

2025 年度 黒柳研究室
SVM 資料

古郷泰至

2025 年 5 月 26 日

1 初めに

SVM とはサポートベクターマシン (Support Vector Machine) の略であり, 分類問題や回帰問題に用いられる教師あり機械学習のアルゴリズムである. SVM は最適化するパラメータが少なく高次元のデータや非線形なデータに対しても高い分類性能を発揮するが, 計算コストが高いためデータが大規模になると計算量が大きくなる [1]. また, SVM は主に 2 クラス分類で用いられるが多クラス分類にも拡張可能である. 本資料では, 初めに SVM の理解に必要な知識であるラグランジュの未定乗数法, KKT 条件, カーネル法について説明しその後線形 SVM, 非線形 SVM について説明する.

2 前提知識

この章では SVM を理解する上で必要となる知識について説明する.

2.1 ラグランジュの未定乗数法

ラグランジュの未定乗数法 (Lagrange multiplier) とは等式制約のもとで多変数関数の最適化問題を解く手法である.

本資料では簡略化のため最適化する関数を 2 変数関数 $f(x, y)$ とし, $\mathbf{x} = [x, y]$ とする. 式 (1) の等式制約のもとで関数 $f(\mathbf{x})$ が点 (a, b) に極値を持つとする.

$$g(x, y) = 0 \quad (1)$$

このとき, \mathbf{x} とラグランジュ乗数 λ を変数に持つラグランジュ関数 L を式 (2) のように定義する.

$$L(\mathbf{x}, \lambda) = f(\mathbf{x}) - \lambda g(\mathbf{x}) \quad (2)$$

このとき式 (3) を満たす点では, 式 (4) を満たす λ' が存在する [2].

$$\frac{\partial g}{\partial x}(a, b) \neq 0$$

または

$$\frac{\partial g}{\partial y}(a, b) \neq 0 \quad (3)$$

$$\begin{aligned} \frac{\partial L}{\partial x}(a, b, \lambda') &= 0 \\ \frac{\partial L}{\partial y}(a, b, \lambda') &= 0 \\ \frac{\partial L}{\partial \lambda}(a, b, \lambda') &= 0 \end{aligned} \quad (4)$$

式 (4) の連立方程式を解くことで極値をとる点 (a, b) を求められる. この (a, b) を求める手法をラグランジュの未定乗数法という [2].

2.2 KKT 条件

KKT 条件 (Karush-Kuhn-Tucker conditions) とは、一般的な制約付き最適化問題の解が満たす条件である。この制約は等式制約のみならずラグランジュの未定乗数法では扱えなかった不等式制約も含めた制約である。KKT 条件は目的としている最適化問題の最適性の必要十分条件であるため、この条件を満たす解は最適解であることが保証される。式 (5) のような d 次元のベクトル \mathbf{x} を変数とする複数の等式・不等式制約付き最適化問題を考える。

$$\begin{aligned} \min_{\mathbf{x}} f(\mathbf{x}) \\ \text{s.t. } g_i(\mathbf{x}) \leq 0 \quad (i = 1, 2, \dots, G) \\ h_j(\mathbf{x}) = 0 \quad (j = 1, 2, \dots, H) \end{aligned} \quad (5)$$

このとき、ラグランジュ関数は以下の式 (6) のように定義される。

$$\begin{aligned} L(\mathbf{x}, \boldsymbol{\lambda}, \boldsymbol{\mu}) &= f(\mathbf{x}) + \sum_{i=1}^G \lambda_i g_i(\mathbf{x}) + \sum_{j=1}^H \mu_j h_j(\mathbf{x}) \\ \text{ただし} \\ \boldsymbol{\lambda} &= [\lambda_1, \lambda_2, \dots, \lambda_G] \\ \boldsymbol{\mu} &= [\mu_1, \mu_2, \dots, \mu_H] \end{aligned} \quad (6)$$

ただし、不等式制約のラグランジュ未定乗数は $\lambda_i \geq 0$ である。

この問題についての式 (8)～(12) が KKT 条件である。データ \mathbf{x} に対して式 (7) を満たす点が少なくとも一つ存在するというスレーター条件を仮定した時に KKT 条件を満たす λ_i, μ_i が存在することが保証される。このとき、KKT 条件を満たすことが問題の最適解であるための必要十分条件となる。式 (8) はラグランジュ関数の微分条件、式 (10)～(12) は等式・不等式制約とラグランジュ未定乗数の非負条件である。また、式 (9) は相補性条件と呼ばれ [3]、この条件により最適解に関係のない不等式制約に対してラグランジュ未定乗数を 0 としている。

$$g_i(\mathbf{x}) < 0, h_j(\mathbf{x}) = 0 \quad (7)$$

$$\frac{\partial}{\partial \mathbf{x}} L(\mathbf{x}, \boldsymbol{\lambda}, \boldsymbol{\mu}) = 0 \quad (8)$$

$$\lambda_i g_i(\mathbf{x}) = 0 (i = 1, 2, \dots, G) \quad (9)$$

$$g_i(\mathbf{x}) \leq 0 (i = 1, 2, \dots, G) \quad (10)$$

$$h_i(\mathbf{x}) = 0 (i = 1, 2, \dots, H) \quad (11)$$

$$\lambda_i \geq 0 (i = 1, 2, \dots, G) \quad (12)$$

2.3 カーネル法

カーネル法とはもとの空間では線形分離不可能なデータ群をより高い次元数の空間 (特徴空間) に写像することで線形分離可能にする手法のことである [4]. このとき, 写像する際の関数を $\phi(\mathbf{x})$ とする. SVM において最適化問題を解く上では $\phi(\mathbf{x})^\top \phi(\mathbf{x}')$ のような内積の形でのみ現れるため, 内積の結果のみを返す新たな関数を考える. この関数をカーネル関数と呼び一般に式 (13) のように表される. カーネル関数を用いることによって $\phi(\mathbf{x})$ を明示的に計算せずに内積を求めることができる. これをカーネルトリックという [4].

$$K(\mathbf{x}, \mathbf{x}') = \phi(\mathbf{x})^\top \phi(\mathbf{x}') \quad (13)$$

カーネル関数は式 (14) のような対称性と式 (15) のような半正定値性を満たす必要がある [5].

$$K(\mathbf{x}, \mathbf{x}') = K(\mathbf{x}', \mathbf{x}) \quad (14)$$

$$\sum_{i=1}^n \sum_{j=1}^n c_i c_j K(\mathbf{x}_i, \mathbf{x}_j) \geq 0 \quad (15)$$

カーネル関数の具体例を表 1 に示す [6]. 利用するカーネル関数やそのハイパーパラメータによって得られる識別境界の様子が変化するため, データの特性に合わせてカーネル関数を選択する必要がある. RBF カーネルの例では γ が大きいほど近い範囲のデータを特徴空間上で近い位置に写像するようになる.

表 1 カーネル関数の例

カーネル関数名	数式	ハイパーパラメータ
線形カーネル	$K(\mathbf{x}, \mathbf{x}') = \mathbf{x}^\top \mathbf{x}'$	無し
多項式カーネル	$K(\mathbf{x}, \mathbf{x}') = (\mathbf{x}^\top \mathbf{x}' + c)^q$	c, q
シグモイドカーネル	$K(\mathbf{x}, \mathbf{x}') = \tanh(a \mathbf{x}^\top \mathbf{x}' + b)$	a, b
RBF カーネル	$K(\mathbf{x}, \mathbf{x}') = \exp(-\gamma \ \mathbf{x} - \mathbf{x}'\ ^2)$	γ

カーネル関数のハイパーパラメータは事前に設定するため、SVM とは別に最適化する必要がある。表 2 にいくつかのハイパーパラメータの最適化手法を示す。通常、パラメータは一度の操作では最適化されているか分からないため、選択されたパラメータを評価する手法が必要である。評価にはしばしばデータをモデルの学習用のトレーニングデータとモデルの評価に使うテストデータに分けるクロスバリデーション (交差検証) が用いられる [7]。ベイズ最適化はアルゴリズム内で獲得関数と呼ばれる関数を用いて次のパラメータを選択するため評価関数を用いない [8]。

表 2 ハイパーパラメータ最適化の手法

手法	説明
グリッドサーチ	事前に候補をそれぞれ複数用意し、それらの組み合わせをすべて試す手法
ランダムサーチ	候補をランダムに選択する操作を繰り返し行う手法
ベイズ最適化	パラメータを確率的に探索する手法

3 線形 SVM

線形 SVM は d 次元データの 2 クラス分類問題において式 (16) のような線形分類境界を導出するアルゴリズムである。

$$f(\mathbf{x}) = \boldsymbol{\omega}^T \mathbf{x} + \beta = 0 \quad (16)$$

図 1,2 のような d 次元のデータ群を考えたときにデータ群を分類する $d-1$ 次元の超平面を分離超平面と呼び、分離超平面とその超平面に最も近いデータとの距離をマージンと呼ぶ。また、分離超平面に最も近いデータのことをサポートベクトルと呼ぶ。マージンには 2 つの種類があり、図 1 のように線形分離可能 (1 つの直線で分けられる) なデータ群を前提としたマージンをハードマージン、図 2 のように線形分離不可能 (1 つの直線で分けられない) なデータ群を前提としたマージンをソフトマージンと呼ぶ。SVM はこのマージンを最大化するという考え方に基づいて分類境界を導出する。

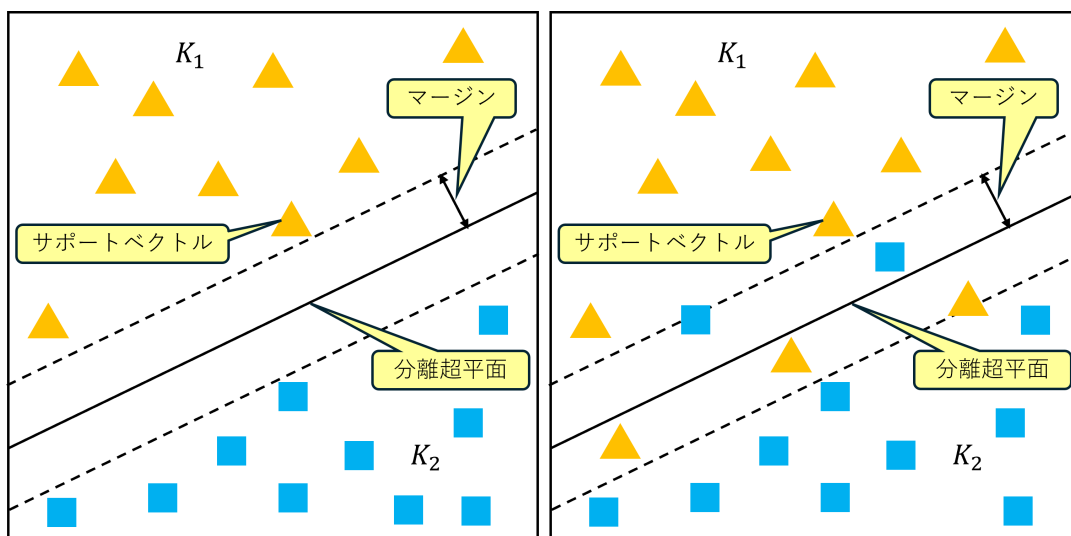


図 1 線形分離可能な問題の例

図 2 線形分離不可能な問題の例

図 1 では三角のデータをクラス K_1 、四角のデータをクラス K_2 に属すると考えると、式 (17) を満たす。

$$\begin{aligned} \boldsymbol{\omega}^T \mathbf{x}_i + \beta &> 0 \quad (\mathbf{x}_i \in K_1) \\ \boldsymbol{\omega}^T \mathbf{x}_i + \beta &< 0 \quad (\mathbf{x}_i \in K_2) \end{aligned} \quad (17)$$

式 (17) をまとめるため、また SVM を設計するためにラベル変数 y_i を式 (18) で定義する。このとき、すべてのデータ \mathbf{x}_i において式 (18) を満たす超平面 $f(\mathbf{x}) = 0$ が存在するとき、線形分離可能という。

$$y_i = \begin{cases} +1 & (f(\mathbf{x}_i) > 0) \\ -1 & (f(\mathbf{x}_i) < 0) \end{cases} \quad (18)$$

式 (18) を用いて式 (17) をまとめると式 (19) のように表すことができる。

$$y_i(\boldsymbol{\omega}^\top \mathbf{x}_i + \beta) > 0 \quad (i = 1, 2 \dots N) \quad (19)$$

3.1 ハードマージン SVM

ハードマージン SVM は線形分離可能なデータを対象とした線形 SVM である。ある \mathbf{x}_i から識別境界までの距離 d_i は式 (20) となる。ただし y_i は式 (18) で定義したラベル関数である。

$$d_i = \frac{y_i(\boldsymbol{\omega}^\top \mathbf{x}_i + \beta)}{\|\boldsymbol{\omega}\|} \quad (i = 1, 2 \dots n) \quad (20)$$

今、線形分離可能なデータを仮定しているため、式 (21) をすべての i で満たす $\boldsymbol{\omega}$, β 及び M が存在するはずである。

$$y_i(\boldsymbol{\omega}^\top \mathbf{x}_i + \beta) \geq M > 0 \quad (21)$$

このとき、マージンは式 (20),(21) とマージン最大化より式 (22) で求めることができる。

$$\begin{aligned} & \max_{\boldsymbol{\omega}, \beta, M} \frac{M}{\|\boldsymbol{\omega}\|} \\ \text{s.t. } & y_i(\boldsymbol{\omega}^\top \mathbf{x}_i + \beta) \geq M \quad (i = 1, 2 \dots n) \end{aligned} \quad (22)$$

ここで式 (23) のように変数を定義しなおすと、式 (22) は式 (24) となる。

$$\tilde{\boldsymbol{\omega}} = \frac{\boldsymbol{\omega}}{M}, \tilde{\beta} = \frac{\beta}{M} \quad (23)$$

$$\begin{aligned} & \max_{\tilde{\boldsymbol{\omega}}, \tilde{\beta}} \frac{1}{\|\tilde{\boldsymbol{\omega}}\|} \\ \text{s.t. } & y_i(\tilde{\boldsymbol{\omega}}^\top \mathbf{x}_i + \tilde{\beta}) \geq 1 \quad (i = 1, 2 \dots n) \end{aligned} \quad (24)$$

また、 $\frac{1}{\|\tilde{\boldsymbol{\omega}}\|}$ の最大化は $\|\tilde{\boldsymbol{\omega}}\|$ の最小化と等価でありかつ、 $\|\tilde{\boldsymbol{\omega}}\|$ の最小化は $\|\tilde{\boldsymbol{\omega}}\|^2$ の最小化と等価なため、式 (24) の変数を $\boldsymbol{\omega}, \beta$ に戻し変形することで、式 (25) が得られる。ただし、後の計算のために $\frac{1}{2}$ を加えている。式 (25) が線形分離可能な場合における最適化問題である。このような二次関数を一次式の制約のもと最適化する問題は二次計画問題と呼ばれる [9]。

$$\begin{aligned} & \min_{\boldsymbol{\omega}, \beta} \frac{1}{2} \|\boldsymbol{\omega}\|^2 \\ \text{s.t. } & y_i(\boldsymbol{\omega}^\top \mathbf{x}_i + \beta) \geq 1 \quad (i = 1, 2 \dots n) \end{aligned} \quad (25)$$

3.2 ソフトマージン SVM

ソフトマージン SVM とは，線形分離不可能なデータを前提とし，誤判別を許容する線形 SVM である．ハードマージン SVM では線形分離不可能な場合に式 (25) の制約条件を満たす ω, β が存在しない．そこで，それぞれのデータについて $\xi_i \geq 0$ だけ制約を違反しても良いこととする．このとき最適化問題は式 (26) となる．ただし $\xi = [\xi_1, \xi_2 \dots \xi_n]$ である． ξ_i はスラック変数と呼ばれ，制約違反の度合を表しているため小さい方が望ましい．スラック変数の性質を図 3，表 3 に示す．また， C は罰則の強さを調整するハイパーパラメータであり大きいほど過学習を起こしやすくなる [10]．また， C が ∞ のときハードマージン SVM に帰着する．

$$\begin{aligned} \min_{\omega, \beta, \xi} \quad & \frac{1}{2} \|\omega\|^2 + C \sum_{i=1}^n \xi_i \\ \text{s.t.} \quad & \xi_i \geq 0, C > 0, y_i(\omega^\top \mathbf{x}_i + \beta) \geq 1 - \xi_i \quad (i = 1, 2 \dots n) \end{aligned} \quad (26)$$

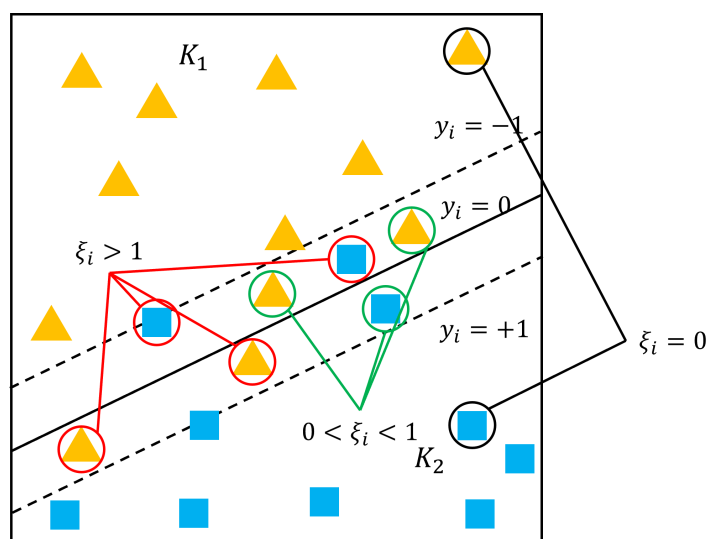


図 3 スラック変数の例

表 3 スラック変数の性質

スラック変数	データ点の位置
$\xi_i = 0$	正しく分類されかつ，マージンの境界線上またはそれより外側に位置する．
$0 < \xi_i < 1$	正しく分類されかつ，マージンの境界の内側で分類境界は越えていない
$\xi_i > 1$	分類境界を越えており誤分類されている．

3.3 最適化問題の主問題と双対問題

式 (25), 式 (26) の最適化問題は SVM における主問題と呼ばれる。これらの問題は単純に傾きを求める勾配法では制約条件を扱えず、扱うことのできる解法は低速であることが多い。そのため、SVM では双対問題と呼ばれる主問題と等価な別の問題を解くことが多い。双対問題とは、主問題に対して補集合となる、裏表の関係にある最適化問題である。主問題と双対問題の間にはどちらか一方が最適解を持つなら、もう一方も最適解を持ち、主問題の最小値と双対問題の最大値が一致するという双対定理 (duality theorem) が成立する。そのため SVM では双対問題を解くことによって最適化問題を簡単に解くことができる [11]。

まず、ハードマージン SVM についての最適化問題である式 (25) についてラグランジュの未定乗数法を用いるとラグランジュ関数 $L(\boldsymbol{\omega}, \beta, \boldsymbol{\lambda})$ は式 (27) となる。ただし $\boldsymbol{\lambda} = [\lambda_1, \lambda_2 \dots \lambda_n]$ である。

$$\begin{aligned} \min_{\boldsymbol{\omega}, \beta} \quad & \frac{1}{2} \|\boldsymbol{\omega}\|^2 \\ \text{s.t.} \quad & y_i(\boldsymbol{\omega}^\top \mathbf{x}_i + \beta) \geq 1 \quad (i = 1, 2 \dots n) \end{aligned} \quad (25)$$

$$L(\boldsymbol{\omega}, \beta, \boldsymbol{\lambda}) = \frac{1}{2} \|\boldsymbol{\omega}\|^2 - \sum_{i=1}^n \lambda_i \{y_i(\boldsymbol{\omega}^\top \mathbf{x}_i + \beta) - 1\} \quad (27)$$

このとき、ラグランジュ関数の微分条件より、式 (28)~(30) の条件を満たす必要がある。

$$\lambda_i \geq 0 \quad (i = 1, 2 \dots n) \quad (28)$$

$$\frac{\partial L}{\partial \boldsymbol{\omega}}(\boldsymbol{\omega}, \beta, \boldsymbol{\lambda}) = \boldsymbol{\omega} - \sum_{i=1}^n \lambda_i y_i \mathbf{x}_i = 0 \quad (29)$$

$$\frac{\partial L}{\partial \beta}(\boldsymbol{\omega}, \beta, \boldsymbol{\lambda}) = - \sum_{i=1}^n \lambda_i y_i = 0 \quad (30)$$

式 (29), (30) はそれぞれ変形すると、式 (31), (32) となる。

$$\boldsymbol{\omega} = \sum_{i=1}^n \lambda_i y_i \mathbf{x}_i \quad (31)$$

$$\sum_{i=1}^n \lambda_i y_i = 0 \quad (32)$$

式 (31) を式 (27) に代入し整理すると、式 (33) となる。これにより、式 (25) を最小化する $\boldsymbol{\omega}, \beta$ を求める主問題は、双対問題は式 (33) を最大化する $\boldsymbol{\lambda}$ を求める問題に帰着することができる。双対問題の制約条件は式 (28), (32) となる。

$$\max_{\boldsymbol{\lambda}} L(\boldsymbol{\omega}, \beta, \boldsymbol{\lambda}) = \sum_{i=1}^n \lambda_i - \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n \lambda_i \lambda_j y_i y_j \mathbf{x}_i^\top \mathbf{x}_j \quad (33)$$

同様にソフトマージン SVM の最適化問題である式 (26) について、ラグランジュ関数 L は式 (34) となる。 ω, β, ξ を主変数、 λ, μ を双対変数といい $\lambda \geq 0, \mu \geq 0$ である。

$$\min_{\omega, \beta, \xi} \frac{1}{2} \|\omega\|^2 + C \sum_{i=1}^n \xi_i \quad (26)$$

$$\text{s.t. } \xi_i \geq 0, C > 0, y_i(\omega^\top \mathbf{x}_i + \beta) \geq 1 - \xi_i \quad (i = 1, 2 \dots n)$$

$$L(\omega, \beta, \xi, \lambda, \mu) = \frac{1}{2} \|\omega\|^2 + C \sum_{i=1}^n \xi_i - \sum_{i=1}^n \lambda_i \{y_i(\omega^\top \mathbf{x}_i + \beta) - 1 + \xi_i\} - \sum_{i=1}^n \mu_i \xi_i \quad (34)$$

このときの KKT 条件は式 (35)～(40) となる。

$$\lambda_i \{1 - \xi_i - y_i(\omega^\top \mathbf{x}_i + \beta)\} = 0 \quad (35)$$

$$\mu_i \xi_i = 0 \quad (36)$$

$$1 - \xi_i - y_i(\omega^\top \mathbf{x}_i + \beta) \leq 0 \quad (37)$$

$$\xi_i \geq 0 \quad (38)$$

$$\lambda_i \geq 0 \quad (39)$$

$$\mu_i \geq 0 \quad (40)$$

式 (26) の双対問題は式 (41) となる。

$$\min_{\omega, \beta, \xi} \frac{1}{2} \|\omega\|^2 + C \sum_{i=1}^n \xi_i \quad (26)$$

$$\max_{\lambda, \mu} \min_{\omega, \beta, \xi} L(\omega, \beta, \xi, \lambda, \mu) \quad (41)$$

このときハードマージン SVM 同様にラグランジュ関数の微分条件より式 (42)～(44) の条件を満たす必要がある。

$$\frac{\partial L}{\partial \omega} = \omega - \sum_{i=1}^n \lambda_i y_i \mathbf{x}_i = 0 \quad (42)$$

$$\frac{\partial L}{\partial \beta} = \sum_{i=1}^n \lambda_i y_i = 0 \quad (43)$$

$$\frac{\partial L}{\partial \xi_i} = C - \lambda_i - \mu_i = 0 \quad (44)$$

式 (34) を整理すると式 (45) となり、さらに式 (42)～(44) を代入すると双対問題は式 (46) となる。

$$L(\omega, \beta, \xi, \lambda, \mu) = \frac{1}{2} \|\omega\|^2 + \sum_{i=1}^n \lambda_i + \sum_{i=1}^n \lambda_i y_i (\omega^\top \mathbf{x}_i + \beta) + \sum_{i=1}^n (C - \lambda_i - \mu_i) \xi_i \quad (45)$$

$$\max_{\lambda, \mu} \min_{\omega, \beta, \xi} -\frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n \lambda_i \lambda_j y_i y_j \mathbf{x}_i^\top \mathbf{x}_j + \sum_{i=1}^n \lambda_i \quad (46)$$

このときの制約条件は双対変数 λ, μ が非負であることと式 (43), (44) より式 (47) となる.

$$\begin{aligned} 0 &\leq \lambda_i \leq C \\ \sum_{i=1}^n \lambda_i y_i &= 0 \end{aligned} \quad (47)$$

3.4 線形分類境界の導出

この章では, 式 (16) で表される線形分類境界を求める手法について説明する.

$$f(\mathbf{x}) = \boldsymbol{\omega}^\top \mathbf{x} + \beta = 0 \quad (16)$$

3.3 節では主問題を双対変数 λ_i のみの関数についての双対問題に帰着させたが, ここで λ_i の性質について表 4 に示す. この性質から, $0 < \lambda_i \leq C$ であるデータのみが境界に関係することが分かる.

表 4 双対変数の性質

データの位置	λ_i	分類境界の決定への影響
マージンの外側 ($y_i(\boldsymbol{\omega}^\top \mathbf{x}_i + \beta) > 1$)	$\lambda_i = 0$	無し
マージン上 ($y_i(\boldsymbol{\omega}^\top \mathbf{x}_i + \beta) = 1$)	$0 < \lambda_i < C$	有り
マージンの内側 ($y_i(\boldsymbol{\omega}^\top \mathbf{x}_i + \beta) < 1$)	$\lambda_i = C$	有り

ここで分類境界の式 (16) の $\boldsymbol{\omega}$ に式 (31) を代入すると式 (48) となる. $\boldsymbol{\omega}$ 及び $f(\mathbf{x})$ は $\lambda_i, y_i, \mathbf{x}_i$ の積の線形結合により表されている. また, 総和を取る項数はデータ数に等しいため, 全データを参照して分類境界を決定していることが分かる.

$$f(\mathbf{x}) = \sum_{i=1}^n \lambda_i y_i \mathbf{x}_i^\top \mathbf{x} + \beta \quad (48)$$

また, β に関して別の表現を考える. KKT 条件のうち相補性条件の式 (36) からラグランジュ関数の微分条件式 (44) により μ_i を消去したものを式 (49) に示す.

$$\mu_i \xi_i = 0 \quad (36)$$

$$\frac{\partial L}{\partial \xi_i} = C - \lambda_i - \mu_i = 0 \quad (44)$$

$$(C - \lambda_i) \xi_i = 0 \quad (49)$$

表 4 より分類境界の決定に関係するのは $0 < \lambda_i \leq C$ のときであり, $0 < \lambda_i < C$ 条件下では $\lambda_i \neq 0, C - \lambda_i \neq 0$ が成立するので, 式 (35),(49) より式 (50),(51) が成り立つ.

$$\lambda_i \{1 - \xi_i - y_i(\omega^\top \mathbf{x}_i + \beta)\} = 0 \quad (35)$$

$$1 - \xi_i - y_i(\omega^\top \mathbf{x}_i + \beta) = 0 \quad (50)$$

$$\xi_i = 0 \quad (51)$$

式 (50),(51) をまとめ, 式 (31) により ω を消去すると式 (52) となる.

$$y_i \left(\sum_{j=1}^n \lambda_j y_j \mathbf{x}_j^\top \mathbf{x}_i + \beta \right) = 1 \quad (52)$$

y_i は 1 または -1 のどちらかなので $y_i^2 = 1$ である. 式 (52) の両辺に y_i をかけて β について整理すると, 式 (53) となる.

$$\beta = y_i - \sum_{j=1}^n \lambda_j y_j \mathbf{x}_j^\top \mathbf{x}_i \quad (53)$$

したがって, あるサポートベクトル \mathbf{x}_i により β を導くことができる. しかしサポートベクトルは複数存在し得るため, 式 (54) のように複数のサポートベクトルについて求めた β の平均を計算することで誤差を小さくすることができる. ただし SV はサポートベクトルのインデックスの集合で $|SV|$ はインデックスの個数である.

$$\beta = \frac{1}{|SV|} \sum_{i \in SV} \left(y_i - \sum_{j=1}^n \lambda_j y_j \mathbf{x}_j^\top \mathbf{x}_i \right) \quad (54)$$

式 (54) により β を求められるが, サポートベクトル以外のデータでは $\lambda_i = 0$ であり, 分類境界の決定に影響しない. このようにデータのうち一部の少ない要素が本質的な情報を持つ性質をスパース性と呼び [12], スパース性のあるデータにおいて $\lambda_i = 0$ となるデータの計算は省略することができる. スパース性を踏まえた上で式 (54) を書き直すと式 (55) となり, 計算量を減らすことができる.

$$\beta = \frac{1}{|SV|} \sum_{i \in SV} \left(y_i - \sum_{j \in SV} \lambda_j y_j \mathbf{x}_j^\top \mathbf{x}_i \right) \quad (55)$$

4 非線形 SVM

線形分離不可能なデータに対してソフトマージン SVM を説明したが, 本章では線形分離不可能なデータに対する非線形 SVM について説明する. 特に今回は線形分離不可能なデータを線形分離可能なデータとなるように元のデータ空間から別のデータ空間に射影する, カーネル SVM について説明する. 図 4 のような元のデータ空間からの射影先の空間を特徴空間という.

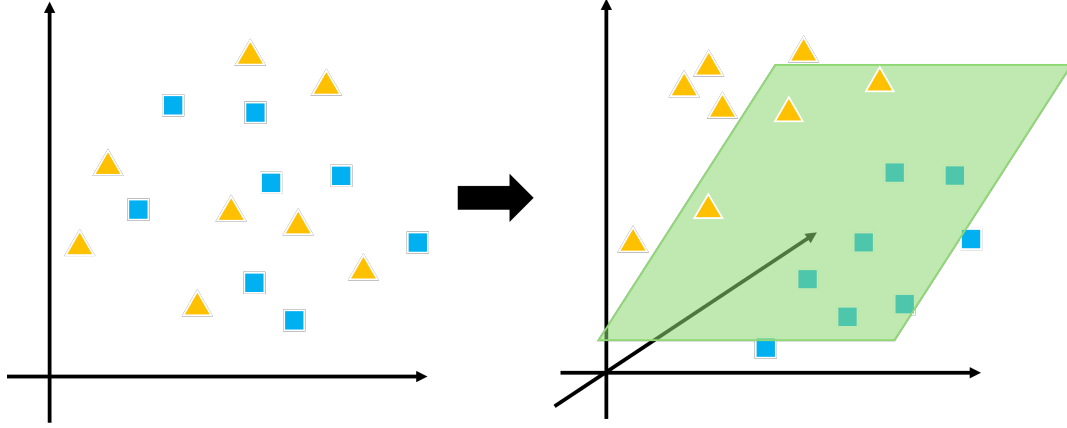


図4 特徴空間への写像

データ \mathbf{x} を射影した $\phi(\mathbf{x})$ の空間で考えた時、 $\phi(\mathbf{x}) = \mathbf{x}$ とすれば線形 SVM に帰着する．したがって、分類境界の式 (16) を書き換えると、式 (56) となる．

$$f(\mathbf{x}) = \boldsymbol{\omega}^\top \phi(\mathbf{x}) + \beta \quad (56)$$

式 (56) に対する最適化問題は 3.3 節で説明した導出方法を用いることができるため、双対問題は式 (57) のように表すことができる．

$$\begin{aligned} \max_{\boldsymbol{\lambda}} \left(\sum_{i=1}^n \lambda_i - \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n \lambda_i \lambda_j y_i y_j \phi(\mathbf{x}_i)^\top \phi(\mathbf{x}_j) \right) \\ \text{s.t. } \sum_{i=1}^n \lambda_i y_i = 0, 0 \leq \lambda_i \leq C \end{aligned} \quad (57)$$

ここで $\phi(\mathbf{x}_i)^\top \phi(\mathbf{x}_j)$ は内積であるため、カーネル関数 K を用いると式 (58) と書き換えられる．

$$\begin{aligned} \max_{\boldsymbol{\lambda}} \left(\sum_{i=1}^n \lambda_i - \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n \lambda_i \lambda_j y_i y_j K(\mathbf{x}_i, \mathbf{x}_j) \right) \\ \text{s.t. } \sum_{i=1}^n \lambda_i y_i = 0, 0 \leq \lambda_i \leq C \end{aligned} \quad (58)$$

また、線形 SVM において分類境界は式 (48) のように書けるため非線形 SVM も同様に式 (59) となり、この式にカーネル関数を用いると式 (60) と書ける．

$$f(\mathbf{x}) = \sum_{i=1}^n \lambda_i y_i \mathbf{x}_i^\top \mathbf{x} + \beta \quad (48)$$

$$f(\mathbf{x}) = \sum_{i=1}^n \lambda_i y_i \phi(\mathbf{x}_i)^\top \phi(\mathbf{x}) + \beta \quad (59)$$

$$f(\mathbf{x}) = \sum_{i=1}^n \lambda_i y_i K(\mathbf{x}, \mathbf{x}_i) + \beta \quad (60)$$

さらに，3.4 節と同様に考えて β は式 (61) で推定される．

$$\beta = \frac{1}{|SV|} \sum_{i \in SV} \left(y_i - \sum_{j \in SV} \lambda_j y_j K(\mathbf{x}_i, \mathbf{x}_j) \right) \quad (61)$$

5 最適化問題を解くアルゴリズム

ここまでの章では求めるべき最適化問題の導出方法を説明したが，本章では求めた最適化問題の解となる λ_i を求めるアルゴリズムについて説明する．

5.1 最急降下法

図 5 のようにある点から最も傾きが急な方向に点を少し動かし，動かした点でまた傾きが急な方向に点を動かすという操作を繰り返すことで局所解を見つける方法を最急降下法という [13]．

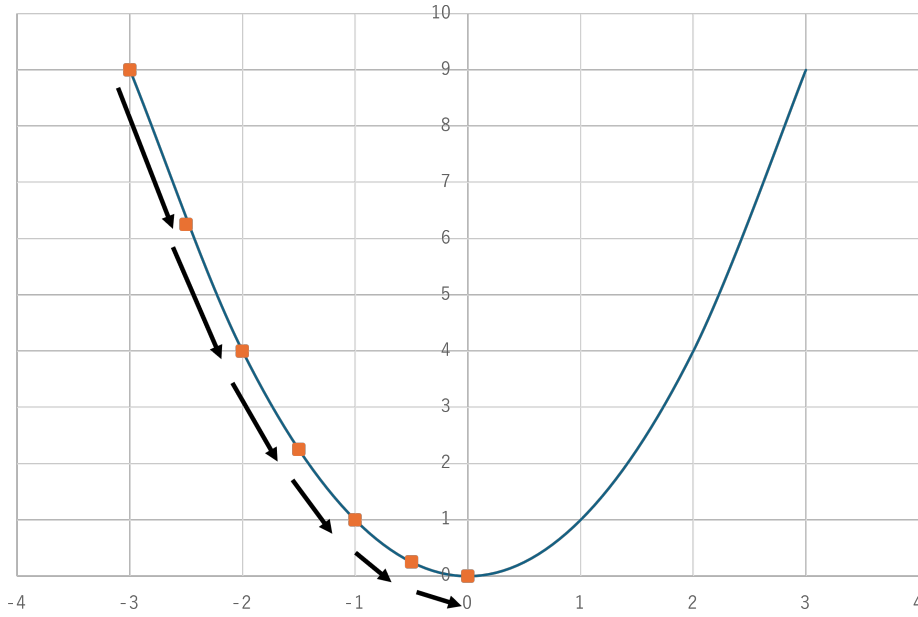


図 5 最急降下法

最急降下法を用いて λ_i を求める．式 (58) の

$$\sum_{i=1}^n \lambda_i - \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n \lambda_i \lambda_j y_i y_j K(\mathbf{x}_i, \mathbf{x}_j)$$

の部分 $L(\boldsymbol{\lambda})$ としこの関数を最大化する $\boldsymbol{\lambda}$ を求めることを考える．今回は最大化する $\boldsymbol{\lambda}$ を求めるため， $L(\boldsymbol{\lambda})$ が増加する方向，即ち勾配ベクトルの方向に $\boldsymbol{\lambda}$ を更新する．更新する際の更新規則

は式 (62) となる．ただし $\lambda_i^{(k)}$ は k 回目の更新時の λ_i である．また，初期値である $\lambda_i^{(0)}$ には適当な値が設定されているものとする．(多くの場合で $\lambda_i = 0$) また， τ は学習率と呼ばれる $0 < \tau \leq 1$ の値を取るハイパーパラメータである．この学習率は別の分野ではステップ幅とも呼ばれ，更新時にどれだけ大きく λ_i を変化させるかを決める．大きすぎると発散してしまい学習が進まず，小さすぎると収束するのに必要な更新回数が増える．そのため適切な学習率を設定する必要がある．式 (62) によりすべての λ_i を求めることができる．

$$\begin{aligned}\lambda_i^{(k+1)} &= \lambda_i^{(k)} + \tau \frac{\partial L(\boldsymbol{\lambda})}{\partial \lambda_i} \\ &= \lambda_i^{(k)} + \tau \left\{ 1 - \sum_{j=1}^n \lambda_i^{(k)} y_i y_j K(\mathbf{x}_i, \mathbf{x}_j) \right\}\end{aligned}\tag{62}$$

5.2 SMO

最急降下法では 1 変数ずつ更新したが，双対問題でこの方法を使うと式 (58) の制約条件を満たさなくなってしまう．そこで，2 変数ずつ更新することで制約条件を満たしながら最適化する．SVM においてこの手法を SMO(Sequential Minimal Optimization) と呼ぶ [14]．まず，表記の簡単のため $\alpha_i = y_i \lambda_i$ ， $k_{ij} = K(\mathbf{x}_i, \mathbf{x}_j)$ とすると，式 (58) は式 (63) と書ける．

$$\begin{aligned}\max_{\boldsymbol{\alpha}} & -\frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n \alpha_i \alpha_j k_{ij} + \sum_{i=1}^n y_i \alpha_i \\ \text{s.t } & 0 \leq \alpha_i \leq C \quad (y_i = +1) \\ & -C \leq \alpha_i \leq 0 \quad (y_i = -1) \\ & \sum_{i=1}^n \alpha_i = 0\end{aligned}\tag{63}$$

ここで二つの変数 α_h と α_k を更新することを考える．このとき更新量を $\Delta\alpha_h$ とすると式 (64) のようにすることで制約条件を満たしながら更新することができる．

$$\begin{aligned}\alpha_h^{(k+1)} &= \alpha_h^{(k)} + \Delta\alpha_h \\ \alpha_k^{(k+1)} &= \alpha_k^{(k)} - \Delta\alpha_h\end{aligned}\tag{64}$$

次に，更新後の $\alpha_h^{(k+1)}$ と $\alpha_k^{(k+1)}$ も式 (65) のように式 (58) の一つ目と二つ目の制約を満たす必要があるため， $\Delta\alpha_h$ にも制約を加える必要がある．

$$\begin{aligned}0 &\leq \alpha_i + \Delta\alpha_h \leq C \quad (i \in i | y_i = +1) \\ -C &\leq \alpha_i + \Delta\alpha_h \leq 0 \quad (i \in i | y_i = -1)\end{aligned}\tag{65}$$

これを y_h, y_k の符号のそれぞれの組み合わせに対して考えると, $\Delta\alpha_h$ は式 (66) を満たす必要がある.

$$\begin{aligned} \max(-\alpha_h, \alpha_k - C) &\leq \Delta\alpha_h \leq \min(C - \alpha_h, \alpha_k) & (y_h = +1, y_k = +1) \\ \max(-\alpha_h, \alpha_k) &\leq \Delta\alpha_h \leq \min(C - \alpha_h, C + \alpha_k) & (y_h = +1, y_k = -1) \\ \max(-C - \alpha_h, \alpha_k - C) &\leq \Delta\alpha_h \leq \min(-\alpha_h, \alpha_k) & (y_h = -1, y_k = +1) \\ \max(-C - \alpha_h, \alpha_k) &\leq \Delta\alpha_h \leq \min(-\alpha_h, C + \alpha_k) & (y_h = -1, y_k = -1) \end{aligned} \quad (66)$$

式 (63) に式 (64) を代入して $\Delta\alpha_h$ についての制約付き最適化問題として表すと式 (67) となる. ただし, L, U は式 (66) の 4 つの不等式のうちのいずれかであり C' は定数項である.

$$\begin{aligned} \max_{\alpha} \frac{1}{2}(-k_{hh} + 2k_{hk} - k_{kk})\Delta\alpha_h^2 + \left(y_h - \sum_{i \in n} \alpha_i k_{ih} - y_k + \sum_{i \in n} \alpha_i k_{ik} \right) \Delta\alpha_h + C' \\ \text{s.t. } L \leq \Delta\alpha_h \leq U \end{aligned} \quad (67)$$

式 (67) は変数が $\Delta\alpha_h$ 1 つの制約付き 2 次関数の最小化問題となるので, $\Delta\alpha_h$ は式 (68) となる. これを用いることで λ_i を求めることができる.

$$\Delta\alpha_h = \begin{cases} L & \text{if } L > \Delta\alpha_h \\ U & \text{if } U < \Delta\alpha_h \\ \frac{-y_h + y_k + \sum_{i \in n} \alpha_i (k_{ih} - k_{ik})}{-k_{hh} + 2k_{hk} - k_{kk}} & \text{otherwise} \end{cases} \quad (68)$$

6 まとめ

最後に, ハードマージン線形 SVM による分類境界の求め方をまとめたものを図 6 に示す [15]. この図のようにして分類問題を解くことができる.

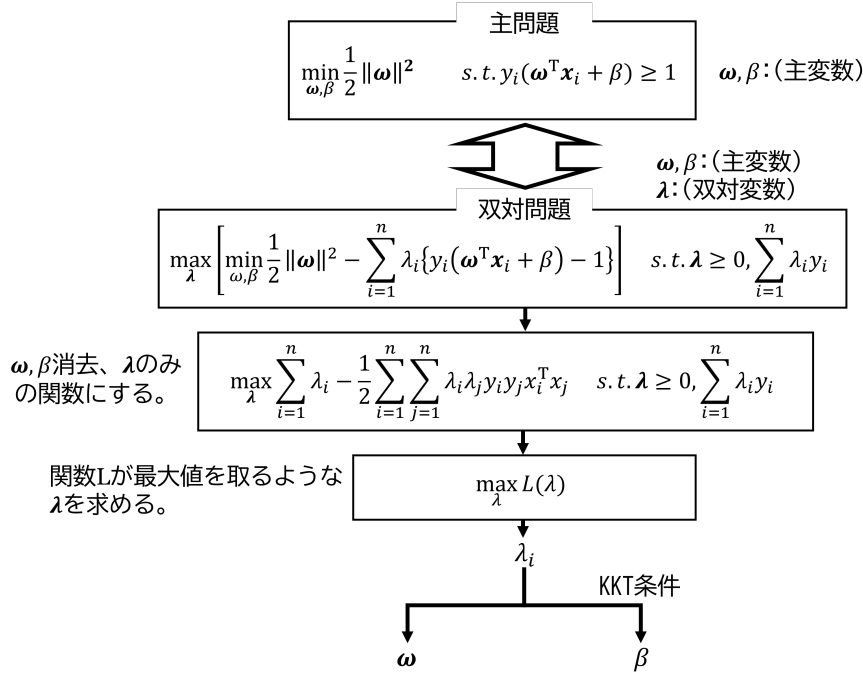


図6 SVMの流れ

カーネル関数を用いた線形 SVM の導出方法をまとめる．求める分類境界の式は式 (69) で表され，パラメータ ω, β を最適化する必要がある．

$$f(\mathbf{x}) = \omega^T \phi(\mathbf{x}) + \beta \quad (69)$$

ω, β を最適化するために式 (70) の最適化問題を解く．

$$\begin{aligned} \min_{\omega, \beta} \quad & \frac{1}{2} \|\omega\|^2 \\ \text{s.t.} \quad & y_i(\omega^T \phi(\mathbf{x}_i) + \beta) \geq 1 \quad (i = 1, 2, \dots, n) \end{aligned} \quad (70)$$

ω についての主問題を解くのは困難であるため λ についての双対問題である式 (57) を考える．

$$\begin{aligned} \max_{\lambda} \quad & \left(\sum_{i=1}^n \lambda_i - \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n \lambda_i \lambda_j y_i y_j K(\mathbf{x}_i, \mathbf{x}_j) \right) \\ \text{s.t.} \quad & \sum_{i=1}^n \lambda_i y_i = 0, 0 \leq \lambda_i \leq C \end{aligned} \quad (58)$$

この λ についての双対問題を解き λ を式 (61) に代入することで，分類境界を決定するパラメータ β が得られる．このときスパース性よりサポートベクトル以外の λ_i は考えなくてよい．

$$\beta = \frac{1}{|SV|} \sum_{i \in SV} \left(y_i - \sum_{i \in SV} \lambda_j y_j K(\mathbf{x}_i, \mathbf{x}_j) \right) \quad (61)$$

また、分類境界の式は式 (60) のように書き換えられるため、 ω を求めることなく分類境界の式を求められる。

$$f(\mathbf{x}) = \sum_{i \in SV} \lambda_i y_i K(\mathbf{x}, \mathbf{x}_i) + \beta \quad (60)$$

得られた分類境界について、未知のデータに対してもスパース性により分類の計算に使う λ_i はサポートベクトルのもののみとなる。

References

- [1] Qiita. SVM(サポートベクターマシン) についてまとめてみました - Qiita. URL: <https://qiita.com/masaya8028/items/cdecf3d8eeb4e02e11ac>.
- [2] risalc.info. ラグランジュ未定乗数法 ～ 基本から解説 ～ - 理数アラカルト -. URL: <https://risalc.info/src/Lagrange-multiplier.html>.
- [3] www.msi.co.jp. KKT 条件 — 数理最適化用語集. URL: <https://www.msi.co.jp/solution/nuopt/docs/glossary/articles/Karush-Kuhn-TuckerCondition.html>.
- [4] 機械学習と情報技術. カーネル法とカーネルトリックを図を用いて解説する / 機械学習と情報技術. 2022. URL: https://disassemble-channel.com/kernel-method-kernel-trick/#google_vignette.
- [5] Qiita. カーネル法のお話 - Qiita. URL: <https://qiita.com/ting/items/cc49b3703a9ea1067041>.
- [6] 化学とインフォマティクスと時々雑記. 【初心者向け】 ガウス過程回帰～カーネル関数についてわかりやすく説明～. 2023. URL: https://boritaso-blog.com/gaussian_process_regression_kernel/.
- [7] AI Academy Media. 交差検証(クロスバリデーション)とは？合わせてグリッドサーチに関しても学ぼう！ - AI Academy Media. 2020. URL: <https://aiacademy.jp/media/?p=263>.
- [8] Qiita. ベイズ最適化 from Scratch - Qiita. 2022. URL: <https://qiita.com/meltyyyyy/items/e67f22f98a96e30e5461>.
- [9] 本谷秀堅. パターン認識講義資料.
- [10] ai.reinforz.co.jp. SVM の C 値とカーネルタイプの選び方：最新研究から学ぶ最適化の秘訣 / Reinforz.ai. 2024. URL: <https://ai.reinforz.co.jp/797>.
- [11] MyEnigma. 最適化問題における双対問題とは？ - MyEnigma. URL: https://myenigma.hatenablog.com/entry/2017/01/30/083441#google_vignette.
- [12] Think IT (シンクイット) . 実感しよう！ スパース性が「潜む」データ. 2020. URL: <https://thinkit.co.jp/article/17713>.
- [13] 米国データサイエンティストのブログ -. 最急降下法を図と数式で理解する (超重要) 【機械学習入門 3】 - 米国データサイエンティストのブログ. URL: https://datawokagaku.com/gradient_descent/.

- [14] nomulog. サポートベクターマシン (*SVM*) の最適化. URL: https://www.nomuyu.com/svm_optimize/#st-toc-h-9.
- [15] Qiita. *SVM* をおさらい (ソフトマージン線形 *SVM* の編) - *Qiita*. URL: <https://qiita.com/s-nakagawa2/items/2845e601180aef0765ef>.