最適化問題

最適化問題 : Given functions f, g_i, $i = 1, \ldots, k$, and h_i, $i = 1, \ldots, m$, defined on a domain $\Omega \subseteq R^n$, optimization problem is formalized as follows:

$$\begin{array}{llll} \text{minimize} & f(\boldsymbol{w}), & \boldsymbol{w} \in \Omega & \textit{objective function} \\ \text{subject to} & g_i(\boldsymbol{w}) \leq 0, & i = 1, \ldots, k & \textit{inequality constraints} \\ & h_i(\boldsymbol{w}) = 0, & i = 1, \ldots, m & \textit{equality constraints} \end{array}$$

凸性 : A real-valued function f is called convex if, $\forall \boldsymbol{w}, \boldsymbol{u} \in R^n$, and for any $\theta \in (0, 1)$,

$$f(\theta \boldsymbol{w} + (1 - \theta)\boldsymbol{u}) \leq \theta f(\boldsymbol{w}) + (1 - \theta) f(\boldsymbol{u})$$

フェルマの定理 : A necessary condition for $\boldsymbol{w}^*$ to be a minimum of a function $f(\boldsymbol{w})$ is

$$\frac{\partial f(\boldsymbol{w}^*)}{\partial \boldsymbol{w}} = \mathbf{0}.$$

This condition, together with convexity of f, is also sufficient.

Lagrange 法

Lagrange 法 : Given an optimization problem with objective function $f(\boldsymbol{w})$, and **equality constraints** $h_i(\boldsymbol{w}) = 0$, $i = 1, ..., m$, we define the ***Lagrangian function*** as

$$L(\boldsymbol{w}, \boldsymbol{\beta}) = f(\boldsymbol{w}) + \sum_{i=1}^{m} \beta_i h_i(\boldsymbol{w})$$

where the coefficients β_i are called the ***Lagrange multipliers***.

Lagrange の定理 : A necessary condition for a point $\boldsymbol{w}^*$ to be a minimum of $f(\boldsymbol{w})$ subject to $h_i(\boldsymbol{w}) = 0$, $i = 1, \ldots, m$, is

$$\frac{\partial L(\boldsymbol{w}^*, \boldsymbol{\beta}^*)}{\partial \boldsymbol{w}} = \mathbf{0}$$

$$\frac{\partial L(\boldsymbol{w}^*, \boldsymbol{\beta}^*)}{\partial \boldsymbol{\beta}} = \mathbf{0}$$

for some values $\boldsymbol{\beta}^*$. The above conditions are also sufficient provided that $L(\boldsymbol{w}, \boldsymbol{\beta}^*)$ is a convex function of $\boldsymbol{w}$.

Karush-Kuhn-Tucker 法

Generalized Lagrangian: Given an optimization problem with domain $\Omega \subseteq R^n$,

$$\begin{array}{llll} \textbf{minimize} & f(\boldsymbol{w}), & \boldsymbol{w} \in \Omega & \textit{objective function} \\ \textbf{subject to} & g_i(\boldsymbol{w}) \leq 0, & i = 1, \ldots, k & \textit{inequality constraints} \\ & h_i(\boldsymbol{w}) = 0, & i = 1, \ldots, m & \textit{equality constraints} \end{array}$$

we define the *generalized Lagrangian function* as

$$\begin{aligned} L(\boldsymbol{w}, \boldsymbol{\alpha}, \boldsymbol{\beta}) &= f(\boldsymbol{w}) + \sum_{i=1}^{k} \alpha_i g_i(\boldsymbol{w}) + \sum_{i=1}^{m} \beta_i h_i(\boldsymbol{w}) \\ &= f(\boldsymbol{w}) + \boldsymbol{\alpha}^T \boldsymbol{g}(\boldsymbol{w}) + \boldsymbol{\beta}^T \boldsymbol{h}(\boldsymbol{w}) \end{aligned}$$

Karush-Kuhn-Tucker の定理: Sufficient conditions for a point $\boldsymbol{w}^*$ to be an optimum are the existence of $\boldsymbol{\alpha}^*, \boldsymbol{\beta}^*$ such that

$$\begin{aligned} \frac{\partial L(\boldsymbol{w}^*, \boldsymbol{\alpha}^*, \boldsymbol{\beta}^*)}{\partial \boldsymbol{w}} &= \mathbf{0}, \\ \frac{\partial L(\boldsymbol{w}^*, \boldsymbol{\alpha}^*, \boldsymbol{\beta}^*)}{\partial \boldsymbol{\beta}} &= \mathbf{0}, \\ \alpha_i^* g_i(\boldsymbol{w}^*) &= 0, \quad i = 1, \ldots, k, \\ g_i(\boldsymbol{w}^*) &\leq 0, \quad i = 1, \ldots, k, \\ \alpha_i^* &\geq 0, \quad i = 1, \ldots, k. \end{aligned}$$

Karush-Kuhn-Tucker 法 (例題)

例題 1

$$\begin{array}{ll}\textbf{maximize} & x + y \\ \textbf{subject to} & x^2 + y^2 \le 1, \quad x \ge 0, \quad y \ge 0\end{array}$$

例題 2

$$\begin{array}{ll}\textbf{maximize} & (x-1)^2 + (y-1)^2 \\ \textbf{subject to} & x + 2y \le 1, \quad x \ge 0, \quad y \ge 0\end{array}$$

Karush-Kuhn-Tucker 法 (例題1)

例題1

$$\begin{array}{ll} \textbf{maximize} & x+y \\ \textbf{subject to} & x^2+y^2 \le 1, \quad x \ge 0, \quad y \ge 0 \end{array}$$

標準形に変換

$$\begin{aligned} f(\boldsymbol{x}) &= -x-y && \text{(最小化問題にするため)} \\ g_1(\boldsymbol{x}) &= x^2+y^2-1 && (\le 0 \text{ の不等式制約 } 1) \\ g_2(\boldsymbol{x}) &= -x && (\le 0 \text{ の不等式制約 } 2) \\ g_3(\boldsymbol{x}) &= -y && (\le 0 \text{ の不等式制約 } 3) \\ L(\boldsymbol{w}, \boldsymbol{\alpha}, \boldsymbol{\beta}) &= -x-y+\alpha_1(x^2+y^2-1)-\alpha_2 x-\alpha_3 y \end{aligned}$$

とおく．また $D = \frac{\partial}{\partial \boldsymbol{x}}$ **と書くことにすると，**

$$\begin{aligned} Df(\boldsymbol{x}) &= (-1,-1) \\ Dg_1(\boldsymbol{x}) &= (2x,2y) \\ Dg_2(\boldsymbol{x}) &= (-1,0) \\ Dg_3(\boldsymbol{x}) &= (0,-1) \end{aligned}$$

を得る．

Karush-Kuhn-Tucker 法 (例題１のつづき)

したがって最適解は

$$\mathbf{0} = \frac{\partial L(\boldsymbol{w}^*, \boldsymbol{\alpha}^*, \boldsymbol{\beta}^*)}{\partial \boldsymbol{w}} = \begin{bmatrix} -1 \\ -1 \end{bmatrix} + \alpha_1 \begin{bmatrix} 2x \\ 2y \end{bmatrix} + \alpha_2 \begin{bmatrix} -1 \\ 0 \end{bmatrix} + \alpha_3 \begin{bmatrix} 0 \\ -1 \end{bmatrix} \quad (1)$$

$$0 = \alpha_1(x^2 + y^2 - 1) \quad (2)$$

$$0 = \alpha_2(-x) \quad (3)$$

$$0 = \alpha_3(-y) \quad (4)$$

$$\alpha_i \geq 0 \quad (5)$$

をみたす．

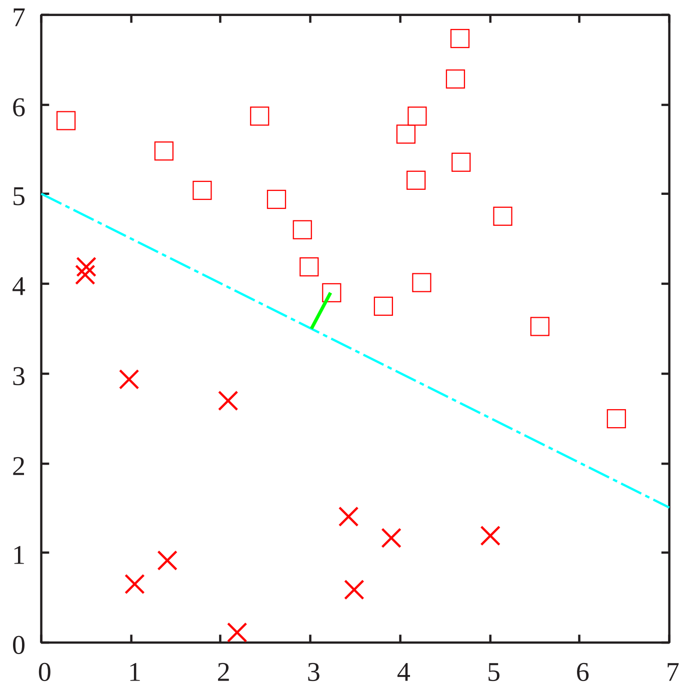
1. $x > 0, y > 0$ のとき： (3)，(4) より $\alpha_2 = \alpha_3 = 0$．(1) に代入し $2\alpha_1 x = 2\alpha_1 y = 1$．つまり $x = y = 1/2\alpha_1$．これを (2) に代入し，$2\alpha_1^2 = 1$．$\alpha_1 > 0$ を考慮すると $\alpha_1 = 1/\sqrt{2}, x = 1/\sqrt{2}, y = 1/\sqrt{2}$ を得る．
2. $x = 0$ のとき： 第１式より $\alpha_2 = -1$ となり不適．
3. $y = 0$ のとき： 第１式より $\alpha_3 = -1$ となり不適．

したがって，求める解は $(x, y) = (1/\sqrt{2}, 1/\sqrt{2})$．

例題２の解答は省略．求める解は $(x, y) = (0, 0)$．

関数的マージン

- **Functional margin**: $\gamma = y(\langle \boldsymbol{w}, \boldsymbol{x} \rangle + b)$



幾何学的マージン

- **Scale $\boldsymbol{w}$ and b so that:** $y_i(\langle \boldsymbol{w}, \boldsymbol{x}_i \rangle + b) \geq 1 \quad \forall i$
- **Support vectors:** $\boldsymbol{x}^+, \boldsymbol{x}^-$

$$\langle \boldsymbol{w}, \boldsymbol{x}^+ \rangle + b = 1, \quad \langle \boldsymbol{w}, \boldsymbol{x}^- \rangle + b = -1$$

- **Geometric margin:** d

$$\begin{aligned} d &= \frac{1}{2}\left(\langle \frac{\boldsymbol{w}}{\|\boldsymbol{w}\|}, \boldsymbol{x}^+ \rangle\right) - \left(\langle \frac{\boldsymbol{w}}{\|\boldsymbol{w}\|}, \boldsymbol{x}^- \rangle\right) \\ &= \frac{1}{2\|\boldsymbol{w}\|}\left(\langle \boldsymbol{w}, \boldsymbol{x}^+ \rangle\right) - \left(\langle \boldsymbol{w}, \boldsymbol{x}^- \rangle\right) = \frac{1}{\|\boldsymbol{w}\|} \end{aligned}$$

最大マージン識別器 (Primal form)

命題: Given a linearly separable training sample

$$S = ((\boldsymbol{x}_1, y_1), \ldots, (\boldsymbol{x}_\ell, y_\ell))$$

the hyperplane $(\boldsymbol{w}, b)$ that solves the optimization problem

$$\begin{array}{ll} \text{minimize}_{\boldsymbol{w},\, b} & \langle \boldsymbol{w}, \boldsymbol{w} \rangle \\ \text{subject to} & y_i(\langle \boldsymbol{w}, \boldsymbol{x}_i \rangle + b) \geq 1, \quad \text{for} \quad i = 1, \ldots, \ell \end{array}$$

realizes the maximal margin hyperplane with geometric margin $\gamma = 1/\|\boldsymbol{w}\|$.

- Lagrangian

$$L(\boldsymbol{w}, b, \alpha) = \frac{1}{2}\langle \boldsymbol{w}, \boldsymbol{w} \rangle - \sum_{i=1}^{\ell} \alpha_i [y_i(\langle \boldsymbol{w}, \boldsymbol{x}_i \rangle + b) - 1]$$

where $\alpha_i \geq 0$ are Lagrange multipliers.

最大マージン識別器 (Primal form to Dual form)

- **Imposing stationarity condition, we have**

$$
\begin{aligned}
\frac{\partial L(\boldsymbol{w}, b, \alpha)}{\partial \boldsymbol{w}} &= \boldsymbol{w} - \sum_{i=1}^{\ell} y_i \alpha_i \boldsymbol{x}_i = \mathbf{0}, \\
\frac{\partial L(\boldsymbol{w}, b, \alpha)}{\partial b} &= \sum_{i=1}^{\ell} y_i \alpha_i = 0.
\end{aligned}
$$

- **Lagrangian**

$$
\begin{aligned}
L(\boldsymbol{w}, b, \alpha) &= \frac{1}{2}\langle \boldsymbol{w}, \boldsymbol{w} \rangle - \sum_{i=1}^{\ell} \alpha_i [y_i(\langle \boldsymbol{w}, \boldsymbol{x}_i \rangle + b) - 1] \\
&= \frac{1}{2}\langle \boldsymbol{w}, \sum_{i=1}^{\ell} y_i \alpha_i \boldsymbol{x}_i \rangle + \sum_{i=1}^{\ell} \alpha_i - \sum_{i=1}^{\ell} \alpha_i y_i (\langle \boldsymbol{w}, \boldsymbol{x}_i \rangle) - b \sum_{i=1}^{\ell} \alpha_i y_i \\
&= \sum_{i=1}^{\ell} \alpha_i - \frac{1}{2}\langle \boldsymbol{w}, \sum_{i=1}^{\ell} y_i \alpha_i \boldsymbol{x}_i \rangle \\
&= \sum_{j=1}^{\ell} \alpha_i - \frac{1}{2} \sum_{i=1}^{\ell} y_i \alpha_i \sum_{j=1}^{\ell} y_j \alpha_j \langle \boldsymbol{x}_i, \boldsymbol{x}_j \rangle
\end{aligned}
$$

- **Note that the Lagrangian does not depend on b nor $\boldsymbol{w}$.**
- **From duality of KKT Lemma, the <u>minimization</u> of the Lagrangian with $\boldsymbol{w}$ should be considered as the <u>maximization</u> of the Lagrangian with $\boldsymbol{\alpha}$**

最大マージン識別器 (Dual form)

命題: Given a linearly separable training sample

$$S = ((\boldsymbol{x}_1, y_1), \ldots, (\boldsymbol{x}_\ell, y_\ell))$$

and suppose the parameters $\boldsymbol{\alpha}^*$ solve the following quadratic optimization problem:

$$\begin{aligned} \textbf{maximize} \quad & W(\boldsymbol{\alpha}) = \sum_{j=1}^{\ell} \alpha_i - \frac{1}{2} \sum_{i,j=1}^{\ell} y_i y_j \alpha_i \alpha_j \langle \boldsymbol{x}_i, \boldsymbol{x}_j \rangle \\ \textbf{subject to} \quad & \sum_{i=1}^{\ell} y_i \alpha_i = 0, \\ & \alpha_i \geq 0 \quad \textbf{for} \quad i = 1, \ldots, \ell \end{aligned}$$

Then the weight vector $\boldsymbol{w} = \sum_{i=1}^{\ell} y_i \alpha_i^* \boldsymbol{x}_i$ realizes the maximal margin hyperplane with geometric margin $\gamma = 1/\|\boldsymbol{w}^*\|$.

サポートベクタの性質

1. **Support Vectors:**

$$\boldsymbol{x}^- = \arg\max_{y_i=-1}(\langle \boldsymbol{w}^*, \boldsymbol{x}_i \rangle), \qquad \boldsymbol{x}^+ = \arg\max_{y_i=1}(\langle \boldsymbol{w}^*, \boldsymbol{x}_i \rangle),$$

$$\langle \boldsymbol{w}, \boldsymbol{x}^+ \rangle + b^* = 1, \qquad \langle \boldsymbol{w}, \boldsymbol{x}^- \rangle + b^* = -1$$

2. **The value of b**

$$b^* = -\frac{1}{2}\left(\max_{y_i=-1}(\langle \boldsymbol{w}^*, \boldsymbol{x}_i \rangle) + \min_{y_i=1}(\langle \boldsymbol{w}^*, \boldsymbol{x}_i \rangle)\right)$$

3. **Karush-Kuhn-Tucker conditions state that the optimal solutions $\boldsymbol{\alpha}^*, (\boldsymbol{w}^*, b^*)$ must satisfy**

$$\alpha_i^*[y_i(\langle \boldsymbol{w}^*, \boldsymbol{x}_i \rangle + b^*) - 1] = 0.$$

Only for support vectors (inputs $\boldsymbol{x}_i$ for which the functional margin is one $(y_i(\langle \boldsymbol{w}^*, \boldsymbol{x}_i \rangle + b^*) = 1)$), the corresponding α_i^* are non-zero. All the other parameters α_i^* are zero.

4. **The optimal hyperplane can be expressed in terms of support vectors**

$$f(\boldsymbol{x}, \boldsymbol{\alpha}^*, b^*) = \sum_{i=1}^{\ell} y_i \alpha_i \langle \boldsymbol{x}_i, \boldsymbol{x} \rangle + b^* = \sum_{i \in \mathbf{SV}} y_i \alpha_i \langle \boldsymbol{x}_i, \boldsymbol{x} \rangle + b^*$$

Points that are not support vectors have no influence.

サポートベクタの性質

5. **Another important consequence of the Karush-Kuhn-Tucker complementarity condition is that for $j \in$ sv,**

$$y_j f(\boldsymbol{x}_j, \boldsymbol{\alpha}^*, b^*) = y_j \left(\sum_{i \in \mathbf{SV}} y_i \alpha_i \langle \boldsymbol{x}_i, \boldsymbol{x}_j \rangle + b^* \right) = 1,$$

and therefore

$$\begin{aligned}
\langle \boldsymbol{w}^*, \boldsymbol{w}^* \rangle &= \sum_{i,j=1}^{\ell} y_i y_j \alpha_i^* \alpha_j^* \langle \boldsymbol{x}_i, \boldsymbol{x}_j \rangle \\
&= \sum_{j \in \mathbf{SV}} \alpha_j^* y_j \sum_{i \in \mathbf{SV}} y_i \alpha_i^* \langle \boldsymbol{x}_i, \boldsymbol{x}_j \rangle \\
&= \sum_{j \in \mathbf{SV}} \alpha_j^* (1 - y_j b^*) \\
&= \sum_{i \in \mathbf{SV}} \alpha_i^*
\end{aligned}$$

i.e.,

$$\gamma = 1/\|\boldsymbol{w}^*\| = \left(\sum_{i \in \mathbf{SV}} \alpha_i^* \right)^{1/2}$$