{2\pi}\sigma_X} \exp\left[-\frac{1}{2}\left(\frac{x-\mu_X}{\sigma_X}\right)^2\right] \frac{1}{\sqrt{2\pi}\sigma_Y} \exp\left[-\frac{1}{2}\left(\frac{y-\mu_Y}{\sigma_Y}\right)^2\right] = f_X(x) f_Y(y)$$



Gaussian Distribution (cont'd)

Multivariate

- Consider an *n*-dimensional *random vector* $\mathbf{X} = [X_1 \quad \dots \quad X_n]^T$.
- The random vector is called an ***n-variate normal*** if its joint pdf is given by

$$\boxed{f_{\mathbf{X}}(\mathbf{x}) = \frac{1}{(2\pi)^{n/2}\,|\det K|^{1/2}} \exp\left[-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})^T K^{-1}(\mathbf{x}-\boldsymbol{\mu})\right]}$$

where

$$\mathbf{x} = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} \qquad \boldsymbol{\mu} = E[X] = \begin{bmatrix} \mu_1 \\ \vdots \\ \mu_n \end{bmatrix} = \begin{bmatrix} E(X_1) \\ \vdots \\ E(X_n) \end{bmatrix} \qquad K = \begin{bmatrix} \sigma_{11} & \cdots & \sigma_{1n} \\ \vdots & \ddots & \vdots \\ \sigma_{n1} & \cdots & \sigma_{nn} \end{bmatrix} \qquad \sigma_{ij} = \mathrm{Cov}(X_i, X_j)$$

- Note that $f_{\mathbf{X}}(\mathbf{x})$ stands for $f_{X_1 \cdots X_n}(x_1, \dots, x_n)$.

※ If X_i's are uncorrelated, then $K = \begin{bmatrix} \sigma_{11} & & 0 \\ & \ddots & \\ 0 & & \sigma_{nn} \end{bmatrix}$ and $|\det K| = \left|\prod_{i=1}^{n} \sigma_{ii}\right|$ and $f_{\mathbf{X}}(\mathbf{x}) = \prod_{i=1}^{n} f_{X_i}(x_i)$.

Gaussian Distribution (cont'd)

http://cs229.stanford.edu/section/gaussians.pdf

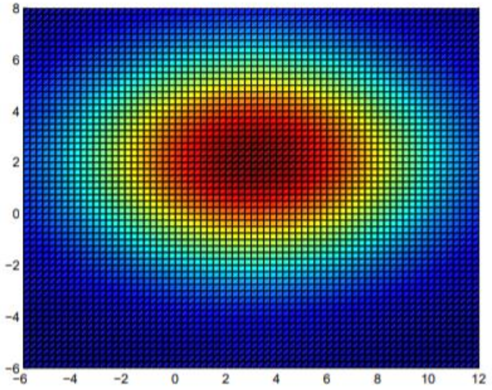
◆ The Diagonal Covariance Matrix (i.e., *Uncorrelated* Gaussian)

- Consider the simple case where n = 2 (i.e., bivariate):

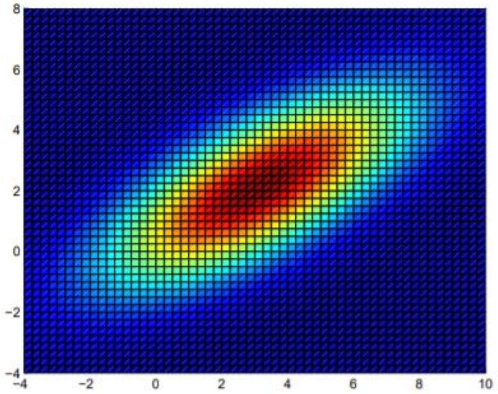
$$x = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \qquad \mu = \begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix} \qquad \Sigma = \begin{bmatrix} \sigma_1^2 & 0 \\ 0 & \sigma_2^2 \end{bmatrix}$$

$$\begin{aligned}
p(x;\mu,\Sigma) &= \frac{1}{2\pi \begin{vmatrix} \sigma_1^2 & 0 \\ 0 & \sigma_2^2 \end{vmatrix}^{1/2}} \exp\left(-\frac{1}{2} \begin{bmatrix} x_1-\mu_1 \\ x_2-\mu_2 \end{bmatrix}^T \begin{bmatrix} \sigma_1^2 & 0 \\ 0 & \sigma_2^2 \end{bmatrix}^{-1} \begin{bmatrix} x_1-\mu_1 \\ x_2-\mu_2 \end{bmatrix} \right) \\
&= \frac{1}{2\pi(\sigma_1^2 \cdot \sigma_2^2 - 0 \cdot 0)^{1/2}} \exp\left(-\frac{1}{2} \begin{bmatrix} x_1-\mu_1 \\ x_2-\mu_2 \end{bmatrix}^T \begin{bmatrix} \frac{1}{\sigma_1^2} & 0 \\ 0 & \frac{1}{\sigma_2^2} \end{bmatrix} \begin{bmatrix} x_1-\mu_1 \\ x_2-\mu_2 \end{bmatrix} \right), \\
&= \frac{1}{2\pi\sigma_1\sigma_2} \exp\left(-\frac{1}{2} \begin{bmatrix} x_1-\mu_1 \\ x_2-\mu_2 \end{bmatrix}^T \begin{bmatrix} \frac{1}{\sigma_1^2}(x_1-\mu_1) \\ \frac{1}{\sigma_2^2}(x_2-\mu_2) \end{bmatrix} \right) \\
&= \frac{1}{2\pi\sigma_1\sigma_2} \exp\left(-\frac{1}{2\sigma_1^2}(x_1-\mu_1)^2 - \frac{1}{2\sigma_2^2}(x_2-\mu_2)^2 \right) \\
&= \frac{1}{\sqrt{2\pi}\sigma_1} \exp\left(-\frac{1}{2\sigma_1^2}(x_1-\mu_1)^2 \right) \cdot \frac{1}{\sqrt{2\pi}\sigma_2} \exp\left(-\frac{1}{2\sigma_2^2}(x_2-\mu_2)^2 \right).
\end{aligned}$$

$$\Sigma = \begin{bmatrix} 25 & 0 \\ 0 & 9 \end{bmatrix}$$



$$\Sigma = \begin{bmatrix} 10 & 5 \\ 5 & 5 \end{bmatrix}$$



- In general, an *n-dimensional Gaussian* with mean $\mu \in \mathbb{R}^n$ & diagonal covariance matrix $\Sigma = \mathrm{diag}(\sigma_1^2, \ldots, \sigma_n^2)$ is the same as the *product of n independent Gaussian* with mean μ_i and variance σ_n^2, respectively.

  ★ Gaussian의 경우는 uncorrelated면 independent!

Q & A