ML and Optimization Intro

Hyerim Bae

Quiz

- What is minimum value of the following function?

$$f(x) = x^2 - 2x + 2$$

$$f(x) = 3\sin(x)\cos(x)\,(6x^2 + 3x^3 + x)\tan(x)$$

- How to get the answer?
  - Mathematically
  - Geographically

Prof. Hyerim Bae (hrbae@pusan.ac.kr)

What is optimization?

- Optimization means 'minimization' or 'maximization' of a (objective) function

$$\min \quad f(x)$$

Prof. Hyerim Bae (hrbae@pusan.ac.kr)

Optimization form

- Unconstrained optimization
  - arg max

$$\underset{x}{\operatorname{argmax}} f(x)$$

- Constrained Optimization

$$\begin{aligned} &\text{minimize } f(x) \\ &\text{subject to } \; c_j(x) = 0 \\ &\qquad\qquad\quad\; c_k(x) \geq 0 \end{aligned}$$

- Black-box optimization
  - Objective function is unknown

$$y = f(\lambda)$$

Prof. Hyerim Bae (hrbae@pusan.ac.kr)

How to solve optimization problems

- Mathematical methods
  - Simplex
  - Lagrangian method
- Geographically
- Meta Heuristics
- Heuristics
- Differentiation
- Gradient descent
- Optimization by search

1. Mathematical methods for constrained optimization

- Linear Programming
- Integer Programming

Linear Programming

- Linear Programming
  - Object function and constraints are all linear
    - Min. (Max.) a function under constraints.
      - ✓ Material Mixing in production system
      - ✓ Job allocation
      - ✓ Transportation ….

A,B 두 상품을 생산하는데 상품 A는 개당 2원의 이익이 나고, B는 개당 5원의 이익이 발생한다. 상품 A를 생산하는 데 9개의 재료와 3시간 동안 기계를 사용해야 하며, B는 5개의 재료와 4시간의 기계를 사용해야 한다. 이때 재료는 총 300개를 사용할 수 있으며, 기계 가동 시간은 최대 200시간이라고 한다. 또 상품 A는 최소 5개 이상을 생산해야만 한다고 한다. 이때 최대의 이익을 산출해 내는 상품 A와 B의 생산량을 결정하라.

결정 변수: 제품 A의 생산량 => x_1
제품 B의 생산량 => x_2

$$
\begin{array}{lrl}
\text{목적함수} & \textit{Maximize}\quad 2x_1 + 5x_2 & \longleftarrow \text{이익의 최대화} \\
\text{제약식} & s.t.\quad 3x_1 + 4x_2 \le 200 & \longleftarrow \text{기계 가동시간 제약} \\
 & 9x_1 + 5x_2 \le 300 & \longleftarrow \text{재료 사용량 제약} \\
 & x_1 \ge 5 & \longleftarrow \text{A의 최소 생산량 제약} \\
 & x_2 \ge 0 & \longleftarrow \text{비음인 해만을 구함}
\end{array}
$$

Prof. Hyerim Bae (hrbae@pusan.ac.kr)

- Minimax 문제

$$\begin{aligned}
&\text{Minimize maximum}\{12x_1 - 21x_2, 17x_1 - 10x_2\} \\
&s.t. \quad 2x_1 - 7x_2 \geq 12 \\
&\qquad\;\; 6x_1 + 11x_2 \geq 41 \\
&\qquad\;\; 9x_1 + 17x_2 \leq 102 \\
&\qquad\;\; x_i \geq 0, \quad \forall i
\end{aligned}
\longrightarrow
\begin{aligned}
&\text{Minimize Z} \\
&s.t. \quad Z \geq 2x_1 - 21x_2 \\
&\qquad\;\; Z \geq 17x_1 - 10x_2 \\
&\qquad\;\; 2x_1 - 7x_2 \geq 12 \\
&\qquad\;\; 6x_1 + 11x_2 \geq 41 \\
&\qquad\;\; 9x_1 + 17x_2 \leq 102 \\
&\qquad\;\; x_i \geq 0, \quad \forall i
\end{aligned}$$

- Minimum Absolute Value
- Goal Programming
- ….

Convex optimization

- In the constrained optimization,

$$\text{minimize } f(x)$$

$$\text{subject to } \; c_j(x) = 0$$
$$c_k(x) \geq 0$$

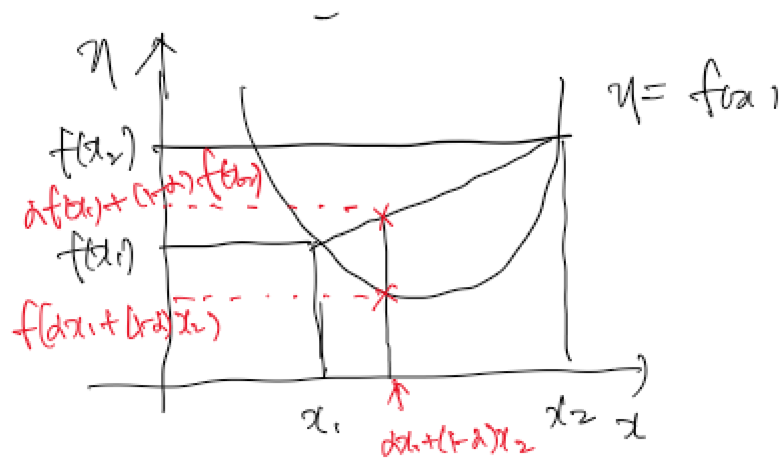
- ① Objective function is convex and ② feasible set is convex!

Prof. Hyerim Bae (hrbae@pusan.ac.kr)

Convex function

- A function $f(x)$ is convex if

$$f(\alpha x_1 + (1-\alpha)x_2) \le \alpha f(x_1) + (1-\alpha) f(x_2), \quad 0 \le \alpha \le 1$$



Prof. Hyerim Bae (hrbae@pusan.ac.kr)

Convex set

- A set F is convex, if

$$x_1, x_2 \in F, \quad 0 \le \alpha \le 1$$

$$\alpha x_1 + (1 - \alpha) x_2 \in F$$

Prof. Hyerim Bae (hrbae@pusan.ac.kr)

Convex optimality

- Local optimum (x^*) is global optimum *① Objective function is convex and ② feasible set is convex!*

If $f(x') < f(x^*)$, (Local optimal exists but there is smaller optimal …)

$$\alpha x^* + (1-\alpha)x' \in F$$ *"F is convex"*

$$f(\underbrace{\alpha x^* + (1-\alpha)x'}) \le \alpha f(x^*) + (1-\alpha) f(x')$$ *"f is convex"*

$$= f(x^*) + (1-\alpha)\,(\underbrace{f(x') - f(x^*)}_{<0})$$

$$< f(x^*)$$

if $\alpha \to 1$ x^* 주변에 $f(x^*)$ 보다 작은 값이 존재!

Feasible set 안에

Necessity condition for optimality

- f‘=0: NC for local optimal

- What will be NC for constrained optimization?
  - Lagrangian!

Prof. Hyerim Bae (hrbae@pusan.ac.kr)

Lagrangian

- Constrained optimization => Unconstrained optimization

$$\min_{\boldsymbol{x}}. f(\boldsymbol{x})$$

$$\text{subject to } h_j(\boldsymbol{x}) = 0\,, j = 1, \ldots k$$

$$\downarrow$$

$$\min_{\boldsymbol{x},\boldsymbol{\lambda}}. L(\boldsymbol{x}, \boldsymbol{\lambda}) = \min_{\boldsymbol{x},\boldsymbol{\lambda}}. L(\boldsymbol{x}, \lambda_1, \lambda_2, \ldots, \lambda_k)$$

$$L(\boldsymbol{x}, \boldsymbol{\lambda}) = f(\boldsymbol{x}) + \sum_{j=1}^{k} \lambda_j h_j(\boldsymbol{x})$$

- Local optimum x* can be obtained where $L'(\boldsymbol{x}, \boldsymbol{\lambda})$=0

- Using gradient

$$\nabla_x L(\boldsymbol{x}, \boldsymbol{\lambda}) = 0$$

$$\nabla_\lambda L(\boldsymbol{x}, \boldsymbol{\lambda}) = 0 \text{ (or } h_j(\boldsymbol{x}) = 0, j\text{=1,..., } k\text{)}$$

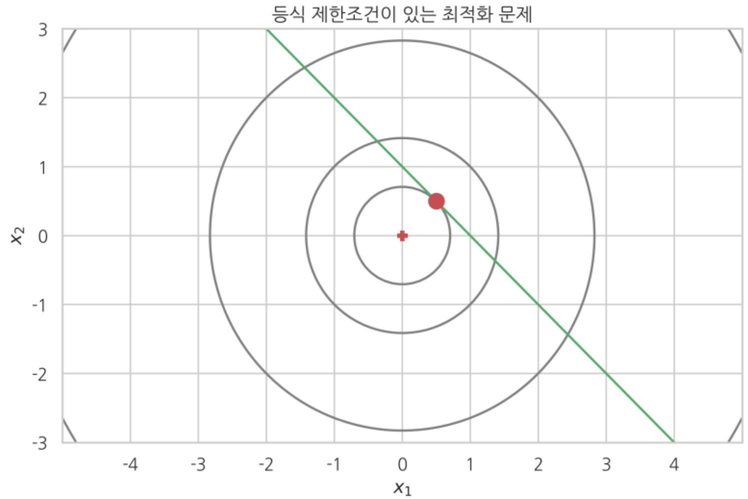
Prof. Hyerim Bae (hrbae@pusan.ac.kr)

- Meaning

$$\nabla L = \nabla f + \lambda \nabla h = 0$$

$$f(x_1, x_2) = x_1^2 + x_2^2$$

$$h(x_1, x_2) = x_1 + x_2 - 1 = 0$$



https://datascienceschool.net

Prof. Hyerim Bae (hrbae@pusan.ac.kr)

Finding solution using Lagrangian

- By solving system of equations

$$\nabla_{x} L(\boldsymbol{x}, \boldsymbol{\lambda}) = 0$$

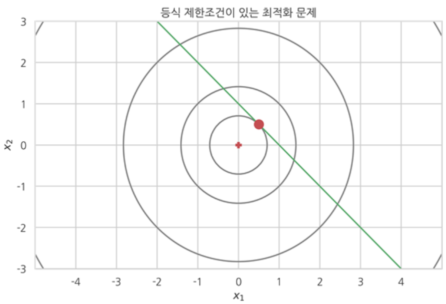
$$\nabla_{\lambda} L(\boldsymbol{x}, \boldsymbol{\lambda}) = 0 \text{ (or } h_j(\boldsymbol{x}) = 0, j=1,\ldots,k)$$

- Another solution
  - Find $arg \min_{x} L(\boldsymbol{x}, \boldsymbol{\lambda})$ by $\nabla_{x} L(\boldsymbol{x}, \boldsymbol{\lambda}) = 0$
  - Use the result (find $L(\boldsymbol{\lambda})$)
  - Find $\boldsymbol{\lambda}$ by $\nabla_{\lambda} L(\boldsymbol{\lambda}) = 0$ (maximize)

Prof. Hyerim Bae (hrbae@pusan.ac.kr)

$$f(x_1, x_2) = x_1^2 + x_2^2$$

$$h(x_1, x_2) = x_1 + x_2 - 1 = 0$$



https://datascienceschool.net

ex) $f(x_1, x_2) = x_1^2 + x_2^2$

M1 $h(x_1, x_2) = x_1 + x_2 - 1 = 0$

$L(\underline{x}, \underline{\lambda}) = x_1^2 + x_2^2 + \lambda(x_1 + x_2 - 1)$

$\nabla_{x_1} L = 2x_1 + \lambda = 0$ $\quad x_1 = -\frac{\lambda}{2}$

$\nabla_{x_2} L = 2x_2 + \lambda = 0$ $\quad x_2 = -\frac{\lambda}{2}$

$\nabla_{\lambda} L = x_1 + x_2 - 1 = 0$

$-\lambda - 1 = 0$ $\quad \lambda = -1$ $\quad x_1 = \frac{1}{2}$ $\quad x_2 = \frac{1}{2}$

$f(\underline{x}) = \frac{1}{4} + \frac{1}{4} = \frac{1}{2}$

M2 $\nabla_x L = 0$ $\quad 2x_1 + \lambda = 0$, $\quad 2x_2 + \lambda = 0$

$x_1 = -\frac{\lambda}{2}$

$L(\lambda) = \frac{\lambda^2}{4} + \frac{\lambda^2}{4} + \lambda\left(-\frac{\lambda}{2} - \frac{\lambda}{2} - 1\right)$

$= \frac{\lambda^2}{2} - \lambda^2 - \lambda$

$= -\frac{\lambda^2}{2} - \lambda$

$\nabla_{\lambda} L = -\lambda - 1 = 0$ $\quad \therefore \lambda = -1$

With inequality in the condition

- With inequality

$$\min_{\boldsymbol{x}}.\, f(\boldsymbol{x})$$

$$\text{subject to } \; g_i(\boldsymbol{x}) \le 0 \,, i = 1, \ldots m$$

$$h_j(\boldsymbol{x}) = 0 \,, j = 1, \ldots k$$

$$L(\boldsymbol{x}, \boldsymbol{\lambda}) = f(\boldsymbol{x}) + \sum_{i=1}^{m} \mu_i g_i(\boldsymbol{x}) + \sum_{j=1}^{k} \lambda_j h_j(\boldsymbol{x})$$

- Need additional things: KKT (Karush-Kuhn-Tucker) conditions

Prof. Hyerim Bae (hrbae@pusan.ac.kr)

KKT

$\nabla_x L(\boldsymbol{x}, \boldsymbol{\lambda}) = 0$

$\nabla_\lambda L(\boldsymbol{x}, \boldsymbol{\lambda}) = 0$ (or $h_j(\boldsymbol{x}) = 0$, j=1,..., k)

- Stationarity:

$$\nabla_x f(\boldsymbol{x}) + \sum_{i=1}^{m} \mu_i \nabla_x g_i(\boldsymbol{x}) + \sum_{j=1}^{k} \lambda_j \nabla_x h_j(\boldsymbol{x}) = 0$$

- Primal constraints:

$$g_i(\boldsymbol{x}) \leq 0\,, i = 1, \dots m$$

$$h_j(\boldsymbol{x}) = 0\,, j = 1, \dots k$$

- Dual constraints:

$$\mu_i \geq 0, i = 1, \dots, m$$

- Complementary slackness:

$$\mu_i g_i(\boldsymbol{x}) = 0, i = 1, \dots, m$$

Prof. Hyerim Bae (hrbae@pusan.ac.kr)

How to solve

- By KKT
- By duality
  - Find $arg\min_{x} L(\boldsymbol{x}, \boldsymbol{\mu}, \boldsymbol{\lambda})$ by $\nabla_{\boldsymbol{x}} L(\boldsymbol{x}, \boldsymbol{\mu}, \boldsymbol{\lambda}) = 0$
  - Find $\boldsymbol{\lambda}$ by $\nabla_{\boldsymbol{\mu},\boldsymbol{\lambda}} \underbrace{\min_{\boldsymbol{x}} L(\boldsymbol{x}, \boldsymbol{\mu}, \boldsymbol{\lambda})}_{q(\mu, \lambda)} = 0$ (maximize)

Duality

- Dual problem as a lower bound to the primal

$$f(x) \geq L(x,\mu,\lambda) \geq min_x L(x,\mu,\lambda) = \underbrace{q(\mu,\lambda)}_{\text{Lower bound}}$$

$$f^* = f(x^*) \geq L(x^*,\mu,\lambda)$$

$$q(\mu,\lambda) \leq \max_{\mu \geq 0,\lambda} q(\mu,\lambda) = q^* \leq f^* \text{(weak duality)}$$

$$q^* = f^* \text{ (strong duality)}$$

- Duality gap

- Optimal
  - Feasible x, μ, λ such that q=f

$$q(\mu,\lambda) \le q^* \le f^* \le f(x)$$

- If q=f

$$q(\mu,\lambda) = q^*$$

$$f(x) = f^*$$

$$\min_{\boldsymbol{x}}.\ f(\boldsymbol{x}) \qquad\qquad \max.\ q(\mu,\lambda)$$

$$\text{s.t. } g(\boldsymbol{x}) \le 0 \quad \longleftrightarrow \quad \text{s.t. } \mu \ge 0$$

$$h(\boldsymbol{x}) = 0$$

Duality in LP

$$
\begin{array}{llcl}
\text{minimize} & 6x_1 + 4x_2 + 2x_3 & & \\
\text{subject to} & 4x_1 + 2x_2 + x_3 & \geq & 5, \\
 & x_1 + x_2 & \geq & 3, \\
 & x_2 + x_3 & \geq & 4, \\
 & x_1, x_2, x_3 & \geq & 0.
\end{array}
$$

$$6x_1 + 4x_2 + 2x_3 \geq 4x_1 + 2x_2 + x_3 \geq 5$$

$$6x_1 + 4x_2 + 2x_3 \geq (4x_1 + 2x_2 + x_3) + 2 \cdot (x_1 + x_2) \geq 11$$

LB

$$6x_1 + 4x_2 + 2x_3 \geq x_1 \cdot (4y_1 + y_2) + x_2 \cdot (2y_1 + y_2 + y_3) + x_3 \cdot (y_1 + y_3)$$

$$6 \geq 4y_1 + y_2$$

$$4 \leq 2y_1 + y_2 + y_3$$

$$2 = y_1 + y_3$$

If we generalize…

$$
\begin{array}{llcl}
\text{minimize} & 6x_1 + 4x_2 + 2x_3 & & \\
\text{subject to} & 4x_1 + 2x_2 + x_3 & \geq & 5, \\
 & x_1 + x_2 & \leq & 3, \\
 & x_2 + x_3 & = & 4, \\
 & x_1 & \geq & 0, \\
 & x_2 & \leq & 0, \\
 & (x_3 & \text{is} & \text{free}).
\end{array}
\longrightarrow
\begin{array}{llcl}
\text{maximize} & 5y_1 + 3y_2 + 4y_3 & & \\
\text{subject to} & y_1 & \geq & 0, \\
 & y_2 & \leq & 0, \\
 & (y_3 & \text{is} & \text{free}), \\
 & 4y_1 + y_2 & \leq & 6, \\
 & 2y_1 + y_2 + y_3 & \geq & 4, \\
 & y_1 + y_3 & = & 2.
\end{array}
$$

$$6x_1 + 4x_2 + 2x_3 \geq y_1 \cdot (4x_1 + 2x_2 + x_3) + y_2 \cdot (x_1 + x_2) + y_3 \cdot (x_2 + x_3)$$

$$y_1 \cdot (4x_1 + 2x_2 + x_3) + y_2 \cdot (x_1 + x_2) + y_3 \cdot (x_2 + x_3) \geq 5y_1 + 3y_2 + 4y_3$$

$$y_1 \geq 0, \quad y_2 \leq 0, \quad y_3 \text{ free}$$

Prof. Hyerim Bae (hrbae@pusan.ac.kr)

- Ex

$$\min x_1 + x_2$$

$$\text{s.t. } x_1^2 + x_2^2 \le 1$$

Prof. Hyerim Bae (hrbae@pusan.ac.kr)

Convex Optimization in ML:SVM

- A separating hyperplane can be written as

  $$\boldsymbol{W} \cdot \boldsymbol{X} + b = 0$$

  where $\boldsymbol{W}=\{w_1, w_2, \ldots, w_n\}$ is a weight vector and b a scalar (bias)

- For 2-D it can be written as

  $$w_0 + w_1 x_1 + w_2 x_2 = 0$$

- The hyperplane defining the sides of the margin:

  H_1: $w_0 + w_1 x_1 + w_2 x_2 \geq 1$ for $y_i = +1$, and

  H_2: $w_0 + w_1 x_1 + w_2 x_2 \leq -1$ for $y_i = -1$

- Any training tuples that fall on hyperplanes H_1 or H_2 (i.e., the sides defining the margin) are **support vectors**

- This becomes a **constrained (convex) quadratic optimization** problem: Quadratic objective function and linear constraints → *Quadratic Programming (QP)* → Lagrangian multipliers

Prof. Hyerim Bae (hrbae@pusan.ac.kr)

- Objective function

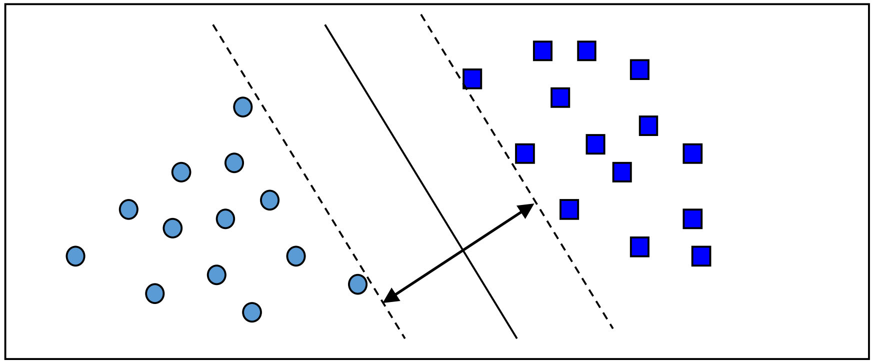
$$b_{i1}: \boldsymbol{w} \cdot \boldsymbol{x} + b = 1$$

$$b_{i2}: \boldsymbol{w} \cdot \boldsymbol{x} + b = -1$$

$$w \cdot (x_1 - x_2) = 2$$

$$\|\boldsymbol{w}\| \times d = 2$$

$$d = \frac{2}{\|\boldsymbol{w}\|}$$



- Constraints

$$\boldsymbol{w} \cdot \boldsymbol{x}_i + b \geq 1 \ (y_i = 1)$$

$$\boldsymbol{w} \cdot \boldsymbol{x}_i + b \leq -1 \ (y_i = -1)$$

$$\boldsymbol{y}_i(\boldsymbol{w} \cdot \boldsymbol{x}_i + b) \geq 1, i = 1,2, \dots, N$$

Prof. Hyerim Bae (hrbae@pusan.ac.kr)

- SVM Learning

$$\min_{\boldsymbol{w}} f(\boldsymbol{w}) = \frac{\|\boldsymbol{w}\|^2}{2}$$

s.t. $$y_i(\boldsymbol{w} \cdot \boldsymbol{x}_i + b) \geq 1, i = 1,2,\dots,N$$

- Convex optimization => Lagrangian

$$L_P = \frac{1}{2}\|\boldsymbol{w}\|^2 - \sum_{i=1}^{N} \lambda_i (y_i(\boldsymbol{w} \cdot \boldsymbol{x}_i + b) - 1)$$

$$\frac{\partial L_p}{\partial \mathbf{w}} = 0 \Longrightarrow \mathbf{w} = \sum_{i=1}^{N} \lambda_i y_i \mathbf{x}_i$$

$$\frac{\partial L_p}{\partial b} = 0 \Longrightarrow \sum_{i=1}^{N} \lambda_i y_i = 0$$

- Dual Lagrangian

$$L_D = \sum_{i=1}^{N} \lambda_i - \frac{1}{2}\sum_{i,j}^{N} \lambda_i \lambda_j y_i y_j \boldsymbol{x}_i \cdot \boldsymbol{x}_j$$

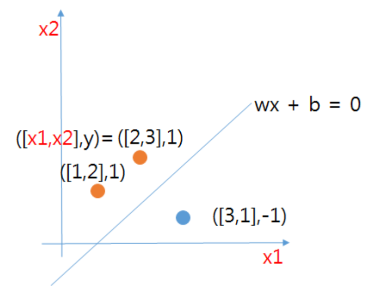
- KKT (Karush-Khun-Tucker) condition for inequality in the constraints: necessary condition

최적해가 되기 위한 조건

$$\lambda_i \geq 0$$

$$\lambda_i(y_i(\boldsymbol{w} \cdot \boldsymbol{x}_i + b) - 1) = 0$$

Prof. Hyerim Bae (hrbae@pusan.ac.kr)



$L = \frac{1}{2}w^2 + \sum \alpha_i[y_i(wx+b)-1]$

Then differentiate the above equations in terms of w and b

$\frac{L}{\partial w} = w - \sum \alpha_i y_i x_i = 0$

$w = \sum \alpha_i y_i x_i$.... (1)

$\frac{L}{\partial b} = \sum \alpha_i y_i = 0$.... (2)

Apply (1) and (2) into L

$L = \frac{1}{2}(\sum \alpha_i y_i x_i)^2 - \sum[\alpha_i y_i(\sum \alpha_i y_i x_i) + \alpha_i y_i b - \alpha_i]$

$= \frac{1}{2}(\sum \alpha_i y_i x_i)^2 - \sum \alpha_i y_i x_i(\sum \alpha_j y_j x_j) - \sum \alpha_i y_i b + \sum \alpha_i$

According to equation(2), $\sum \alpha_i y_i b = 0$

$= \frac{1}{2}(\sum \alpha_i y_i x_i)^2 - (\sum \alpha_i y_i x_i)(\sum \alpha_j y_j x_j) + \sum \alpha_i$

$= -\frac{1}{2}(\sum \alpha_i y_i x_i \sum \alpha_j y_j x_j) + \sum \alpha_i$

$= \sum \alpha_i - \frac{1}{2}(\sum \alpha_i y_i x_i \sum \alpha_j y_j x_j)$

$= \sum \alpha_i - \frac{1}{2}(\sum\sum \alpha_i \alpha_j y_i y_j x_i x_j)$

$x_1 = (1,2), y_1 = 1$
$x_2 = (2,3), y_2 = 1$
$x_3 = (3,1), y_3 = -1$

By applying into the equations,

$\text{L} = \sum \alpha_i - \frac{1}{2}\sum \alpha_i y_i x_i(\alpha_1 y_1 x_1 + \alpha_2 y_2 x_2 + \alpha_3 y_3 x_3)$

$= \sum \alpha_i - \frac{1}{2}\sum \alpha_i y_i x_i(\alpha_1 x_1 + \alpha_2 x_2 - \alpha_3 x_3)$

$= \sum \alpha_i - \frac{1}{2}(\alpha_1 y_1 x_1(\alpha_1 x_1 + \alpha_2 x_2 - \alpha_3 x_3) - \frac{1}{2}(\alpha_2 y_2 x_2(\alpha_1 x_1 + \alpha_2 x_2 - \alpha_3 x_3)) - \frac{1}{2}(\alpha_3 y_3 x_3(\alpha_1 x_1 + \alpha_2 x_2 - \alpha_3 x_3))$

$= \sum \alpha_i - \frac{1}{2}(\alpha_1 x_1(\alpha_1 x_1 + \alpha_2 x_2 - \alpha_3 x_3) - \frac{1}{2}(\alpha_2 x_2(\alpha_1 x_1 + \alpha_2 x_2 - \alpha_3 x_3)) + \frac{1}{2}(\alpha_3 x_3(\alpha_1 x_1 + \alpha_2 x_2 - \alpha_3 x_3))$

$= \sum \alpha_i - \frac{1}{2}(5\alpha_1^2 + 13\alpha_2^2 + 10\alpha_3^2 + 16\alpha_1\alpha_2 - 10\alpha_1\alpha_2 - 18\alpha_2\alpha_3)$

$= \alpha_1 + \alpha_2 + \alpha_3 - \frac{1}{2}(5\alpha_1^2 + 13\alpha_2^2 + 10\alpha_3^2 + 16\alpha_1\alpha_2 - 10\alpha_1\alpha_2 - 18\alpha_2\alpha_3)$

$\frac{L}{\partial \alpha_1} = 1 + \frac{1}{2}(-10\alpha_1 - 16\alpha_2 + 10\alpha_3) = 0$

Remember we know that $\sum \alpha_i y_i = 0$, so $\alpha_1 + \alpha_2 = \alpha_3$

$1 + \frac{1}{2}(-10\alpha_1 - 16\alpha_2 + 10(\alpha_1 + \alpha_2)) = 0$

$\alpha_2 = \frac{1}{3}$

$\frac{L}{\partial \alpha_3} = 1 - \frac{1}{2}(26\alpha_2 + 16\alpha_1 - 18\alpha_3) = 0$

$\frac{L}{\partial \alpha_3} = 1 - (13\alpha_2 + 8\alpha_1 - 9\alpha_3) = 0$

$8\alpha_1 - 9\alpha_3 = -\frac{10}{3}$

We know that $\alpha_1 + \alpha_2 = \alpha_3$, that is, $\alpha_1 + \frac{1}{3} = \alpha_3$

Finally we can find $(\alpha_1, \alpha_2, \alpha_3) = (\frac{1}{3}, \frac{1}{3}, \frac{2}{3})$

By applying α into $w = \sum \alpha_i y_i x_i$, we can get w

$w = (-1, 1)$
Our plane is $y = wx + b$, therefore $y = (-a, a)(x_1, x_2) + b = -ax_1 + ax_2 + b$

Apply $x_1 = (1,2), y_1 = 1$, $x_2 = (3,1), y_3 = -1$ and get $a = \frac{2}{3}, b = \frac{1}{3}$

Prof. Hyerim Bae (hrbae@pusan.ac.kr)

- For test case z,

$$f(\boldsymbol{z}) = sign(\boldsymbol{w} \cdot \boldsymbol{z} + b) = sign(\sum_{i=1}^{N} \lambda_i y_i x_i \cdot \boldsymbol{z} + b)$$

  - If f=1, z will be classified as positive

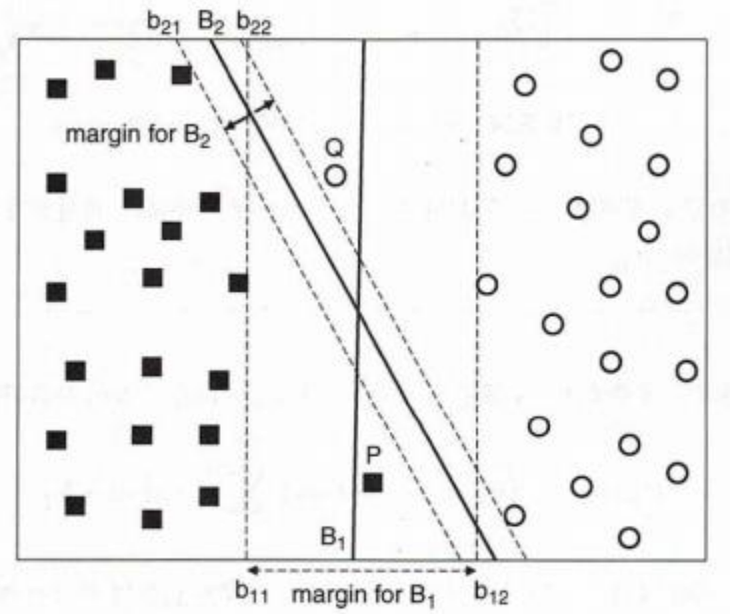
Why Is SVM Effective on High Dimensional Data?

- The **complexity** of trained classifier is characterized by the # of support vectors rather than the dimensionality of the data
- The **support vectors** are the essential or critical training examples —they lie closest to the decision boundary (MMH)
- If all other training examples are removed and the training is repeated, the same separating hyperplane would be found
- The number of support vectors found can be used to compute an (upper) bound on the expected error rate of the SVM classifier, which is independent of the data dimensionality
- Thus, an SVM with a small number of support vectors can have good generalization, even when the dimensionality of the data is high

Prof. Hyerim Bae (hrbae@pusan.ac.kr)

Soft Margin

$$\boldsymbol{w} \cdot \boldsymbol{x}_i + b \geq 1 - \xi_i \ (y_i = 1)$$
$$\boldsymbol{w} \cdot \boldsymbol{x}_i + b \leq -1 + \xi_i \ (y_i = -1)$$

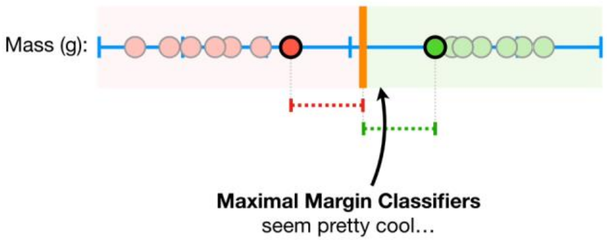


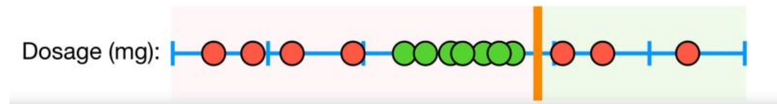
$$\min_{\boldsymbol{w}} f(\boldsymbol{w}) = \frac{\|\boldsymbol{w}\|^2}{2} + C\left(\sum_{i=1}^{N} \xi_i\right)^k$$

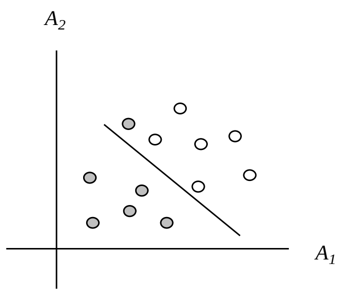
$$L_p = \frac{1}{2}\|\boldsymbol{w}\|^2 - C\sum_{i=1}^{N} \xi_i - \sum_{i=1}^{N} \lambda_i \{\boldsymbol{y}_i(\boldsymbol{w} \cdot \boldsymbol{x}_i + b) - 1 + \xi_i\} - \sum_{i=1}^{N} \mu_i \xi_i$$

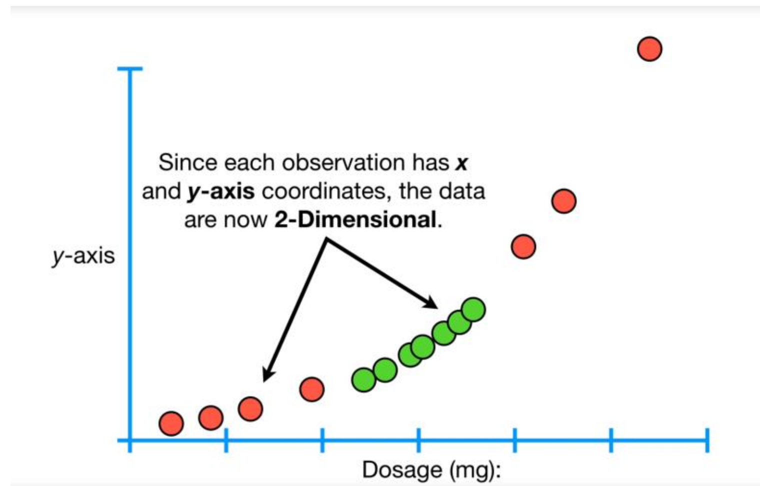
SVM—Linearly Inseparable

- Transform the original input data into a higher dimensional space









Example 6.8 Nonlinear transformation of original input data into a higher dimensional space. Consider the following example. A 3D input vector $\mathbf{X} = (x_1, x_2, x_3)$ is mapped into a 6D space Z using the mappings $\phi_1(X) = x_1, \phi_2(X) = x_2, \phi_3(X) = x_3, \phi_4(X) = (x_1)^2, \phi_5(X) = x_1x_2$, and $\phi_6(X) = x_1x_3$. A decision hyperplane in the new space is $d(\mathbf{Z}) = \mathbf{WZ} + b$, where $\mathbf{W}$ and $\mathbf{Z}$ are vectors. This is linear. We solve for $\mathbf{W}$ and b and then substitute back so that we see that the linear decision hyperplane in the new ($\mathbf{Z}$) space corresponds to a nonlinear second order polynomial in the original 3-D input space,

$$\begin{aligned} d(Z) &= w_1x_1 + w_2x_2 + w_3x_3 + w_4(x_1)^2 + w_5x_1x_2 + w_6x_1x_3 + b \\ &= w_1z_1 + w_2z_2 + w_3z_3 + w_4z_4 + w_5z_5 + w_6z_6 + b \end{aligned} \blacksquare$$

- Search for a linear separating hyperplane in the new space

Kernel Method – in svm

- **Kernel: Higher degree mapping, inner product together**

$$\max L(\alpha_i) = \sum_{i=1}^{n}\alpha_i - \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n}\alpha_i\alpha_j t_i t_j \Phi(x_i)^T\Phi(x_j)$$
$$= \sum_{i=1}^{n}\alpha_i - \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n}\alpha_i\alpha_j t_i t_j K(x_i, x_j)$$

$$K(x_i,\ x_j) = \Phi(x_i)^T\Phi(x_j)$$

- **Kernel trick: $\boldsymbol{\Phi(x_i)^T, \Phi(x_j)}$를 함수로 구현**

$$\underbrace{\Phi(x)}_{n\times1} = \underbrace{A}_{n\times m}\,\underbrace{x}_{m\times1}$$: m차 feature를 n차로

$$K(x_i, x_j) = \underbrace{\Phi(x_i)^T}_{1\times n}\,\underbrace{\Phi(x_j)}_{n\times1} = x_i^T A^T A x_j$$ ⇒ 1×1 (scalar)

- 변환을 해서 내적을 하는것이 아닌, 변환된 곳에서의 내적을 위처럼 표현이 가능하다.
  - 데이터 -> 변환 -> 내적 순서대로 진행하는것이 아닌, 변환을 건너뛰고 바로 내적을 이끌어 내는 과정이 kernel이다.
  - 즉, Kernel은 **"높은 차원에서의 내적을 계산하는 함수"** 이다.

Prof. Hyerim Bae (hrbae@pusan.ac.kr)

SVM: Different Kernel functions

- Instead of computing the dot product on the transformed data, it is math. equivalent to applying a kernel function K($\mathbf{X}_i$, $\mathbf{X}_j$) to the original data, i.e., K($\mathbf{X}_i$, $\mathbf{X}_j$) = Φ($\mathbf{X}_i$) Φ($\mathbf{X}_j$)
- Typical Kernel Functions

$$\textbf{Polynomial kernel of degree } h: \quad K(\boldsymbol{X}_i, \boldsymbol{X}_j) = (\boldsymbol{X}_i \cdot \boldsymbol{X}_j + 1)^h$$

$$\textbf{Gaussian radial basis function kernel}: \quad K(\boldsymbol{X}_i, \boldsymbol{X}_j) = e^{-\|\boldsymbol{X}_i - \boldsymbol{X}_j\|^2 / 2\sigma^2}$$

$$\textbf{Sigmoid kernel}: \quad K(\boldsymbol{X}_i, \boldsymbol{X}_j) = \tanh(\kappa \boldsymbol{X}_i \cdot \boldsymbol{X}_j - \delta)$$

- SVM can also be used for classifying multiple (> 2) classes and for regression analysis (with additional parameters)

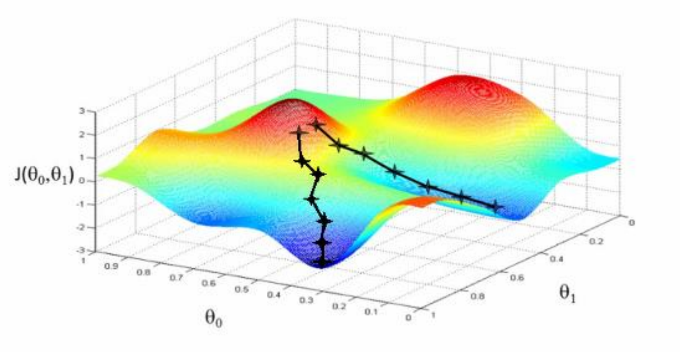
Prof. Hyerim Bae (hrbae@pusan.ac.kr)

2. Using differentiation for unconstrained optimization

- Minimum (Maximum) value is on the point that

$$f'(x) = 0$$

  - Finding x such that $f'(x) = 0$



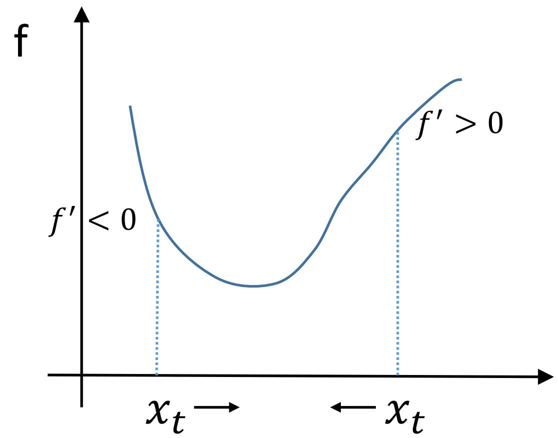
출처: http://blog.datumbox.com/tuning-the-learning-rate-in-gradient-descent/

Prof. Hyerim Bae (hrbae@pusan.ac.kr)

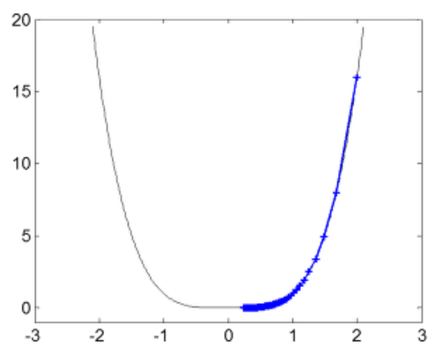
Gradient descent

- Update x value starting from x_0

$$x_{t+1} = x_t - \lambda f'(x_t)$$



$f(x)=x^4$

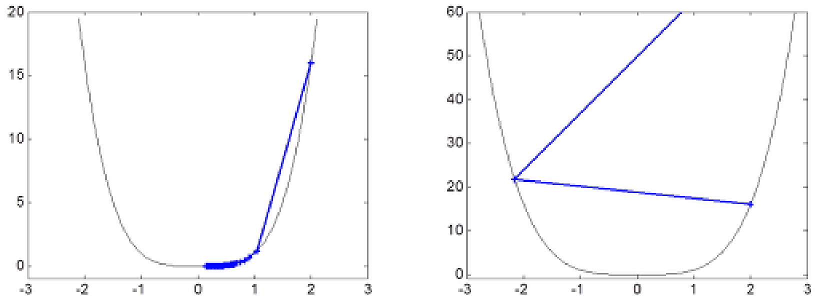


$x_0 = 2.0$
$x_1 = x_0 - \lambda f'(x_0) = 1.6800$
$x_2 = x_1 - \lambda f'(x_1) = 1.4903$
$x_3 = x_2 - \lambda f'(x_2) = 1.3579$
$x_4 = x_3 - \lambda f'(x_3) = 1.2578$
$x_5 = x_4 - \lambda f'(x_4) = 1.1782$
$x_6 = x_5 - \lambda f'(x_5) = 1.1128$
$x_7 = x_6 - \lambda f'(x_6) = 1.0576$
$x_8 = x_7 - \lambda f'(x_7) = 1.0103$
$x_9 = x_8 - \lambda f'(x_8) = 0.9691$
...

https://darkpgmr.tistory.com/149

Prof. Hyerim Bae (hrbae@pusan.ac.kr)

- Problems
  - Even after 200 times of updating, it is not close to x=0
- If we use larger value of λ



$\lambda = 0.03$ $\lambda = 0.13$

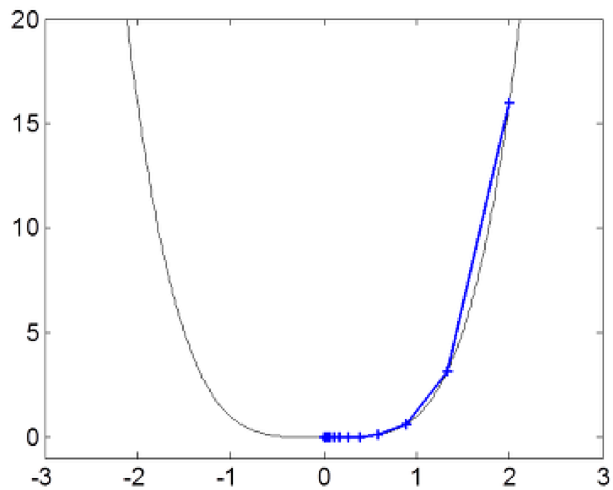
Using secondary derivatives: Newton's method

- Update using the equation: 보다 빠르게 수렴

$$x_{t+1} = x_t - \frac{f'(x_t)}{f''(x_t)}$$

- Problems: 변곡점에서는 정의되지 않음

$f''(x_t)$=0



$x_0 = 2.0$
$x_1 = 1.3333$
$x_2 = 0.8889$
$x_3 = 0.5926$
$x_4 = 0.3951$
$x_5 = 0.2634$
$x_6 = 0.1756$
$x_7 = 0.1171$
$x_8 = 0.0780$
$x_9 = 0.0520$
…

Using random sampling

- Sampling base inference

Black-box optimization

- Bayesian optimization

Conclusions

- ML for opt.
- Opt. for ML

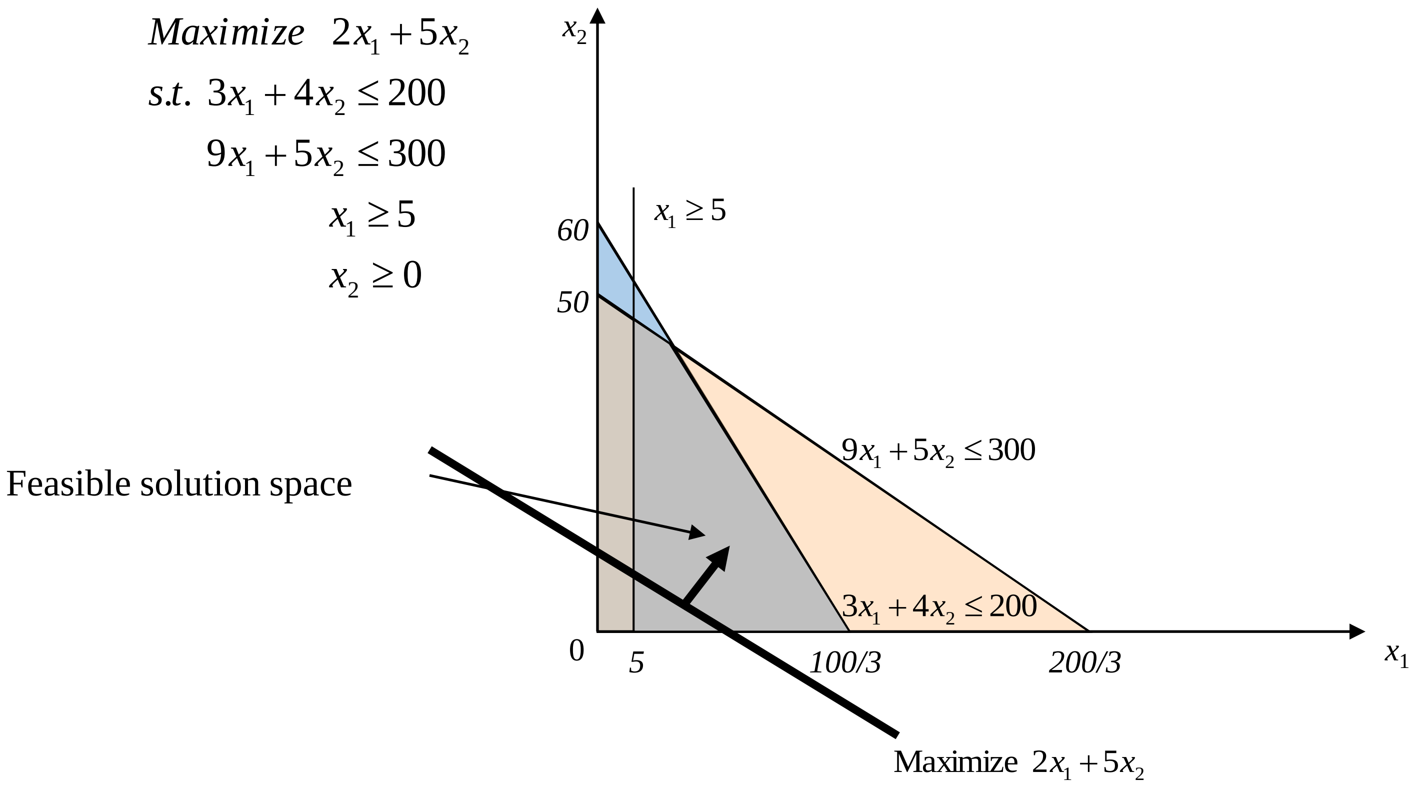
references

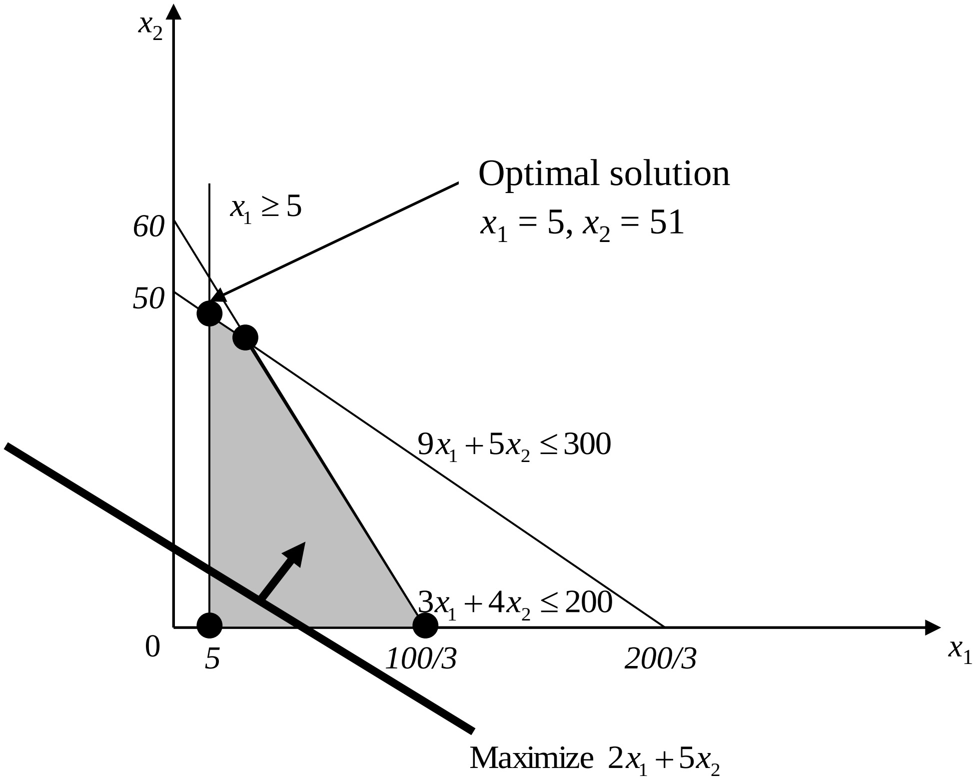
- https://developers.google.com/optimization/lp/lp_example#python_7
- https://darkpgmr.tistory.com/149

Prof. Hyerim Bae (hrbae@pusan.ac.kr)

Geometric way

$$
\begin{aligned}
&\textit{Maximize} \quad 2x_1 + 5x_2 \\
&s.t. \quad 3x_1 + 4x_2 \le 200 \\
&\qquad\ \ 9x_1 + 5x_2 \le 300 \\
&\qquad\qquad\quad x_1 \ge 5 \\
&\qquad\qquad\quad x_2 \ge 0
\end{aligned}
$$





x_2
x_1
60
50
0
5
100/3
200/3
$x_1 \geq 5$
Optimal solution
$x_1 = 5, x_2 = 51$
$9x_1 + 5x_2 \leq 300$
$3x_1 + 4x_2 \leq 200$
Maximize $2x_1 + 5x_2$

[Ex] Production plan for maximization of profit

A factory produces three products: A, B, and C. One unit of product A requires 2.4 minutes of rolling time and 5.0 minutes of assembly process. Profit is $600. One unit of product B requires 3.0 minutes of rolling time, 2.5 minutes of welding process, and $700 of profit is generated. One unit of product C requires 2.0 minutes of rolling time, 1.5 minutes of welding time, 2.5 minutes of assembly time, and generates a profit of $500.
The production time of the rolling process is 1,200 minutes per week, the welding process is 600 minutes per week, and the assembly process can be 1,500 minutes per week.

What is the production of products A, B, and C that can generate the greatest profit?

$$
\begin{array}{lll}
\textit{Maximize} \;\; 600x_1 + 700x_2 + 500x_3 & \longleftarrow & \text{Profit maximization} \\
s.t. \;\; 2.4x_1 + 3.0x_2 + 2.0x_3 \le 1200 & \longleftarrow & \text{Rolling time constraints} \\
\qquad 0.0x_1 + 2.5x_2 + 1.5x_3 \le 600 & \longleftarrow & \text{Welding time constraints} \\
\qquad 5.0x_1 + 0.0x_2 + 2.5x_3 \le 1500 & \longleftarrow & \text{Assembly time constraints} \\
\qquad x_1, x_2, x_3 \ge 0 & \longleftarrow & \text{Non-negative}
\end{array}
$$

- Geometric approach?
  - Solutions are located at Vertexes
  - Within 2 or 3 dimension
    - Not easy for more than 3-dimensional problems
    - Not easy for a problem with a large number of constraints

=> Need algorithmic approach

- Simplex Method
  - Dantzig
  - Exploring the optimal solution among the intersections of constraints.
- Karmarkar Method

Prof. Hyerim Bae (hrbae@pusan.ac.kr)

Simplex Method

- Simplex method(單體法)
  - Using simultaneous linear equations
    - Matrix operation: Gauss-Jordan elimination
  - Easy to understand, easy to use
  - Find the opt. sol. among vertexes in a feasible space
    - Initial basic solutions => [repetitive improvement] => optimal solution

$$\begin{cases} \text{Maximize} & 12x_1 + 15x_2 \\ \text{s.t.} & \left.\begin{array}{l} 4x_1 + 3x_2 \le 12 \\ 2x_1 + 5x_2 \le 10 \end{array}\right\} \\ & x_1, x_2 \ge 0 \end{cases} \xrightarrow{\textbf{To equality}} \begin{cases} \text{Maximize} & 12x_1 + 15x_2 \\ \text{s.t.} & \left.\begin{array}{l} 4x_1 + 3x_2 + x_3 \qquad = 12 \\ 2x_1 + 5x_2 \qquad + x_4 = 10 \end{array}\right\} \\ & x_1, x_2, x_3, x_4 \ge 0 \end{cases}$$

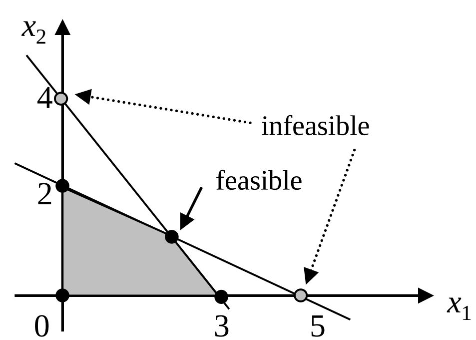
System of equations

$$\left.\begin{cases} 4x_1 + 3x_2 + x_3 \qquad = 12 \\ 2x_1 + 5x_2 \qquad + x_4 = 10 \end{cases}\right\} \text{Infinite number of solutions}$$

6 solutions by pairs of solutions ($_4C_2$)

기저	해 (x_1, x_2, x_3, x_4)
(x_1, x_2)	(15/7, 8/7, 0, 0)
(x_1, x_3)	(5, 0, -8, 0)
(x_1, x_4)	(3, 0, 0, 4)
(x_2, x_3)	(0, 2, 6, 0)
(x_2, x_4)	(0, 4, 0, -10)
(x_3, x_4)	(0, 0, 12, 10)

$\Leftrightarrow$



Prof. Hyerim Bae (hrbae@pusan.ac.kr)

Object function

$$\text{Maximize } 12x_1 + 15x_2$$

A single solution that maximizes the O-function => optimal solution

Basis	Sol. (x_1, x_2, x_3, x_4)	Values of O-function
(x_1, x_2)	(15/7, 8/7, 0, 0)	300/7
(x_1, x_3)	(5, 0, -8, 0)	60
(x_1, x_4)	(3, 0, 0, 4)	36
(x_2, x_3)	(0, 2, 6, 0)	30
(x_2, x_4)	(0, 4, 0, -10)	60
(x_3, x_4)	(0, 0, 12, 10)	0

Infeasible

Optimal solution

Infeasible

A solution using simply a system of equations of constraints should find ${}_nC_m$ solutions if it is a problem with m constraints consisting of n determinants. Depending on the values of *n* and *m*, the search space increases exponentially.

=> There is a need for a method that does not search for infeasible solutions, but only within the feasible, but steadily increases the value of the objective function.

Simplex method

$$
\begin{aligned}
\text{Maximize} \quad & 4x_1 + 5x_2 + 8x_3 + 11x_4 \\
\text{s.t.} \quad & x_1 + x_2 + x_3 + x_4 \le 15 \\
& 3x_1 + 5x_2 + 10x_3 + 15x_4 \le 105 \\
& 7x_1 + 5x_2 + 3x_3 + 2x_4 \le 120 \\
& x_j \ge 0, \ \forall j
\end{aligned}
$$

Adding slack var.

Min. Coef. **= -11**

Init. Sol.

$$
\begin{aligned}
z - 4x_1 - 5x_2 - 8x_3 - 11x_4 &= 0 \\
x_1 + x_2 + x_3 + x_4 + x_5 &= 15 \\
3x_1 + 5x_2 + 10x_3 + 15x_4 + x_6 &= 105 \\
7x_1 + 5x_2 + 3x_3 + 2x_4 + x_7 &= 120 \\
x_j \ge 0, \ \forall j &
\end{aligned}
$$

15/1 = 15

105/15 = 7

120/2 = 60

Min. ratio

pivoting

$$
\begin{aligned}
z - \frac{9}{5}x_1 - \frac{4}{3}x_2 - \frac{2}{3}x_3 + \frac{11}{15}x_6 &= 77 \\
\frac{4}{5}x_1 + \frac{2}{3}x_2 + \frac{1}{3}x_3 + x_5 - \frac{1}{15}x_6 &= 8 \\
\frac{1}{5}x_1 + \frac{1}{3}x_2 + \frac{2}{3}x_3 + x_4 + \frac{1}{15}x_6 &= 7 \\
\frac{33}{5}x_1 + \frac{13}{3}x_2 + \frac{5}{3}x_3 - \frac{2}{15}x_6 + x_7 &= 106 \\
x_j \ge 0, \ \forall j &
\end{aligned}
$$

A new solution (0, 0, 0, 7, 8, 0, 106)

A new value of O-function 77

Prof. Hyerim Bae (hrbae@pusan.ac.kr)

$$
\begin{aligned}
z - \boxed{\tfrac{9}{5}x_1} - \tfrac{4}{3}x_2 - \tfrac{2}{3}x_3 \qquad\qquad + \tfrac{11}{15}x_6 &= 77 \\
\boxed{\tfrac{4}{5}x_1 + \tfrac{2}{3}x_2 + \tfrac{1}{3}x_3 \qquad + x_5 - \tfrac{1}{15}x_6 = 8} & \\
\tfrac{1}{5}x_1 + \tfrac{1}{3}x_2 + \tfrac{2}{3}x_3 + x_4 \qquad + \tfrac{1}{15}x_6 &= 7 \\
\tfrac{33}{5}x_1 + \tfrac{13}{3}x_2 + \tfrac{5}{3}x_3 \qquad - \tfrac{2}{15}x_6 + x_7 &= 106 \\
x_j \ge 0,\ \forall j &
\end{aligned}
$$

Pivoting

All the coef. >= 0

$$
\begin{aligned}
\boxed{z \quad + \tfrac{1}{6}x_2 + \tfrac{1}{12}x_3 \qquad + \tfrac{9}{4}x_5 + \tfrac{7}{12}x_6 = 95} & \\
x_1 + \tfrac{5}{6}x_2 + \tfrac{5}{12}x_3 \qquad + \tfrac{5}{4}x_5 - \tfrac{1}{12}x_6 &= 10 \\
\tfrac{1}{6}x_2 + \tfrac{7}{12}x_3 + x_4 - \tfrac{1}{4}x_5 + \tfrac{1}{12}x_6 &= 5 \\
-\tfrac{7}{6}x_2 - \tfrac{13}{12}x_3 \qquad - \tfrac{33}{4}x_5 + \tfrac{5}{12}x_6 + x_7 &= 40 \\
x_j \ge 0,\ \forall j &
\end{aligned}
$$

Opt. Sol. $x^* = (10,0,0,5,0,0,40)$

Opt. Val. $z^* = 95$

단체표의 이용 [요약정리]

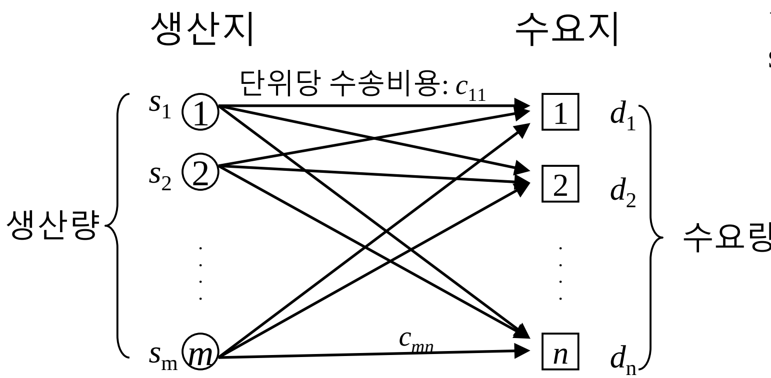
z	x_1	x_2	x_3	x_4	x_5	x_6	x_7	
1	-4	-5	-8	-11	0	0	0	0
0	1	1	1	1	1	0	0	15
0	3	5	10	15	0	1	0	105
0	7	5	3	2	0	0	1	120
1	-9/5	-4/3	-2/3	0	0	11/15	0	77
0	4/5	2/3	1/3	0	1	-1/15	0	8
0	1/5	1/3	2/3	1	0	1/15	0	7
0	33/5	13/3	5/3	0	0	-2/15	1	106
1	0	1/6	1/12	0	9/4	7/12	0	95
0	1	5/6	5/12	0	5/4	-1/12	0	10
0	0	1/6	7/12	1	-1/4	1/12	0	5
0	0	-7/6	-13/12	0	-33/4	5/12	1	40

Prof. Hyerim Bae (hrbae@pusan.ac.kr)

Other Linear Problems

- Transportation Problem

  m개의 생산지로부터 n개의 수요지로 수요량을 만족시키면서, 최소의 비용으로 전달하는 문제



$$
\begin{aligned}
\text{Minimize} \quad & c_{11}x_{11} + c_{12}x_{12} + \ldots + c_{mn}x_{mn} \\
\text{s.t.} \quad & x_{11} + x_{12} + \ldots + x_{1n} \le s_1 \\
& \ldots\ldots \\
& x_{m1} + x_{m2} + \ldots + x_{mn} \le s_m \\
& x_{11} + x_{21} + \ldots + x_{m1} \ge d_1 \\
& \ldots\ldots \\
& x_{1n} + x_{2n} + \ldots + x_{mn} \ge d_n \\
& x_{ij} \ge 0, \quad \forall i, j
\end{aligned}
$$

Prof. Hyerim Bae (hrbae@pusan.ac.kr)

Optimizing transportation

A 건설회사에서 3곳의 야산으로부터 모래를 운반하여 4곳의 아파트 부지에 공급한다.
모래의 운반과 관련한 비용 및 생산량과 수요량이 다음의 행렬에 정리되어 있다.
최소의 운반 비용을 얻을 수 있는 수송 경로를 구하여라.

야산 \ 아파트 부지	d_1	d_2	d_3	d_4	공급량
s_1	2	3	11	7	6
s_2	1	0	6	1	1
s_3	5	8	15	9	10
수요량	7	5	3	2	합: 17

단위: 100만원/톤

$$\text{Minimize} \quad 2x_{11} + 3x_{12} + 11x_{13} + 7x_{14} + x_{21} + 6x_{23} + x_{24} + 5x_{31} + 8x_{32} + 15x_{33} + 9x_{34}$$

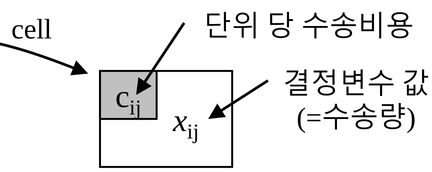
$$\begin{aligned}
\text{s.t.} \quad & x_{11} + x_{12} + x_{13} + x_{14} \le 6 \\
& x_{21} + x_{22} + x_{23} + x_{24} \le 1 \\
& x_{31} + x_{32} + x_{33} + x_{34} \le 10 \\
& x_{11} + x_{21} + x_{31} \ge 7 \\
& x_{12} + x_{22} + x_{32} \ge 5 \\
& x_{13} + x_{23} + x_{33} \ge 3 \\
& x_{14} + x_{24} + x_{34} \ge 2 \\
& x_{ij} \ge 0, \quad \forall i, j
\end{aligned}$$

Prof. Hyerim Bae (hrbae@pusan.ac.kr)

Step 1) 초기해를 여러 가지 방법을 활용하여 설정한다.
Step 2) 초기해를 개선시킬 수 있는 수송비용의 음환(negative cycle)을 찾는다.
Step 3) 더 이상 해를 개선시킬 수 있는 음환을 찾을 수 없을 때가 최적이다.

<u>1) 최소가법을 이용하여 설정된 초기해</u>

수요지 / 공급지	d_1	d_2	d_3	d_4	공급량
s_1	[2] 6	[3]	[11]	[7]	6
s_2	[1]	[0] 1	[6]	[1]	1
s_3	[5] 1	[8] 4	[15] 3	[9] 2	10
수요량	7	5	3	2	총합: 17



$$총비용 = 2\cdot 6 + 0\cdot 1 + 5\cdot 1 + 8\cdot 4 + 15\cdot 3 + 9\cdot 2 = 112$$

<u>2) 음환을 찾는다.</u>

$$x_{12}\uparrow = +3-2+5-8 = -2 < 0$$
$$x_{13}\uparrow = +11-2+5-15 = -1 < 0$$
$$x_{14}\uparrow = +7-2+5-9 = 1 > 0$$
$$x_{21}\uparrow = +1-5+8-0 = 4 > 0$$
$$x_{23}\uparrow = +6-0+8-15 = -1 < 0$$
$$x_{24}\uparrow = +1-0+8-9 = 0$$

x_{12} 진입

<u>3) 개선된 해</u>

수요지 / 공급지	d_1	d_2	d_3	d_4	공급량
s_1	[2] 2	[3] 4	[11]	[7]	6
s_2	[1]	[0] 1	[6]	[1]	1
s_3	[5] 5	[8]	[15] 3	[9] 2	10
수요량	7	5	3	2	총합: 17

$$총비용 = 2\cdot 2 + 3\cdot 4 + 0\cdot 1 + 5\cdot 5 + 15\cdot 3 + 9\cdot 2 = 104$$

Diet Problem

- Diet

  여러 가지 영양분을 지닌 음식들로부터 필수 영양분을 최소의 음식값으로 섭취하는 문제

[예] 주위에서 흔히 볼 수 있는 음식 재료에 포함된 영양분이 다음 표에 정리되어 있다. 최소의 비용으로 식단을 마련해 보고자 한다. 단 하루의 식단에서 쌀은 20포, 쇠고기는 1근, 우유는 2통, 계란은 3개, 배추는 3단을 넘지 않기로 한다.

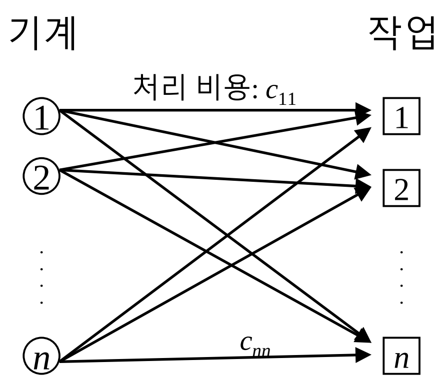
재료 / 영양	쌀(포)	쇠고기(근)	우유(통)	계란(12 개)	배추(단)	1 일 필요량
열량(Kcal)	340	1080	362	1040	17	2200
단백질(g)	6.5	167	19	78	1.3	70
비타민(I.U)	0	97	758	7080	255	5000
철분(mg)	0.4	11	0.3	13	0.3	12.5
탄수화물(g)	52	30	25	0	5	
콜레스테롤(u)	0	22	11	120	0	
값(원)	75	1640	370	550	110	

$$
\begin{aligned}
\text{Minimize}\quad & 75x_1 + 1640x_2 + 370x_3 + 550x_4 + 110x_5 \\
\text{s.t.}\quad & 340x_1 + 1080x_2 + 362x_3 + 1040x_4 + 17x_5 \ge 2200 \\
& 6.5x_1 + 167x_2 + 19x_3 + 78x_4 + 1.3x_5 \ge 70 \\
& 97x_2 + 758x_3 + 7080x_4 + 255x_5 \ge 5000 \\
& 0.4x_1 + 11x_2 + 0.3x_3 + 13x_4 + 0.3x_5 \ge 12.5 \\
& x_1 \le 20,\quad x_2 \le 1,\quad x_3 \le 2,\quad x_4 \le 0.25,\quad x_5 \le 3 \\
& x_{ij} \ge 0,\quad \forall i, j
\end{aligned}
$$

Assignment Problem

배정 문제(Assignment Problem): n개의 작업을 n개의 기계에 각기 하나씩 최소의 비용이 되도록 할당하는 문제

공급량과 수요량이 각기 1씩 발생하는 특별한 경우의 수송문제



만약 기계 i가 작업 j에 할당되면 $x_{ij} = 1$

아니면 $x_{ij} = 0$

$$\text{Minimize} \sum_{i=1}^{n} \sum_{j=1}^{n} c_{ij} x_{ij}$$

$$\text{s.t.} \quad \sum_{j=1}^{n} x_{ij} = 1, \quad i = 1, 2, \ldots, n$$

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

$$x_{ij} \geq 0, \quad \forall i, j$$

Prof. Hyerim Bae (hrbae@pusan.ac.kr)

Knapsack Problem

- Knapsack Problem

  한정된 배낭의 용량에 맞게 각각의 용량을 가지는 물건을 최대의 효용을 얻도록 채우는 문제

  – 대표적인 정수계획법 문제 중의 하나

[예] 갑돌이는 등산을 계획하고 있는데, 가면서 먹을 음식을 결정해야만 한다. 배당에는 총 1.6 kg까지만 음식을 담기로 결정했다고 한다. 각각의 음식의 무게와 그 음식을 가져 감으로써 얻을 수 있는 만족도가 다음과 같을 때, 가장 큰 만족도를 얻을 수 있는 음식의 조합을 결정하시오.

물건	고기	쌀	라면	과일	빵	
만족도	20	48	14	18	20	배낭의 무게
무게(100g)	8	6	2	3	2	16

$$
\begin{aligned}
&\text{Maximize } 20x_1 + 48x_2 + 14x_3 + 18x_4 + 20x_5 \\
&\text{s.t.} \qquad 8x_1 + 6x_2 + 2x_3 + 3x_4 + 2x_5 \le 16 \\
&\qquad\quad x_i = 1 \text{ or } 0, \quad \forall i
\end{aligned}
$$

Prof. Hyerim Bae (hrbae@pusan.ac.kr)

ORTools

```
!pip install ortools
```

```
from ortools.linear_solver import pywraplp


def LinearProgrammingExample():
    """Linear programming sample."""
    # Instantiate a Glop solver, naming it LinearExample.
    solver = pywraplp.Solver.CreateSolver('GLOP')

    # Create the two variables and let them take on any non-negative value.
    x1 = solver.NumVar(0, solver.infinity(), 'x1')
    x2 = solver.NumVar(0, solver.infinity(), 'x2')
    print('Number of variables =', solver.NumVariables())

    # Constraint 0: 4x1 + 3x2 <= 12.
    solver.Add(4*x1 + 3 * x2 <= 12)

    # Constraint 1: 2x1 + 5x2 <= 10.
    solver.Add(2*x1 + 5*x2 <= 10)


    print('Number of constraints =', solver.NumConstraints())

    # Objective function: 12x1 + 15x2.
    solver.Maximize(12 * x1 + 15 * x2)

    # Solve the system.
    status = solver.Solve()

    # print_solution
    if status == pywraplp.Solver.OPTIMAL:
        print('Solution:')
        print('Objective value =', solver.Objective().Value())
        print('x1 =', x1.solution_value())
        print('x2 =', x2.solution_value())
    else:
        print('The problem does not have an optimal solution.')


LinearProgrammingExample()
```

Prof. Hyerim Bae (hrbae@pusan.ac.kr)

Home work

- If f is convex, prove that q is concave

Prof. Hyerim Bae (hrbae@pusan.ac.kr)