

성균프레시맨세미나2

Department of Systems Management Engineering
(Industrial Engineering)

시스템경영공학과 (산업공학)

성균관대학교 공과대학 시스템경영공학과 이 종 석

(jongseok@skku.edu)

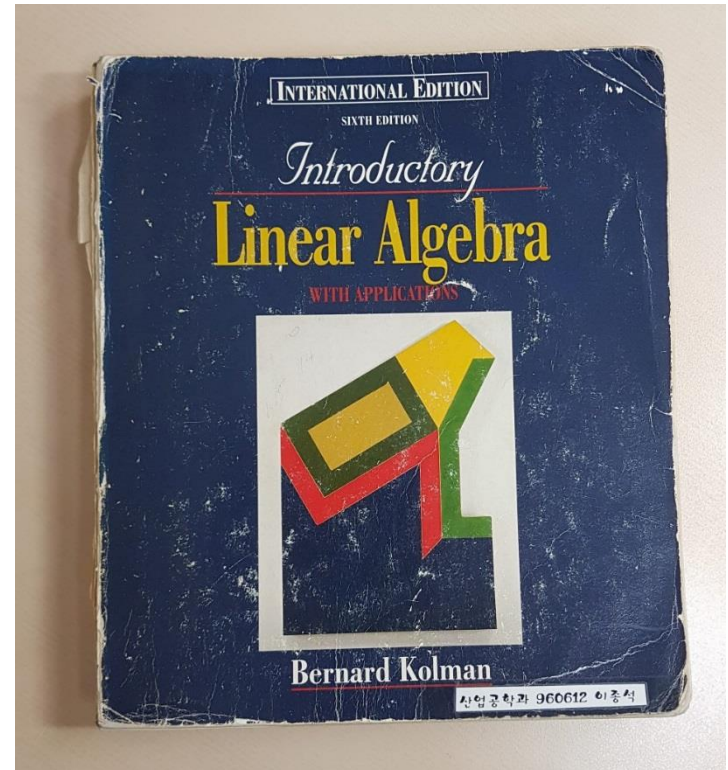
학부 4학년...



- ‘수업은 나름 열심히 들은 것 같은데...’
- ‘응용통계1(확률), 응용통계2(통계), 통계적공정관리(SPC), 실험계획법... 통계관련 관목도 많이 들었는데...’
- ‘선형계획법 (Linear Programming), 선형대수 (Linear Algebra) 둘 다 선형(Linear)인데.. so what??’
- ‘비선형계획법 (Non-Linear Programming) 은 비선형이고, 정수계획법 (Integer Programming) 은 정수형이겠지 뭐..’
- ‘컴퓨터 프로그래밍 수업도 몇 개 들었는데... 비디오 대여 프로그램 만들기도 어렵고..’
- ‘정보시스템이라.. 흠.. DB가 이렇게 생긴 것이구나...’

학부 4학년...

- 응용통계1(확률), 응용통계2(통계)
- 통계적 공정관리
- 실험계획법(Design of Experiments)
- 선형대수 (음.. 당시 선배님들은..)
- OR(Operations Research) 1 & 2
- 경영정보시스템 (MIS)
- C 프로그래밍
- 자료구조 (Data Structure)
- 생산관리, 품질관리, and more



학부 4학년...



- ‘그나저나 난 졸업 후 무엇을 할 수 있을까?’
- ‘에라 모르겠고, 연봉 많이 주는데 취업하자!!’
- ‘아.. IMF’
- 고민.. 고민.. 고민..
- 어느 교수님께서 “대학원에 가 보는 게 좋을 것 같은데!?”
- (IMF 때문에 결정한 것은 아닙니다!!!)

학부 4학년...

- ‘그나저나 난 졸업 후 무엇을 할 수 있을까?’
- 어쩌면 당연한 의문..



산업공학 CEO 시대 열리나...

고동진·구현모·권봉석·김범수 활약

기술·산업 이해도 높고 문제해결 능력 뛰어나
4차 산업혁명 시대 디지털전환 이끌 책임자

〈중략〉



왼쪽부터 고동진 삼성전자 사장, 구현모 KT 사장, 권봉석 LG 전자 사장, 김범수 카카오 의장 / 각 사 제공 · 조선일보DB

IT업계에 ‘산업공학 전공 CEO’ 시대가 열리고 있다. 전기·전자, 컴퓨터, 기계 등에 비해 공대에서 비주류로 불렸던 산업공학 전공자들이 ‘4차 산업혁명 시대’를 맞아 두각을 나타내고 있는 것이다. 이들은 기술과 산업의 특성을 파악, 문제해결 능력이 뛰어나다는 공통점을 갖고 있다. (출처: 조선일보 https://biz.chosun.com/site/data/html_dir/2020/01/05/2020010501293.html 2020.1.6)

Tim Cook



Cook in September 2021

Born	Timothy Donald Cook November 1, 1960 (age 61) Mobile, Alabama, U.S.
Citizenship	American
Education	Auburn University (BSc) Duke University (MBA)
Occupation	CEO of Apple Inc.
Parents	Donald Cook (father) Geraldine Cook (mother)
Website	Apple Leadership Profile 

Signature

A handwritten signature of Tim Cook in black ink. The signature is stylized, starting with a large 'T' and ending with a long, sweeping horizontal stroke.

Early life and education

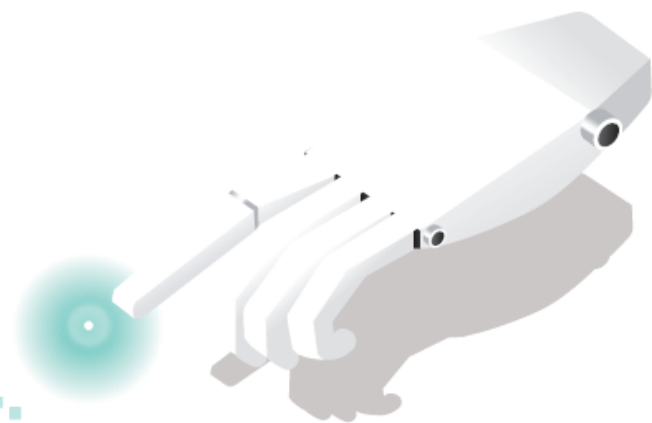
Cook was born in [Mobile, Alabama, United States](#).^{[11][12]} He was baptized in a [Baptist church](#)^[13] and grew up in nearby [Robertsdale](#). His father, Donald, was a [shipyard worker](#), and his mother, Geraldine, worked at a pharmacy.^{[11][14]}

Cook graduated from [Robertsdale High School](#) in 1978.^[15] He earned a [Bachelor of Science](#) in industrial engineering from [Auburn University](#) in 1982,^[16] and his [Master of Business Administration \(MBA\)](#) from [Duke University's Fuqua School of Business](#) in 1988.^[17]

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



산업공학은 어떤 학문인가요?



**ENGINEERS MAKE THINGS;
INDUSTRIAL ENGINEERS MAKE THINGS BETTER.**

산업공학이란 자연과학과 사회과학, 공학적 원리를 결합하여 최적의 프로세스와 시스템을 설계하고 운영함으로써 “무엇인가를 더 잘 할 수 있도록” 지원하는 학문입니다. 따라서 산업공학도들은 제품 및 서비스의 품질을 개선하거나, 생산공정의 효율성을 향상시키거나, 신기술을 더욱 효과적으로 개발하거나, 업무를 보다 쉽고 가치있게 만드는 것 등에 기여해 왔습니다.



오케스트라를 지휘하는...

산업 시스템을 구성하는 모든 분야를
조화롭게 지휘하고 조절하는 방법에 대한 학문



나무와 숲을 동시에 보는...

산업 시스템의 개별 구성 요소에 대한 지식과 함께 해당
구성요소를 시스템에 효율적으로 통합하는 방법에 대한
학문



모든 첨단산업을 선도하는...

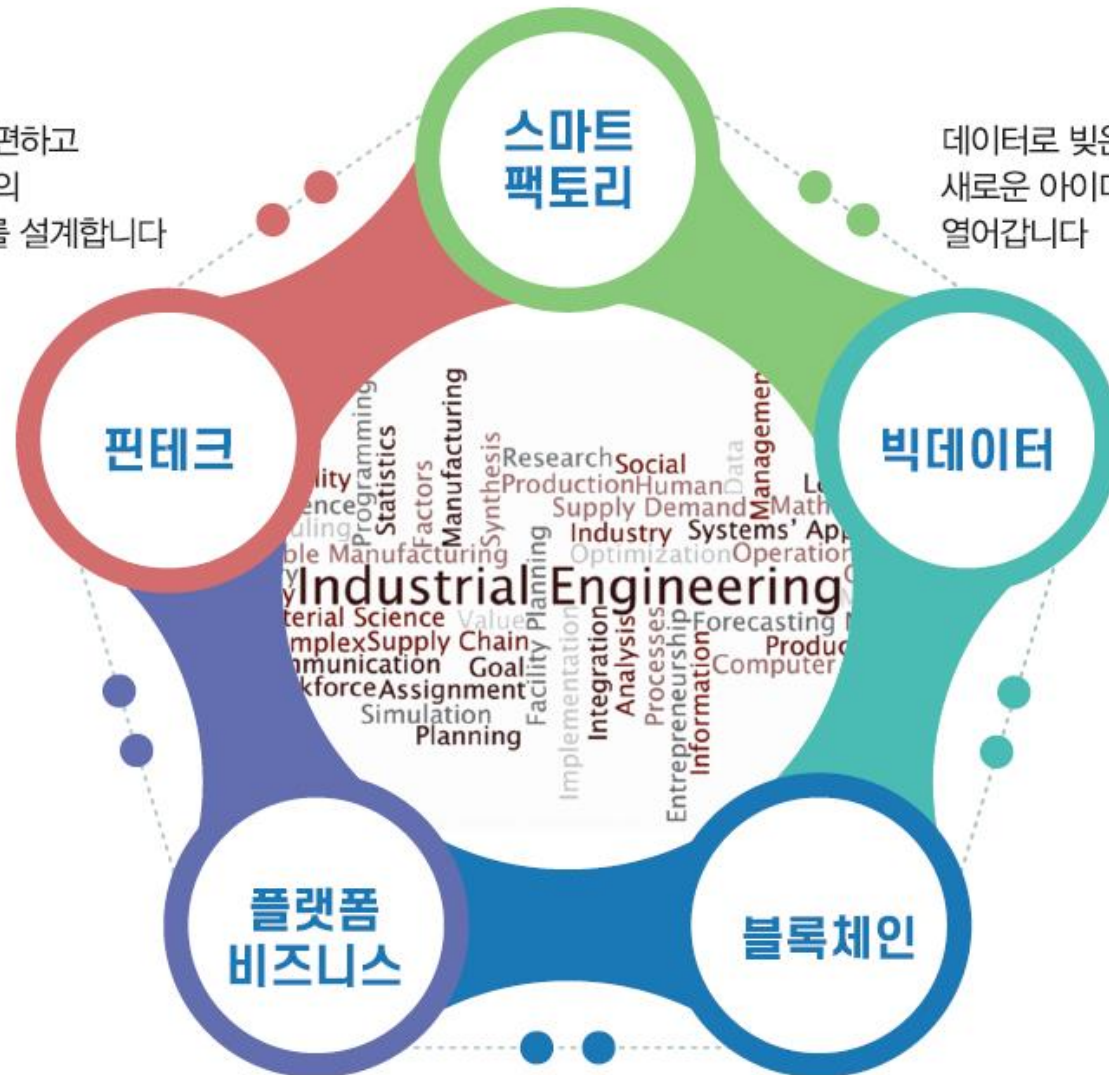
제조, ICT, 물류·유통, 의료·바이오, 국방, 금융 등
모든 산업 분야에 적용될 수 있는 포괄적 융복합 학문

IE Designs the Future

산업공학은 4차 산업혁명을 위한 가장 핵심적인 학문입니다

빠르고 쉽고 편하고
안전한 손 안의
금융 서비스를 설계합니다

데이터로 빛은 불빛이
새로운 아이디어의 길을
열어줍니다



온 세상을 열어
기술과 수요가 만나는 시장을
설계합니다

스마트 트랜잭션을 위한
모두가 참여하는 상호 신뢰의
사슬을 설계합니다

엘리베이터의 최적 대수, 전기차 충전소 위치 등
합리적 의사결정을 위한 모형

스마트폰에 최신 기술을 모두 적용하는게
가장 경제적인가?

개인맞춤형 자산관리 서비스를 통해
개인의 성향에 맞는 투자목표를 달성

기술을 적기에 개발하고 적절한 상품에
구현하여 고객가치를 실현

데이터마이닝: 빅데이터 세상에서 숨겨진
가치를 찾다

왜 택배는 옥천으로 가는가?
택배 배송을 위한 물류창고, 배송 허브,
최단경로 관리

경영
과학

경제성
공학

금융
공학

기술
경영

빅데이터

물류
관리

산업공학
어떤

내용을 배우나요?

생산 경영

자동차 공장을 어떻게 체계적으로 관리할까?

서비스 사이언스

인천국제공항은 어떻게 세계에서 가장 빠른 출입국 서비스를 제공하게 되었을까?

스마트 제조

누구든지 원하는 대로 만든다
3D프린팅 기술을 이용한 개인맞춤형 제품 생산

인간 공학

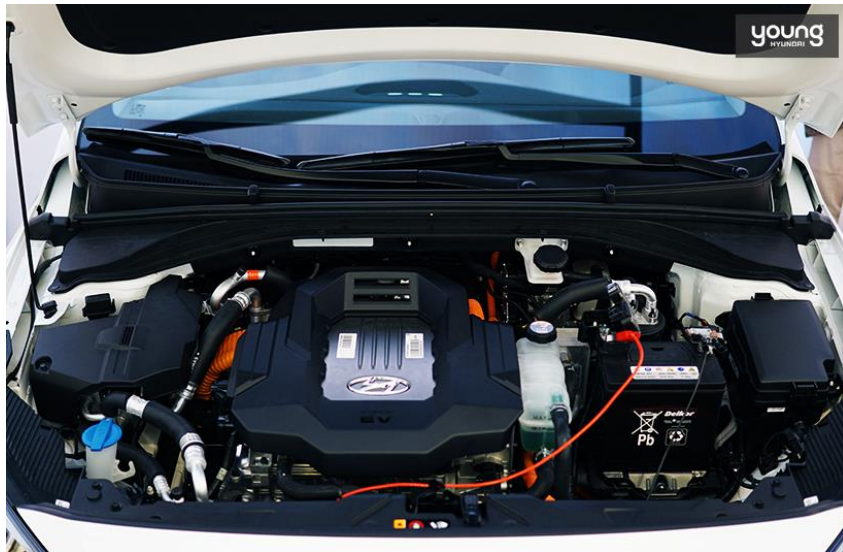
아이폰과 안드로이드폰 중에서 어떤 인터페이스가 편할까?

정보 경영

IT기술을 이용하여 기업 운영에 필요한 정보와 데이터를 전달

품질 공학

인공지능을 이용하여 제품과 공정의 품질을 어떻게 개선할 수 있을까?



도요타, 2022년까지 ‘소프트웨어 퍼스트’로 조직개편

도요타자동차가 2022년까지 ‘소프트웨어 퍼스트(제일주의)’로 전사 조직을 개편, 자동차의 컴퓨터·소프트웨어화에 대응하고 테슬라 타도에 나선다.

일본 기술전문매체 닛케이크로스테크는 지난 15일 “도요타가 2022년까지 ‘소프트웨어 퍼스트’ 체제로 이행하며, 차량 탑재용 전자아키텍처(architecture·기반)도 완전히 새로 개발하기로 했다”고 전했다. 매체에 따르면, 도요타는 최근 사내 간부들을 모아놓고 이 같은 개발 체제 전환 방침을 설명했다.

지금까지 도요타에서 소프트웨어는 어디까지나 차량 자체 개발의 일부 혹은 들러리 성격이었다. 그러나 앞으로 소프트웨어와 하드웨어 개발체제를 분리함으로써 하드웨어 개발에 앞서 전사적 관점에서 소프트웨어를 개발하고 차량 개발을 주도할 수 있는 조직을 만들겠다는 것이다.

소프트웨어 인재 영입 박차... 현대차·기아·모비스 총력전

현대자동차그룹이 본격적인 소프트웨어 인재 모시기 전쟁에 들어갔다. 전통의 제조업체에서 벗어나 자율주행, 통신기술 등으로 중심을 옮겨가기 시작한 것이다. 소프트웨어를 얼마나 잘 개발해 적용하느냐가 수십 년간 주도권 향배를 가를 열쇠가 될 것 이란 전망이 나온다.

10일 업계에 따르면 현대차그룹은 전사적으로 소프트웨어 인재 발굴 및 영입에 나섰다. 수시채용 방식을 도입한 뒤 이뤄지는 채용치고는 이례적인 규모다.

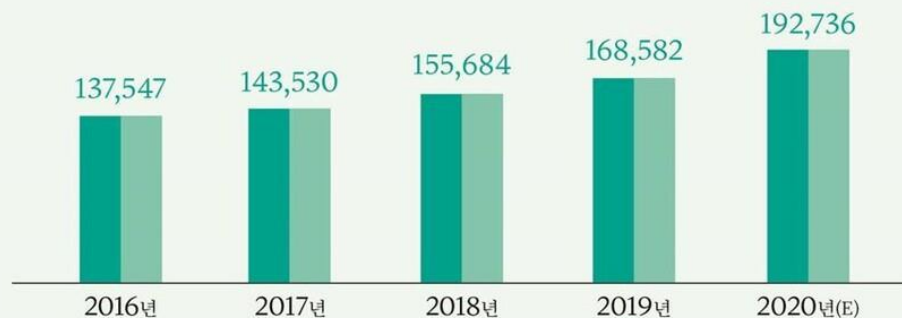
<https://biz.newdaily.co.kr/site/data/html/2021/05/10/2021051000100.html>

데이터 경제 가속화...지난해 데이터산업 14.3% 성장했다

01 데이터산업 시장규모

2020년(E) 기준 데이터산업 시장은 2019년 대비 14.3% 성장한 19조 2,736억원 규모

(단위:억원)



02 데이터산업 부문별 시장규모

2020년(E) 기준 데이터산업 시장의 부문별 규모는

데이터 판매 및 제공 서비스업

9조 3,752억원(48.6%),

데이터 구축 및 컨설팅 서비스업

7조 4,361억원(38.6%),

데이터 처리 및 관리 솔루션 개발·공급업

2조 4,623억원(12.8%)을 차지

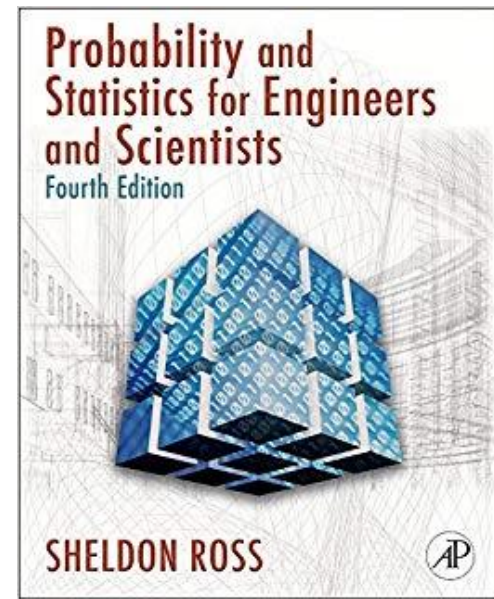
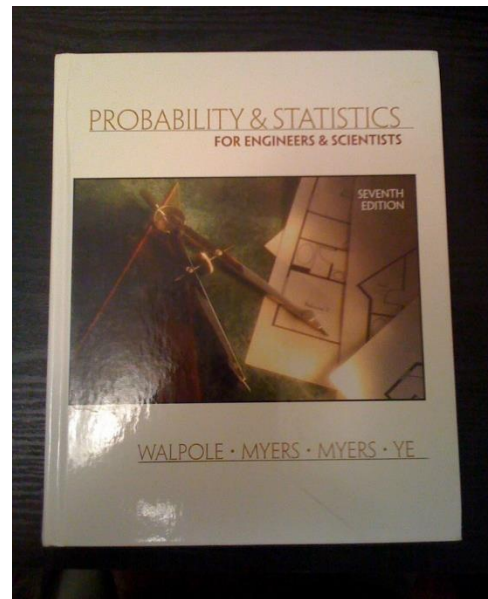
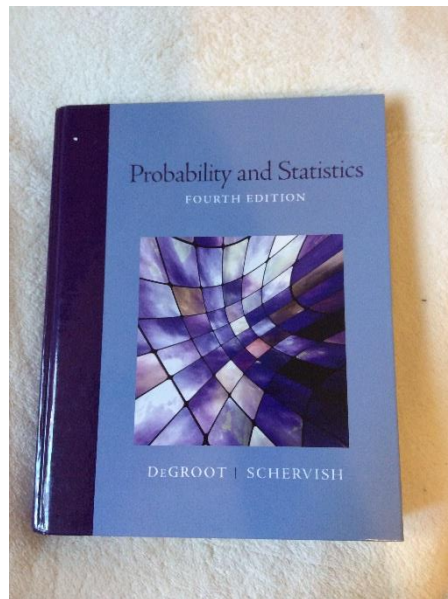


2021-03-16 11:16

IE as a data science incubator

산업공학 & 데이터사이언스

Probability and Statistics



Probability and Statistics

- 문제) 동전을 두 번 던질 때 적어도 한 번 앞면이 나올 확률은?

- Empirical Probability: 100만번 던져서 확률을 계산한다.

$$\text{Probability} = \frac{\text{number of successes}}{1,000,000} = 0.7485$$

앞	뒤	O
앞	앞	O
뒤	뒤	X
		.
		.
		.

- Theoretical Probability

- 모든 경우를 늘어 놓고 {(앞,앞), (앞,뒤), (뒤,앞), (뒤,뒤)}
- 원하는 경우를 고른다. {(앞,앞), (앞,뒤), (뒤,앞)}
- 각 경우가 $\frac{1}{4}$ 의 확률을 가지므로, 구하는 확률은 $\frac{3}{4}$

Probability and Statistics

- 앞이 한번만 나올 확률을 알고 싶다면?
- 100번 던져서 앞이 70번 이상 나올 확률을 알고 싶다면?
- Empirical 과 Theoretical 사이에서 어떤 방법이 더 좋을까?
- Theoretical Probability 측면에서 우리에게 필요한 것은...
 - 확률을 부여 받을 수 있는 대상(사상, event)들을 체계적으로 구성하는 통일된 방법이 필요하다. (예. **확률변수(random variable)**)
 - "사상"들에 확률 값을 체계적으로 부여하는 통일된 방법이 필요하다.(예. **확률분포(probability distribution)**)

Modern Probability

Probability and Statistics

- 동전의 앞면이 나올 확률이 정말 $\frac{1}{2}$ 일까??
- Only God knows...
- 우리가 할 수 있는 최선은?

Probability and Statistics

- 데이터를 수집하는 과정이 필요하고, 이들을 잘 요약하는 과정이 필요하다.
- Statistics

Collection and analysis of data with inference as objective

- 추정 (Estimation)
 - 데이터로부터 우리가 모르는 모집단에 대한 정보를 최대한 사실(true)에 가깝게 짐작하는 것
 - Bayesian vs Non-Bayesian

확률 문제

- 문제) 항아리에 동전이 1000개 있다. 이 중 999개는 앞면과 뒷면이 있는 정상 동전이고, 1개는 앞면만 있는 동전이다. 이 항아리에서 동전을 한 개 선택해서 (정상/이상을 확인하지 않고) 10번 던졌더니 10번 모두 앞면이 나왔다. 이 동전이 앞면만 있는 동전일 확률은?

베이즈 정리 (Bayes' Theorem)

- 결합 확률 (Joint Probability): 사건 X와 Y가 동시에 발생할 확률

$$P(X, Y)$$

- 조건부 확률 (Conditional Probability): 사건 X가 발생했을 때, 사건 Y가 발생할 확률

$$P(Y|X) = \frac{P(X, Y)}{P(X)}, P(X|Y) = \frac{P(X, Y)}{P(Y)}$$

- 베이즈 정리 (Bayes' Theorem): 사건 X가 발생하였는데, 그것의 **원인이 Y일 확률** (조건부 확률, 사후확률: posterior probability)

$$P(Y|X) = \frac{P(X|Y)P(Y)}{P(X)}$$

베이즈 정리 (Bayes' Theorem)

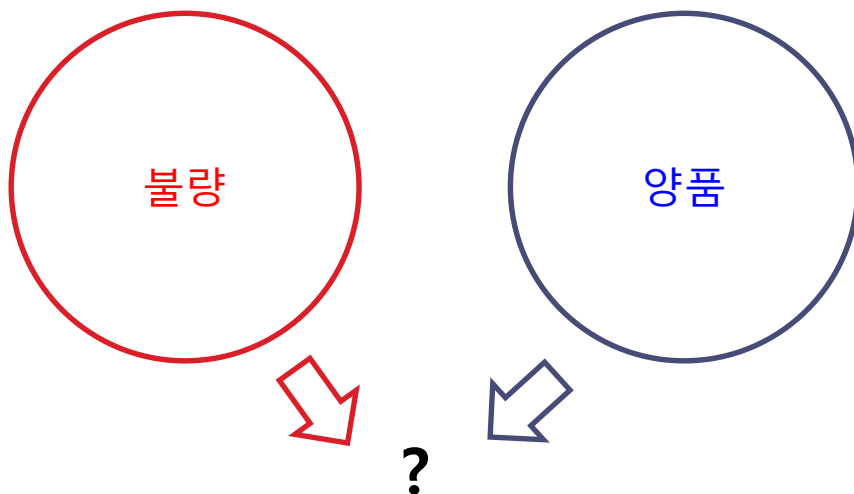
- 문제) 항아리에 동전이 1000개 있다. 이 중 999개는 앞면과 뒷면이 있는 정상 동전이고, 1개는 앞면만 있는 동전이다. 이 항아리에서 동전을 한 개 선택해서 (정상/이상을 확인하지 않고) 10번 던졌더니 10번 모두 앞면이 나왔다. 이 동전이 앞면만 있는 동전일 확률은?
- 발생한 사건 (X): 10번 던져서 모두 앞면
- 가능한 원인 1 (Y1): 앞면만 있는 동전을 선택
- 가능한 원인 2 (Y2): 앞뒷면이 있는 동전을 선택

$$P(Y1|X) = \frac{P(X|Y1)P(Y1)}{P(X)} = \frac{P(X|Y1)P(Y1)}{P(X|Y1)P(Y1) + P(X|Y2)P(Y2)}$$

베이지스 분류 (Bayesian Classification)

- 베이지스 정리를 분류 문제에 적용하면
- “ $\{X_1 = x_1, X_2 = x_2, \dots, X_p = x_p\}$ 의 값이 나온 원인은 이것이 불량군에서 나왔기 때문일까? 아니면 양품군에서 나왔기 때문일까?” (즉, 불량일까? 양품일까?)

$$P(Y|X_1, X_2, \dots, X_p) = \frac{P(X_1, X_2, \dots, X_p|Y)P(Y)}{P(X_1, X_2, \dots, X_p)}$$



$\{X_1 = x_1, X_2 = x_2, \dots, X_p = x_p\}$: 발생한 사실

베이지스 분류 (Bayesian Classification)

- 아래 식이 최대가 되는 Y값(class)으로 분류한다.

$$P(Y|X_1, X_2, \dots, X_p) = \frac{P(X_1, X_2, \dots, X_p|Y)P(Y)}{P(X_1, X_2, \dots, X_p)}$$

- 아래 식이 최대가 되는 Y값(class)으로 분류하는 것과 같다.

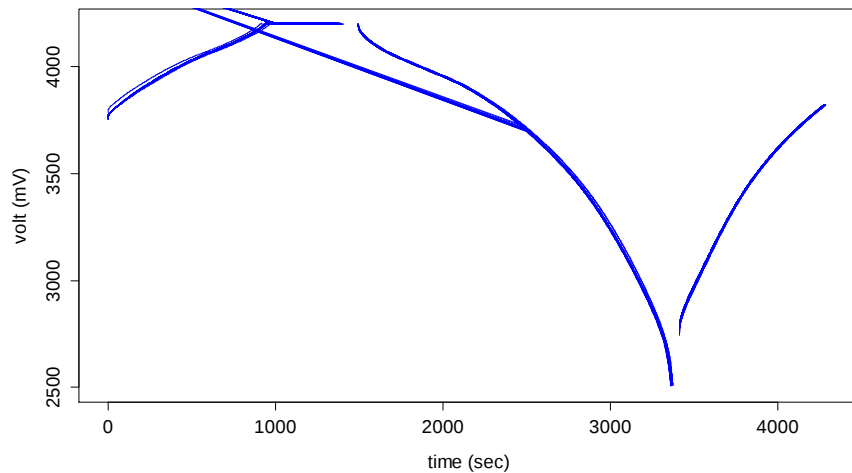
$$P(X_1, X_2, \dots, X_p|Y)P(Y)$$

- 이 확률을 데이터로부터 계산하려면? (참고. 결합확률을 계산하는 것이 쉽지 않음)

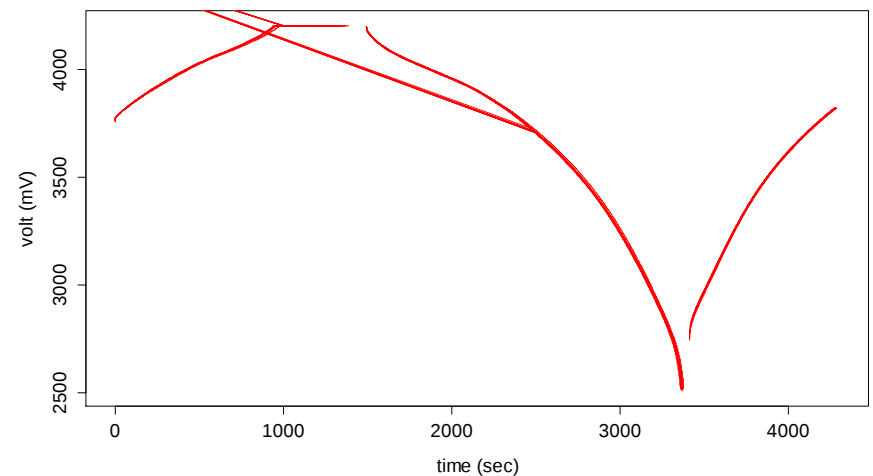
불량 배터리 탐지

- 전지(Cell) 생산과정에서 측정된 전압(mV) 데이터
- 전압 시계열 데이터를 이용한 양품과 불량률의 예측 모델링

양품



불량



불량 배터리 탐지

- 불량 배터리 탐지 모형을 만들기 위한 기본 아이디어
 - "이런 전압변화를 보이는 것은 양품군에서 나왔기 때문일까, 아니면 불량군에서 나왔기 때문일까?"

		예측	
실제	학습 (train)	양품	불량
	양품	4054	458
	사내불량	0	66

		예측	
실제	평가 (test)	양품	불량
	양품	3734	779
	사외불량	0	3
	사내불량	0	65

양품 정분류율: 82.7%

불량 정분류율: 100%

최적화 (Optimization)

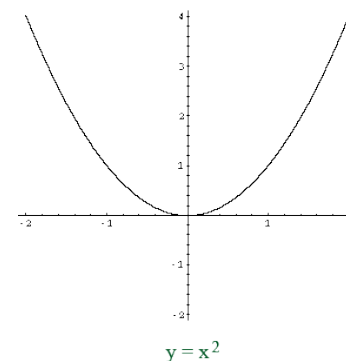
최적화 문제 (Optimization Problem)

Given $\mathbf{x} = (x_1, x_2, \dots, x_n)^\top \rightarrow$ **Decision variable**

minimize $f(\mathbf{x}) \rightarrow$ **Objective function**

subject to $g_1(\mathbf{x}) \leq a_1$
 $g_2(\mathbf{x}) \leq a_2$
 $\vdots$
 $g_m(\mathbf{x}) \leq a_m \rightarrow$ **A set of constraints**

(예) $\arg \min_x y = x^2$

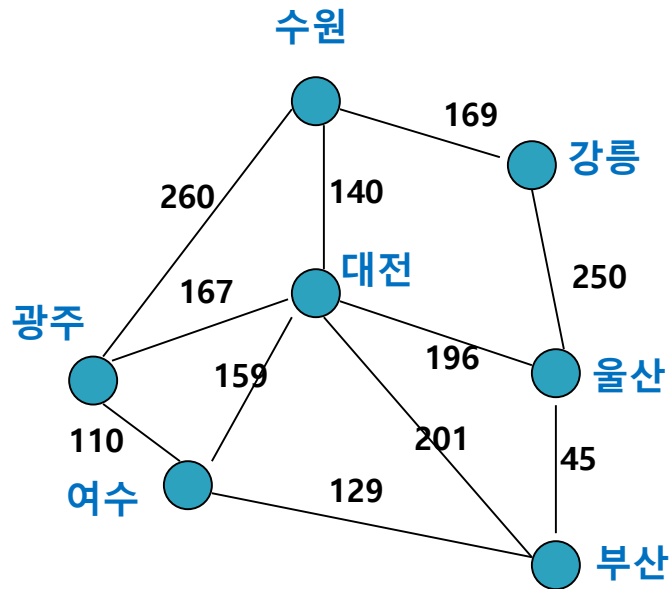


참고. $\arg \min :=$ arguments of the minima

최적화 (Optimization)

외판원 문제 (Salesman Travelling Problem)

이번 출장에는 광주, 여수, 대전, 강릉, 울산, 부산을 거쳐서 다시 수원으로 돌아와야 하는데 어떻게 이동하는 것이 좋을까?



최적화 (Optimization)

외판원 문제 (Salesman Travelling Problem)

$$x_{ij} = \begin{cases} 1, & \text{if path goes from city } i \text{ to city } j \\ 0, & \text{otherwise} \end{cases}$$

$$\text{minimize} \quad \sum_{i=1}^n \sum_{j=1, j \neq i}^n d_{ij} x_{ij}$$

$$\text{subject to} \quad \sum_{i=1, i \neq j}^n x_{ij} = 1, \quad j = 1, 2, \dots, n$$

$$\sum_{j=1, j \neq i}^n x_{ij} = 1, \quad i = 1, 2, \dots, n$$

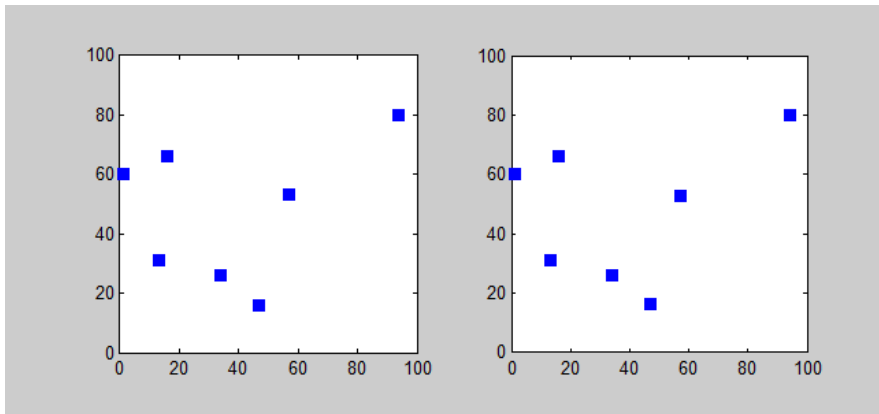
$$u_i - u_j + nx_{ij} \leq n - 1, \quad 2 \leq i \neq j \leq n$$

$$0 \leq u_i \leq n - 1, \quad 2 \leq i \leq n$$

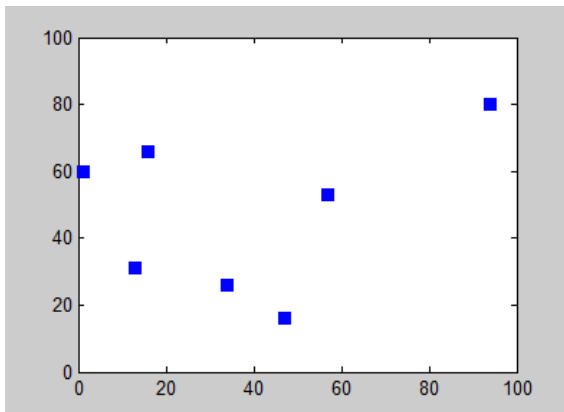
최적화 (Optimization)

외판원 문제 (Salesman Travelling Problem)

- 문제를 푸는 방법은 아래와 같이 여러가지 (그림출처. Wikipedia)



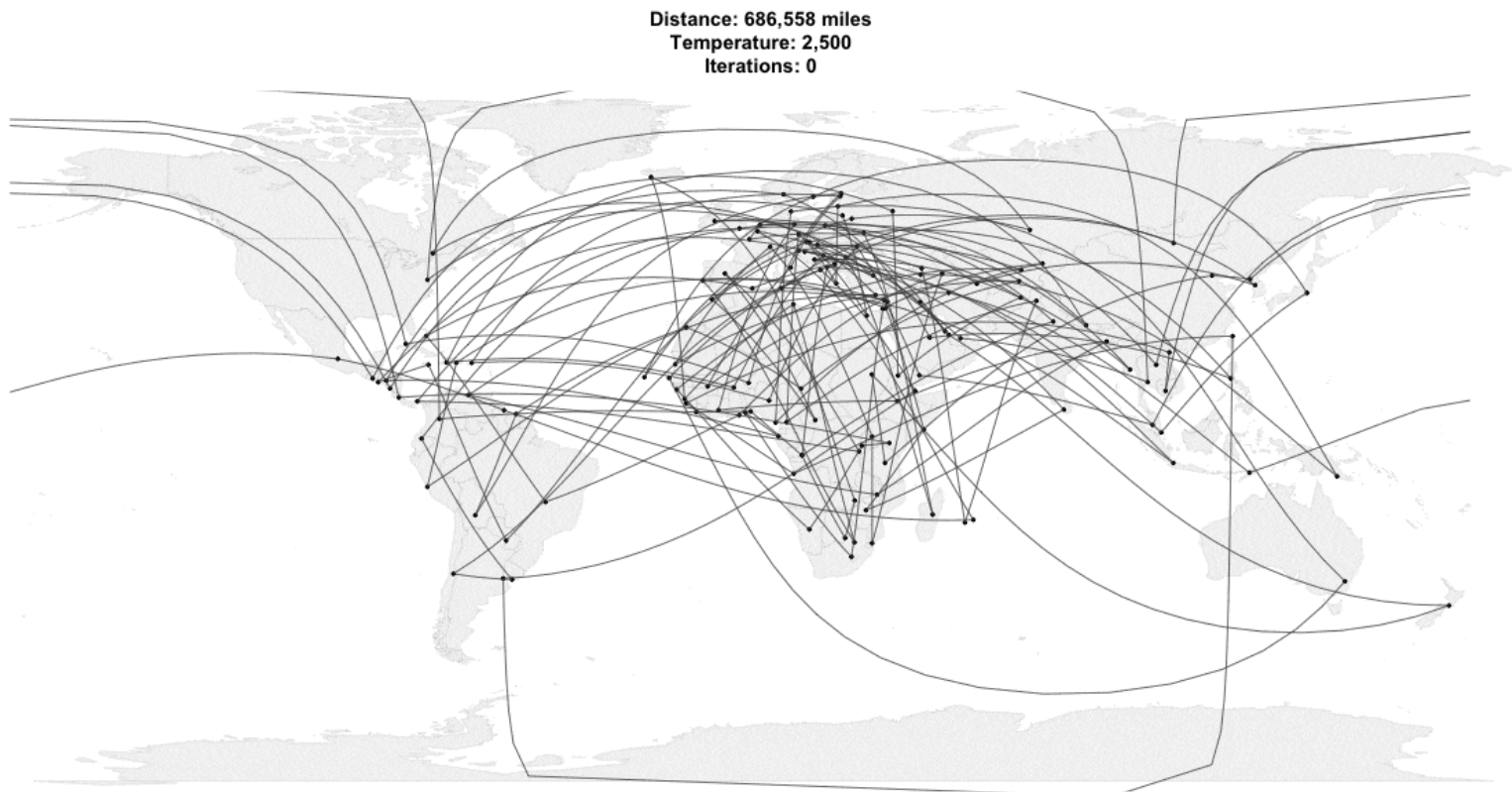
(Branch and Bound Algorithm)



(Nearest Neighbor Algorithm)

최적화 (Optimization)

외판원 문제 (Salesman Travelling Problem)



(그림출처. <http://toddschneider.com/page/6/>)

최적화 (Optimization)

다른 최적화 문제들

- 공항의 항공기 주기장 배정 문제
 - 승객의 이동거리 최소, 항공기 대기시간 최소, 주기장 이용률 최대
- 고속도로 톨게이트 운영 문제
 - 차량의 대기시간 최소, 게이트 수 (비용) 최소
- 전기차 충전소 배치 문제
 - 소모 전력 밸런싱, 차량의 유동량 고려
- 병원의 병상 배정 문제
 - 병상 가동률 최대, 급에 따른 병실료 최소
- 생산공정의 스케줄링 문제
 - 작업 종료시점 최소, 납기일 맞추기, 설비 가동률 최대

최적화 (Optimization)

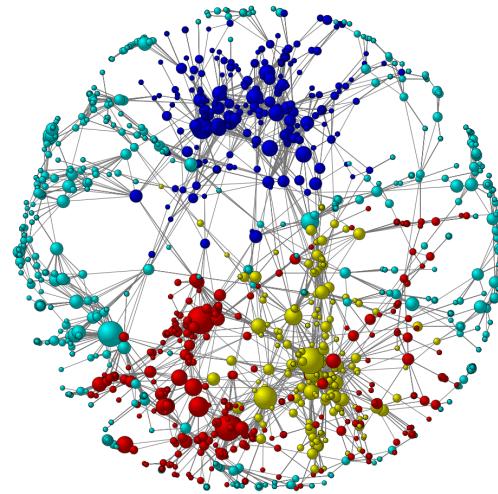
최적화 문제의 분류

- Continuous Optimization vs Discrete Optimization
 - Linear Programming, Nonlinear Programming
 - Integer Programming, Mixed Integer Programming
- Unconstrained Optimization vs Constrained Optimization
 - None (feasibility), One, or Many (Conflicting) Objectives
- Deterministic Optimization vs Stochastic Optimization

k-Means Clustering

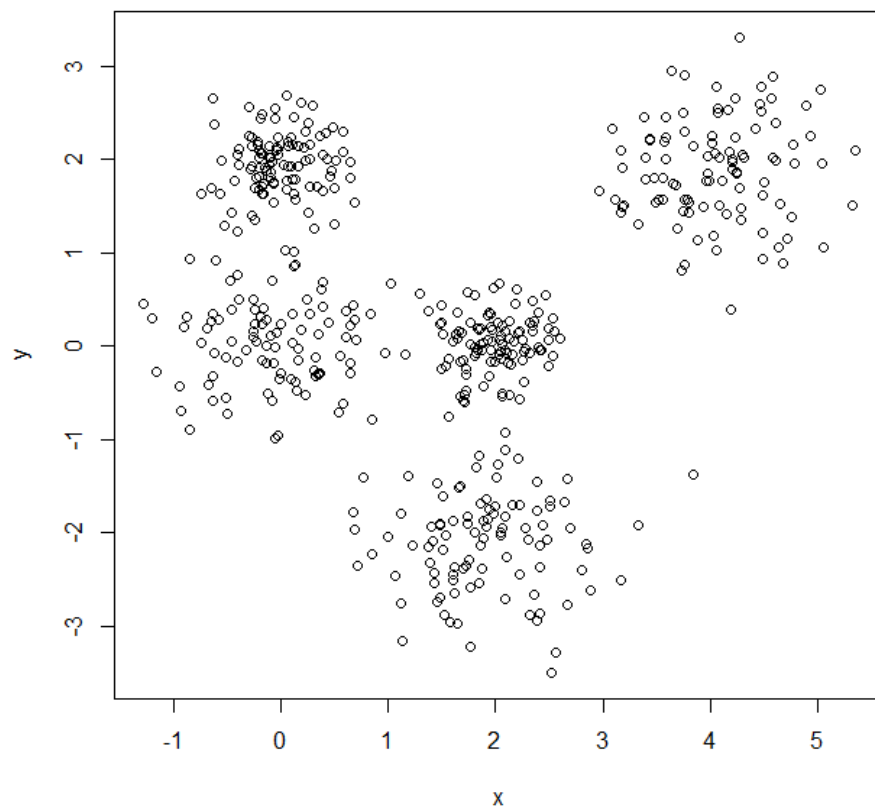
Clustering (or Unsupervised Learning)

- 군집분석 또는 비지도학습
- 목적: 가까운 개체(object)들끼리 잘 묶자.
- 개체: 도시, 고객, 제품 등등 여러분이 정의할 수 있는 거의 모든 것



k-Means Clustering

Clustering (or Unsupervised Learning)



k-Means Clustering

Optimization formulation

- k-means clustering 은 최적화 문제 $a_{ij} = \begin{cases} 1, & \text{if object } i \text{ belongs to cluster } j \\ 0, & \text{otherwise} \end{cases}$

$$\text{minimize} \quad \sum_{i=1}^n \sum_{j=1}^k d(\mathbf{x}_i, \mathbf{c}_j) a_{ij}$$

$$\text{subject to} \quad \mathbf{c}_j = \frac{\sum_{i=1}^n \mathbf{x}_i a_{ij}}{\sum_{i=1}^n a_{ij}}, \quad j = 1, 2, \dots, k$$

$$\sum_{j=1}^k a_{ij} = 1, \quad i = 1, 2, \dots, n$$

$$\sum_{i=1}^n a_{ij} \geq 1, \quad j = 1, 2, \dots, k$$

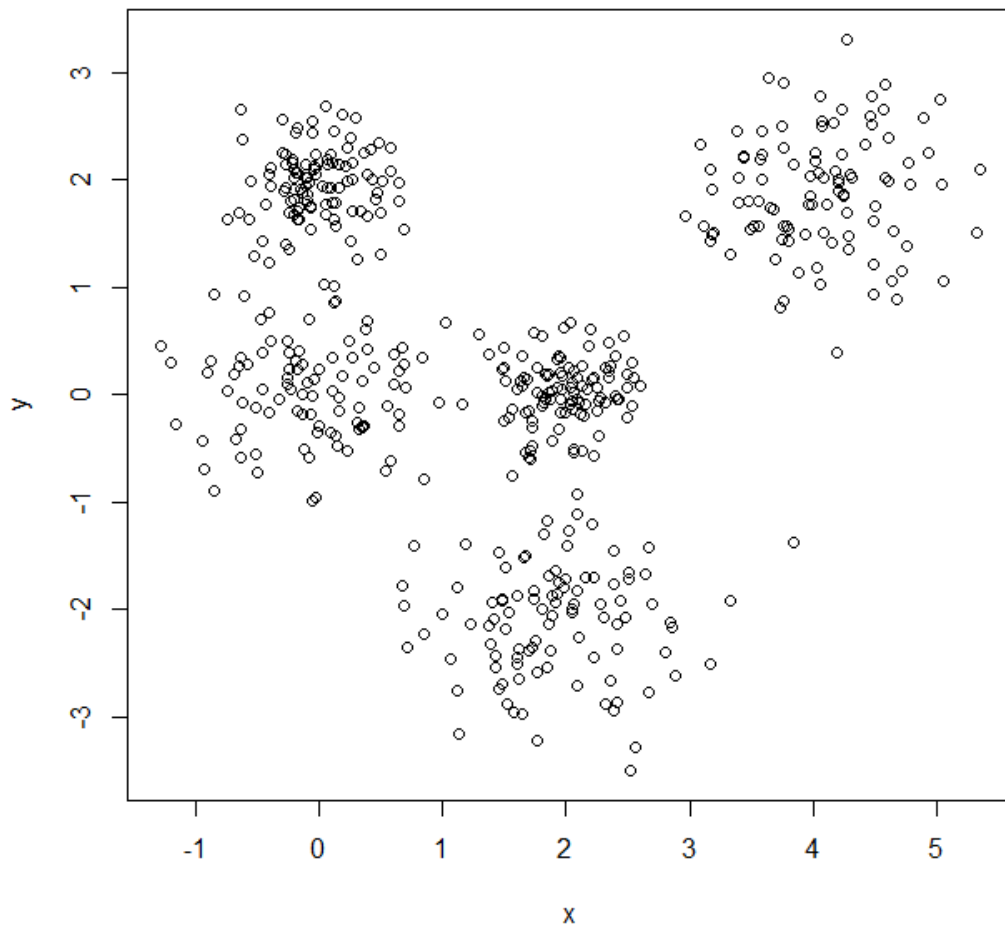
k-Means Clustering

Algorithm

- 단계 0
 - 초기 k개의 중심개체를 무작위로 선택한다.
- 단계 1 (할당)
 - 중심개체가 아닌 다른 개체들과 중심개체 사이의 거리를 계산한다.
 - 각 개체에 대해서 가장 가까운 중심개체로 할당한다. (군집 형성)
- 단계2 (새로운 중심개체)
 - 각 군집에 대해서 새로운 중심개체를 계산한다.
- 단계3 (반복 또는 수렴)
 - 만약 이전 군집과 결과가 동일하면 계산을 멈춘다.
 - 그렇지 않다면 단계 1로 이동한다.

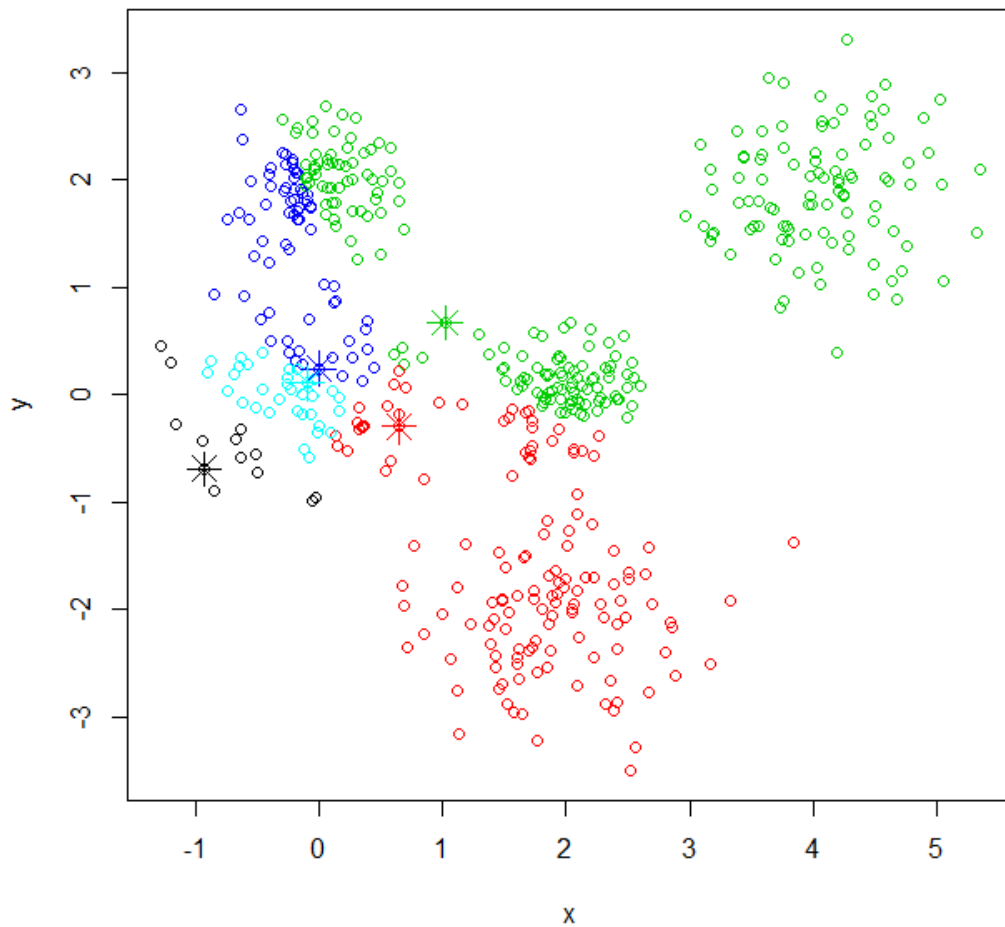
k-Means Clustering

- 데모 예제 ($k=5$)



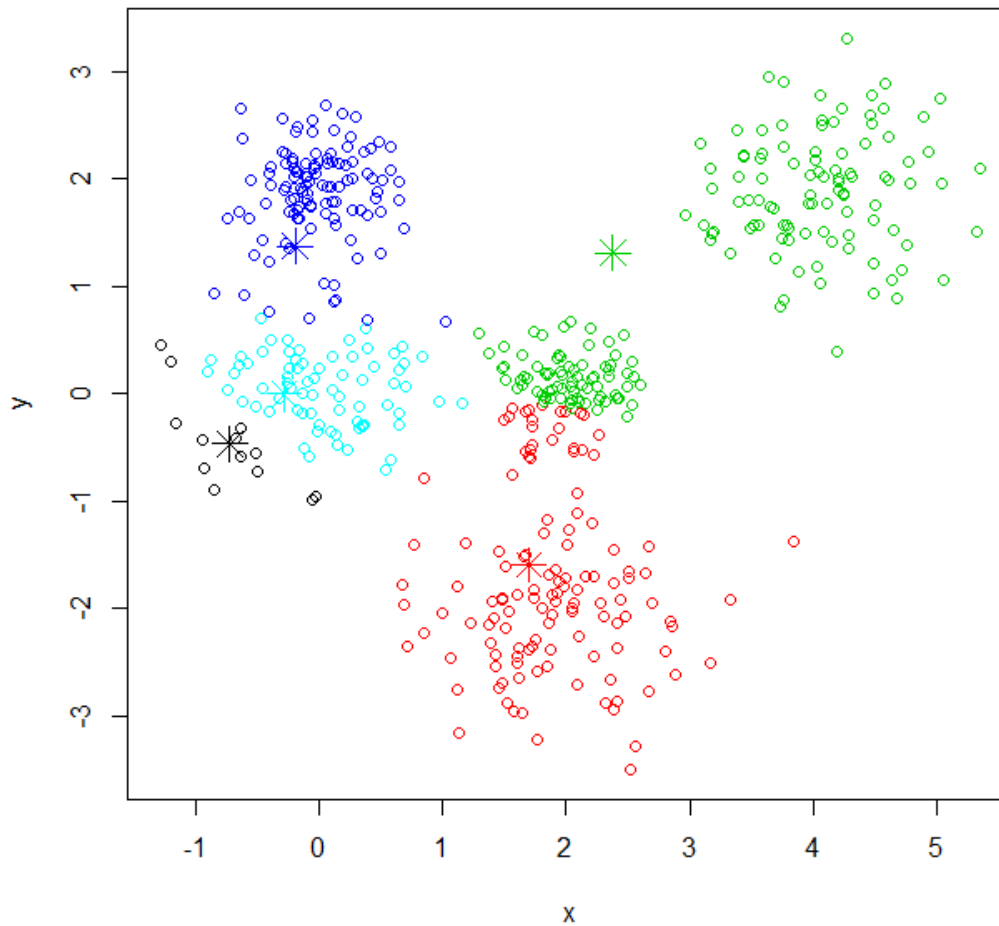
k-Means Clustering

- 초기 군집



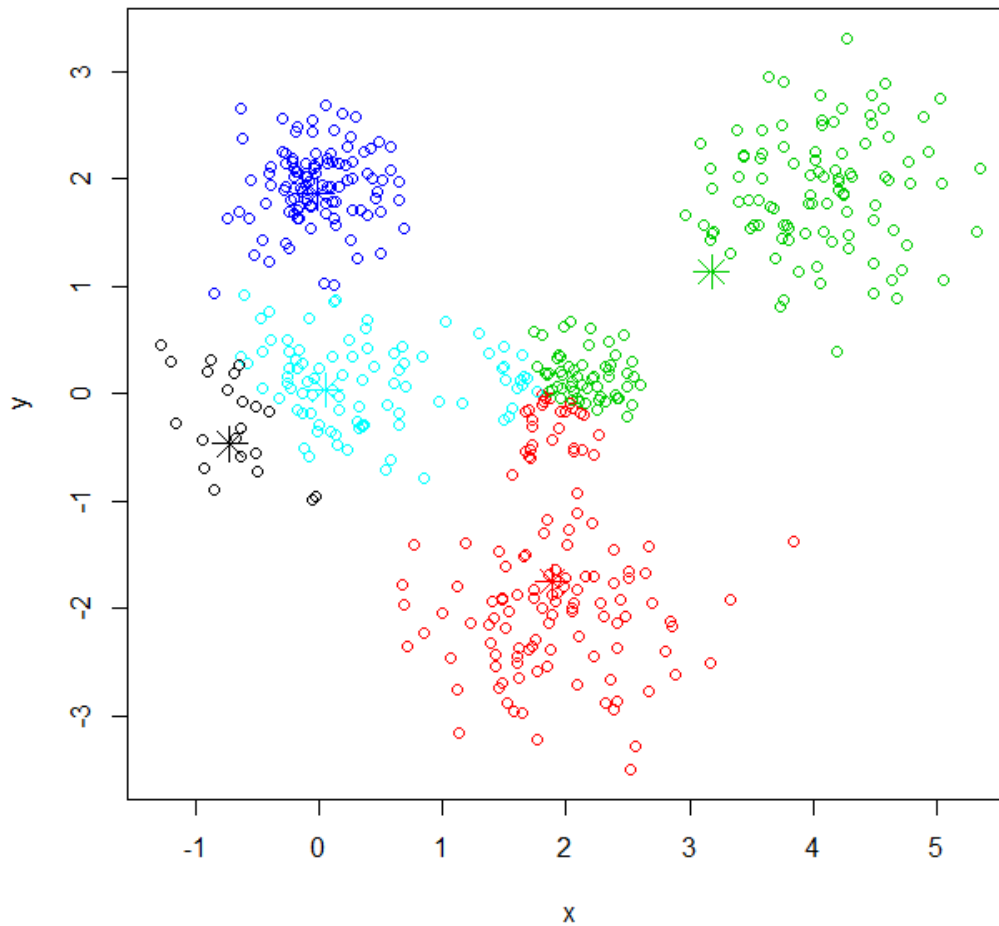
k-Means Clustering

- 반복 1



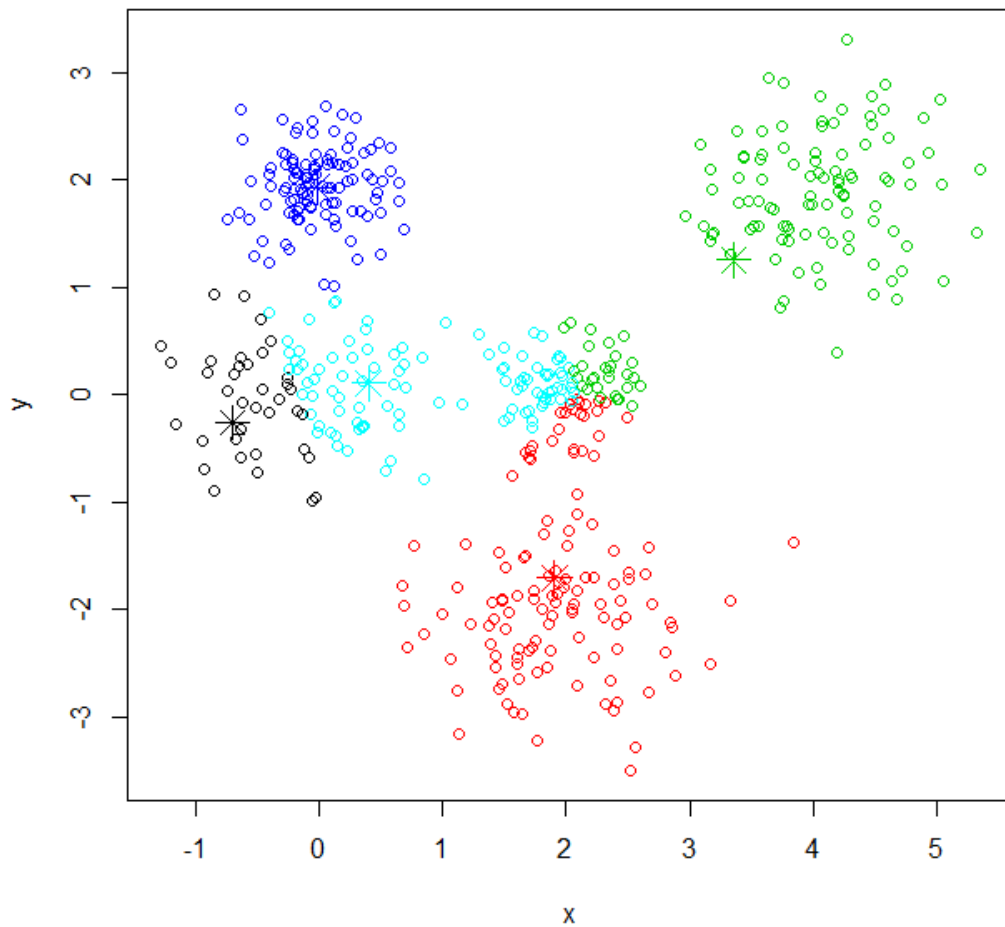
k-Means Clustering

- 반복 2



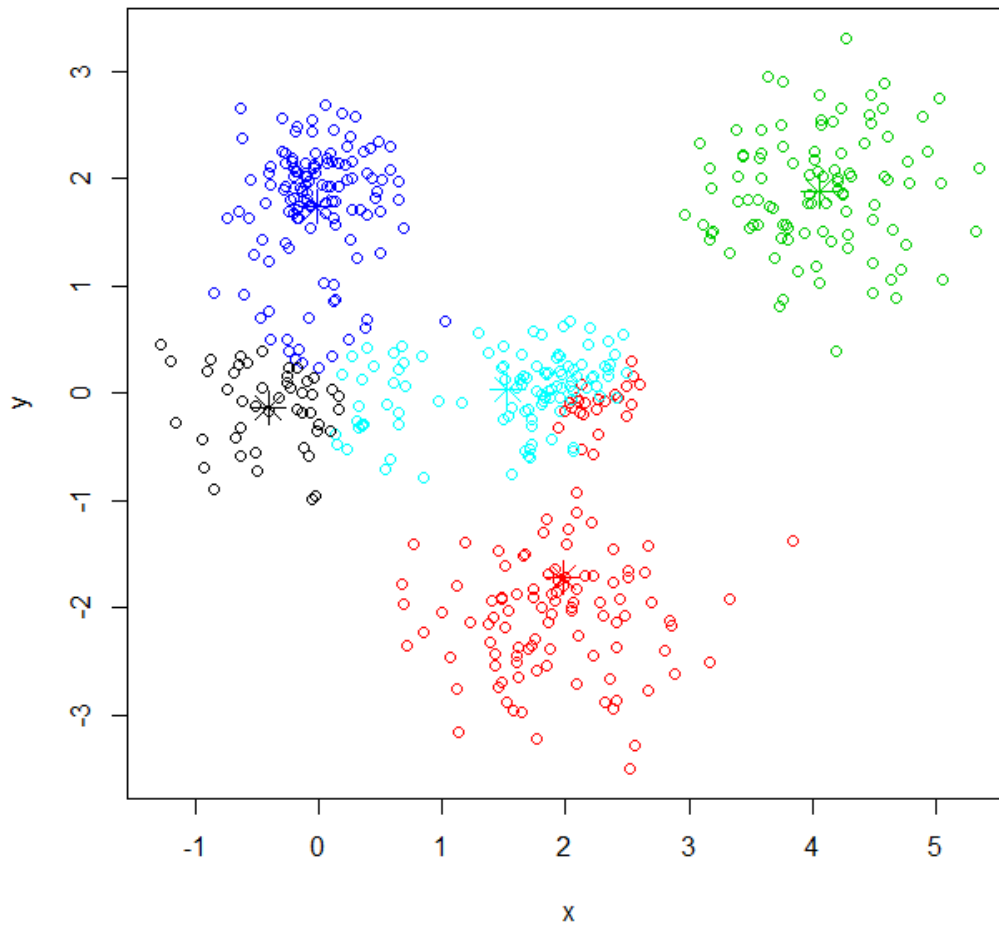
k-Means Clustering

- 반복 3



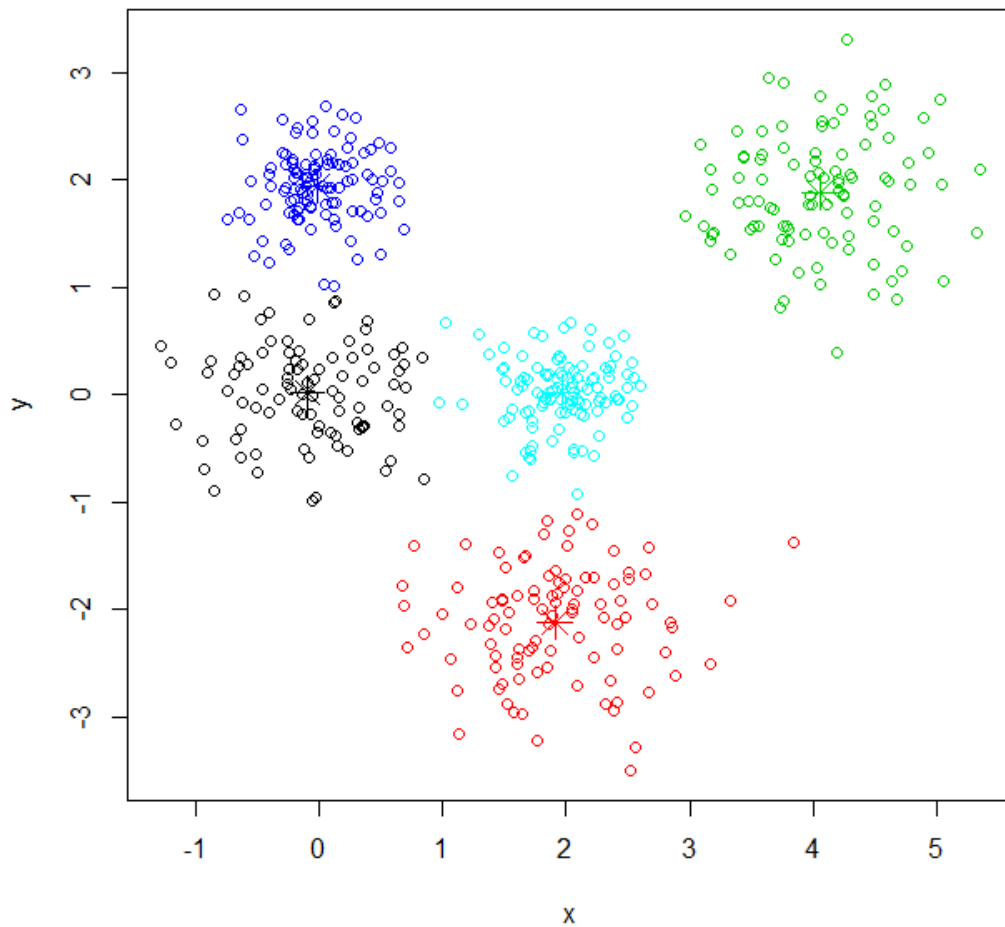
k-Means Clustering

- 반복 4

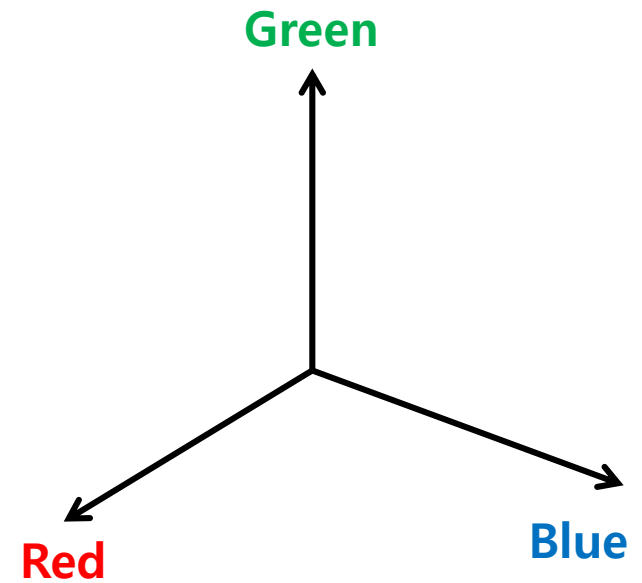
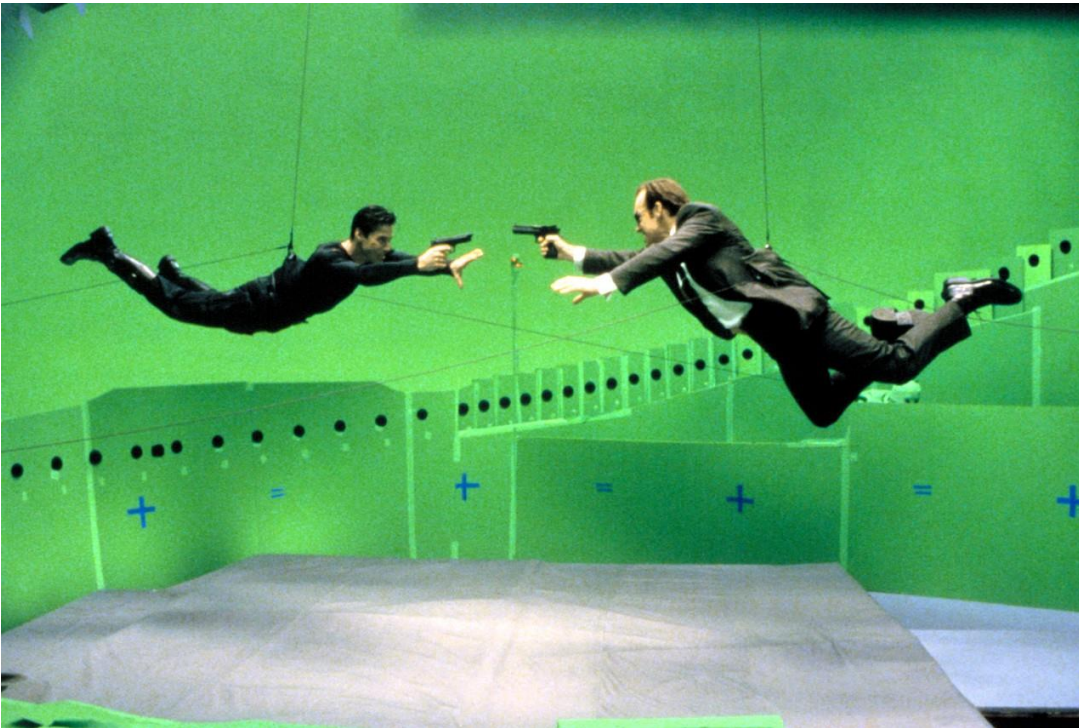


k-Means Clustering

- 반복 5 (수렴)

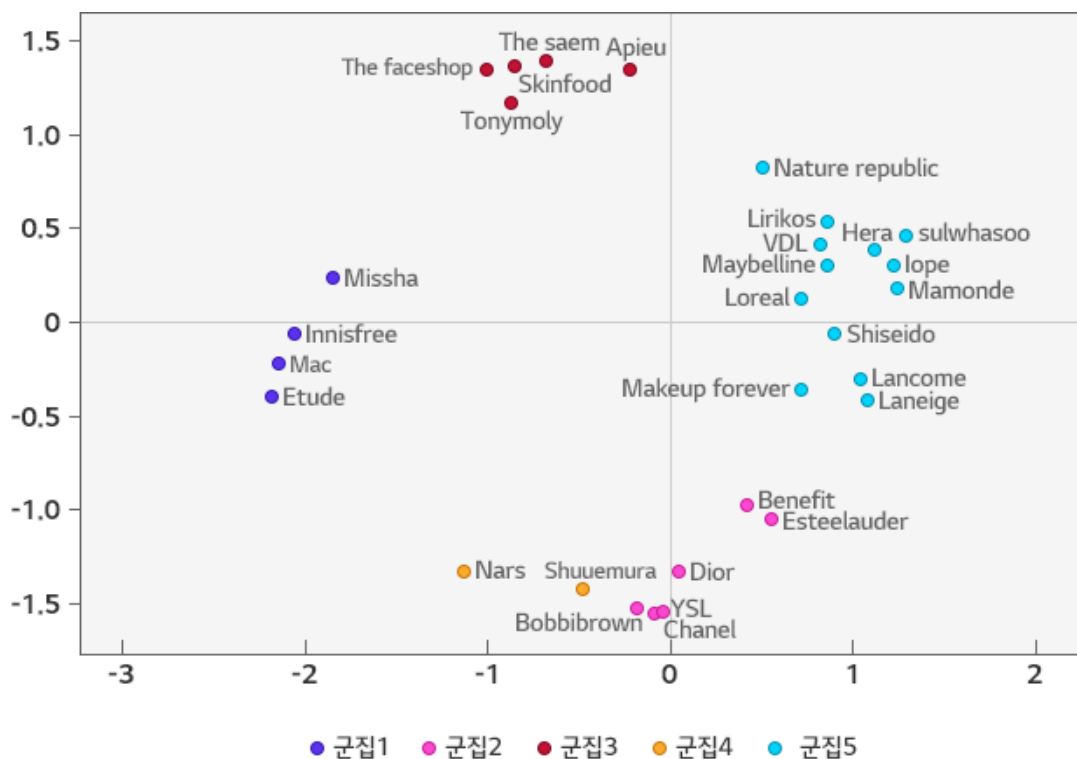


Why Clustering?



Why Clustering?

SNS 데이터 활용

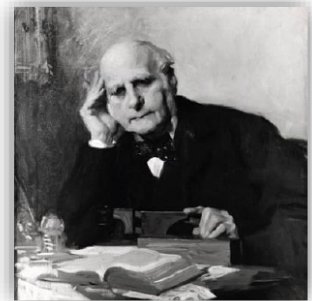
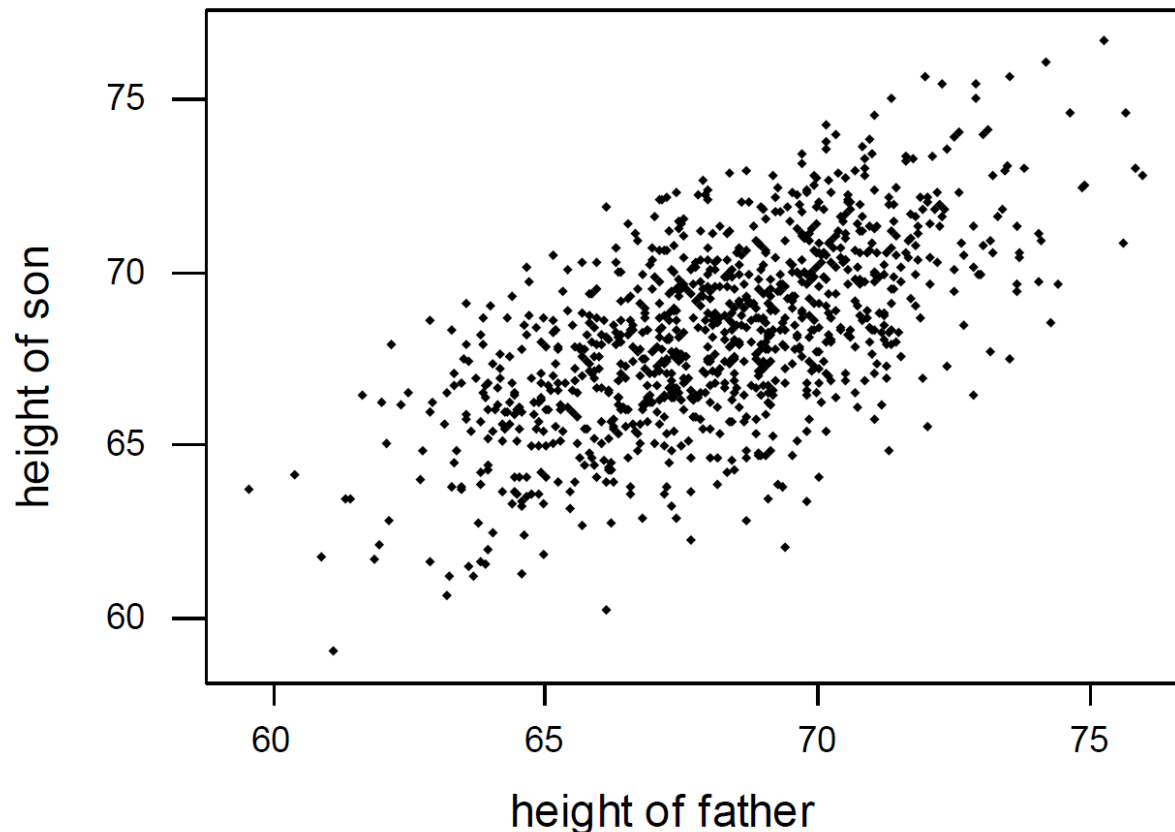


<https://blog.lgcns.com/1305>

- 군집 1 & 3
 - 저가 브랜드, 10대 후반 ~ 20대 초반
- 군집 2
 - 해외 명품 브랜드
 - 색조 화장품과 향수 제품
- 군집 4
 - 한 가지 제품에 특화된 브랜드
 - 나스-립스틱, 슈에무라-아이브로우
- 군집 5
 - 고가 브랜드, 기능성 제품

회귀분석: 변수 사이의 관계

$$y = f(x) \quad \text{vs.} \quad y = f(x) + \epsilon, \quad \epsilon \sim N(0, \sigma^2)$$



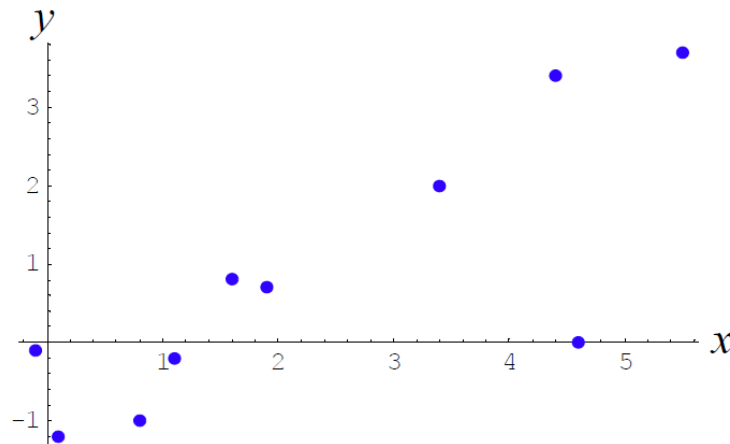
Francis Galton
(1822~1911)

직선 찾기

- 데이터가 주어졌다. (질문) 직선을 어떻게 찾을 것인가?
- (예) 10개의 관측치들 $(x_1, y_1), \dots, (x_{10}, y_{10})$

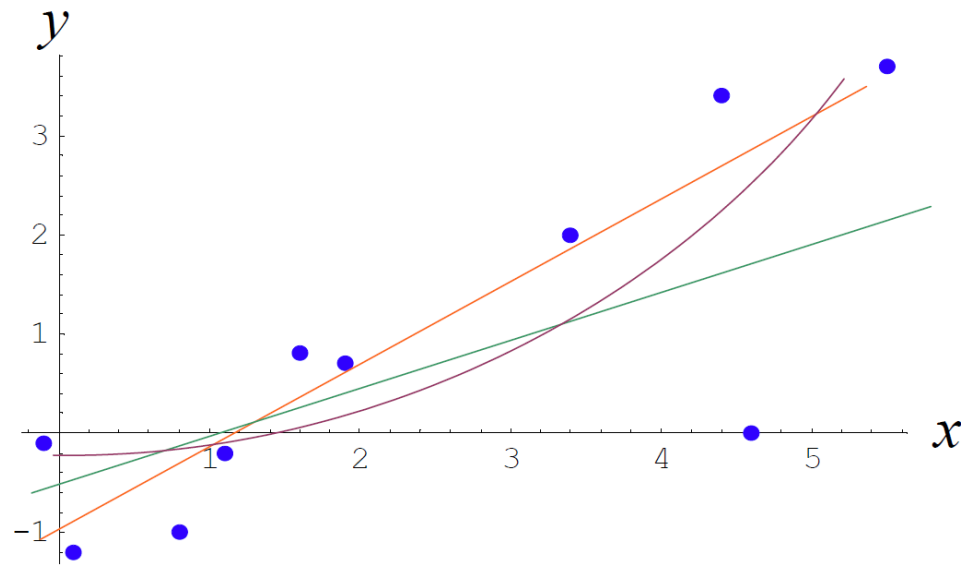
x_i	1.9	0.8	1.1	0.1	-0.1	4.4	4.6	1.6	5.5	3.4
y_i	0.7	-1.0	-0.2	-1.2	-0.1	3.4	0.0	0.8	3.7	2.0

- 위 데이터를 2차원 공간에 점으로 표시하면 아래와 같다. (질문) 이 점들을 대표하는 직선은?



직선 찾기

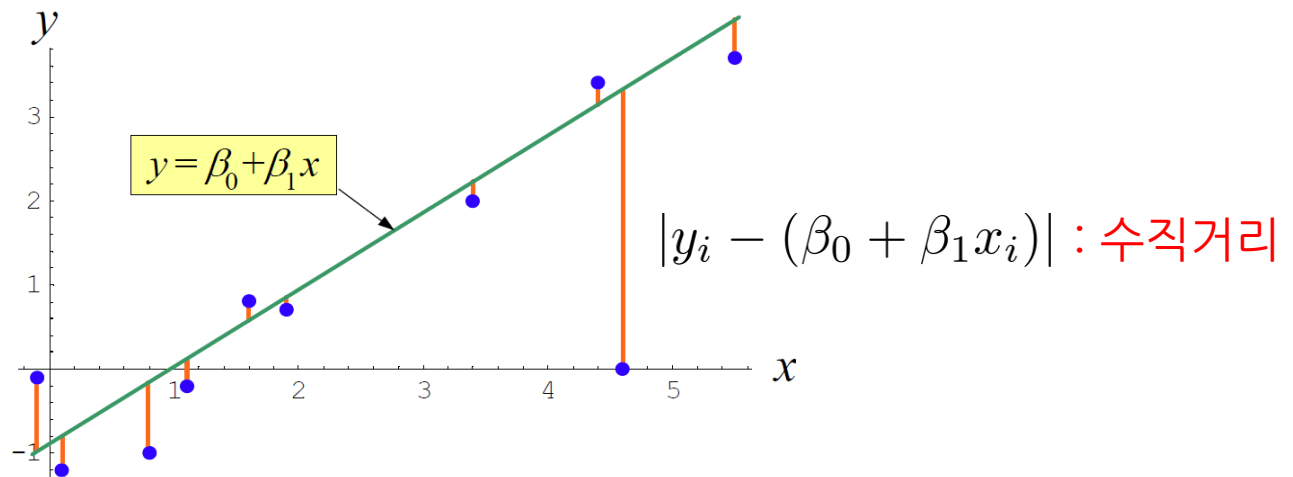
- 가능한 직선은 무수히 많다. (참고: 가능한 곡선도 무수히 많다.)



- (질문) 어떤 직선이 가장 좋은가?

직선 찾기

- 점들을 “잘” 관통하는 직선이 가장 좋은 선이 아닐까?
- “잘 관통하는 직선”을 “**직선과 점들 사이의 거리가 최소가 되는 직선**”으로 정의하자.
- 최소자승법 (Method of Least Squares)



$$Q = \sum_{i=1}^n [y_i - (\beta_0 + \beta_1 x_i)]^2$$

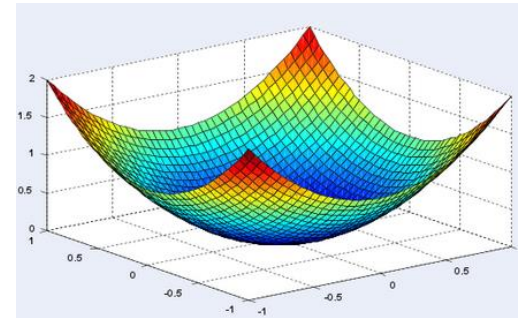
직선 찾기

- (풀이) minimize $Q = \sum_{i=1}^n [y_i - (\beta_0 + \beta_1 x_i)]^2$

$$\frac{\partial Q}{\partial \beta_0} = -2 \sum_{i=1}^n (y_i - \beta_0 - \beta_1 x_i) = 0$$

$$\frac{\partial Q}{\partial \beta_1} = -2 \sum_{i=1}^n (y_i - \beta_0 - \beta_1 x_i) x_i = 0$$

} Normal equations



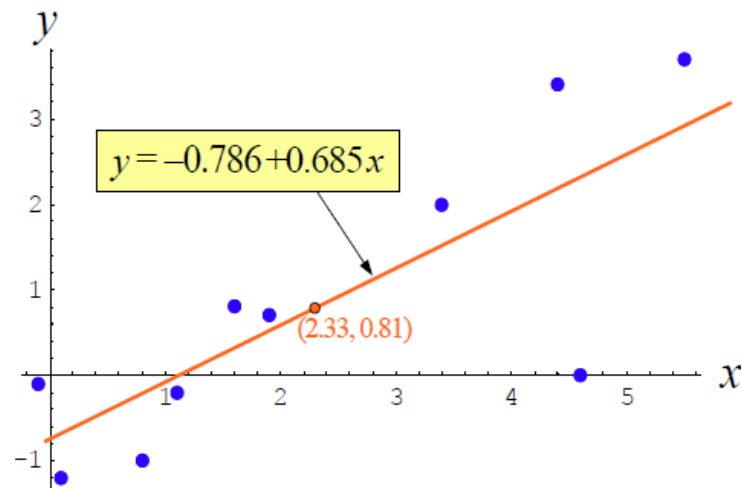
- (답)

$$\beta_1 = \frac{\sum_{i=1}^n (x_i - \bar{x}_n)(y_i - \bar{y}_n)}{\sum_{i=1}^n (x_i - \bar{x}_n)^2} = \frac{\sum_{i=1}^n x_i y_i - n \bar{x}_n \bar{y}_n}{\sum_{i=1}^n x_i^2 - n \bar{x}_n^2}$$

$$\beta_0 = \bar{y}_n - \beta_1 \bar{x}_n$$

직선 찾기

- 앞의 데이터에 대해서



$$\beta_0 = -0.786, \beta_1 = 0.685 \Rightarrow y = -0.786 + 0.685x$$

다중 회귀 분석

- x 가 2개인 경우에는? $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2$

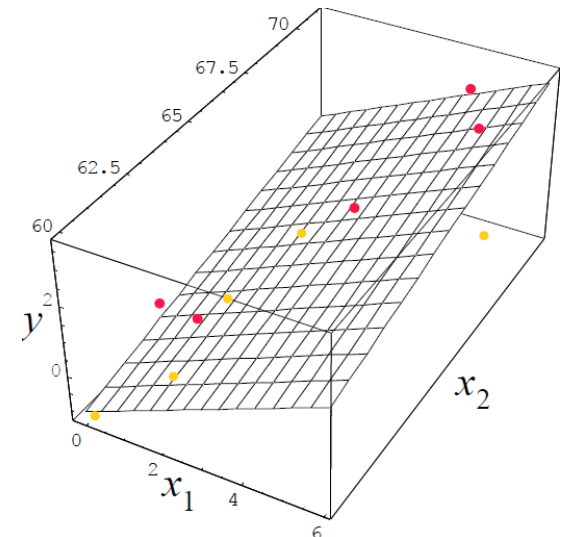
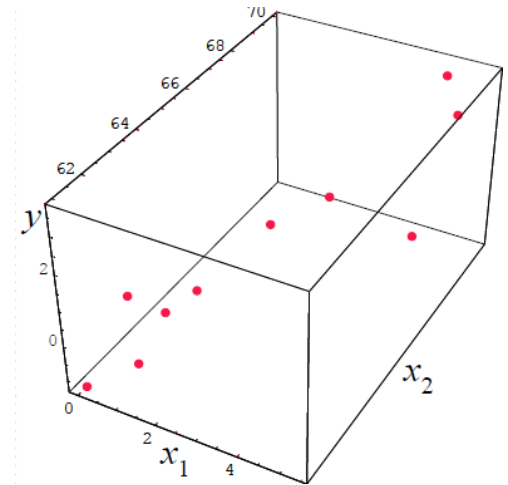
$$\text{minimize } Q = \sum_{i=1}^n [y_i - (\beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i})]^2$$

- 미지수가 3개인 연립방정식

$$\beta_0 n + \beta_1 \sum_{i=1}^n x_{i1} + \beta_2 \sum_{i=1}^n x_{i2} = \sum_{i=1}^n y_i$$

$$\beta_0 \sum_{i=1}^n x_{i1} + \beta_1 \sum_{i=1}^n x_{i1}^2 + \beta_2 \sum_{i=1}^n x_{i1} x_{i2} = \sum_{i=1}^n x_{i1} y_i$$

$$\beta_0 \sum_{i=1}^n x_{i2} + \beta_1 \sum_{i=1}^n x_{i2} x_{i1} + \beta_2 \sum_{i=1}^n x_{i2}^2 = \sum_{i=1}^n x_{i2} y_i$$



다중 회귀 분석

- x 가 2개 이상인 경우에는? $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_p x_p$

$$\text{minimize } Q = \sum_{i=1}^n [y_i - (\beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \cdots + \beta_p x_{pi})]^2$$

- 아... 쓰기 힘들어 

다중 회귀 분석

- 벡터와 행렬을 이용해서 표현해 보자.

$$\begin{pmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{pmatrix} = \begin{pmatrix} 1 & x_{11} & \cdots & x_{p1} \\ 1 & x_{12} & \cdots & x_{p2} \\ \vdots & \vdots & \vdots & \vdots \\ 1 & x_{1n} & \cdots & x_{pn} \end{pmatrix} \begin{pmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_p \end{pmatrix} + \begin{pmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_n \end{pmatrix} \Rightarrow \mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$$

다중 회귀 분석

- 이제 쓰기가 다소 편해진 듯...

$$\text{minimize } Q = \sum_{i=1}^n [y_i - (\beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \cdots + \beta_p x_{pi})]^2$$

$$= (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})^\top (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})$$

$$= \mathbf{y}^\top \mathbf{y} - \mathbf{y}^\top \mathbf{X}\boldsymbol{\beta} - \boldsymbol{\beta}^\top \mathbf{X}^\top \mathbf{y} + \boldsymbol{\beta}^\top \mathbf{X}^\top \mathbf{X}\boldsymbol{\beta}$$

$$= \mathbf{y}^\top \mathbf{y} - 2\boldsymbol{\beta}^\top \mathbf{X}^\top \mathbf{y} + \boldsymbol{\beta}^\top \mathbf{X}^\top \mathbf{X}\boldsymbol{\beta}$$

$$\frac{\partial Q}{\partial \boldsymbol{\beta}} = -2\mathbf{X}^\top \mathbf{y} + 2\mathbf{X}^\top \mathbf{X}\boldsymbol{\beta} = 0$$

↓

$$(\mathbf{X}^\top \mathbf{X})\boldsymbol{\beta} = \mathbf{X}^\top \mathbf{y} \rightarrow \hat{\boldsymbol{\beta}} = (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top \mathbf{y}$$

(Normal equation)

선형대수

$$\hat{\beta} = (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top \mathbf{y}$$

- 최소한 전치행렬, 역행렬이 무엇인지는 알아야겠군...
- 역행렬이라면...

$$A = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \quad A^{-1} = \frac{1}{ad - bc} \begin{pmatrix} d & -b \\ -c & a \end{pmatrix}$$

- 역행렬을 구할 수 있으려면...

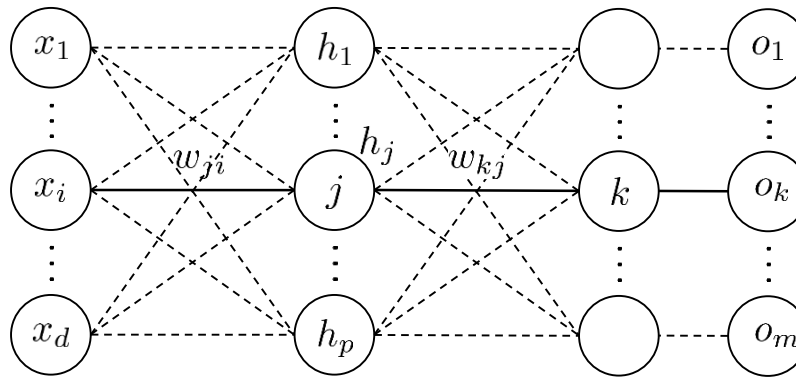
$$ad - bc \neq 0$$

선형대수

- $\mathbf{X}$ 는 $(n \times (p + 1))$ matrix. $\rightarrow \mathbf{X}^\top \mathbf{X}$ 는 $((p + 1) \times (p + 1))$ symmetric matrix.
- Vectors $\{\mathbf{1}, \mathbf{x}_1, \dots, \mathbf{x}_p\}$ are linearly independent. (기본가정)
- Therefore, $rank(\mathbf{X}) = p + 1$ (full column rank)
- Therefore, $\mathbf{X}^\top \mathbf{X}$ is a positive definite matrix.
- Therefore, $rank(\mathbf{X}^\top \mathbf{X}) = p + 1$ (full rank)
- Therefore, $(\mathbf{X}^\top \mathbf{X})^{-1}$ 이 반드시 존재한다.

신경망 (Neural Networks)

- 아래와 같은 신경망 구조를 생각해 보자.



- 주어진 데이터 즉, 학습을 위한 데이터가 아래와 같다고 하면

$$\begin{array}{lll}
 D_1 = (x_{11}, x_{12}, \dots, x_{1d}, t_{11}, t_{12}, \dots, t_{1m}) & & (o_{11}, o_{12}, \dots, o_{1m}) \\
 D_2 = (x_{21}, x_{22}, \dots, x_{2d}, t_{21}, t_{22}, \dots, t_{2m}) & & (o_{21}, o_{22}, \dots, o_{2m}) \\
 \vdots & & \vdots \\
 D_N = (\underbrace{x_{N1}, x_{N2}, \dots, x_{Nd}}_{\text{독립변수 값 (inputs)}}, \underbrace{t_{N1}, t_{N2}, \dots, t_{Nm}}_{\text{종속변수 값 (targets)}}) & & (\underbrace{o_{N1}, o_{N2}, \dots, o_{Nm}}_{\text{예측값 (outputs)}})
 \end{array}$$

신경망 (Neural Networks)

- 우리는 아래 정의된 에러를 최소로 하는 $\mathbf{w}$ 를 찾아야 한다.

$$E(\mathbf{w}) = \sum_{n=1}^N E_n(\mathbf{w}) , E_n(\mathbf{w}) = \frac{1}{2} \sum_{k=1}^m (t_{nk} - o_{nk})^2$$

- 어떻게? 비선형 최적화!!
- 그런데 사실...

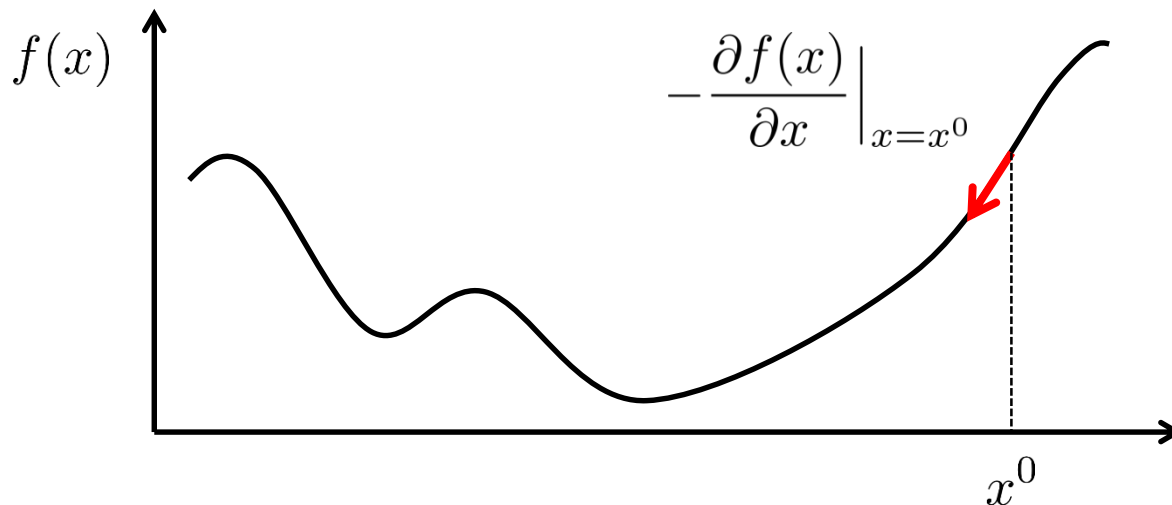
$$E_n(\mathbf{w}) = \frac{1}{2} \sum_{k=1}^m \left(t_k - \frac{1}{1 + \exp \left(- \left(w_{k0} + \sum_{j=1}^p w_{kj} \left(\frac{1}{1 + \exp \left(- (w_{j0} + \sum_{i=1}^d w_{ji} x_i) \right) } \right) \right) \right) \right)^2$$

- 미분이 어려워... 

신경망 (Neural Networks)

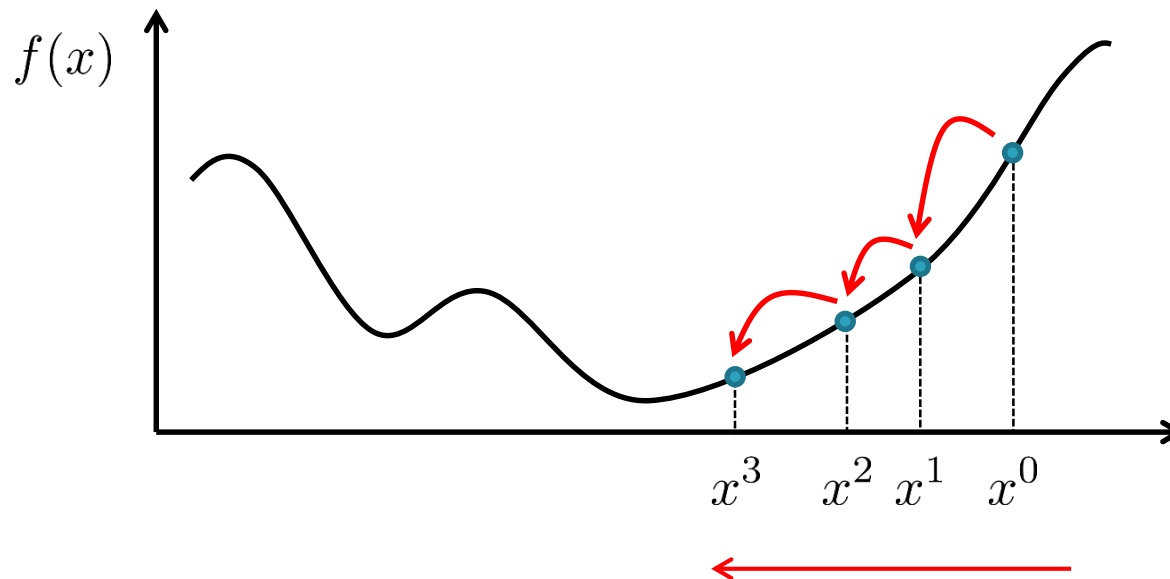
- 그림에도 불구하고, 방법은 있다. (Discovery is waiting for you)
- 여하튼, 기본 아이디어는 비선형 최적화!
- 경사하강법 (Gradient Descent Method)
 - 비선형 최적화 문제(nonlinear optimization problem)를 해결하는 가장 기본적인 방법 중 하나

$$\arg \min_x f(x)$$



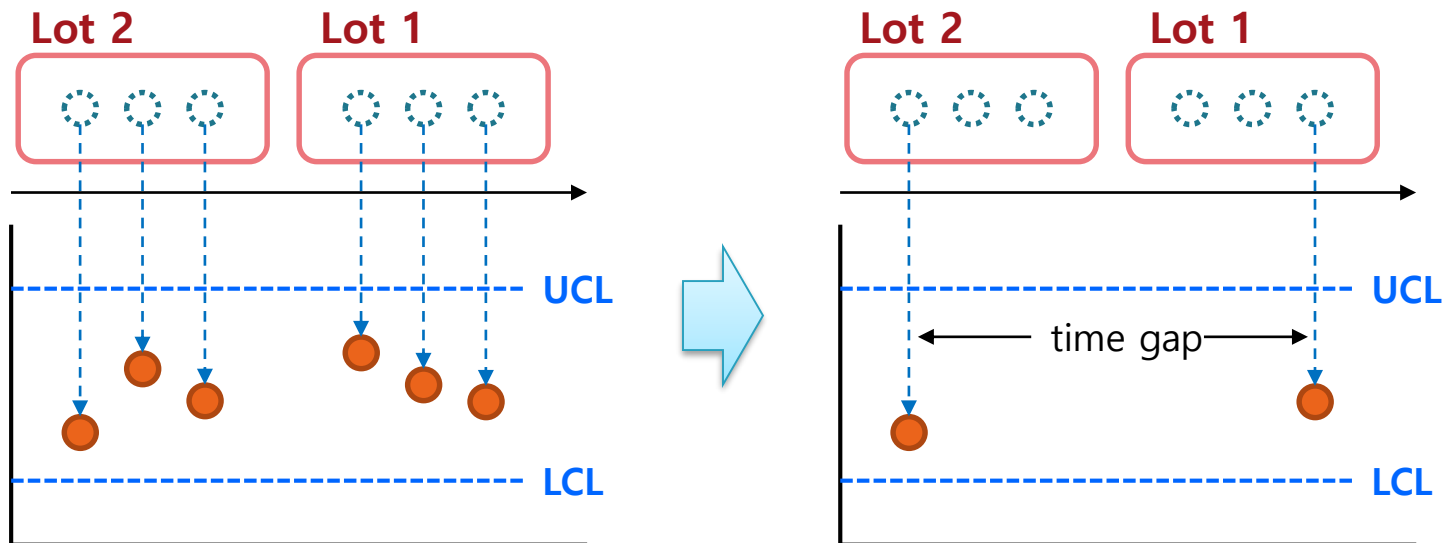
신경망 (Neural Networks)

- 경사하강법 (Gradient Descent Method)



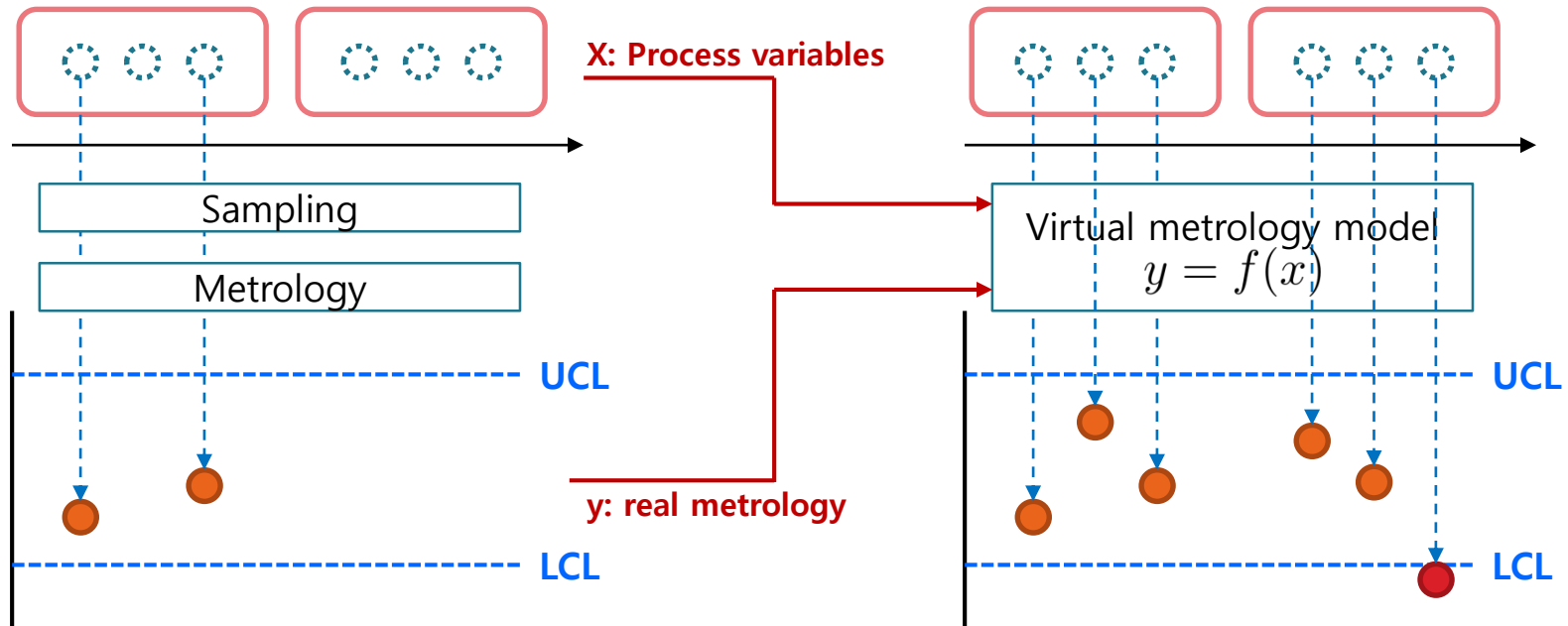
전수품질검사 vs 샘플링을 통한 품질검사

- 모든 제품을 검사하는 것은 시간과 비용이 많이 소요되어 현실적으로 불가능 함
- 일반적으로 1개의 Lot에서 1개 또는 2개의 제품을 샘플링하여 품질검사
- 문제점
 - Lot 기반의 품질검사는 시간 차이로 인하여 불량률 낮게 탐지할 가능성이 있음
 - 실제 불량률이 검사 과정을 거치지 못하여 불량률이 고객에게 출하될 가능성이 있음

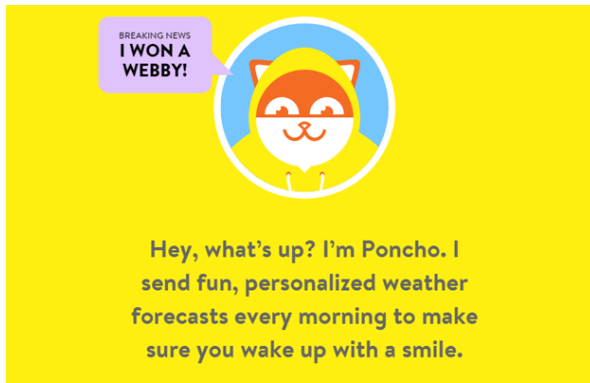


가상계측 (Virtual Metrology)

- VM 모델은 샘플링 된 제품의 품질변수(y)와 이에 대응되는 설비변수(x)를 이용한 예측 모델
- 구축된 VM 모델을 이용하여 샘플링 되지 않은 제품의 품질변수 값을 예측
- 전수 품질 검사의 효과를 얻을 수 있음 (비용절감, 품질개선)



인공지능 (Artificial Intelligence)



씻물 튀는 제철소 톡톡 튀는 AI 활용 스마트하게 변신

세계 최초 인공지능 제철소로 거듭나는 포스코



빅데이터 실시간 분석
도금량 자동제어 통해
차량용 강판 품질 개선
생산원가 크게 줄이고
작업능률도 끌어올려

인공지능 점차 확대해
세계시장 신기술 선도

WP(World Premium) 제품인 자동차용 도금강판은 현재 세계 800여 개 철강회사 가운데 20곳가량만 생산할 수 있을 정도로 고도의 기술력을 필요로 하는 고급 제품이다. 포스코는 지난해 약 900만 t의 자동차 강판을 판매하여 전 세계 자동차 강판의 10%를 공급했다.

권오준 포스코 회장은 취임 이후부터 스마트팩토리 모델 구축 등을 통해 4차 산업혁명의 흐름을 선도하기 위한 준비를 해왔다. 여기에 알파고의 출현으로 AI에 세간의 이목이 집중되자 권 회장은 이를 현장에서 적극적으로 활용해 보자고 제안했다.

포스코 기술연구원은 지난해 6월 도금량 제어 자동화에 대한 니즈를 발굴한 후 조업, 정비, EIC(전기 및 설비 제어) 등 여러 부서로부터 의견을 수합하여 관련 데이터를 수집했다. 이후 데이터 마이닝과 머신러닝 분야 전문가인 이종석 성균관대 시스템경영공학과 교수에게 인공지능 도금량 예측모델 알고리즘 개발을 위탁하고, 이 교수는 포스코 연구원과의 협업을 통하여 도금공정에 대한 이해를 바탕으로 도금량 예측 프로그램을 개발했다.

지난해 7월부터 2개월간 기본적인 프로그램이 개발됐다. 여기에 포스코 기술연구원이 조업 노하우를 반영해 현장설비 및 조업조건 변경 시에도 잘 운용되도록 하는 제어기술을 융합한 현장 맞춤형 프로그램을 추가해 '인공지능 기반 도금량 제어 자동화 솔루션'이 완성됐다.



인공지능 (Artificial Intelligence)

예측, 인지

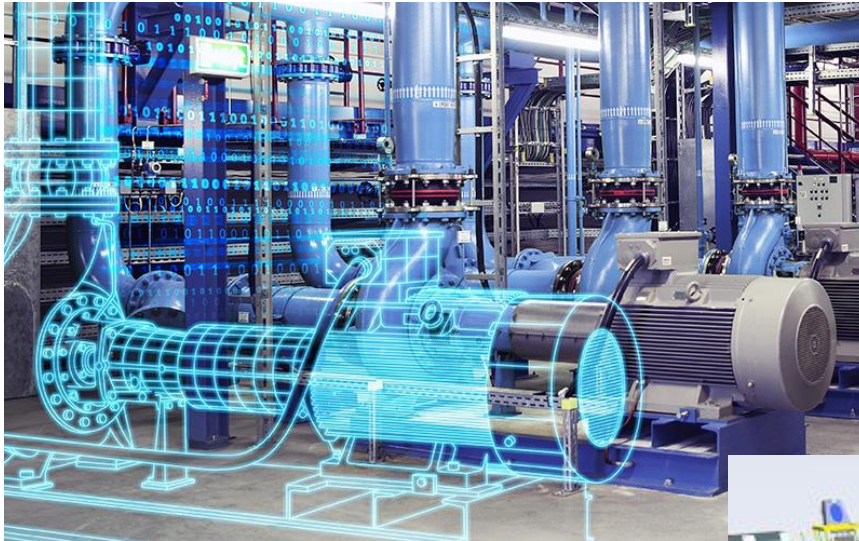


최적화

- Deep Neural Networks
- Neural Networks
- Linear Regression
- Support Vector Machines
- Decision Tree
- Random Forests
- Bayesian Methods
- Bagging
- Boosting (AdaBoost, XGBoost)
- K-Nearest Neighbors

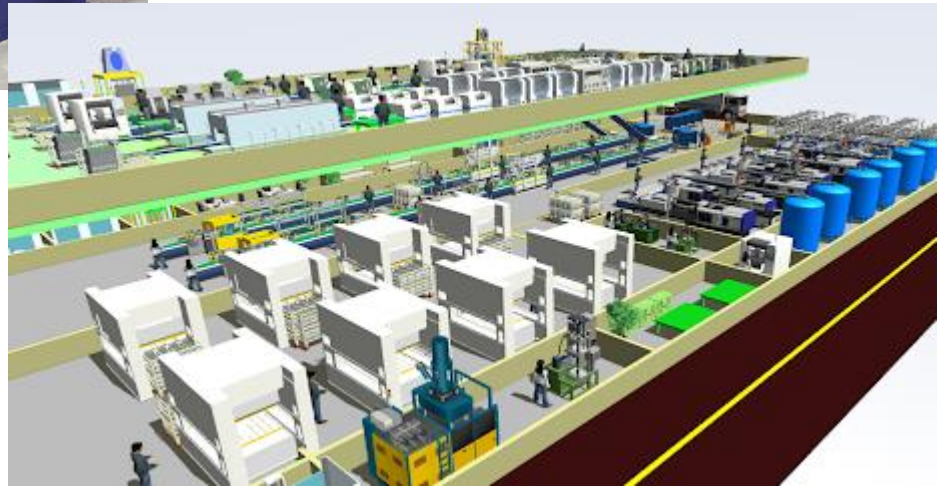
- Reinforcement Learning
- Simplex Method
- Gradient Descent Method
- Stochastic Gradient Descent
- Newton Method
- Trust Region Method
- Tabu Search Algorithm
- Genetic Algorithm
- Particle Swarm Optimization
- Nested Partitions Method

디지털 트윈 (Digital Twin)



새로운 설비를 추가하면 생산
효율이 얼마나 좋아질까?

신제품 생산 라인을 어떻게
배치해야 가장 효율적일까?



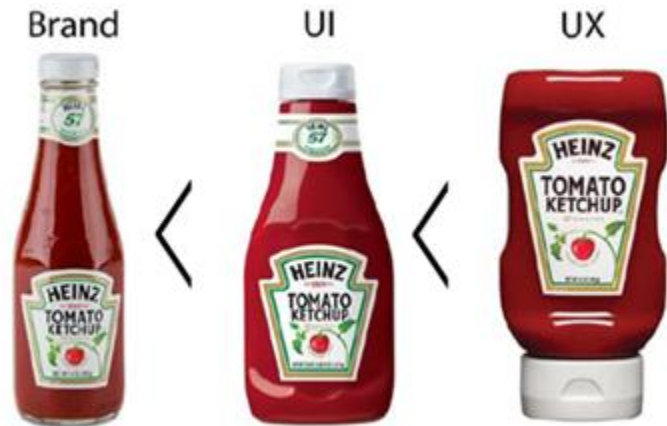
인간공학, 사용자경험 (Ergonomics, UX)



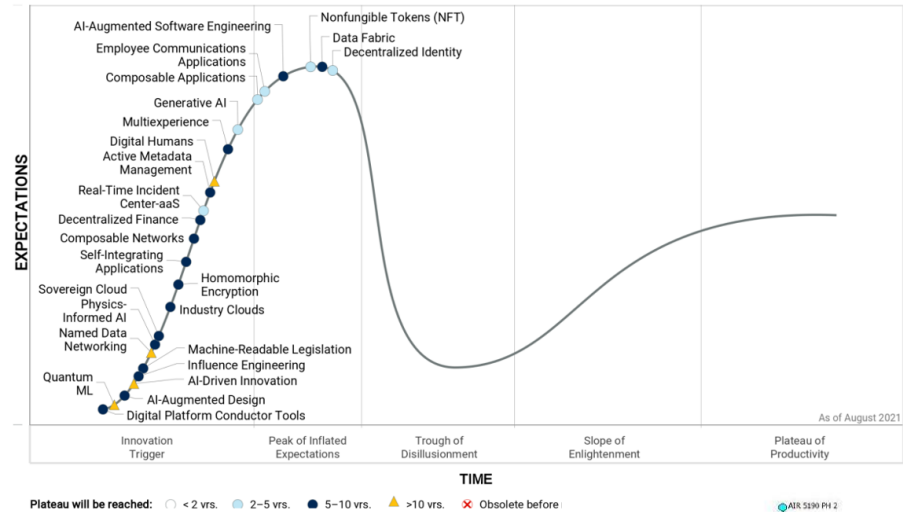
<https://www.sedaily.com/NewsView/1S264T3GLJ>

현대인이 원하는 침대의 기능은?
가장 편안함을 느끼는 구조는?

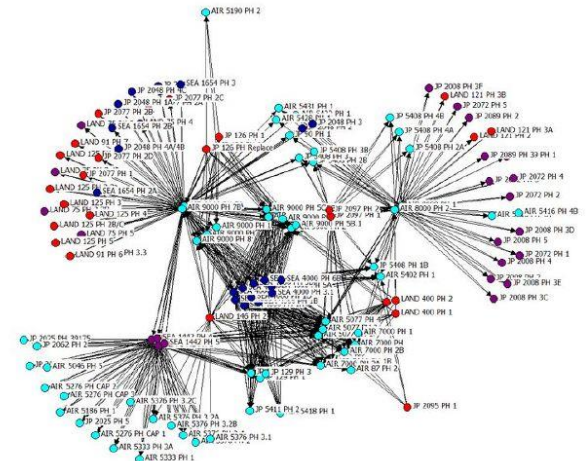
사용자 경험에 기반한
케첩 용기의 변천



기술경영 (Technology Management)



Source: Gartner (August 2021)
747576



미래를 선도할 기술은?

어떤 기술을 확보하고 운용해야 경쟁 우위를
선점할 수 있을까?

산업공학을 전공하면 어떤 분야에 진출하나요?



고대운 LG CNS AI 빅데이터 담당 제조빅데이터 1팀

- 머신러닝 알고리즘 활용 빅데이터 기반 예측 경영관리 체계 구축
- **산업공학은?** 다양한 산업에 대한 빠른 이해, 공학적 해석능력



최수환 Boston Consulting Group 서울 오피스

- 다양한 산업분야의 전략 컨설팅 프로젝트 수행
- **산업공학은?** 기술 · 공학적 지식과 비즈니스 실용적 관점의 결합



김대현 두나무 데이터 밸류팀

- 암호화폐거래소 업비트 및 증권플러스 서비스 런칭 및 운영
- **산업공학은?** 커뮤니케이션에 강점을 가진 조화로운 연결자



배영목 SK 하이닉스 NANO 개발총괄 NANO PE그룹

- 미래 Test solution 준비, 특성 및 수율분석을 통한 완성도 제고
- **산업공학은?** 회사의 문제점을 정의하고 해결을 위한 방법론 제시

산업공학을 전공하면 어떤 분야에 진출하나요?



권력환 현대카드 금융 Price 최적화팀

- 카드금융 상품 고객 반응 예측 실험 설계 및 모델링
- 산업공학은? 비즈니스 현장에서 검증된 공학적 방법론



임현숙 원스토어 서비스 혁신센터 서비스혁신팀

- 앱마켓 서비스 UX 디자인 및 Product Management
- 산업공학은? 협업을 통해 아이디어를 결과로 이끄는 능력

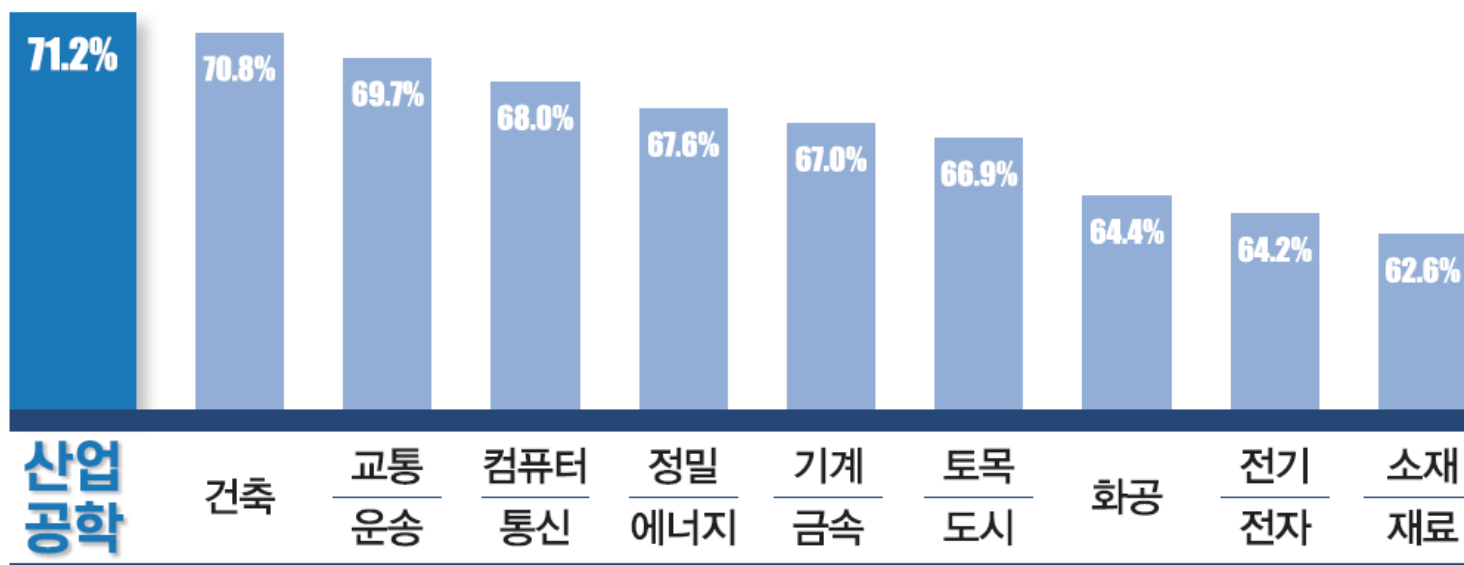


방선주 아산병원 의생명 연구소 신경과

- 뇌졸중 환자의 MRI 이미지 등 의생명 데이터 분석
- 산업공학은? 폭넓은 분야의 지식을 통한 종합적 이해능력

공학분야 취업률

출처: 2019 취업통계연보



감사합니다.

기본에 충실한 성균관대학교 구성원 되시길 바랍니다.