Constrained Optimization for SVM (Lagrange Multiplier)

Data Intelligence and Learning (DIAL) Lab

Prof. Jongwuk Lee

Lagrange Multiplier Basics

Unconstrained Optimization

- Given a differentiable objective function,

$$f(\mathbf{x}) = f(x_1, \dots, x_d)$$

- If we have no constraints, the extrema must necessarily satisfy the following system of equations.

$$\nabla f(\mathbf{x}) = 0 \quad \Leftrightarrow \quad \frac{\partial f}{\partial x_i} = 0 \ \text{ for } \ i = 1, \dots, d$$

- What if finding the extrema under additional constraints?

Lagrange Multipliers

- **It is an optimization method for finding the maxima or minima of a function subject to equality constraints.**
  - It converts a **constrained problem** into **the derivative test of an unconstrained problem**.

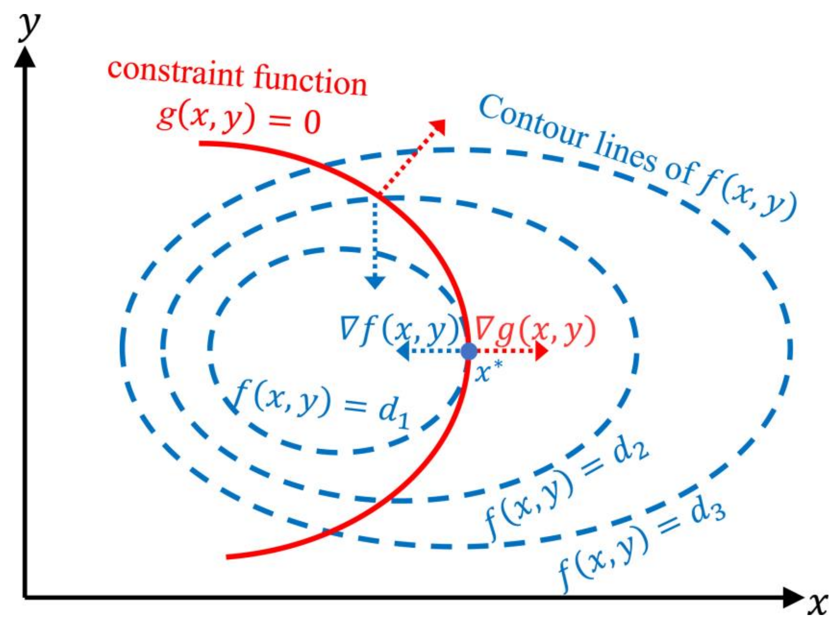
$$\min_{x,y} f(x,y) \text{ subject to } g(x,y)=0$$

The largest value of c such that $f(x,y)=c$ intersects $g(x,y)=k$.

⬇

It happens when the lines are **parallel**.

$$\nabla f(x,y)=\lambda \nabla g(x,y)$$



Idea of Lagrange Multipliers

- Find the minima of $f(\mathbf{x})$ subject to m constraints $g_j(\mathbf{x})$.

$$\min_{\mathbf{x}} f(\mathbf{x}) \text{ subject to } g_j(\mathbf{x}) = 0 \quad \text{for} \quad j = 1, \dots, k$$

$$\Downarrow$$

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

- At the minimum of $L(\mathbf{x}, \boldsymbol{\lambda})$, $\nabla_{\mathbf{x},\boldsymbol{\lambda}} L = 0$.

$$\frac{\partial L}{\partial x_i} = 0 \text{ for } i = 1, \dots, d \text{ and } \frac{\partial L}{\partial \lambda_j} = 0 \text{ for } g = 1, \dots, k$$

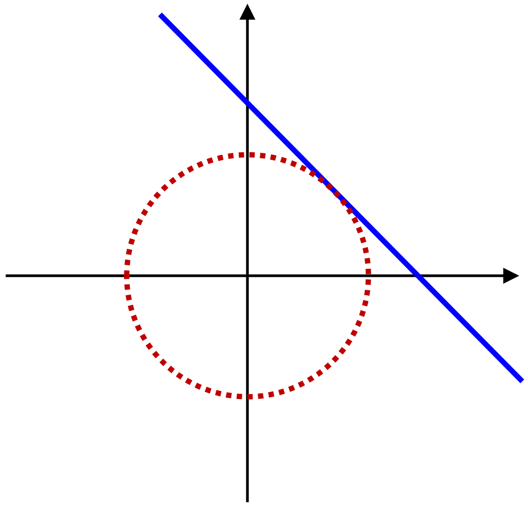
This makes the curves be parallel.

Example 1

$$\min_{x,y} x^2 + y^2 \quad \text{subject to} \quad x + y = 1$$

$$L(x, y, \lambda) = (x^2 + y^2) + \lambda(x + y - 1)$$

$$\left.\begin{aligned} \frac{\partial L}{\partial x} &= 2x + \lambda = 0 \\ \frac{\partial L}{\partial y} &= 2y + \lambda = 0 \\ \frac{\partial L}{\partial \lambda} &= x + y - 1 = 0 \end{aligned}\right\} \quad x = 0.5, y = 0.5, \lambda = -1$$



Example 2

$$\min_{x,y,z} x^2 + y^2 + z^2 \quad \text{subject to} \;\; x + y = 2 \text{ and } x + z = 3$$

$$L(x, y, z, \lambda_1, \lambda_2) = (x^2 + y^2 + z^2) + \lambda_1(x + y - 2) + \lambda_2(x + z - 3)$$

$$\frac{\partial L}{\partial x} = 2x + \lambda_1 + \lambda_2 = 0$$

$$\frac{\partial L}{\partial y} = 2y + \lambda_1 = 0$$

$$\frac{\partial L}{\partial z} = 2z + \lambda_2 = 0$$

$$\frac{\partial L}{\partial \lambda_1} = x + y - 2 = 0$$

$$\frac{\partial L}{\partial \lambda_2} = x + z - 3 = 0$$

$$x = \frac{5}{3}, y = \frac{1}{3}, z = \frac{4}{3}$$

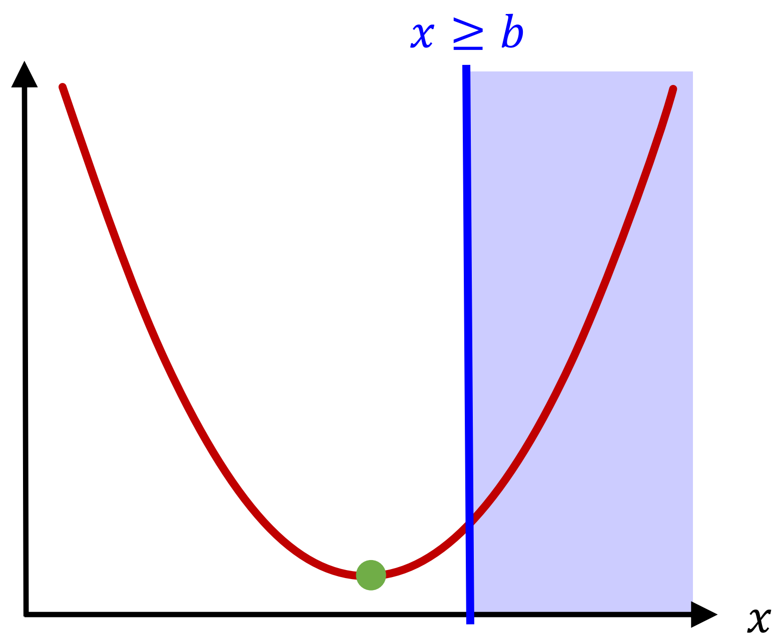
$$\lambda_1 = -\frac{2}{3}, \lambda_2 = -\frac{8}{3}$$

Generalizing Lagrange Multipliers

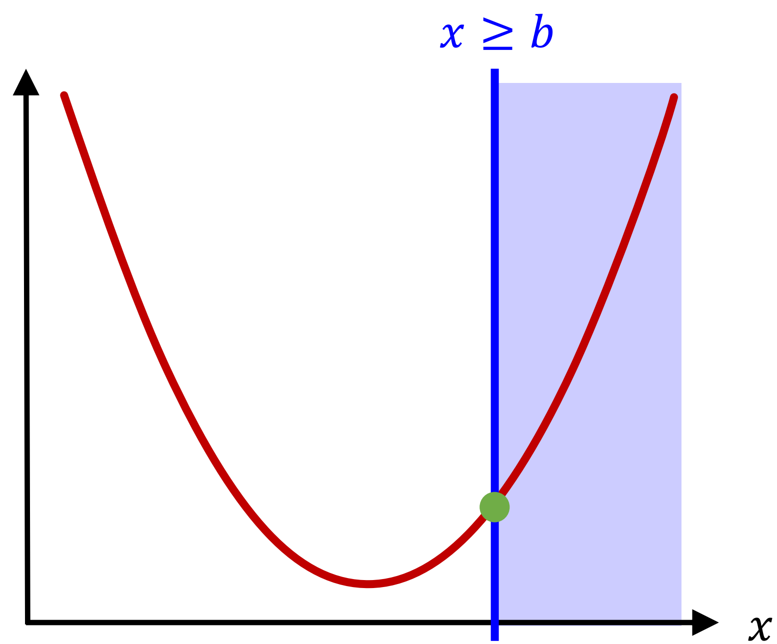
How Inequality Constraints?

- **The constraint condition may affect finding the optima.**

$$\min_x x^2 \ \text{ subject to } x \geq b$$



Inactive constraint



Active constraint

Handling Inequality Constraints

- **Objective function**

$$\min_{\mathbf{x}} f(\mathbf{x}) \text{ subject to } h(\mathbf{x}) \leq 0$$

Inequality constraint

$$L(\mathbf{x}, \mu) = f(\mathbf{x}) + \mu h(\mathbf{x})$$

- **Solve $\nabla_{\mathbf{x},\mu} L = 0$.**
- **Also, consider $\mu h(\mathbf{x}) = 0$ and $\boldsymbol{\mu} \geq 0$ and $h(\mathbf{x}) \leq 0$.**
  - If $\mu > 0$, the constraint $h(\mathbf{x})$ is **active**, i.e., Lagrange multiplier.
  - If $\mu = 0$, the constraint $h(\mathbf{x})$ is **not active**.

- **It is called Karush-Kuhn-Tucker (KKT) conditions.**

Karush-Kuhn-Tucker (KKT) Conditions

- **Assume that functions are convex.**

$$\min_{\mathbf{x}} f(\mathbf{x}) \quad \text{subject to} \begin{cases} g_i(\mathbf{x}) = 0 \ \text{ for } \ i = 1, \dots, k \\ h_j(\mathbf{x}) \le 0 \ \text{ for } \ j = 1, \dots, m \end{cases}$$

$$\Downarrow$$

$$\min_{\mathbf{x}} \max_{\boldsymbol{\lambda},\boldsymbol{\mu}} L(\mathbf{x}, \boldsymbol{\lambda}, \boldsymbol{\mu}) \ \text{subject to} \begin{cases} g_i(\mathbf{x}) = 0 \ \text{ for } \ i = 1, \dots, k \\ h_j(\mathbf{x}) \le 0 \ \text{ for } \ j = 1, \dots, m \end{cases}$$

where

$$L(\mathbf{x}, \boldsymbol{\lambda}, \boldsymbol{\mu}) = f(\mathbf{x}) + \sum_{i=1}^{k} \lambda_i \, g_i(\mathbf{x}) + \sum_{j=1}^{m} \mu_j \, h_j(\mathbf{x})$$

$\boldsymbol{\lambda} = (\lambda_1, \dots, \lambda_k)$, $\quad \boldsymbol{\mu} = (\mu_1, \dots, \mu_m)$

Lagrange multiplier **KKT multiplier**

Karush-Kuhn-Tucker (KKT) Conditions

$$\min_{\mathbf{x}} \max_{\boldsymbol{\lambda},\boldsymbol{\mu}} L(\mathbf{x}, \boldsymbol{\lambda}, \boldsymbol{\mu}) \text{ subject to } \begin{cases} g_i(\mathbf{x}) = 0 \text{ for } i = 1, \dots, k \\ h_j(\mathbf{x}) \leq 0 \text{ for } j = 1, \dots, m \end{cases}$$

where

$$L(\mathbf{x}, \boldsymbol{\lambda}, \boldsymbol{\mu}) = f(\mathbf{x}) + \sum_{i=1}^{k} \lambda_i g_i(\mathbf{x}) + \sum_{j=1}^{m} \mu_j h_j(\mathbf{x})$$

$$\boldsymbol{\lambda} = (\lambda_1, \dots, \lambda_k), \qquad \boldsymbol{\mu} = (\mu_1, \dots, \mu_m)$$

- **The solution satisfies the following conditions.**

$$\frac{\partial L}{\partial x_i} = 0 \text{ for } i = 1, \dots, d \text{ and } \frac{\partial L}{\partial \lambda_j} = 0 \text{ for } g = 1, \dots, k$$

$$\lambda_i g_i(\mathbf{x}) = 0 \text{ for } i = 1, \dots, k$$

$$\mu_j h_j(\mathbf{x}) = 0 \text{ and } \mu_j \geq 0 \text{ and } h_j(\mathbf{x}) \leq 0 \text{ for } j = 1, \dots m$$

Karush-Kuhn-Tucker (KKT) Conditions

- **How to solve this problem?**

$$\frac{\partial L}{\partial x_i} = 0 \text{ for } i = 1, \ldots, d \text{ and } \frac{\partial L}{\partial \lambda_j} = 0 \text{ for } g = 1, \ldots, k$$

$$\lambda_i g_i(\mathbf{x}) = 0 \text{ for } i = 1, \ldots, k$$

$$\mu_j h_j(\mathbf{x}) = 0 \text{ and } \mu_j \geq 0 \text{ and } h_j(\mathbf{x}) \leq 0 \text{ for } j = 1, \ldots m$$

- **Step 1: Divide into subproblems using KKT multipliers.**
- **Step 2: Solve equations for each subproblem.**
- **Step 3: Check whether the solution is feasible or not.**

Example 1

$$\min_{x,y} x^2 + y^2 \quad \text{subject to} \;\; x + y = 1 \;\; \text{and}\; x \geq 2$$

$$L(x, y, \lambda, \mu) = (x^2 + y^2) + \lambda(x + y - 1) + \mu(-x + 2)$$

$$\left.\begin{aligned} &\frac{\partial L}{\partial x} = 2x + \lambda - \mu = 0 \\ &\frac{\partial L}{\partial y} = 2y + \lambda = 0 \\ &\frac{\partial L}{\partial \lambda} = x + y - 1 = 0 \\ &\mu(-x + 2) = 0 \text{ and } -x + 2 \leq 0 \end{aligned}\right\}$$

Case 1: $\boldsymbol{\mu = 0}$

$$2x + \lambda = 0$$
$$2y + \lambda = 0$$
$$x + y - 1 = 0$$
$$-x + 2 \leq 0$$

Case 2: $\boldsymbol{\mu \neq 0}$

$$2x + \lambda - \mu = 0$$
$$2y + \lambda = 0$$
$$x + y - 1 = 0$$
$$-x + 2 = 0$$

Example 1

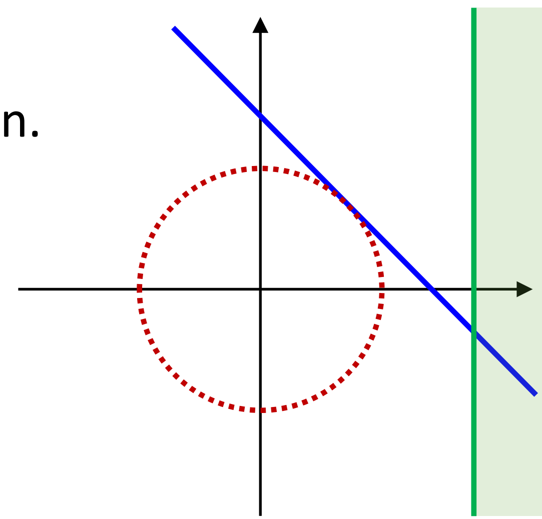
- **For the Case 1,**
  - $x = 0.5, y = 0.5, \lambda = -1$
  - Since $x \geq 2$, it is infeasible.
- **For the Case 2,**
  - $x = 2, y = -1$
  - $\lambda = -2, \mu = 2$
  - It is a feasible solution.
- **Finally, we choose the case 2.**

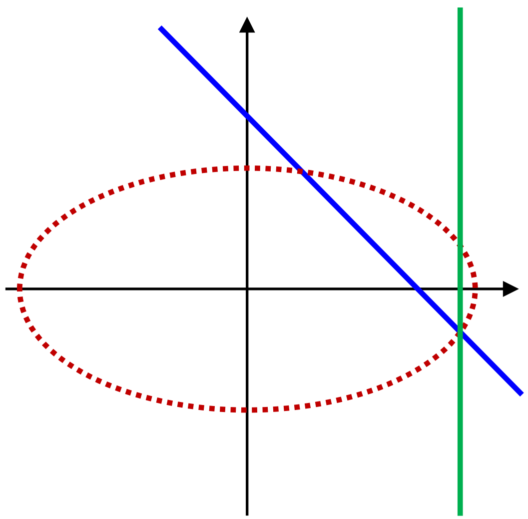
Case 1: $\boldsymbol{\mu = 0}$

$$2x + \lambda = 0$$
$$2y + \lambda = 0$$
$$x + y - 1 = 0$$
$$-x + 2 \leq 0$$

Case 2: $\boldsymbol{\mu \neq 0}$

$$2x + \lambda - \mu = 0$$
$$2y + \lambda = 0$$
$$x + y - 1 = 0$$
$$-x + 2 = 0$$





Example 2

$$\min_{x,y} x^2 + y^2 \quad \text{subject to } \; x + y = 1 \; \text{ and } x \geq 2 \text{ and } y \leq 2$$

$$L(x, y, \lambda, \mu) = (x^2 + y^2) + \lambda(x + y - 1) + \mu_1(-x + 2) + \mu_2(y - 2)$$

$$\frac{\partial L}{\partial x} = 2x + \lambda - \mu_1 = 0$$

$$\frac{\partial L}{\partial y} = 2y + \lambda + \mu_2 = 0$$

$$\frac{\partial L}{\partial \lambda} = x + y - 1 = 0$$

$$\mu_1(-x + 2) = 0 \text{ and } -x + 2 \leq 0$$

$$\mu_2(y - 2) = 0 \text{ and } y - 2 \leq 0$$

We need to solve four cases.

Case 1: $\mu_1 = 0, \mu_2 = 0$

Case 2: $\mu_1 = 0, \mu_2 \neq 0$

Case 3: $\mu_1 \neq 0, \mu_2 = 0$

Case 4: $\mu_1 \neq 0, \mu_2 = 0$

Example 2

$$\frac{\partial L}{\partial x} = 2x + \lambda - \mu_1 = 0$$
$$\frac{\partial L}{\partial y} = 2y + \lambda + \mu_2 = 0$$
$$\frac{\partial L}{\partial \lambda} = x + y - 1 = 0$$

$\mu_1(-x + 2) = 0$ and $-x + 2 \leq 0$

$\mu_2(y - 2) = 0$ and $y - 2 \leq 0$

$\Downarrow$

C1: $\mu_1 = 0, \mu_2 = 0$

$$2x + \lambda = 0$$
$$2y + \lambda = 0$$
$$x + y - 1 = 0$$
$$-x + 2 \leq 0$$
$$y - 2 \leq 0$$

C2: $\mu_1 = 0, \mu_2 \neq 0$

$$2x + \lambda = 0$$
$$2y + \lambda + \mu_2 = 0$$
$$x + y - 1 = 0$$
$$-x + 2 = 0$$
$$y - 2 \leq 0$$

C3: $\mu_1 \neq 0, \mu_2 = 0$

$$2x + \lambda - \mu_1 = 0$$
$$2y + \lambda = 0$$
$$x + y - 1 = 0$$
$$-x + 2 = 0$$
$$y - 2 \leq 0$$

C4: $\mu_1 \neq 0, \mu_2 \neq 0$

$$2x + \lambda - \mu_1 = 0$$
$$2y + \lambda + \mu_2 = 0$$
$$x + y - 1 = 0$$
$$-x + 2 = 0$$
$$y - 2 = 0$$

Example 2

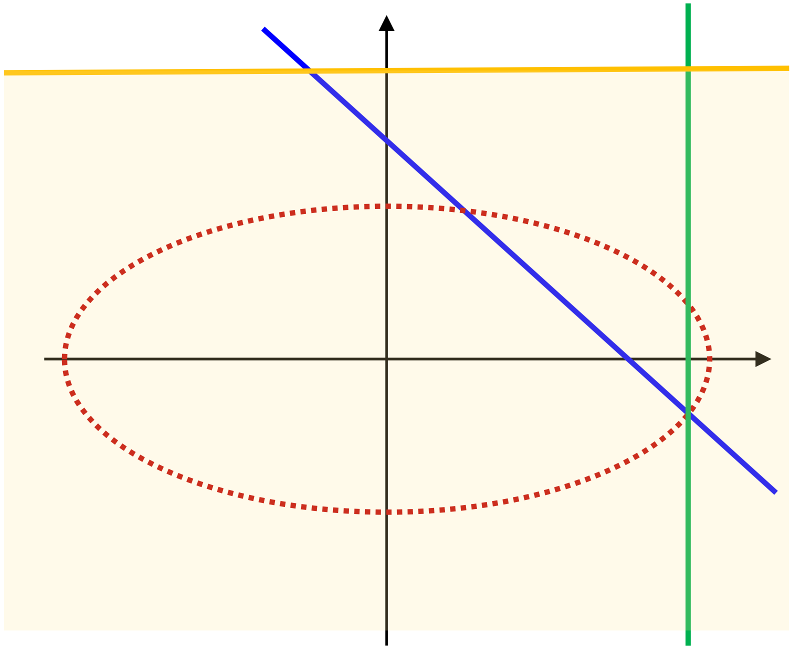
- Solve each subproblem and check the feasible solution.

C1: $\boldsymbol{\mu_1 = 0, \mu_2 = 0}$	**C2:** $\boldsymbol{\mu_1 = 0, \mu_2 \neq 0}$	**C3:** $\boldsymbol{\mu_1 \neq 0, \mu_2 = 0}$	**C4:** $\boldsymbol{\mu_1 \neq 0, \mu_2 \neq 0}$
$2x + \lambda = 0$	$2x + \lambda = 0$	$2x + \lambda - \mu_1 = 0$	$2x + \lambda - \mu_1 = 0$
$2y + \lambda = 0$	$2y + \lambda + \mu_2 = 0$	$2y + \lambda = 0$	$2y + \lambda + \mu_2 = 0$
$x + y - 1 = 0$	$x + y - 1 = 0$	$x + y - 1 = 0$	$x + y - 1 = 0$
$-x + 2 \leq 0$	$-x + 2 \leq 0$	$-x + 2 = 0$	$-x + 2 = 0$
$y - 2 \leq 0$	$y - 2 = 0$	$y - 2 \leq 0$	$y - 2 = 0$

C1	C2	C3	C4
$x = 0.5$	$x = -1$	$x = 2$	No solution
$y = 0.5$	$y = 2$	$y = -1$	
$\lambda = -1$	$\lambda = 2$	$\lambda = 2$	
	$\mu_2 = -6$	$\mu_1 = 6$	
Infeasible!	**Infeasible!**		

Example 2

- **When $\mu_1 \neq 0, \mu_2 = 0$, it satisfies the condition.**
  - The solution is $x = 2$ and $y = -1$.



- **Note: For m inequality conditions, we need to solve 2^m equations in the worst case.**

Dual Form to Solve the Problem

- **If the primal problem is convex and exists at least one strictly feasible solution, two optimization problems are equivalent.**
  - The dual form is used for solving the objective function of SVM.

(Primal)

$$\min_{\mathbf{w},b} \max_{\boldsymbol{\mu}} L(\mathbf{w}, b, \boldsymbol{\mu}) \ \text{subject to}\ \mu_i \geq 0 \text{ for } i = 1, \cdots, k$$

Swap min and max

(Dual)

$$\max_{\boldsymbol{\mu}} \min_{\mathbf{w},b} L(\mathbf{w}, b, \boldsymbol{\mu}) \ \text{subject to}\ \mu_i \geq 0 \text{ for } i = 1, \cdots, k$$

First, compute the derivative of $\mathbf{w}$ and b, and represent $L(\mathbf{w}, b, \boldsymbol{\mu})$ as the function of $\boldsymbol{\mu}$.

Q&A