カーネル法入門
4. SVMの最適化

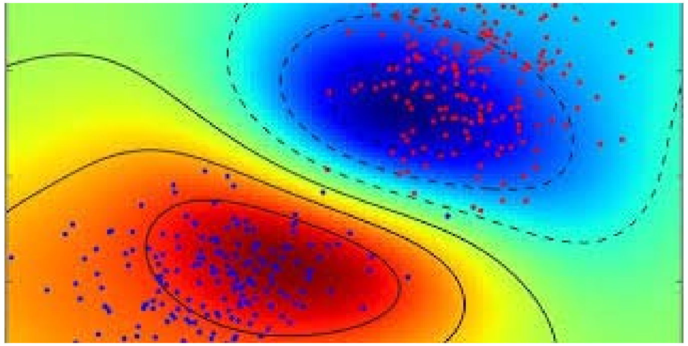
福水健次

統計数理研究所／総合研究大学院大学

大阪大学大学院基礎工学研究科・集中講義

2014 September

- 最適化問題の双対問題の一般論
- サポートベクターマシンの最適化とサポートベクター

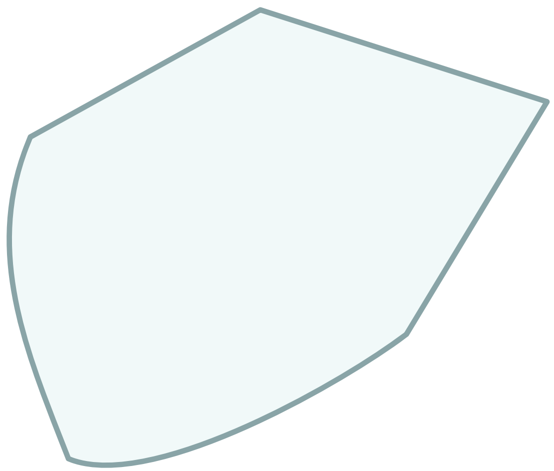


双対問題の一般論

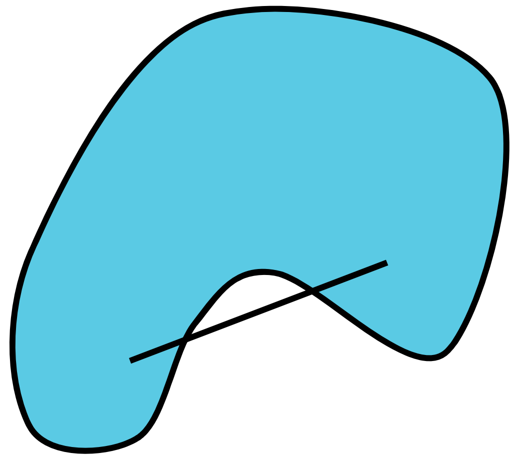
凸集合と凸関数

■ 凸集合

定義．　ベクトル空間 V の部分集合 E が凸（convex）であるとは，任意の2点 $x, y \in E$と　$t \in [0,1]$ に対し，$tx + (1-t)y \in E$ となることをいう.



convex set



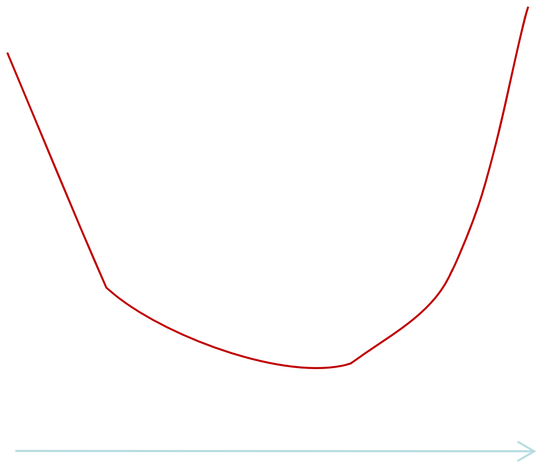
non-convex set

■ 凸関数／凹関数

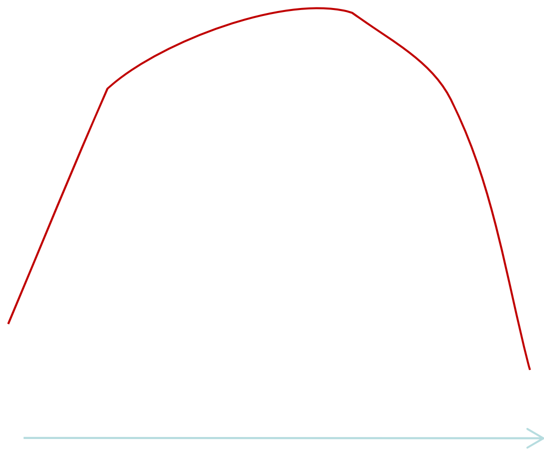
定義　ベクトル空間 V の凸部分集合 E 上定義された関数 $f\colon E \to R$ が 凸(convex)であるとは，任意の任意の2点 $x, y \in E$と　$t \in [0,1]$ に対し，

$$f(tx + (1-t)y) \le tf(x) + (1-t)f(y)$$

となることをいう. $-f$ が凸関数の場合，f を凹関数という.



convex function
凸関数



concave function
凹関数

Lagrange双対問題

– 主問題（凸性は仮定しない）

$$\min_{x \in D} f(x) \qquad \text{subject to} \begin{cases} h_i(x) \le 0 & (1 \le i \le \ell), \\ r_j(x) = 0 & (1 \le j \le m). \end{cases}$$

– Lagrange双対関数

$$g(\lambda, \nu) \coloneqq \min_{x \in D} L(x; \lambda, \nu) \qquad \lambda \ge 0, \nu \in \mathbf{R}^m$$

where

$$L(x; \lambda, \nu) \coloneqq f(x) + \sum_{i=1}^{\ell} \lambda_i h_i(x) + \sum_{j=1}^{m} \nu_j r_j(x).$$

λ, ν は Lagrange乗数と呼ばれる.

– Fact: $g(\lambda, \nu)$ は凹関数である.

– 双対問題

$$\max g(\lambda, \nu) \qquad \text{subj. to } \lambda \geq 0.$$

– 最適化(特に凸最適化)では，双対問題を考えると便利なことが多い.

<u>定理4．1</u>(弱双対性)

p_*: 主問題の最適値(infimum)

d_*: 双対問題の最適値(supremum)

このとき，

$$d_* \leq p_*.$$

∵) For $x \in D$, $\lambda \geq 0$,

$$L(x;\lambda,\nu) = f(x) + \sum_{i=1}^{\ell} \lambda_i h_i(x) + \sum_{j=1}^{m} \nu_j r_j(x) \leq f(x)$$

Take infimum $\Longrightarrow \quad g(\lambda,\nu) \leq p_*$.

強双対性

– 凸最適化問題
  - 目的関数 f と 不等式制約 h_i が凸関数.
  - 等式制約 r_j が線形関数.

– 十分条件： Slater条件
  「ある $\tilde{x} \in \mathrm{relint} D$ があって，$h_i(\tilde{x}) < 0$ かつ $r_j(\tilde{x}) = 0$」

定理4.2（強双対性）

主問題が凸最適化問題で，Slater条件を満たすとき，ある $\lambda_* \geq 0, \nu_* \in \mathbf{R}^m$ があって，

$$d_* = g(\lambda_*, \nu_*) = p_*.$$

証明は，福水B.3，または Boyd 2004 を見よ.

＊relint$D$（相対的内点）：Dを含む最小のアフィン部分空間の相対位相としての内点.

相補性条件

(凸とは限らない)主問題

$$\min_x f(x) \quad \text{subject to} \begin{cases} h_i(x) \le 0 & (1 \le i \le \ell) \\ r_j(x) = 0 & (1 \le j \le m) \end{cases}$$

仮定:

x_*: 主問題の解,　λ_*, ν_*: 双対問題の解($\lambda_* \ge 0$)

強双対性を仮定:　$f(x_*) = g(\lambda_*, \nu_*)$.

このとき

$$\begin{aligned} f(x_*) = g(\lambda_*, \nu_*) &= \inf_x L(x; \lambda_*, \nu_*) \\ &\le L(x_*; \lambda_*, \nu_*) \\ &= f(x_*) + \underbrace{\lambda_*^T h(x_*)}_{\le 0} + \underbrace{\nu_*^T r(x_*)}_{= 0} \\ &\le f(x_*) \end{aligned}$$

2つの不等号は実は等号

– Fact 1:

$$f(x_*) = \inf_x L(x: \lambda_*, \nu_*)$$

主問題の解は，Lagrange関数（最適なλ_*, ν_*のもと）の最適解

– Fact 2:

$$\lambda_{*i} h_i(x_*) = 0 \qquad \forall i$$

[相補性条件]

Complementary slackness

$h_i(x_*) < 0$ ならば $\lambda_{*i} = 0$

$\lambda_{*i} > 0$ ならば $h_i(x_*) = 0.$

KKT条件

- 凸最適化． D が開集合．主問題の目的関数 $f(x)$，不等式制約 $h_i(x)$ が微分可能とする.

- x_*: 主問題の解
  λ_*, ν_*: 双対問題の解（$\lambda_* \geq 0$）
  強双対性を仮定

- Lagrange双対： $x_* = \min_x L(x; \lambda_*, \nu_*)$
  微分可能関数の最小化問題として以下を満たす.

$$\nabla f(x_*) + \sum_{i=1}^{\ell} \lambda_{*i} \nabla h_i(x_*) + \sum_{j=1}^{m} \nu_{*j} \nabla r_j(x_*) = 0.$$

定理4．3 (KKT条件)

微分可能な目的関数，不等式制約からなる凸最適化問題において，強双対性が成り立つとする．x_*が主問題の解，λ_*, ν_*が双対問題の解となるための必要十分条件は，これらが以下のKKT条件を満たすことである.

(1) $h_i(x) \leq 0 \quad (1 \leq i \leq \ell)$ [主問題の制約]

(2) $r_j(x) = 0 \quad (1 \leq j \leq m)$ [主問題の制約]

(3) $\lambda_* \geq 0$ [双対問題の制約]

(4) $\lambda_{*i} h_i(x_*) = 0 \quad \forall i$ [相補性条件]

(5) $\nabla f(x_*) + \sum_{i=1}^{\ell} \lambda_{*i} \nabla h_i(x_*) + \sum_{j=1}^{m} \nu_{*j} \nabla r_j(x_*) = 0$

証明は，福水B.3，またはBoyd 2004

– 相補性条件は強双対性から導かれる.

サポートベクターマシンの最適化

SVMの最適化問題（復習）

$(X_1, Y_1), \dots, (X_n, Y_n)$: ２値分類問題 $Y_i \in \{+1, -1\}$.

– SVMの最適化問題(QP) : 主問題

$$\min_{w,b,\xi} \sum_i w_i w_j K_{ij} + C \sum_i \xi_i$$

$$\text{subj. to} \quad Y_i\left(\sum_j w_j K_{ij} + b\right) \geq 1 - \xi_i$$

$$\xi_i \geq 0.$$

$$K_{ij} = k\left(X_i, X_j\right)$$

$$f(x) = h(x) + b = \sum_{j=1}^{n} w_j \, k\left(x, X_j\right) + b$$

SVMの双対問題

■ SVMの双対問題

$$\max_{\alpha} \sum_{i=1}^{n} \alpha_i - \sum_{i,j=1}^{n} \alpha_i \alpha_j Y_i Y_j K_{ij}$$

$$\text{subject to} \quad \begin{array}{c} 0 \le \alpha_i \le C, \\ \sum_{i=1}^{n} Y_i \alpha_i = 0 \end{array}$$

[演習問題] Langrange双対を書き下すことによって，上の形を導け.

– 双対問題もQP.
– 制約は主問題より簡単（区間制約と1個の等式制約）
  双対問題を解くことが多い.

SVMのKKT条件

■ SVMのKKT条件

(1) $1 - Y_i\,(h^*(X_i) + b^*) - \xi_i^* \leq\ 0\ (\forall\ i)$ [primal constraint]

(2) $\xi_i^* \geq\ 0\ (\forall\ i)$ [primal constraint]

(3) $0 \leq \alpha_i^* \leq C, (\forall\ i)$ [dual constraint]

(4) $\alpha_i^*(1 - Y_i\,(h^*(X_i) + b^*)\ - \xi_i^*)\ =\ 0\ (\forall\ i)$ [complementary]

(5) $\xi_i^*(C - \alpha_i^*)\ =\ 0\ (\forall\ i),$ [complementary]

(6) $\sum_{j=1}^{n} K_{ij} w_j^* - \sum_j^n \alpha_j^* Y_j K_{ij} = 0,$

(7) $\sum_{j=1}^{n} \alpha_j^* Y_j = 0,$

– SVMの解： (6)の条件より, $w_j^* = Y_j \alpha_j^*$.

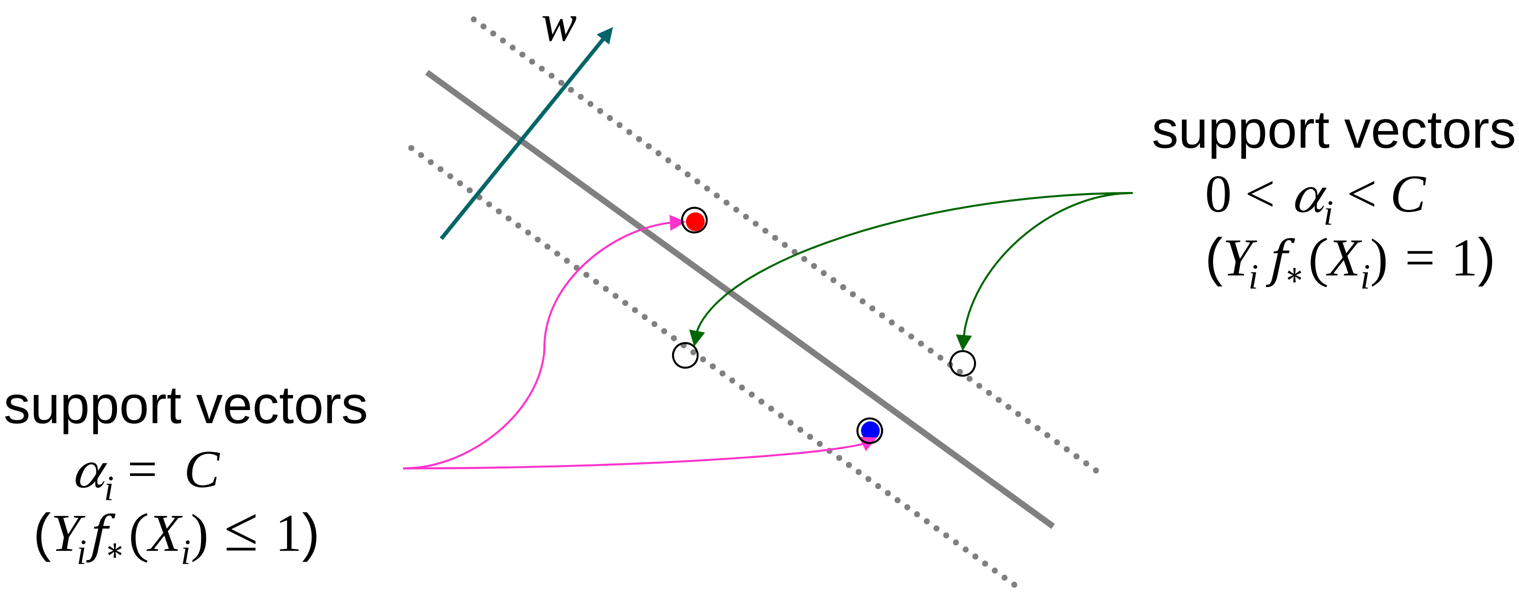
– (4)の相補性条件より, $\alpha_i^* \neq 0$ となるのは, $Y_i\,(h^*(X_i) + b^*) = 1 - \xi_i^*$ の場合に限られる.

サポートベクター

■ SVMの解のスパース性

$$f_*(x) = h_*(x) + b_* = \sum_{X_i:\ \text{support vectors}} \alpha_{i*} Y_i K(x, X_i) + b_*$$

– サポートベクター($\alpha_i^* \neq 0$なるもの)には2種類ある(条件(5)参照).



b の求め方

– 最適な b は条件(4)により求められる.
  $0 < \alpha_i^* < C$ なる i に対して,
  $$Y_i \left(\sum_j Y_j \alpha_j^* k\left(X_i, X_j\right) + b^* \right) = 1$$
  を解く.

– または, 上のすべての i に対する b を平均する.

SVMの解法いろいろ

実際には，大規模な問題には標準的QPソルバーは困難

– SMO（Sequential Minimal Optimization）
２変数の最適化を繰り返す．（Platt 1999)

– 主問題を解く

– オンライン学習

– 並列アルゴリズム

などなど.

多クラス識別への拡張

- マージン最大化規準の多クラスへの拡張

- ２値識別器の組み合わせ
  - One-vs-rest：　各クラスとその他
  - One-vs-one：　すべてのペア
  - ECOC（Error correcting output code）　符号化する

class	f_1	f_2	f_3	f_4	f_5	f_6
C_1	-1	-1	-1	1	1	1
C_2	-1	1	1	-1	-1	1
C_3	1	-1	1	-1	1	-1
C_4	1	1	-1	-1	1	1

汎化誤差

SVMに関しては，Vapnikらが発展させた「統計的学習理論」の枠組みで，マージンなどを用いた汎化誤差に関する理論研究が多く行われている.

ただし，本講義では扱わないので，福水５章，Cristianini & Shawe-Taylor (2000) などを見ていただきたい.

References

福水 「カーネル法入門」 4章　朝倉書店 2010

Schölkopf, B., A.J. Smola. Learning with Kernels. MIT Press 2001

Cristianini, N., J. Shawe-Taylor. An Introduction to Support Vector Machines. Cambridge Univ. Press. 2000

Platt, J. Fast training of support vector machines using sequential minimal optimization. In B. Schölkopf, C. Burges, and A. Smola, editors, *Advances in Kernel Methods - Support Vector Learning*, pages 185–208. MIT Press, 1999.

Boyd S. and L. Vandenberghe. *Convex Optimization*. Cambridge University Press, 2004. http://www.stanford.edu/ boyd/cvxbook/.

– Books on Application domains:

Lee, S.-W., A. Verri (Eds.) Pattern Recognition with Support Vector Machines: First International Workshop, SVM 2002, Niagara Falls, Canada, August 10, 2002. Proceedings. Lecture Notes in Computer Science 2388, Springer, 2002.

Joachims, T. Learning to Classify Text Using Support Vector Machines: Methods, Theory and Algorithms. Springer, 2002.

Schölkopf, B., K. Tsuda, J.-P. Vert (Eds.). Kernel Methods in Computational Biology. MIT Press, 2004.