Lecture 01
Introduction to Course

Oran Kittithreerapronchai[1]

[1]Department of Industrial Engineering, Chulalongkorn University
Bangkok 10330 THAILAND

last updated: August 7, 2023

Outline

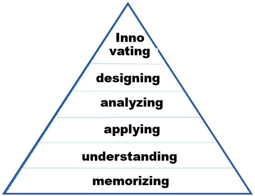
source: General references [NC20, TSK16, BG19, Pat14]

Syllabus: Before we start

Course Description

Industrial engineering problem solving using computational methods; data mining and visualizing; algorithms for inventory models, production planning, production scheduling, production analysis, and optimization

Learning Objectives



- Aware of a general computation method and optimization techniques in IE problems *[a]*
  - Optimization (Unconstraint and Heuristic)
  - Algorithm & Data Mining, Visualization
- Applying suitable algorithm to IE problems *[e]*
- Analyzing data for using large project *[e]*

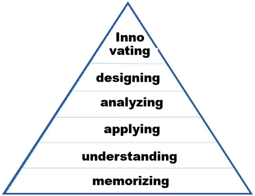
Requirement

Syllabus: Before we start

Course Description

Industrial engineering problem solving using computational methods; data mining and visualizing; algorithms for inventory models, production planning, production scheduling, production analysis, and optimization

Learning Objectives



- Aware of a general computation method and optimization techniques in IE problems ***[a]***
  - Optimization (Unconstraint and Heuristic)
  - Algorithm & Data Mining, Visualization
- Applying suitable algorithm to IE problems ***[e]***
- Analyzing data for using large project ***[e]***

Requirement

- Good background in math modeling and optimization
- Decent programming skills, logic thinking, and perseverance

Still hesitate. Why this one? What for me?

Why do need this course?

OKC's advisees

- **M.Eng:** computational and tool for thesis
- **B.Eng:** senior project, data analyst, advance comp skill

What are benefits of this course?

- **Knowledge:** unconstraint optimization, geom-calcu-algo, data mining, database
- **R/RStudio:** stand-alone/web app & service, wrapper, motivation for modern language
- **Project:** integration of thesis and senior project for team project

Contact Information

Name:	Oran Kittithreerapronchai, PhD
Office:	Room 603, Engineering Building 4
Office Hour:	Tuesday, 16:00-17:00 or by appointment
Email:	oran.k@chula.ac.th
Tel:	02-218-7781
BG & Expr:	Academic: PhD dissertation, course, training, Professional: analysis, stand-alone app, visualization
LMS:	CourseVille (**passwd:** `compMeth<<`*year*`>>`)
WWW:	http://ie.eng.chula.ac.th/~oran https://sites.google.com/site/oranclasses/home

Grading Policy

Grading

- Homework (40%)
- Midterm Exam (30%)
- Term Project: Data Mining (30%)

Grading & Scores

85 and above: final grade id definitely 'A'

between 50 and 85: A, B^+, B, C^+, ... , D

50 and below: final grade is possibly 'F'

Class Rules & Agreements

- No class attendance
- Don't interrupt others
- Focus on your term project
- Be responsible, especially meeting time and assignment
- Participate during class; this is Master level course

- Exam is designed to test student basic knowledge
- Term Project evaluates student performance

Code of honors

- Education with ethic standards and social responsibilities
- Trust as integral and essential part of learning process
- Self-discipline necessity
- Dishonesty hurts the entire community

adapted from: Georgia Institute of Technology –The Honor Code

Any violation to code of honors will severely punished, especially cheating and plagiarism

Textbook and References

Textbook

[NC20] Nwanganga, F. and Phapple, M. 2020. *Practical Machine Learning In R* Wiley

References

[Tor16] Torgo, L. 2011 *Data Mining with R: Learning with Case Studies.* Chapman & Hall.

[Pat14] Pathak, M. 2014 *Beginning Data Science with R.* Springer.

[RS17] Ramasubramanian, K. and Singh, A. 2017 *Machine Learning Using R: A comprehensive Guide to Machine Learning.* Apress

[TSK16] Tan P. *at el* 2016 *Introduction to data mining.* Pearson Education

[?] Chong, E. and Żak, S. 2013 *An introduction to optimization* John Wiley & Sons

Organization

Materials

- **Main Tool:** R/RStudio http://www.rstudio.com/ide/download/desktop
- **Class website:**
  - **LMS:** https://www.mycourseville.com
  - **OKC:** https://tinyurl.com/OKCWebpage/classes/CompMeth
    http://www.ie.eng.chula.ac.th/~oran/classes/CompMeth
- **Other:** Stack Overflow, GitHub, Datacamp

How to turn-in your assignment

- **Medium:** R markdown file and html

  `<ID>_hw<#homework>.Rmd` (must be self-contain with RDS file)
- **Where & When:** through courseville and before midnight of due date

Computational Methods in IE

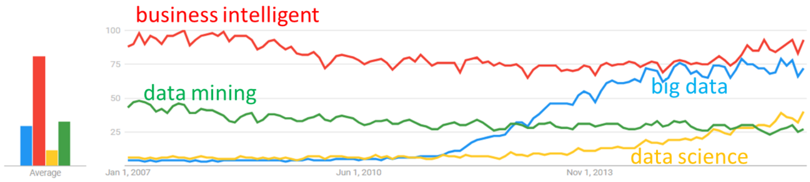
Industrial and System Engineering (ISyE)

- **Core:** applying systematic thinking to make "better" decisions
- **Decisions:** location, inventory, routing, slotting, planning
- **Role of Data Mining in IE:** understanding historical data
  - Target Micro Marketing: customizing promotion for each household (leaded)
  - Amazon Dynamic Pricing: setting price based on number of views (conceded)
  - Multi-Health System: a software company to decide whether a prisoner should get parol
  - Making a hit TV show: data mining approach for creating TV series (Ted Talks)
    - **Amazon** make a competition of single episode of 8 candidate movie (comedy on four Rep. Senators) → Alpha house
    - **Netflix** analyze viewers, actors, producers, history (drama on single Rep. Senator) → House of cards
- **Role of Optimization in IE:** achieving goals (cost, time, risk, regulation, and policies)

Contents and Expectation

Solve IE problem with R

- **Optimization:** → familiar with R, explain algorithms
- **Data Mining:** → IE problem, Business analytic



Source. google.com/trend

- Experience, Time, and Effort are need for skills
- Homework is required to attend this class
- Contents is developing and evolving

Maximizing Revenue

If a company charges a price p USD for a product, then it can sell $3000e^{-p}$ units of the product. What is the revenue function?

- When does it decreasing/increasing in terms of p

- Is there any constraint?

- Find p that maximize the revenue

source: Winston. 2003 Chapter 11 Ex03 [Win22]

Maximizing Revenue

If a company charges a price p USD for a product, then it can sell $3000e^{-p}$ units of the product. What is the revenue function?

$$f(p) = 3000\, p\, e^{-p}$$

- When does it decreasing/increasing in terms of p
- Is there any constraint?
- Find p that maximize the revenue

source: Winston. 2003 Chapter 11 Ex03 [Win22]

Maximizing Revenue

If a company charges a price p USD for a product, then it can sell $3000e^{-p}$ units of the product. What is the revenue function?

$$f(p) = 3000\, p\, e^{-p}$$

- When does it decreasing/increasing in terms of p

  $$\frac{\partial}{\partial p} f(p) = -3000\, p\, e^{-p} + 3000\, e^{-p} = 3000\, e^{-p}(1 - p)$$

  increasing: $0 < p < 1$

  decreasing: $1 < p < \infty$

- Is there any constraint?

- Find p that maximize the revenue

source: Winston. 2003 Chapter 11 Ex03 [Win22]

Maximizing Revenue

If a company charges a price p USD for a product, then it can sell $3000e^{-p}$ units of the product. What is the revenue function?

$$f(p) = 3000\, p\, e^{-p}$$

- When does it decreasing/increasing in terms of p

  $$\frac{\partial}{\partial p} f(p) = -3000\, p\, e^{-p} + 3000\, e^{-p} = 3000\, e^{-p}(1-p)$$

  increasing: $0 < p < 1$

  decreasing: $1 < p < \infty$

- Is there any constraint?

  $$p \geq 0$$

- Find p that maximize the revenue

source: Winston. 2003 Chapter 11 Ex03 [Win22]

Maximizing Revenue

If a company charges a price p USD for a product, then it can sell $3000e^{-p}$ units of the product. What is the revenue function?

$$f(p) = 3000\, p\, e^{-p}$$

- When does it decreasing/increasing in terms of p

  $$\frac{\partial}{\partial p} f(p) = -3000\, p\, e^{-p} + 3000\, e^{-p} = 3000\, e^{-p}(1 - p)$$

  increasing: $0 < p < 1$

  decreasing: $1 < p < \infty$

- Is there any constraint?

  $$p \geq 0$$

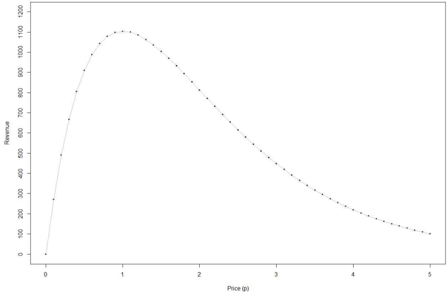
- Find p that maximize the revenue

  $\frac{\partial}{\partial p} f(p) = 0$ and solve for p

  $$f(p = 1) = 1103.638$$

source: Winston. 2003 Chapter 11 Ex03 [Win22]

Multiple ways to solve

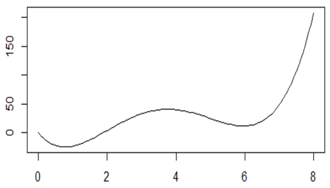


- **Trial & Error:** time consuming & solution quality → always work
- **Graphic:** dimensionality → provide understanding
- **Gradient:** differentiable → exact solution
- **Algorithm:** combine all above all above

Example

$$\min \; f_1(x) = x^4 - 14x^3 + 60x^2 - 70x, \quad x \in \mathbb{R}$$

Grid Search



Calculus

$$f_1(x) = x^4 - 14x^3 + 60x^2 - 70x$$

$$\text{Then } \frac{\partial}{\partial x} f_1(x) \equiv \qquad \#\#[5.957, 3.762, 0.781]$$

$$\frac{\partial^2}{\partial x^2} f_1(x) \equiv$$

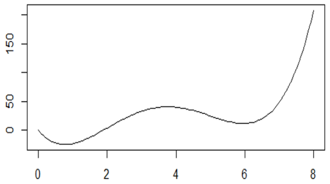
Algorithm

```
tempFn <- function(x){x^4 - 14*x^3 +60*x^2-70*x}
optimize(tempFn,interval = c(0,2),maximum = F)
```

Example

$$\min \ f_1(x) = x^4 - 14x^3 + 60x^2 - 70x, \quad x \in \mathbb{R}$$

Grid Search



Calculus

$$\begin{aligned} f_1(x) &= x^4 - 14x^3 + 60x^2 - 70x \\ \text{Then } \frac{\partial}{\partial x} f_1(x) &\equiv 4x^3 - 42x^2 + 120x - 70 = 0 \qquad \#\#[5.957, 3.762, 0.781] \\ \frac{\partial^2}{\partial x^2} f_1(x) &\equiv \end{aligned}$$

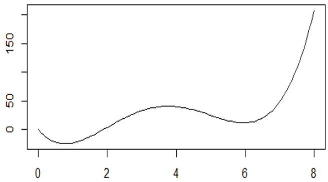
Algorithm

```
tempFn <- function(x){x^4 - 14*x^3 +60*x^2-70*x}
optimize(tempFn,interval = c(0,2),maximum = F)
```

Example

$$\min \ f_1(x) = x^4 - 14x^3 + 60x^2 - 70x, \quad x \in \mathbb{R}$$

Grid Search



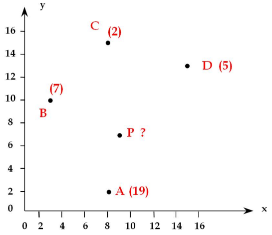
Calculus

$$\begin{aligned} f_1(x) &= x^4 - 14x^3 + 60x^2 - 70x \\ \text{Then } \frac{\partial}{\partial x} f_1(x) &\equiv 4x^3 - 42x^2 + 120x - 70 = 0 \qquad \#\#[5.957, 3.762, 0.781] \\ \frac{\partial^2}{\partial x^2} f_1(x) &\equiv 12x^2 - 84x + 120 \end{aligned}$$

Algorithm

```
tempFn <- function(x){x^4 - 14*x^3 +60*x^2-70*x}
optimize(tempFn,interval = c(0,2),maximum = F)
```

Single Facility Location Problem

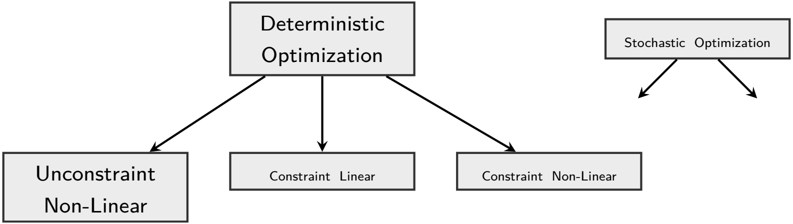


Source. MIT and James Orlin. 2003

Find the location of single warehouse that minimize the total distances from the following location. (**assume** all distances are 'Euclidean'.)

No.	customer	location ($\mathbf{p_i}$)	# shipments (w_i)
1	A	(8,2)	19
2	B	(3,10)	7
3	C	(8,15)	2
4	D	(14,13)	5

Optimization Topics



- Univariate
- Multivariate

Optimization Topics

Optimization Class and Algorithm

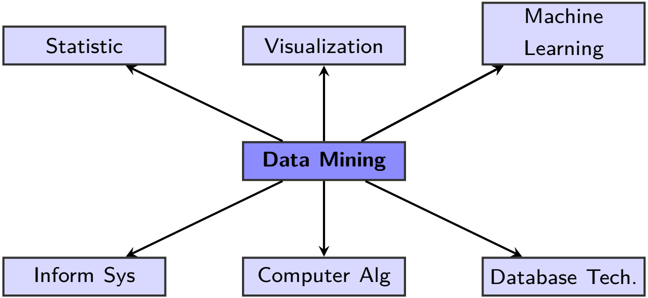
Dimensionality	One-dimensional	Multi-dimensional	Multi-dimensional	Multi-dimensional
Category	Non-gradient based	Gradient based	Hessian based	Non-gradient based
Algorithms	Golden Section Search	Gradient descent methods	Newton and quasi-Newton methods	Golden Section Search, Nelder-Mead
Package	stats	optimx	optimx	optimx
Functions	optimize()	CG	BFGS L-BFGS-B	Nelder-Mead

Optimization in R: https://cran.r-project.org/web/views/Optimization.html

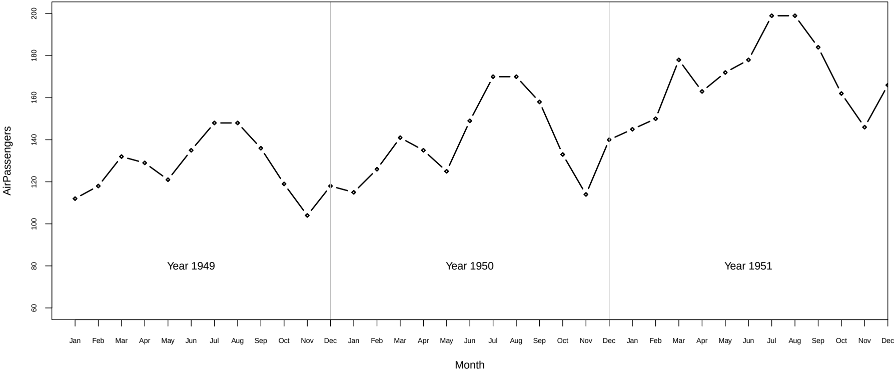
Supplement: math model cycle (see page 34), non-parametric (see page 41)

What is Data Mining?

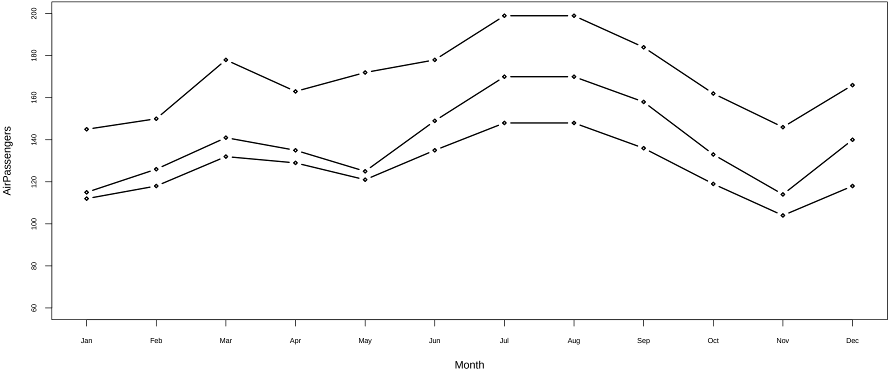
- **What:** process to discover interesting knowledge from data
- **Tasks:** prediction and descriptive (clustering)
- **Why:**
  - IT generates many data
  - computer power is cheap
  - business is high competitive



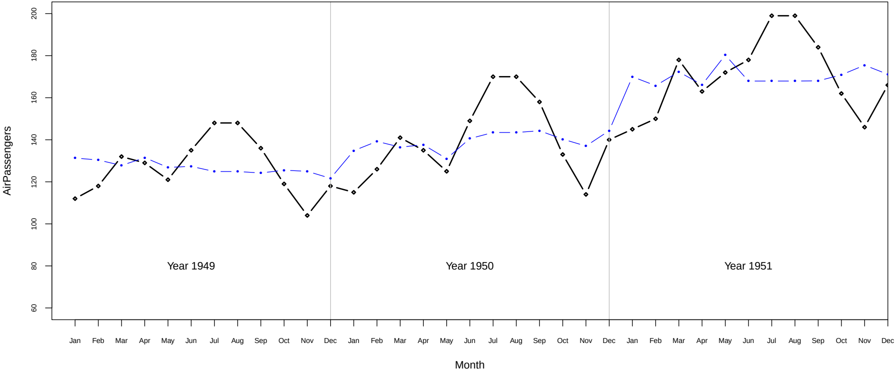
US Air Passengers 1949-1951



US Air Passengers 1949-1951



Air Passengers De-Seasonality



Fraud Transaction Identification

- A company allows its to set price independently
- There are many products and saleperson
- Few percentage can be manually detect fraud

```
> data(sales)
> str(sales)
'data.frame':	401146 obs. of  5 variables:
 $ ID   : Factor w/ 6016 levels "v1","v2","v3",..: 1 2 3 4 3 5 6 7 8 9 ...
 $ Prod : Factor w/ 4548 levels "p1","p2","p3",..: 1 1 1 1 1 2 2 2 2 2 ...
 $ Quant: int  182 3072 20393 112 6164 104 350 200 233 118 ...
 $ Val  : num  1665 8780 76990 1100 20260 ...
 $ Insp : Factor w/ 3 levels "ok","unkn","fraud": 2 2 2 2 2 2 2 2 2 2 ...
```

Can you build a computer algorithm to identify the fraud?

Fraud Transaction Identification

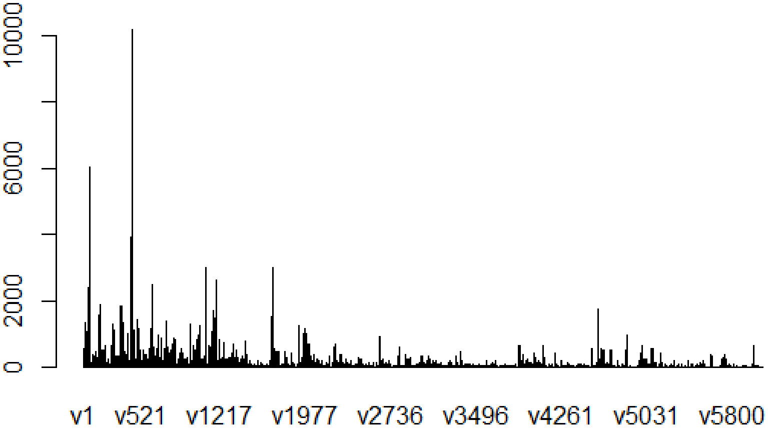
- A company allows its to set price independently
- There are many products and saleperson
- Few percentage can be manually detect fraud

```
> data(sales)
> str(sales)
'data.frame':	401146 obs. of  5 variables:
 $ ID   : Factor w/ 6016 levels "v1","v2","v3",..: 1 2 3 4 3 5 6 7 8 9 ...
 $ Prod : Factor w/ 4548 levels "p1","p2","p3",..: 1 1 1 1 1 2 2 2 2 2 ...
 $ Quant: int  182 3072 20393 112 6164 104 350 200 233 118 ...
 $ Val  : num  1665 8780 76990 1100 20260 ...
 $ Insp : Factor w/ 3 levels "ok","unkn","fraud": 2 2 2 2 2 2 2 2 2 2 ...
```

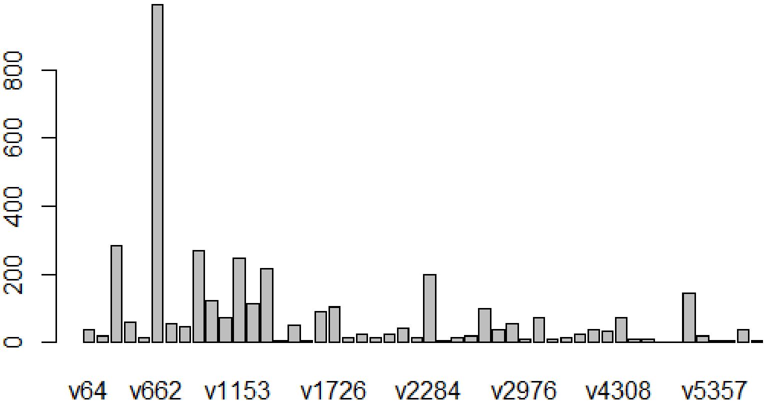
Can you build a computer algorithm to identify the fraud?

- What are differences between "ok" and "fraud"?
- Are unit price difference given `ID,Prod,Insp`?

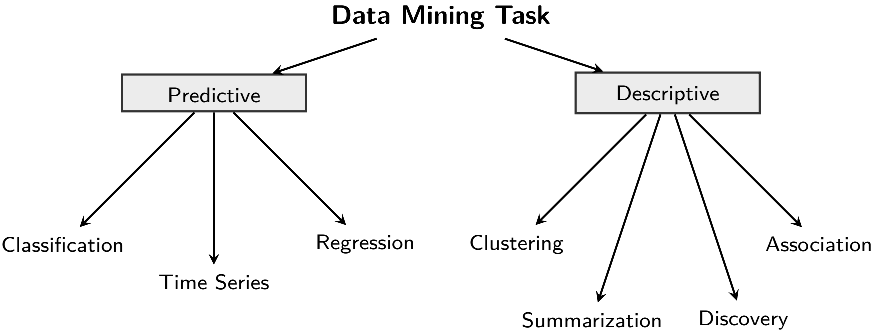
Overwhelming by Big Data



Overwhelming by Big Data



Data Mining Task



Data Mining in R: https://cran.r-project.org/web/views/MachineLearning.html

Supplement: knowledge discovery, data mining cycle (see page 38)

What is parametric univariate?

- **Idea:** known objective function single DV and no constraint
- **Why:** easiest problem
- **Technique:** Grid search, Stochastic search, First order condition, Newton method

Example I

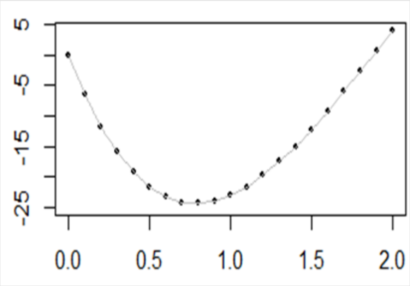
$$\min \; f_1(x) = x^4 - 14x^3 + 60x^2 - 70x, \quad x \in \mathbb{R}$$

Example II

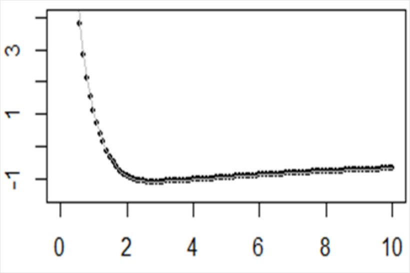
$$\min \; f_2(x) = 3\,x\,e^{-x^2} - 3\frac{1}{x}\ln(x), \quad x \geq 0$$

Grid & Stochastic Search

- **idea:** systematically random/check domain and sort quality of solution



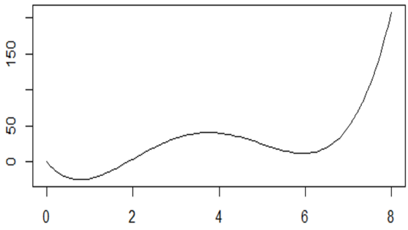
$f_1(x) = x^4 - 14x^3 + 60x^2 - 70x, \quad x \in \mathbb{R}$



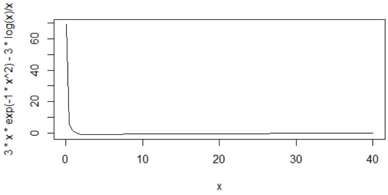
$f_2(x) = 3\,x\,e^{-x^2} - 3\,\frac{1}{x}\,\ln(x), \quad x \geq 0$

Grid & Stochastic Search

- **idea:** systematically random/check domain and sort quality of solution



$f_1(x) = x^4 - 14x^3 + 60x^2 - 70x, \quad x \in \mathbb{R}$



$f_2(x) = 3\,x\,e^{-x^2} - 3\,\frac{1}{x}\,\ln(x), \quad x \geq 0$

First Order Condition

$$f_1(x) = x^4 - 14x^3 + 60x^2 - 70x$$

$$\text{Then } \frac{\partial}{\partial x} f_1(x) =$$

and,

$$\frac{\partial^2}{\partial x^2} f_1(x) =$$

$$f_2(x) = 3\,x\,e^{-x^2} - 3\frac{1}{x}\ln(x)$$

$$\text{Then } \frac{\partial}{\partial x} f_2(x) =$$

and,

$$\frac{\partial^2}{\partial x^2} f_2(x) =$$

First Order Condition

$$f_1(x) = x^4 - 14x^3 + 60x^2 - 70x$$

Then $$\frac{\partial}{\partial x} f_1(x) = 4x^3 - 42x^2 + 120x - 70$$

and,

$$\frac{\partial^2}{\partial x^2} f_1(x) = 12x^2 - 84x + 120$$

$$f_2(x) = 3\,x\,e^{-x^2} - 3\frac{1}{x}\ln(x)$$

Then $$\frac{\partial}{\partial x} f_2(x) =$$

and,

$$\frac{\partial^2}{\partial x^2} f_2(x) =$$

$$+\frac{1}{x^3} - 2\frac{\ln(x)}{x^3}$$

First Order Condition

$$\begin{aligned} f_1(x) &= x^4 - 14x^3 + 60x^2 - 70x \\ \text{Then } \frac{\partial}{\partial x} f_1(x) &= 4x^3 - 42x^2 + 120x - 70 \\ \text{and,} & \\ \frac{\partial^2}{\partial x^2} f_1(x) &= 12x^2 - 84x + 120 \end{aligned}$$

$$\begin{aligned} f_2(x) &= 3\,x\,e^{-x^2} - 3\frac{1}{x}\ln(x) \\ \text{Then } \frac{\partial}{\partial x} f_2(x) &= 3\,e^{-x^2} - 6x^2 e^{-x^2} - 3\frac{1}{x^2} + \frac{\ln(x)}{x^2} \\ \text{and,} & \\ \frac{\partial^2}{\partial x^2} f_2(x) &= -6x\,e^{-x^2} - 12x\,e^{-x^2} + 12x^3 e^{-x^2} + 6\frac{1}{x^3} \\ & \quad + \frac{1}{x^3} - 2\frac{\ln(x)}{x^3} \end{aligned}$$

Taylor Series for Univariate

$$q(x) \quad = \quad f(x_k) + f'(x_k)\,(x - x_k) + \frac{1}{2} f''(x_k)\,(x - x_k)^2$$

where, [Win22]

$f(x)$ = Original objective function

$q(x)$ = Taylor's approximated objective function

First Order Condition

Taylor Series for Univariate

$$q(x) = f(x_k) + f'(x_k)(x - x_k) + \frac{1}{2} f''(x_k)(x - x_k)^2$$

where, [Win22]

$f(x)$ = Original objective function

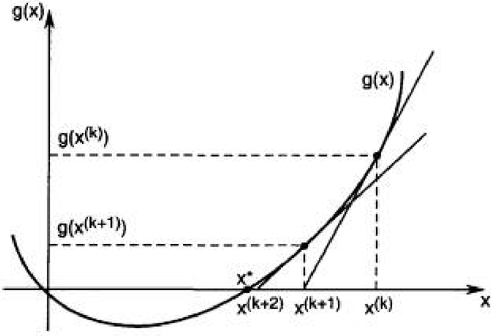
$q(x)$ = Taylor's approximated objective function

First Order Condition

$$\frac{\partial}{\partial x} q(x) = f'(x_k) + (x - x_k) f''(x_k) = 0$$

$$x_{k+1} = x_k - \frac{f'(x_k)}{f'' x_k}$$

What is Newton Method?



Source. Chong & Zak. 2001 pp 106

- **What:** iterative method by estimating a objective function with Taylor's series
- **Advantage:** efficient, basic for multiple decision variable
- **Requirement:** explicit function, twice continuity

Example: Newton Method I

Using Newton Method to find solution of this following function, if $x_0 = 0.0$.

$$f(x) = x^4 - 14x^3 + 60x^2 - 70x$$

Source. Chong & Zak. 2001 pp 93

- $f_1'(x) = \; 4x^3 - 42x^2 + 120x - 70$
- $f_1''(x) = \; 12x^2 - 84x + 120$

i	x_i	$f(x)$	$f'(x_i)$	$f''(x_i)$	x_{i+1}
0	0.0000	0.0000	-70	120.	$0 - \frac{-70}{120}$
1	0.5833	-23.079	-13.5	75.08	$0.5833 - \frac{-13.5}{75.08}$
2	0.7630	-24.360	-1.109	62.888	$0.7630 - \frac{-1.109}{62.888}$
3	0.7807	-24.369	-0.010	61.734	$0.7807 - \frac{-0.010}{61.734}$

When is it Time To Quit: Tolerance

What do we observe?

- **Solution:** virtually unchange, particularly $x_3 = 0.7807$ and $x_4 = 0.7808$
- **Objective Value:** virtually unchange, particularly $f(x_3) = -24.369$ and $f(x_4) = -24.370$
- **Slope:** virtually zero, particularly particularly $f'(x_3) = -1.109$ and $f'(x_4) = -0.010$
- **Iterations:** take too many iterations

Tolerance: ϵ

- **Solution:** $|x_{k+1} - x_k| < \epsilon_x$
- **Objective Value:** $|f(x_{k+1}) - f(x_k)| < \epsilon_{f(x)}$
- **Slope:** $f'(x_{k+1}) < \epsilon_{f'(x)}$
- **Iterations:** $t > N$

Example II: Newton Method

Using Newton Method to find solution of this following function using only the first derivative of $f_2(\cdot)$ start at $x = 1.0$.

$$f_2(x) = 3\,x\,e^{-x^2} - 3\frac{1}{x}\ln(x)$$

- $f_2{}'(x) = 3\,e^{-x^2} - 6x^2e^{-x^2} - 3\frac{1}{x^2} + \frac{\ln(x)}{x^2}$
- $f_2{}''(x) = -6x\,e^{-x^2} - 12x\,e^{-x^2} + 12x^3e^{-x^2} + 6\frac{1}{x^3} + \frac{1}{x^3} - 2\frac{\ln(x)}{x^3}$

i	x_i	$f(x)$	$f'(x_i)$	$f''(x_i)$	x_{i+1}
0	1.0000	1.104	-4.104	6.793	$1.0000 - \frac{-4.104}{6.793}$
1	1.6041	-.517	-1.564	3.069	$1.6041 - \frac{-1.564}{3.069}$
2	2.1136	-.989	-0.442	1.341	$2.1136 - \frac{-0.442}{1.341}$
3	2.4432	-1.078	-0.137	0.584	$2.4432 - \frac{-0.137}{0.584}$

Summary of class

2104529: R programming for some IE works

- **IE Works:** visualization, regression, decision tree, facility location
- **who should take it?** grad student, data cracking senior project
- **why R, not python?:** easy GUI and better help & support

Two Interrelated of Disciplines

- **PART I: Unconstraint Optimization:** theory and practical tools for solving optimization, e.g. FOC, Linear Algebra, Brent, Gradient method, Newton & Quasi-Newton methods
- **PART II: Data Mining:** computation tools using R for IE work, e.g., Decision Tree, Regression, Time Series

Reference

[BG19] Brad Boehmke and Brandon M Greenwell.
Hands-on machine learning with R.
CRC press, 2019.

[NC20] Fred Nwanganga and Mike Chapple.
Practical machine learning in R.
John Wiley & Sons, 2020.

[Pat14] Manas A Pathak.
Beginning data science with R.
Springer, 2014.

[RS17] Karthik Ramasubramanian and Abhishek Singh.
Machine learning using R.
Number 1. Springer, 2017.

[Tor16] Luis Torgo.
Data mining with R: learning with case studies.
CRC press, 2016.

[TSK16] Pang-Ning Tan, Michael Steinbach, and Vipin Kumar.
Introduction to data mining.
Pearson Education India, 2016.

Generalization of Math Model

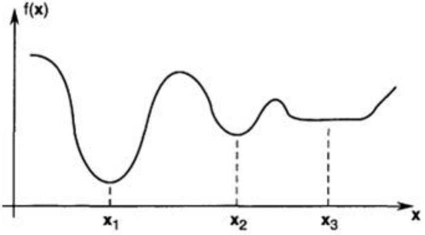
Let the problem have the general mathematical programming (MP) form

$$
\begin{aligned}
(P) \quad & \text{minimize} && f(\mathbf{x}) \\
& \text{subject to} && \\
& && \mathbf{x} \in \mathbb{F}.
\end{aligned}
$$

Mathematical programming

- Yield a solution satisfying constraints $\rightarrow$ feasible solution ($\mathbb{F}$)
- Solve for 'good' solution minimal value $\rightarrow$ optimal solution ($f^*(\mathbf{x})$)

Local and Global Minimizer



Source. Chong & Zak. 2001 pp 72

Local minimizer

Suppose that $f : \mathbb{R}^n \to \mathbb{R}$ is a function defined on $\mathcal{F} \subset \mathbb{R}^n$. A point $\mathrm{x}^* \in \mathcal{F}$ is a *local minimizer* of $f(\cdot)$ if $\exists\ \epsilon > 0$ such that $f(\mathrm{x}) \geq f(\mathrm{x}^*)$ for all $\mathrm{x} \in \mathcal{F} \backslash \mathrm{x}^*$ and $\|\mathrm{x} - \mathrm{x}^*\| < \epsilon$

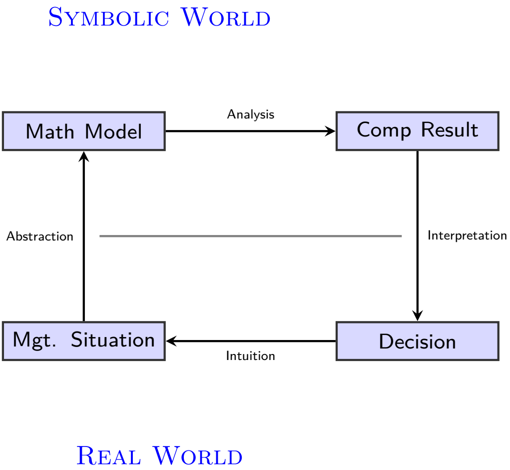
Global minimizer

A point $\mathrm{x}^* \in \mathcal{F}$ is a *global minimizer* of $f(\cdot)$ if $f(\mathrm{x}) \geq f(\mathrm{x}^*)$ for all $\mathrm{x} \in \mathcal{F} \backslash \mathrm{x}^*$

Components of Optimization Model

- **Decision Variables:** What are we interested in?
- **Objective Function:** How do we measure the “best” solution?
- **Direction:** minimize or maximize
- **Constraints:** What are the solutions?
- **Parameters:** Data needed to describe relationship

Optimization Cycle



Pattern Discovery and Feather Selection

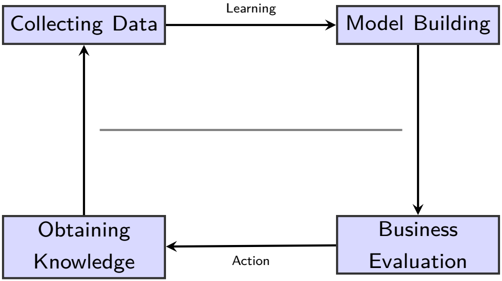
∃ thousands of patterns, yet none of them matter!

- **Pattern Discovery Approach:** better understanding, faster, cleaner
  - **Suggested approach:** Human-centered, query-based, focused mining
  - **Data approach (Feature):** filter method, wrapping method, embedded method
- **interestingness measure**
  - **Objective:** defined by statistics and structures of patterns, e.g., support, confidence
  - **Subjective:** defined by user belief in the data, e.g., unexpectedness, novelty.

Type of Data Mining

- **Data mined:** data warehouse, transactional, stream, spatial, time-series, text, multi-media, heterogeneous, WWW
- **Knowledge discovered (Task):** characterization, association, classification, clustering, trend/deviation, outlier
- **Techniques utilized:** machine learning, statistics, visualization
- **Application adapted:** retail, telecommunication, banking, fraud analysis

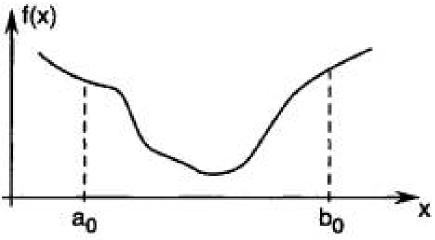
Data Mining Cycle



Background

- **What:** smart trial-and-error of single variable
- **Idea:** reducing the search space/ interval
- **Issues:** unimodal function, eliminate space/ interval
- **Benefits:** no objective function & computational expensive
- **Advantage:** solution quality, 'Smooth' function
- **Examples:** Grid Search, Golden section search, Fibonacci search

Unimodal Function

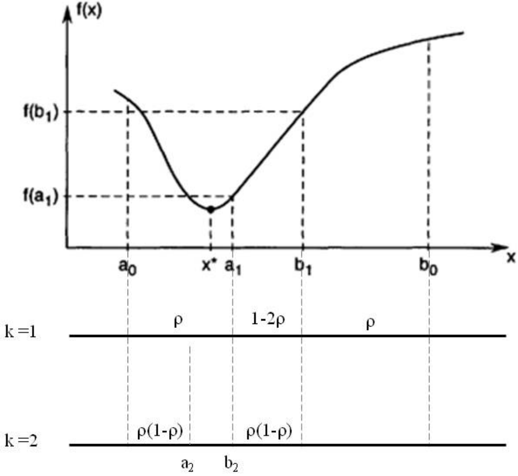


Source. Chong & Zak. 2001 pp 92

Properties

- Function has a single optima (minimal or maximal)
- Evaluating function require 3 points

Golden section search



Source. Chong & Zak. 2001 pp 92

Example: Golden section search

Using Golden section search (3 iterations) to find solution between $[0, 2]$ of this following function.

$$f(x) = x^4 - 14x^3 + 60x^2 - 70x$$

Source. Chong & Zak. 2001 pp 93

	LB		L.center		R.center		UB	
i	a_i	$f(a_i)$	l_i	$f(l_i)$	r_i	$f(r_i)$	b_i	$f(b_i)$
1	0.000	0.000	.7639	-24.36	1.236	-18.96	2.000	4.000

Example: Golden section search

Using Golden section search (3 iterations) to find solution between $[0, 2]$ of this following function.

$$f(x) = x^4 - 14x^3 + 60x^2 - 70x$$

Source. Chong & Zak. 2001 pp 93

	LB		L.center		R.center		UB	
i	a_i	$f(a_i)$	l_i	$f(l_i)$	r_i	$f(r_i)$	b_i	$f(b_i)$
1	0.000	0.000	.7639	-24.36	1.236	-18.96	2.000	4.000
2	0.000	0.000	.4721	-21.10	.7639	-24.36	1.236	-18.96

Example: Golden section search

Using Golden section search (3 iterations) to find solution between $[0, 2]$ of this following function.

$$f(x) = x^4 - 14x^3 + 60x^2 - 70x$$

Source. Chong & Zak. 2001 pp 93

	LB		L.center		R.center		UB	
i	a_i	$f(a_i)$	l_i	$f(l_i)$	r_i	$f(r_i)$	b_i	$f(b_i)$
1	0.000	0.000	.7639	-24.36	1.236	-18.96	2.000	4.000
2	0.000	0.000	.4721	-21.10	.7639	-24.36	1.236	-18.96
3	.4721	-21.10	.7639	-24.36	.9443	-23.59	1.236	-18.96

Example: Golden section search

Using Golden section search (3 iterations) to find solution between $[0, 2]$ of this following function.

$$f(x) = x^4 - 14x^3 + 60x^2 - 70x$$

Source. Chong & Zak. 2001 pp 93

	LB		L.center		R.center		UB	
i	a_i	$f(a_i)$	l_i	$f(l_i)$	r_i	$f(r_i)$	b_i	$f(b_i)$
1	0.000	0.000	.7639	-24.36	1.236	-18.96	2.000	4.000
2	0.000	0.000	.4721	-21.10	.7639	-24.36	1.236	-18.96
3	.4721	-21.10	.7639	-24.36	.9443	-23.59	1.236	-18.96
4	.4721	-21.10	.6525	-23.84	.7639	-24.36	.9443	-23.59