

머신러닝을 위한 수학 - 최적화

이규영

소개

08 홍익대학교 산업공학과(학사)

14 POSCO C&C 품질기술그룹

16 한양대학교 산업공학과(석사)

#산업공학도
#최적화
#제조데이터
#R



18.1 ~ 18.9 NHN ACE Data Scientist

18.9 ~ 11번가 Machine Learning Engineer

#11번가 워킹데일로 11일째
#자연어처리
#머신러닝
#파이썬

이규영

mineatte@gmail.com

오늘 소개할 내용

#최적화의 분류


#최적화의 적용법

#머신러닝과 최적화

최적화?

최적화?

국어사전

최적화 (最適化) [최 : 저과, 최 : 저과] 

[명사] 어떤 조건 아래에서 주어진 함수를 가능한 최대 또는 최소로 하는 일.

최적화?

최적화(最適化, 문화어: 최량화)는 다음을 가리키는 말이다:

- 최적화 문제 (수학)
- 최적화 (정보 공학)
 - 컴파일러 최적화
- 검색 엔진 최적화
- 프로세스 최적화
- 최적화 이론
- 최적화 (음악 그룹)
- 디스크 최량화: 디스크 조각모음을 가리키는 문화어

경제의 최적화에 대해 논하려면, 효용 또는 효율성을 참조하라.

수학에서의 최적화?

제약 조건 하에서 x 에 대한 목적함수 $f(x)$ 의
최대치와 최소치를 찾는 것

minimization or maximization of an objective
function $f(x)$ depending on variables x subject to
constraints

Objective Function : $\min_x f(x)$ or $\max_x f(x)$

Constraints : $h_i(x) = 0$ $g_i(x) \geq 0$

사실
우리는 고등학교 때 모두
최적화를 배웠다.

사실
우리는 고등학교 때 모두
최적화를 배웠다.

[고1] 부등식의 영역에서의 최대/최소
[고3] 미적분의 최대/최소

[고1] 부등식의 영역에서의 최대/최소

점 (x, y) 가 다음 연립부등식의 영역에서 움직일 때, 일차식 $x+y$ 의 최대값과 최소값을 구하여라.

$$x \geq 0, y \geq 0, x+2y \leq 10, 5x+4y \leq 32$$

[풀이] 연립부등식의 영역을 D 라 하면, D 는 오른쪽 그림의 색칠한 부분이다.

이 때, $x+y=k$ (k 는 상수)로 놓으면

$$y = -x + k$$

이므로 이 직선은 $y = -x$ 와 평행하게 움직인다.

직선 $y = -x + k$ 가 영역 D 에서 움직일 때, y 절편인 k 가 최대인 것은 두 직선

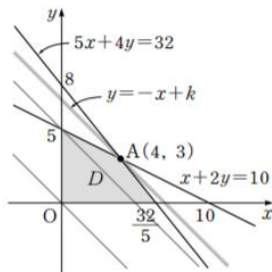
$$x+2y=10, \quad 5x+4y=32$$

의 교점인 점 $A(4, 3)$ 을 지나는 경우이고, 최소인 것은 원점을 지나는 경우이다.

따라서, $x=4, y=3$ 일 때, $x+y$ 는 최대값 7

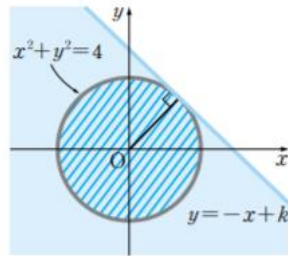
$x=0, y=0$ 일 때, $x+y$ 는 최소값 0

을 가진다.



답 최대값 7, 최소값 0

ㄱ A 는 원 $x^2+y^2=4$ 의 내부(경계선 포함), B 는 직선 $y=-x+k$ 의 아랫부분(경계선 포함)이므로 다음과 같은 경우에 $A \subset B$ 가 성립하면서 k 가 최소값을 갖는다.



직선 $y = -x + k$ 가 원 $x^2+y^2=4$ 에 접하는 경우이므로 원의 중심에서 직선 $y = -x + k$ 에 이르는 거리가 반지름의 길이인 2와 같다.

$$\frac{|-k|}{\sqrt{1+1}} = 2$$

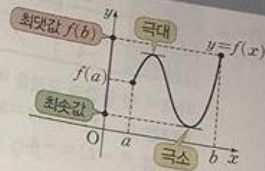
$k > 0$ 이므로 $k \geq 2\sqrt{2}$ 이어야 한다.

따라서, 최소값은 $2\sqrt{2}$ 이다.

[고3] 미적분의 최대/최소

02 최대, 최소를 구하는 방법

- 닫힌 구간 $[a, b]$ 에서 연속함수 $f(x)$ 의 최댓값과 최솟값을 구하려면
- (1) 주어진 구간에서 $f(x)$ 의 극댓값과 극솟값을 구한다.
 - (2) 양 끝점에서의 함수값 $f(a), f(b)$ 를 구한다.
 - (3) 위에서 구한 값들의 크기를 비교한다.
- 최댓값 $\rightarrow \{\text{극댓값}, f(a), f(b)\}$ 중 최대인 것
- 최솟값 $\rightarrow \{\text{극솟값}, f(a), f(b)\}$ 중 최소인 것



참고 최대·최소 정리에 의하여 닫힌 구간에서 연속함수는 반드시 최댓값과 최솟값을 갖는다.
이때 최댓값과 최솟값을 구하려면
(극댓값, 극솟값, 양 끝값)
을 모두 구한 후 크기를 비교하면 된다. 이 중에서 가장 큰 값이 최댓값이고, 가장 작은 값이 최솟값이다.

Deterministic(확정적)

deterministic optimization

최적화 모델을 구성하고 있는 모든
파라미터들이 고정되어 있다고 보는 것

Deterministic(확정적)

deterministic optimization

최적화 모델을 구성하고 있는 모든
파라미터들이 고정되어 있다고 보는 것

→ 머신러닝에서 사용되는 대부분의
최적화 방법들

Table 2 Approaches from Deterministic Optimization

Approach	Key Features
Gradient search	Move locally in most promising direction, according to gradient
Random search	Move randomly to new point, no information used in search
Simulated annealing	Sometimes move in locally worse directions, to avoid being trapped in local extrema
Genetic algorithms and scatter search	Population based, generates new members using (local) operations on attributes of current members
Tabu search	Use memory (search history) to avoid tabu moves
Neural networks	(Nonlinear) function approximation
Math programming	Powerful arsenal of rigorously tested software

Deterministic(확정적) VS

deterministic optimization

최적화 모델을 구성하고 있는 모든
파라미터들이 고정되어 있다고 보는 것

Stochastic(확률적)

stochastic optimization

각 파라미터들은 고정된 값이 아닌
확률분포를 갖는다고 가정하는 것

Deterministic(확정적) VS

deterministic optimization

최적화 모델을 구성하고 있는 모든
파라미터들이 고정되어 있다고 보는 것

Stochastic(확률적)

stochastic optimization

각 파라미터들은 고정된 값이 아닌
확률분포를 갖는다고 가정하는 것

→대표적인 문제가 MDP & 강화학습

Deterministic(확정적) VS

deterministic optimization

최적화 모델을 구성하고 있는 모든
파라미터들이 고정되어 있다고 보는 것

Stochastic(확률적)

stochastic optimization

각 파라미터들은 고정된 값이 아닌
확률분포를 갖는다고 가정하는 것

최적화 분류

Variables	Functions	Classification
All Continuous	All Linear	Linear Programming
	One or more nonlinear	Non-linear Programming

최적화 분류

Variables	Functions	Classification
All Continuous	All Linear	Linear Programming
	One or more nonlinear	Non-linear Programming
All discrete	All Linear	Integer Programming
	One or more nonlinear	Integer Non-linear Programming

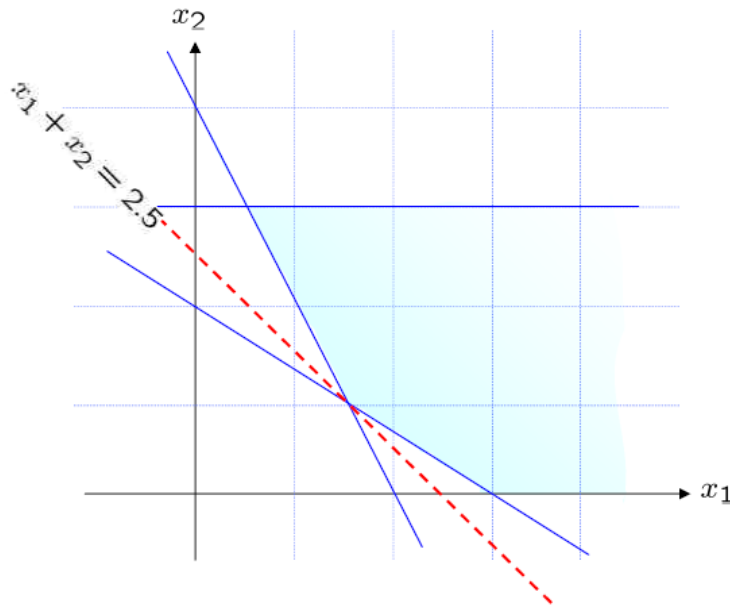
최적화 분류

Variables	Functions	Classification
All Continuous	All Linear	Linear Programming
	One or more nonlinear	Non-linear Programming
All discrete	All Linear	Integer Programming
	One or more nonlinear	Integer Non-linear Programming
Discrete & Continuous	All Linear	Mixed Integer Programming
	One or more nonlinear	Mixed Integer Non-linear Programming

선형최적화(Linear Programming)

목적 함수와 제약식 모두가 선형인 최적화

$$\begin{array}{ll}\min & x_1 + x_2 \\ \text{s.t.} & 2x_1 + x_2 \geq 4 \\ & x_1 + 1.5x_2 \geq 3 \\ & x_2 \leq 3 \\ & x_1, x_2 \geq 0\end{array}$$



비선형최적화(Non-linear Programming)

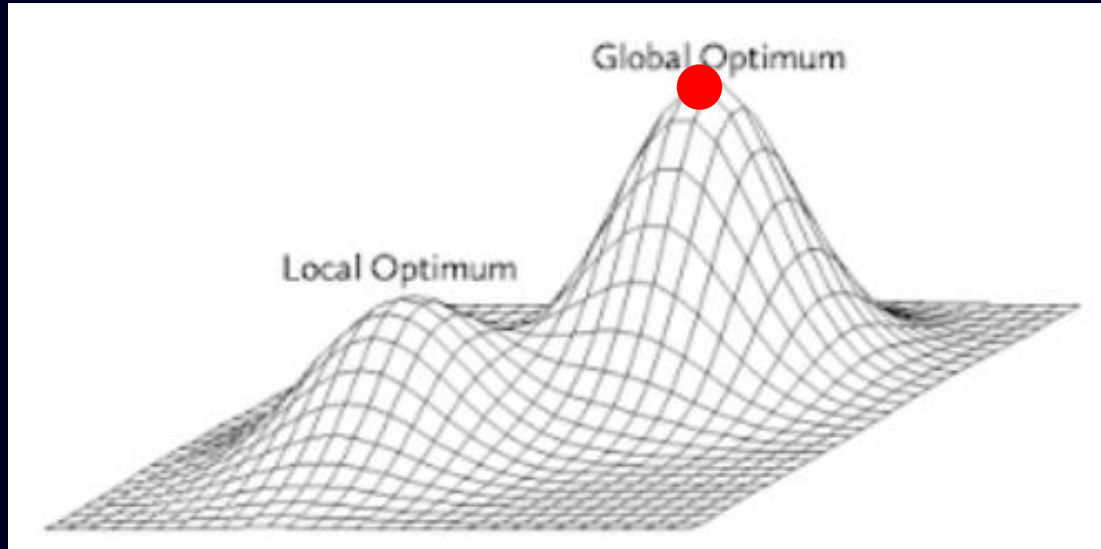
$\min \quad f(\mathbf{x}) \quad \leftarrow \text{Objective function}$

Subject to $h_i(\mathbf{x}) = 0, \quad i = 1, 2, \dots, m \quad \leftarrow \text{Equality constraint}$

$g_j(\mathbf{x}) \geq 0, \quad j = 1, 2, \dots, p \quad \leftarrow \text{Inequality constraint}$

- $\mathbf{x} = (x_1, x_2, \dots, x_n)$ is an n-dim vector with n decision variables.
- $f(\mathbf{x})$, $g_i(\mathbf{x})$, and $h_j(\mathbf{x})$ are real valued twice continuously differentiable functions.

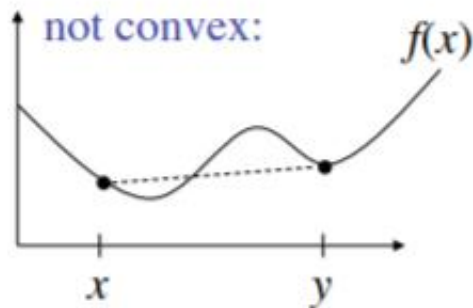
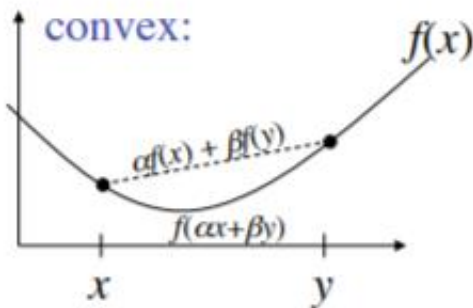
Global vs Local Optimum



Convex or not

All the functions f_i ($i=0\dots m$) are *convex*:

$$f_i(\alpha x + \beta y) \leq \alpha f_i(x) + \beta f_i(y) \quad \text{where} \quad \begin{matrix} \alpha + \beta = 1 \\ \alpha, \beta \in [0,1] \end{matrix}$$



만약 Convex 함수가 아니라면, 여러 개의 local optimal 중 하나가 global optimal

Convex programming

1. If the objective function is strictly convex and a global minimum exists, then the global minimum is unique.
2. A local minimum = a global minimum.

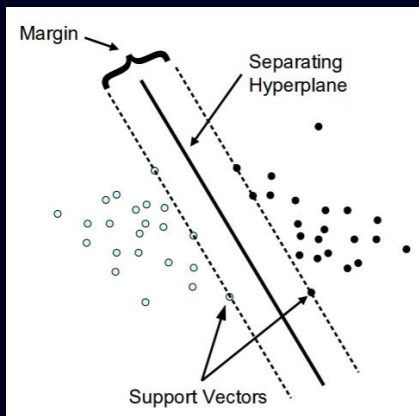
Linear Programming
Quadratic Programming

...

→ Karush-Kuhn-Tucker(KKT) Condition

Convex programming

Support Vector Machine

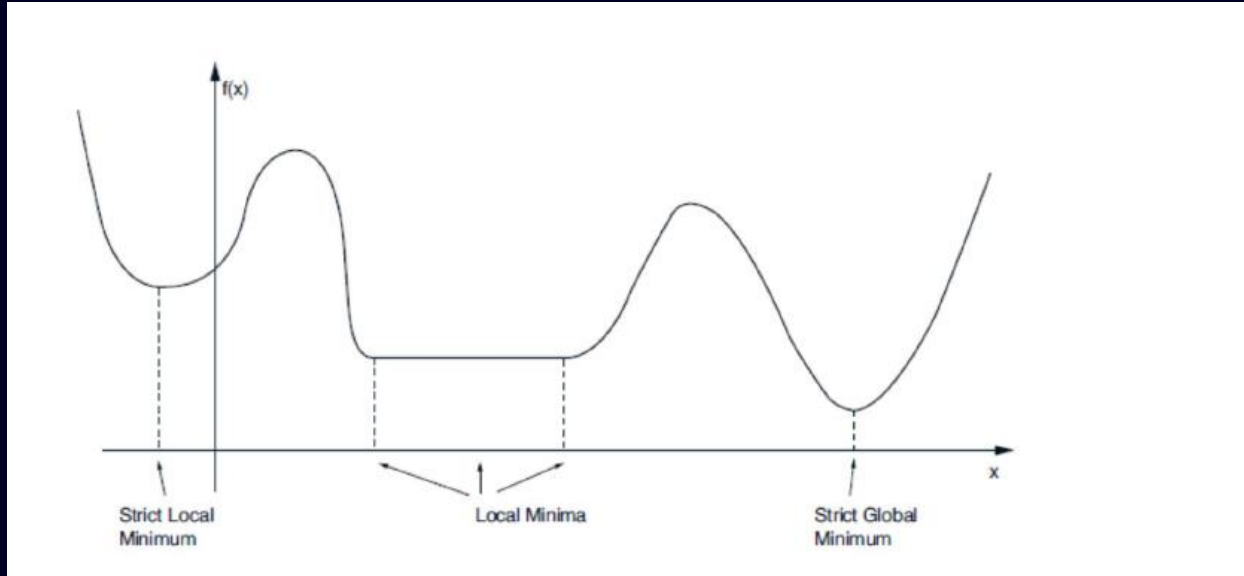


$$\text{minimize } \frac{1}{n} \sum_{i=1}^n \zeta_i + \lambda \|w\|^2$$

subject to $y_i (w \cdot x_i - b) \geq 1 - \zeta_i$ and $\zeta_i \geq 0$, for all i .

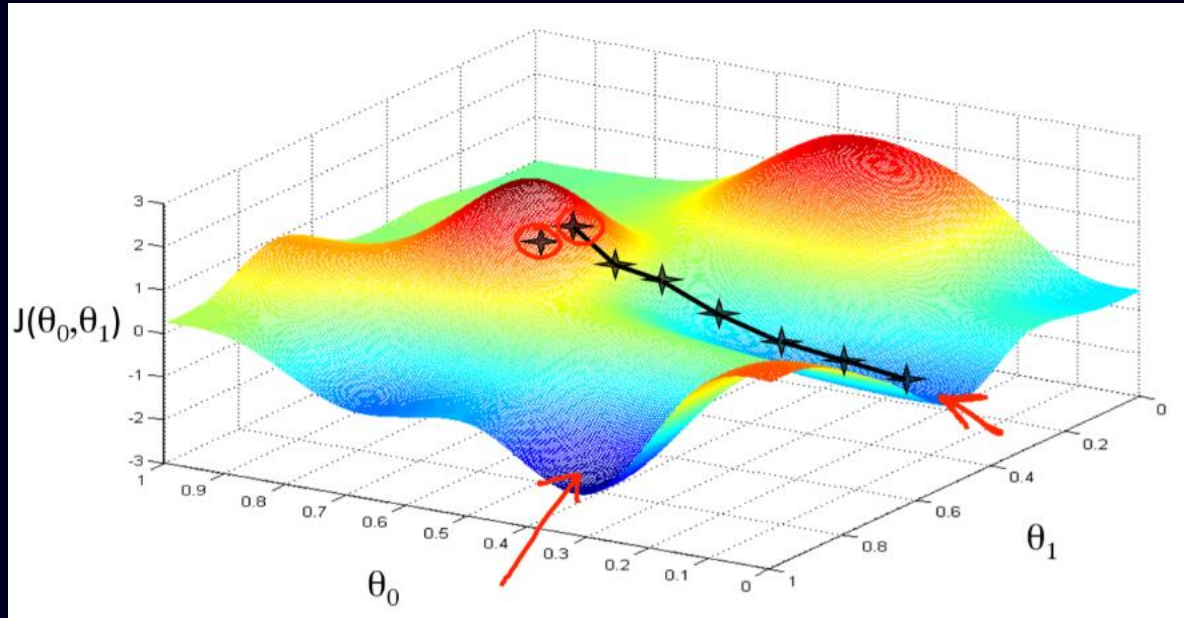
- Quadratic programming
- KKT 조건으로 Optimal을 찾음

Non-convex programming



Non-convex programming

Gradient Descent Algorithm



Non-convex programming

Neural network

신경망 모형은 Cost function $J(\theta)$ 을 minimize하는 parameter θ 를 찾는 문제

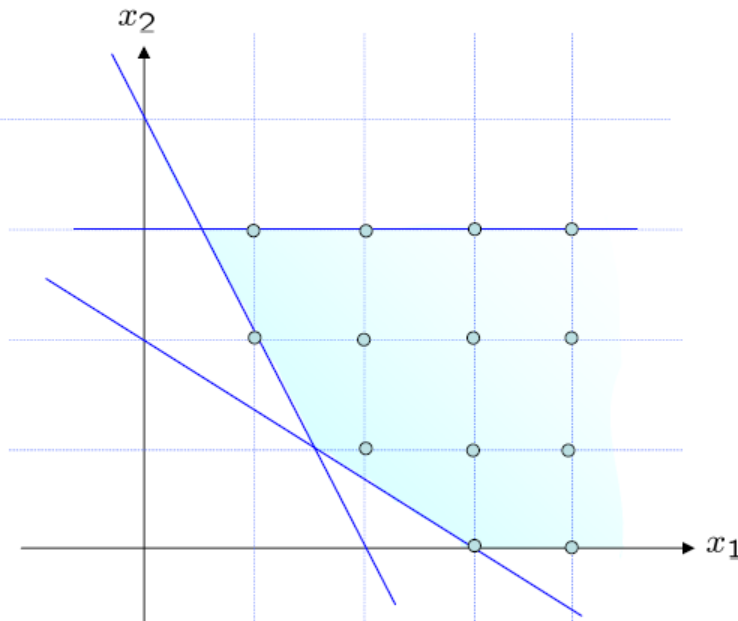
- Gradient Descent Algorithm
- Random search
- Model-based Search

...

정수최적화(Integer Programming)

정수 제약 조건

$$\begin{array}{ll}\min & x_1 + x_2 \\ \text{s.t.} & 2x_1 + x_2 \geq 4 \\ & x_1 + 1.5x_2 \geq 3 \\ & x_2 \leq 3 \\ & x_1, x_2 \geq 0 \text{ and integer}\end{array}$$



확률적 최적화(Stochastic Optimization)

Objective function

or

Constraints

← Random Variables

대표적인 문제 : 강화학습

확률적 최적화(Stochastic Optimization)

#Stochastic Gradient Descent

#Heuristic

- Genetic Algorithm
- Simulated Annealing
- Random Search

...

Genetic Algorithm (유전자 알고리즘)??

Genetic Algorithm (유전자 알고리즘)??

휴리스틱 최적화

휴리스틱(Heuristic)?

휴리스틱의 어원은 라틴어의 'heuristicus' 와 그리스어 'heuriskein' 에서부터 시작되었으며, “찾아내다(find out)” 그리고 발견하다(discover)”라는 의미

한정된 시간에서 최적의 답을 찾는 것이 아니라
만족할 만한 수준의 해법을 찾는 것

NP- Completeness 문제

예시) TSP 문제

여러 도시들이 있고 한 도시에서 다른 도시로 이동하는 비용이 모두 주어졌을 때, 모든 도시들을 단 한 번만 방문하고 원래 시작점으로 돌아오는 최소 비용의 이동 순서를 구하는 문제

P=NP 문제

#수학의 미해결 문제

#컴퓨터 과학의 미해결 문제

#밀레니엄 문제 중 하나

결국, 계산복잡도의 문제로
휴리스틱 기법이 많은 곳에 활용됨

휴리스틱 최적화

#유전알고리즘 : 적자 생존의 아이디어

#시뮬레이티드 어닐링 : 물리적인 담금질 기법

#안티콜로니시스템 : 개미 생태계

...

휴리스틱의 특징 : 가능한 해의 영역을 탐색하면서 지역
최적해(local optimal)에서 탈출할 수 있음

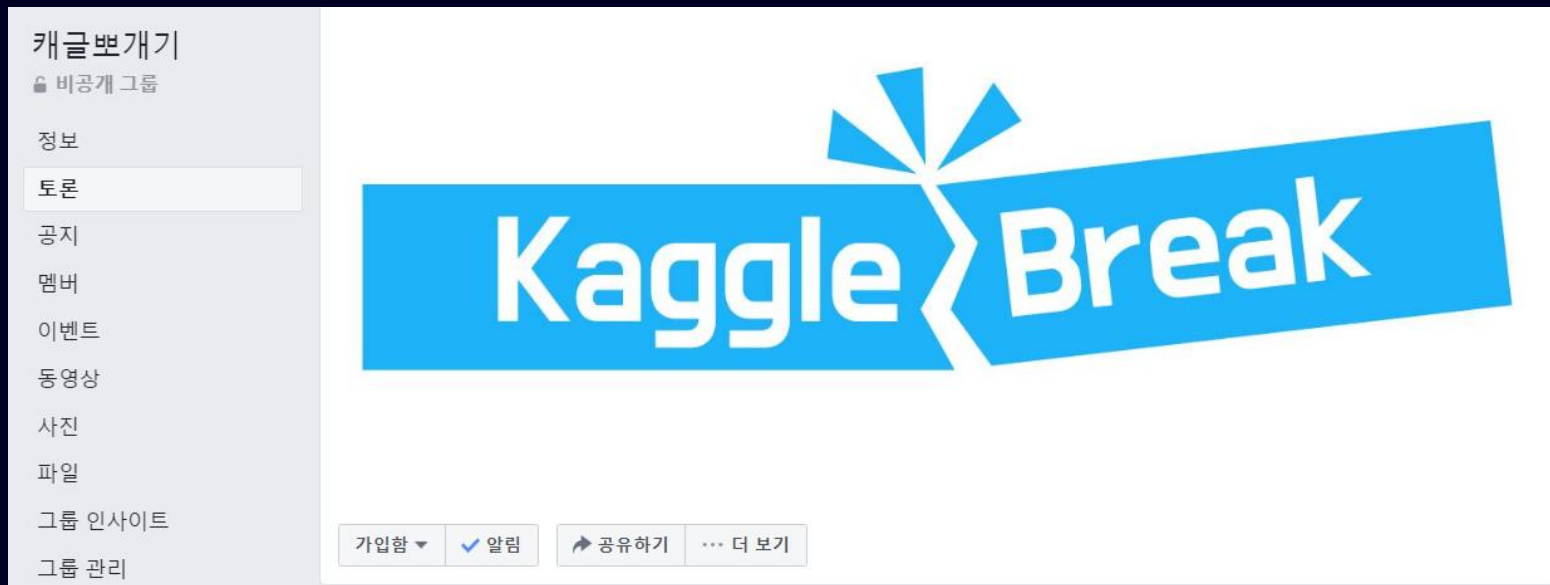
휴리스틱 최적화의 적용

Deterministic or Stochastic

Discrete or Continuous

캐글보개기 커뮤니티

<https://www.facebook.com/groups/kagglebreak/>



캐글보개기 커뮤니티

#캐글보개기 #워킹캐글 #파트4

워킹캐글-DS반 : Part4 - 4회차

일시 : 2018/7/7(토) 10:00~13:00

장소 : 강남컨퍼런스센터 (서울특별시 서초구 강남대로 441 (서초동, 서산빌딩) 9층)

워킹캐글 파트 4 주말반은 "Kaggle" 플랫폼에서 데이터 사이언스를 공부하는 스터디입니다. 파트 4는 기존과 다르게 강제로 나누었던 조별 형태로 운영되지 않습니다. 좀 더 자유도를 높여서 각자 코딩하고, 공유하는 형태로 스터디가 운영될 계획입니다. 그리고 대회 만이 아닌 "흥미로운 데이터셋" 공유 하기도 있습니다.

스터디는 격주 토요일 오전에 진행합니다.

#워킹캐글 스터디

스터디 형태

- 주제 중심으로 모여서 각자 공부하고 코딩하기 (2시간)
- 그 외 슬랙에 그날 공부한 내용 정리하고 공유하기 (1시간)

1. 대회 중심으로 자유롭게 모여서 진행하고 공유하기
2. 발표보다 코딩을 하는 시간
3. 흥미로운 데이터셋 발굴하기 (<https://www.kaggle.com/datasets>)
4. 과거대회 우승자 인터뷰 및 커널 따라하기
5. 데이터 사이언스 Learn (<https://www.kaggle.com/learn/overview>)

현재까지 논의된 주제

- avito 대회, credit Risk 대회
- 과거 커널 따라하기 : Bag of Words Meets Bags of Popcorn
- Zillow 2차 대회
- iMaterialist Challenge(Fashion) 대회

캐글보개기 커뮤니티

원하는 스터디를 열어주세요

Q&A

mineatte@gmail.com

REFERENCE

Michael C. Fu, (2002) Feature Article: Optimization for simulation: Theory vs. Practice. *INFORMS Journal on Computing* 14(3):192-215.

Nelson, B. L. (2010). Optimization via simulation over discrete decision variables. In *Risk and Optimization in an Uncertain World* (pp. 193-207). Informs.

Osman, I. H., & Kelly, J. P. (1996). Meta-heuristics: an overview. In *Meta-heuristics* (pp. 1-21). Springer, Boston, MA.

Kiranyaz, S., Ince, T., & Gabbouj, M. (2014). Optimization Techniques: An Overview. In *Multidimensional Particle Swarm Optimization for Machine Learning and Pattern Recognition* (pp. 13-44). Springer, Berlin, Heidelberg.