Lecture 01: Introduction and Basics

[AIX7021] Computer Vision

Seongsik Park (s.park@dgu.edu)

AI Department, Dongguk University

Course Introduction

Notice

- 강사: 박성식
  - e-mail: s.park@dgu.edu
  - web: spark-lab.com
- 조교: 첫 학기이므로 강사가 직접 맡음
- 시간: 매주 화13:30–16:30
- 장소: e-Class를 통한 webex 수업 (비대면 100%)
- O ice Hour: 수업 후 30분, 별도 시간은 이메일
- Textbook
  - R. Szeliski, Computer Vision: Algorithms and Applications (link)
  - Forsyth and Ponce, Computer Vision: A Modern Approach (link)
- Pre-requisite: Linear Algebra, Probability
- 강의 홈페이지: e-Class (archive: spark-lab.com)

Tentative schedule

week	topic	date
1	Introduction and image formation	09.01
2	Image processing I	09.08
3	Image processing II	09.15
4	Feature detection and matching	09.22
5	Clustering and segmentation I	09.29
6	Clustering and segmentation II	10.06
7	Tracking	10.13
8	**Mid-term exam**	**10.20**
9	Image classification I	10.27
10	Image classification II	11.03
11	Detection	11.10
12	Object recognition I	11.17
13	Object recognition II	11.24
14	Advanced topic	12.01
15	**Final exam**	**12.08**

Policy

- Grading
  - Mid-term (30%): 10월 20일 화요일, 문제별 점수 합계 / 만점 * 30
  - Final exam (30%): 12월 8일 화요일, 문제별 점수 합계 / 만점 * 30
  - Homework (40%): 각 숙제별 점수 합계 / 모든 숙제의 만점 * 40
  - 모든 점수를 합산하여 평가
- 시험방식
  - 시험시작 전까지 모두 webex로 접속
  - 상반신과 책상이 전부 나올 수 있도록 각자 카메라 설치
  - 확인 후 시험문제 공지와 함께 동시에 시작
- 숙제방식
  - 숙제가 있는 경우, 강의 후 공지(화요일 안에)
  - 기한은 다음 월요일 23시 59분까지
  - Discrete 배점: 숙제당 만점 100%, 불량시 20%, 미제출/delay 0%
  - 문제 풀이, Python 프로그래밍, 논문 요약/구현 등
  - e-Class 통한 온라인 제출

Policy

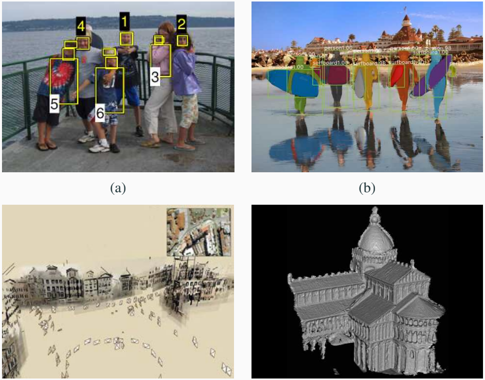
- Misc.
  - 강의 후 다시보기 제공 (e-Class)
  - 강의 다시보기는 종강까지 항상 공개
  - 보강자료나 보강수업 등은 e-Class에 업로드할 계획
  - 대학원 수업은 출석체크 안함
  - 익명 가능한 피드백 https://forms.gle/yYACwh2ikwru7wNJ6

Introduction to Computer Vision

What is Computer Vision?

- 컴퓨터 비전이란?
  - 사람이 시각으로 하는 일들을 컴퓨터로 하는 것
  - 시각적 데이터를 사람의 도움 없이/최소한의 도움으로 컴퓨터가 이해하는 것
- 강의 목표
  - 컴퓨터 비전의 기본 알고리즘 핵심 이해
  - 인공지능 기반의 인식 알고리즘 이해
  - 컴퓨터 비전 알고리즘의 구현 및 응용

Some examples for algorithms and applications



(a) (b) (c) (d)

Figure 1.2 *Some examples of computer vision algorithms and applications. (a)* Face detection *algorithms, coupled with color-based clothing and hair detection algorithms, can locate and recognize the individuals in this image (Sivic, Zitnick, and Szeliski 2006)* © *2006 Springer. (b)* Object instance segmentation *can delineate each person and object in a complex scene (He, Gkioxari* et al. *2017)* © *2017 IEEE. (c)* Structure from motion *algorithms can reconstruct a sparse 3D point model of a large complex scene from hundreds of partially overlapping photographs (Snavely, Seitz, and Szeliski 2006)* © *2006 ACM. (d)* Stereo matching *algorithms can build a detailed 3D model of a building façade from hundreds of differently exposed photographs taken from the Internet (Goesele, Snavely* et al. *2007)* © *2007 IEEE.*

Industrial applications

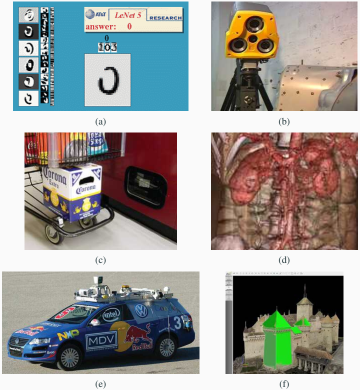


Figure 1.4 *Some industrial applications of computer vision: (a) optical character recognition (OCR), http://yann.lecun.com/exdb/lenet/; (b) mechanical inspection, http://www.cognitens.com/; (c) retail. http://www.evoretail.com/; (d) medical imaging , http://www.clarontech.com/; (e) self-driving cars, (Montemerlo, Becker* et al. *2008) © 2008 Wiley;* [TODO: get copyright approval] *(f) drone-based photogrammetry, https://www.pix4d.com/blog/mapping-chillon-castle-with-drone.* [TODO: get copyright approval]

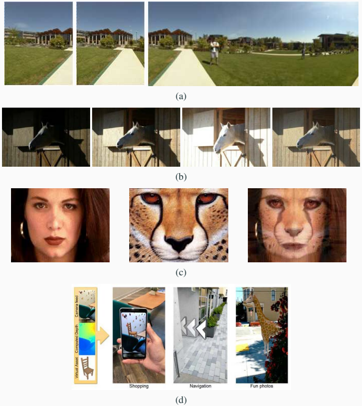
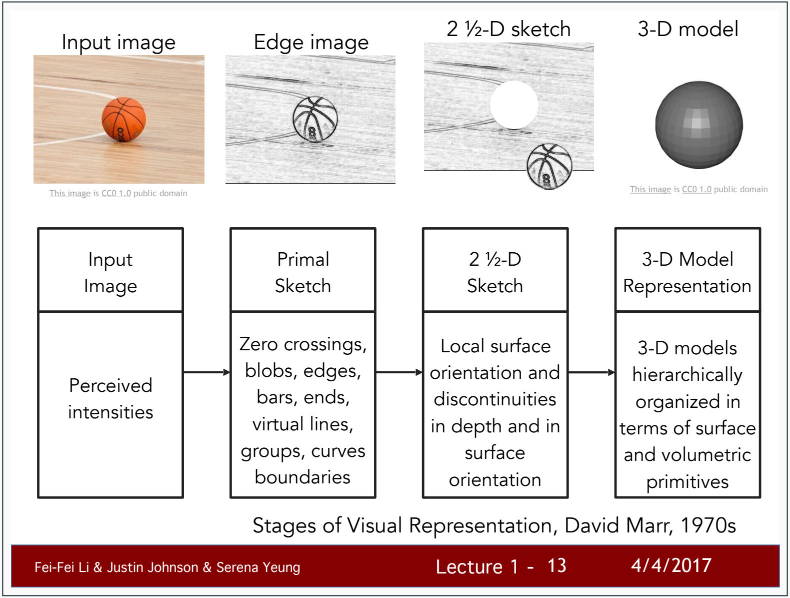


Figure 1.5 *Some consumer applications of computer vision: (a) image stitching: merging different views (Szeliski and Shum 1997) © 1997 ACM; (b) exposure bracketing: merging different exposures; (c) morphing: blending between two photographs (Gomes, Darsa* et al. *1999) © 1999 Morgan Kaufmann; (d) smartphone augmented reality showing real-time depth occlusion effects (Valentin, Kowdle* et al. *2018) © 2018 ACM.*

Stages of visual representation



Major conferences and journals

- Conference
  - Int'l Conf. on Computer Vision (ICCV)
  - IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)
  - European Conf. on Computer Vision (ECCV)
  - Int'l Conf. on Machine Learning (ICML)
  - Conf. on Neural Information Processing Systems (NIPS or NeurIPS)
- Journal
  - IEEE Transactions on Pattern Analysis and Machine Intelligence (T-PAMI)
  - Int'l Journal of Computer Vision

Basics: linear algebra

Vector

- Notation for a column vector

$$\mathbf{x} = (x_1, x_2, ..., x_n) = \begin{bmatrix} x_1 \\ x_2 \\ \cdots \\ x_n \end{bmatrix} \in \mathcal{R}^n \tag{1}$$

- Transpose and a row vector

$$\mathbf{x}^T = (x_1, x_2, ..., x_n)^T = \begin{bmatrix} x_1 & x_2 & \cdots & x_n \end{bmatrix} \tag{2}$$

- Magnitude of a vector (2-norm)

$$\|\mathbf{x}\| = \sqrt{x_1^2 + x_2^2 + \cdots + x_n^2} \tag{3}$$

$$\|\mathbf{x}\| = 0 \iff \mathbf{x} = 0 \tag{4}$$

Vector operation

- Inner product

$$\mathbf{x} = (x_1, x_2, ..., x_n) \quad \mathbf{y} = (y_1, y_2, ..., y_n) \tag{5}$$

$$\mathbf{x} \cdot \mathbf{y} = \mathbf{x}^T \mathbf{y} = \sum_{i=1}^{n} x_i y_i = \|\mathbf{x}\| \|\mathbf{y}\| \cos\theta \tag{6}$$

- 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: $\forall \mathbf{x} \neq 0, \ \mathbf{x} \cdot \mathbf{x} > 0$ and $\mathbf{x} \iff \mathbf{x} \cdot \mathbf{x} = 0$

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 $\mathbf{x} = 0$
- Examples of norm
  - 1-norm: $\|\mathbf{x}\|_1 = \sum_{i=1}^{n} |x_i|$
  - 2-norm: $\|\mathbf{x}\|_2 = \left(\sum_{i=1}^{n} |x_i|^2\right)^{\frac{1}{2}}$
  - p-norm: $\|\mathbf{x}\|_p = \left(\sum_{i=1}^{n} |x_i|^p\right)^{\frac{1}{p}}$
  - Infinity norm: $\|\mathbf{x}\|_\infty = \max_i |x_i|$

Vector Norm

- Inner product revisited

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

$$\mathbf{x} \cdot \mathbf{y} = \mathbf{x}^T \mathbf{y} = \sum_{i=1}^{n} x_i y_i = \|\mathbf{x}\| \|\mathbf{y}\| \cos\theta \tag{8}$$

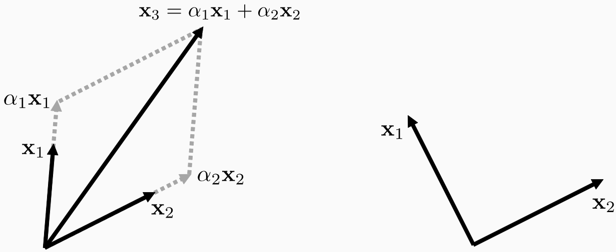
$$\|\mathbf{x}\|_2^2 = \sum_{i=1}^{n} |x_i|^2 = \mathbf{x} \cdot \mathbf{x} = \mathbf{x}^T \mathbf{x} \tag{9}$$

Linear dependency

- Linear combination of vectors $\mathbf{x}_1, \mathbf{x}_2, ..., \mathbf{x}_n$

$$\alpha_1\mathbf{x}_1 + \alpha_2\mathbf{x}_2 + \cdots + \alpha_n\mathbf{x}_n \tag{10}$$

- A set of vectors $\mathcal{X} = \{\mathbf{x}_1, \mathbf{x}_2, ..., \mathbf{x}_n\}$ is *linearly dependent*, if $\mathbf{x}_i \in \mathcal{X}$ can be written as a linear combination of $\mathcal{X} \setminus \{\mathbf{x}_i\}$



Linearly dependent

Linearly independent

- Notation for a 2D array of scalars

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

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

$$\mathbf{I}_m\mathbf{A} = \mathbf{A}\mathbf{I}_n = \mathbf{A} \quad (12)$$

Matrix operation

- Addition
  - Communicative: $\mathbf{A} + \mathbf{B} = \mathbf{B} + \mathbf{A}$
  - Associative: $(\mathbf{A} + \mathbf{B}) + \mathbf{C} = \mathbf{A} + (\mathbf{B} + \mathbf{C})$
- Multiplication
  - Associative: $(\mathbf{AB})\mathbf{C} = \mathbf{A}(\mathbf{BC})$
  - Distributive: $\mathbf{A}(\mathbf{B} + \mathbf{C}) = \mathbf{AB} + \mathbf{AC}$
  - Not communicative: $\mathbf{AB} \neq \mathbf{BA}$
- Transpose: $(\mathbf{A}^T)_{ij} = (\mathbf{A})_{ji}$
  - $(\mathbf{A}^T)^T = \mathbf{A}$
  - $(\mathbf{AB})^T = \mathbf{B}^T\mathbf{A}^T$

Rank of Matrix

- Definition: the number of linearly independent rows or columns in a matrix
- Property
  - $\mathbf{A} \in \mathcal{R}^{m \times n}, \ \mathrm{rank}(\mathbf{A}) \leq \min(m, n)$
  - $\mathbf{A}$ is full rank, if $\mathrm{rank}(\mathbf{A}) = \min(m, n)$
  - $\mathrm{rank}(\mathbf{A}^T) = \mathrm{rank}(\mathbf{A})$
  - $\mathrm{rank}(\mathbf{AB}) \leq \min(\mathrm{rank}(\mathbf{A}), \mathrm{rank}(\mathbf{B}))$
  - $\mathrm{rank}(\mathbf{A} + \mathbf{B}) \leq \mathrm{rank}(\mathbf{A}) + \mathrm{rank}(\mathbf{B})$
  - A matrix with full rank is non-singular; otherwise, the matrix is singular

Matrix inversion

- Inverse matrix
  - Condition to have inverse matrix: square and non-singular
  - Inverse matrix of $\mathbf{A}$ is $\mathbf{A}^{-1}$

$$\mathbf{A}\mathbf{A}^{-1} = \mathbf{A}^{-1}\mathbf{A} = \mathbf{I}_n \text{ where non-singular } \mathbf{A} \in \mathcal{R}^{n\times n} \tag{13}$$

- Property
  - $(\mathbf{AB})^{-1} = \mathbf{B}^{-1}\mathbf{A}^{-1}$
  - $(\mathbf{A}^T)^{-1} = (\mathbf{A}^{-1})^T$
  - For an orthonormal matrix, $\mathbf{A}^{-1} = \mathbf{A}^T$
  - For a diagonal matrix, $\mathbf{D} = \mathrm{diag}(d_1, d_2, ..., d_n)$

$$\mathbf{D}^{-1} = \mathrm{diag}(d_1^{-1}, d_2^{-1}, ..., d_n^{-1}) \tag{14}$$

Determinant

- Recursive formula (Laplace expansion)

$$\det(\mathbf{A}) = \sum_{j=1}^{n} (-1)^{i'+j} a_{i'j} \det(\mathbf{M}_{i'j}) = \sum_{i=1}^{n} (-1)^{i+j'} a_{ij'} \det(\mathbf{M}_{ij'}) \quad (15)$$

where $i', j' \in \{1, 2, ..., n\}$ and $\mathbf{M}_{ij}$ is submatrix of $\mathbf{A}$ removing i-th row and j-th column

- Property
  - $\det(\mathbf{A}) = 0$ iff $\mathbf{A}$ is singular
  - $\det(\mathbf{AB}) = \det(\mathbf{A})\det(\mathbf{B})$
  - $\det(\mathbf{A}) = \det(\mathbf{A})^{-1}$
  - $\det(\lambda\mathbf{A}) = \lambda^n \det(\mathbf{A})$

Basics: probability

Axioms of probability

Axioms for events

1. Ω is an event.
2. For every sequence of events $A_1, A_2, \ldots$, the union $\bigcup_{n=1}^{\infty} A_n$ is an event.
3. For every event A, the complement A^c is an event.

Axioms for probability

1. $\Pr\{\Omega\} = 1$.
2. For every event A, $\Pr\{A\} \geq 0$.
3. The probability of the union of any sequence $A_1, A_2, \ldots$ of disjoint events is given by

$$\Pr\left\{\bigcup_{n=1}^{\infty} A_n\right\} = \sum_{n=1}^{\infty} \Pr\{A_n\}$$

Axioms of probability theory (cont'd)

Corollaries

$$\Pr\{\emptyset\} = 0$$

$$\Pr\left\{\bigcup_{n=1}^{m} A_n\right\} = \sum_{n=1}^{m} \Pr\{A_n\} \qquad \text{for } A_1, \ldots, A_m \text{ disjoint}$$

$$\Pr\{A^c\} = 1 - \Pr\{A\} \qquad \text{for all } A$$

$$\Pr\{A\} \leq \Pr\{B\} \qquad \text{for all } A \subseteq B$$

$$\sum_{n} \Pr\{A_n\} \leq 1 \qquad \text{for all } A_1, \ldots \text{ disjoint}$$

$$\Pr\left\{\bigcup_{n=1}^{\infty} A_n\right\} = \lim_{m\to\infty} \Pr\left\{\bigcup_{n=1}^{m} A_n\right\}$$

$$\Pr\left\{\bigcup_{n=1}^{\infty} A_n\right\} = \lim_{m\to\infty} \Pr\{A_n\} \qquad \text{for } A_1 \subseteq A_2 \subseteq \cdots .$$

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

$$\Pr\{A \mid B\} = \Pr\{AB\}/\Pr\{B\}$$

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

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

Statistical independence

Definition The events $A_1, ..., A_n, n > 2$ are statistically independent if for each collection S of two or more of the integers 1 to n.

$$\Pr\left\{\bigcap_{i \in S} A_i\right\} = \prod_{i \in S} \Pr\{A_i\}$$

This includes the entire collection $S = \{1, ..., n\}$, so one necessary condition for independence is that

$$\Pr\left\{\bigcap_{i=1}^{n} A_i\right\} = \prod_{i=1}^{n} \Pr\{A_i\}.$$

IID condition

To be more specific, given an original sample space Ω, the sample space of an n-repetition model is the Cartesian product

$$\Omega^{\times n} = \{(\omega_1, \omega_2, \cdots, \omega_n) : \omega_i \in \Omega \text{ for each } i, 1 \leq i \leq n\}.$$

More precisely, we assume that for each extended event $\{(A_1 A_2 \cdots A_n)\}$ contained in $\Omega^{\times n}$, we have

$$\Pr\{(A_1 A_2 \cdots A_n)\} = \prod_{i=1}^{n} \Pr\{A_i\},$$

where $\Pr\{A_i\}$ is the probability of event A_i in the original model. We refer to this as the probability model for n independent identically distributed (IID) trials of a given experiment.

Random variables

Definition A random variable (rv) is essentially a function X from the sample space Ω of a probability model to the set of real number $\Re$. Three modifications are needed to make this precise.

First, X might be undefined or infinite for a subset of Ω that has 0 probability.

Second, the mapping $X(\omega)$ must gave the property that $\{\omega \in \Omega : X(\omega) \leq x\}$ is an event for each $x \in \Re$.

Third, every finite set of rv's $X_1, \cdots X_n$ has the property that $\{\omega : X_1(\omega) \leq x_1, \cdots, X_n(\omega) \leq x_n\}$ is an event for each $x_1 \in \Re, \cdots, x_n \in \Re$.

We try to control that confusion here by using $X, X(\omega)$, and x, respectively, to refer to the rv, the sample value taken for a given sample point ω, and a generic sample value.

Distribution function

Definition The distribution function $F_X(x)$ of a random variable (rv) X is a function $\Re \to \Re$, defined by $F_X(x) = \Pr\{\omega \in \Omega : X(\omega) \le x\}$. The argument ω is usually omitted for brevity, so $F_X(x) = \Pr\{X \le x\}$.

We will often work with the distribution function here. This is partly because it is always defined, partly to avoid saying everything thrice, and partly because the distribution function is often most important in limiting arguments such as steady-state time-average arguments.

Conditional probability

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

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

Discrete rv's

$$p_{XY}(x_i, y_j) = p_X(x_i)p_Y(y_j)$$
$$p_{X|Y}(x_i \mid y_j) = p_X(x_i)$$

Continuous rv's

$$f_{XY}(x, y) = f_X(x)f_Y(y)$$
$$f_{X|Y}(x \mid y) = f_X(x)$$

Conditional probability (cont'd)

If we look at the derivatives defining these densities, the conditional density looks at the probability that $\{x \leq X \leq x + \delta\}$ given that $\{y \leq Y \leq y + \epsilon\}$ in the limit, $\delta, \epsilon \to 0$. At some level, this is a very technical point and the intuition of conditioning on $\{Y = y\}$ works very well.

More generally the probability of an arbitrary event A conditional on a given value of a continuous rv Y is given by

$$\Pr\{A \mid Y = y\} = \lim_{\delta \to 0} \frac{\Pr\{A, Y \in [y, y + \delta]\}}{\Pr\{Y \in [y, y + \delta]\}}.$$

Conditional probability (cont'd)

We next generalize the above results to the case of n rv's $X = X_1, \cdots, X_n$.

$$F_X(x_1, \cdots, x_n) = \prod_{i=1}^{n} \Pr\{X_i \leq x_i\} = \prod_{i=1}^{n} F_{X_i}(x_i).$$

For the IID case,

$$F_X(x_1, \cdots, x_n) = \prod_{i=1}^{n} F_X(x_i)$$

Expectation

The expected value $\mathrm{E}[X]$ 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

$$\mathrm{E}[X] = \sum_{x} x p_X(x).$$

If X has a finite number of possible sample values, the above sum must be finite since each sample value must be finite. On the other hand, if X has a countable number of nonnegative sample values, the sum might be either finite or infinite.

The expectation is said to *exist* only if the sum is finite (i.e., if the sum converges to a real number). If the sum is infinite, we say that $\mathrm{E}[X]$ does not exist, but also say that $\mathrm{E}[X] = \infty$.

Variance and standard deviation

The moments $\mathrm{E}[X^n]$ and the central moments $\mathrm{E}[(X-\bar{X})^n]$ of X, where $\bar{X}$ is the mean $\mathrm{E}[X]$. The second central moment is called the *variance*, denoted by σ_X^2 or $\mathrm{Var}[X]$. It is given by

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

The *standard deviation* σ_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 σ_X is a measure of the typical difference between X and $\bar{X}$.

If the rv's $X_1, \cdots, X_n$ are independent, the variance of $S_n = X_1 + \cdots + X_n$ is given by

$$\sigma_{S_n}^2 = \sum_{i=1}^{n} \sigma_{X_i}^2.$$

In the IID case, $\sigma_{S_n}^2 = n\sigma_X^2$.

Conditional expectations

Let X be a positive discrete rv and let y be a sample value of another discrete rv Y such that $p_Y(y) > 0$. Then the conditional expectation of X given $Y = y$ is defined to be

$$\mathrm{E}[X \mid Y = y] = \sum_x x p_{X|Y}(x \mid y)$$

More generally yet, let X be an arbitrary rv and let y be a sample value of a discrete rv Y with $p_Y(y) > 0$, then

$$F_{X|Y}(x \mid y) = \frac{\Pr\{X \leq x, Y = y\}}{\Pr\{Y = y\}}$$

$$\mathrm{E}[X \mid Y = y] = -\int_{-\infty}^{0} F_{X|Y}(x \mid y) dx + \int_{0}^{\infty} F_{X|Y}(x \mid y) dx.$$

Conditional expectations as a rv

The conditional expectation of X conditional on a discrete rv Y can also be viewed as a rv. With the possible exception of a set of zero probability, each $\omega \in \Omega$ maps to $\{Y = y\}$ for some y with $p_Y(y) > 0$ and $\mathrm{E}[X \mid Y = y]$ is defined for that y. Thus, we can define $\mathrm{E}[X \mid Y]$ as a rv that is a function of Y, mapping ω to a sample value, say y of Y, and mapping that y to $\mathrm{E}[X \mid Y = y]$. Regarding a conditional expectation as a rv that is a function of the conditioning rv is a powerful tool both in problem solving and in advanced work.

The unconditional mean of X as

$$\begin{aligned} \mathrm{E}[X] &= \mathrm{E}[\mathrm{E}[X \mid Y]] \\ &= \mathrm{E}_Y\left[\mathrm{E}_{X|Y}[X \mid Y]\right] \end{aligned}$$

Conditional expectations as a rv (cont'd)

Theorem (Total expectation) Let X and Y be discrete rv's. If X is nonnegative, then

$$\mathrm{E}[X] = \mathrm{E}\left[\mathrm{E}[X \mid Y]\right] = \sum_{y} p_Y(y)\mathrm{E}[X \mid Y = y].$$

If X has both positive and negative values, and if at most one of $\mathrm{E}[X^+]$ and $\mathrm{E}[X^-]$ is infinite, then the above is valid.

Appendix

Reference and further reading

- "Chap 1" of R. Szeliski, Computer Vision: Algorithms and Applications
- "Lecture 1" of [CS231n] Introduction to Convolutional Neural Networks for Visual Recognition, Stanford (https://youtu.be/vT1JzLTH4G4)
- "Lecture 2" of [CSED441] Introduction to Computer Vision, POSTECH
- "Chap 1" of R. G. Gallager, Discrete Stochastic Processes (link)

Due: 9월 7일 월요일, 23시 59분까지

- pdf로 업로드하세요.
- 손으로 작성한 파일을 스캔앱(Adobe scan, Office lens 등)을 써서 pdf로 저장해주세요.
- 컴퓨터로 작성(latex, word, ppt, 한글 등)한 파일도 가능합니다. pdf로 저장해주세요.
- 가독성이 떨어지는 파일도 불량처리 합니다.

1. 0이 아닌 두 column vector $\mathbf{x} \in \mathcal{R}^m$ 와 $\mathbf{y} \in \mathcal{R}^n$ 에 대하여 다음을 설명하세요.
    1.1 유도과정을 포함한 $\text{rank}(\mathbf{x}\mathbf{y}^T)$ 의 값 (5점)
    1.2 위의 결과에 대한 물리적 의미 (5점)
2. 두 개의 continuous random variable X, Y에 대해서 $\mathrm{E}[X] = \mathrm{E}\left[\mathrm{E}[X \mid Y]\right]$ 임을 보이세요. (10점)