双対性を利用した解析，モデル

京都大学大学院情報学研究科
山下信雄

紹介すること

1. 最適値関数と感度解析
2. サポートベクターマシン
3. ロバスト最適化
4. Performance Estimation Problem

感度解析

パラメータ u を含む次の問題を考える.

$$\begin{aligned} &\min\ f(x,u) \\ &\text{s.t.}\quad h_i(x,u)=0,\ i=1,\dots,m \\ &\qquad\ \ g_j(x,u)\le 0,\ j=1,\dots,r \end{aligned}$$

この最適値を $\theta(u)$ とする.

この θ の微分を考えるのが,**感度解析(sensitivity analysis)**である.

双対問題と関連を考えるときは,次の簡単な問題を考える.

$$\begin{aligned} &\min\ f(x) \\ &\text{s.t.}\quad g_j(x)\le u_j,\ j=1,\dots,r \end{aligned}$$

まずは最適値関数の一般論

パラメータ u を持つ問題：

$$
\begin{aligned}
\mathrm{P}(u):\ \min\ & f(x,u) \\
\text{s.t.}\ \ & g_j(x,u) \le 0,\ j=1,\ldots,r \\
& x \in S
\end{aligned}
$$

問題 $\mathrm{P}(u)$ の最適値 $\theta(u)$ を**最適値関数**という.

$$
\theta(u) = \min_{x\in S}\left\{ f(x,u) \,\middle|\, g_j(x,u) \le 0,\ j=1,\ldots,r \right\}
$$

最適値関数の凸性

Fiacco, A.V., Kyparisis, J. Convexity and concavity properties of the optimal value function in parametric nonlinear programming. *J Optim Theory Appl* **48**, 95–126 (1986).

例： $f(x,u)=\|x-u\|^2$

$f(x,u)=\phi(x)+\psi(u)$,　ϕ,ψは凸関数

Jointly convex:

$$f(\alpha x+(1-\alpha)y, \alpha u+(1-\alpha)v) \le \alpha f(x,u)+(1-\alpha)f(y,v)$$

点集合写像(実行可能集合)の凸性：$\Omega(u)=\left\{x\in S \,\middle|\, g_j(x,u)\le 0,\ j=1,\dots,r\right\}$

$$\alpha\Omega(u)+(1-\alpha)\Omega(v)\subseteq\Omega(\alpha u+(1-\alpha)v)$$

点集合写像の凸性の十分条件

$g_j(x,u)$ はjointly convex.　S は凸集合

証明.

任意の $\alpha \in [0,1]$ と $z \in \alpha\Omega(u)+(1-\alpha)\Omega(v)$ に対して

$z=\alpha x+(1-\alpha)y$ となる $x \in \Omega(u), y \in \Omega(v)$ が存在する.

$g_j(x,u) \leq 0, g_j(y,v) \leq 0$ より,

$$0 \geq \alpha g_j(x,u)+(1-\alpha)g_j(y,u) \geq g_j(\alpha x+(1-\alpha)y, \alpha u+(1-\alpha)v) = g_j(z, \alpha u+(1-\alpha)v)$$

よって,　$z \in \Omega(\alpha u+(1-\alpha)v)$

最適値関数の凸性

次の仮定が成り立つとき，最適値関数θ は凸関数.
- f はjointly convex.
- 実行可能領域の点集合写像 Ω は凸

証明：
$$\begin{aligned}\theta(\alpha u+(1-\alpha)v) &= \min_{z\in\Omega(\alpha u+(1-\alpha)v)} f(z,\alpha u+(1-\alpha)v)\\ &\le \min_{\substack{x\in\Omega(u)\\ y\in\Omega(v)}} f(\alpha x+(1-\alpha)y,\alpha u+(1-\alpha)v)\\ &\le \min_{\substack{x\in\Omega(u)\\ y\in\Omega(v)}} \alpha f(x,u)+(1-\alpha)f(y,v)\\ &= \min_{x\in\Omega(u)} \alpha f(x,u)+\min_{y\in\Omega(v)}(1-\alpha)f(y,v)\\ &=\alpha\theta(u)+(1-\alpha)\theta(v)\end{aligned}$$

最適値関数の微分①

$$\theta(u) = \min_{x \in S} \left\{ f(x,u) \middle| g_j(x,u) \le 0,\ j = 1, \ldots, r \right\}$$

u を固定したときの最適化問題 $\min_{x \in S} \left\{ f(x,u) \middle| g_j(x,u) \le 0,\ j = 1, \ldots, r \right\}$ のKKT条件：

$$\nabla_x f(x^*(u),u) + \sum_{j=1}^{r} \mu^*(u)_j \nabla_x g_j(x^*(u),u) = 0$$

$$g_j(x^*(u),u) \le 0,\ \mu^*(u)_j \ge 0,\ \mu^*(u)_j g_j(x^*(u),u) = 0,\ j = 1, \ldots, r$$

を満たす点 $(x^*(u), \mu^*(u))$ が唯一で，２次の十分条件を満たすとき，

$$\nabla \theta(u) = \nabla_u f(x^*(u),u) + \sum_{j=1}^{r} \mu^*(u)_j \nabla_u g_j(x^*(u),u)$$

最適値関数の劣微分

$f, g_j, j=1,\ldots,m$ はjointly convexとする.

$L(u,x,\mu)=f(x,u)+\sum_{j=1}^{r}\mu_j g_j(x,u)$ とすると,

$$\nabla_u L(u,x^*(u),\lambda^*(u))=\nabla_u f(x^*(u),u)+\sum_{j=1}^{r}\mu^*(u)_j\nabla_u g_j(x^*(u),u)\in\partial\theta(u)$$

ただし, $(x^*(u),\mu^*(u))$ は問題 $\min_{x\in S}\left\{f(x,u)\middle|g_j(x,u)\leq 0,\ j=1,\ldots,r\right\}$ のKKT点である.

$\nabla_u L(u, x^*(u), \lambda^*(u)) \in \partial_u \omega(u, \lambda^*(u))$ の証明

$\lambda^*(u) \geq 0$ のとき, $L(\cdot, \cdot, \lambda^*(u))$ は凸関数であるから,

$$
\begin{aligned}
L(v, x, \lambda^*(u)) &\geq L(u, x^*(u), \lambda^*(u)) + \left\langle \nabla_x L(u, x^*(u), \lambda^*(u)), x - x^*(u) \right\rangle + \left\langle \nabla_u L(u, x^*(u), \lambda^*(u)), v - u \right\rangle \\
&= L(u, x^*(u), \lambda^*(u)) + \left\langle \nabla_u L(u, x^*(u), \lambda^*(u)), v - u \right\rangle
\end{aligned}
$$

を得る．任意の x に対して成り立つので

$$
\omega(u, \lambda^*(u)) + \left\langle \nabla_u L(u, x^*(u), \lambda^*(u)), v - u \right\rangle \leq \inf_x L(v, x, \lambda^*(u)) = \omega(v, \lambda^*(u))
$$

最適値関数の微分と双対問題の最適解

次の最適値関数を考える．

$$\theta(u) = \min\left\{ f(x) \middle| g_j(x) \le u_j, j = 1, \ldots, r \right\}$$

さらに f と g_j は凸関数とする．

このとき，θ は凸関数であり，

$$-\mu^* \in \partial\theta(0)$$

が成り立つ．ここで μ^* は次の双対問題の最適解である．

$$\begin{aligned} &\max\ \omega(\mu) := \inf\left\{ f(x) + \sum_{j=1}^{r} \mu_j g_j(x) \right\} \\ &\text{s.t.} \quad \mu \ge 0 \end{aligned}$$

μ_j^* を制約 $g_j(x) \le 0$ の**潜在価格(shadow price)**という．

紹介すること

1. 最適値関数と感度解析
2. **サポートベクターマシン**
3. ロバスト最適化
4. Performance Estimation Problem

サポートベクターマシン：モデル化

入力データ

$\boldsymbol{x}^i$ ：因子データ

y^i ：クラスデータ（+1 or -1）

未知の因子データ　x が与えられたとき

$f(x) > 0$ ならクラス+1, $f(x) < 0$ならクラス- 1

と識別する識別関数 fを求めたい.

例：病気の診断

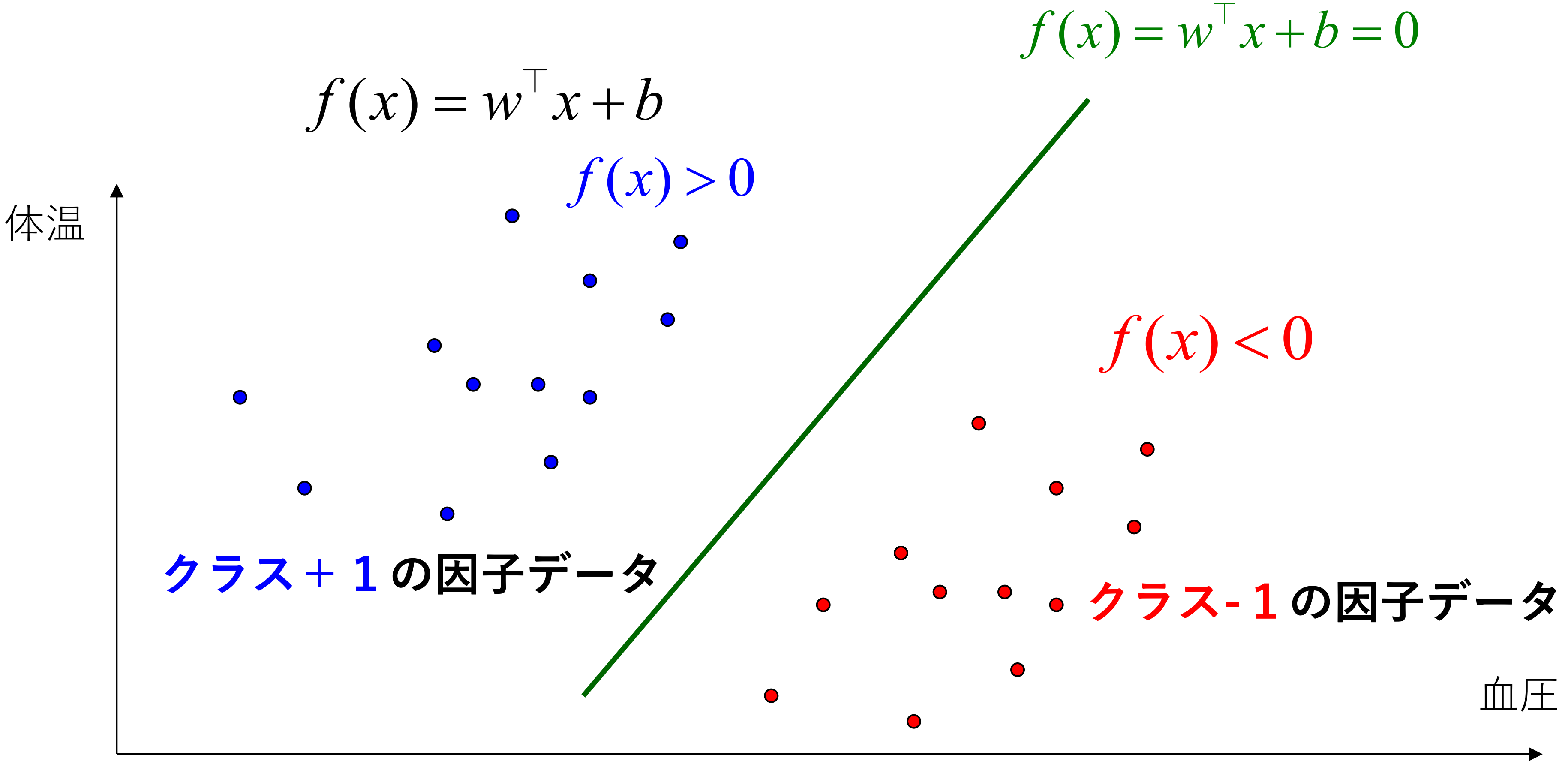
因子データ：　体温，血圧，体重，…

クラスデータ：　陽性(+ 1)，陰性(− 1)

線形サポートベクタマシン（SVM）

識別関数を線形関数とする．

$$f(x) = w^\top x + b$$



SVMの制約条件

線形な識別関数を求めること　⇔　w, b を求めること.

制約条件

正のデータに対しては

$$w^\top x^i + b > 0$$

負のデータに対しては

$$w^\top x^i + b < 0$$

正規化

正のデータに対しては

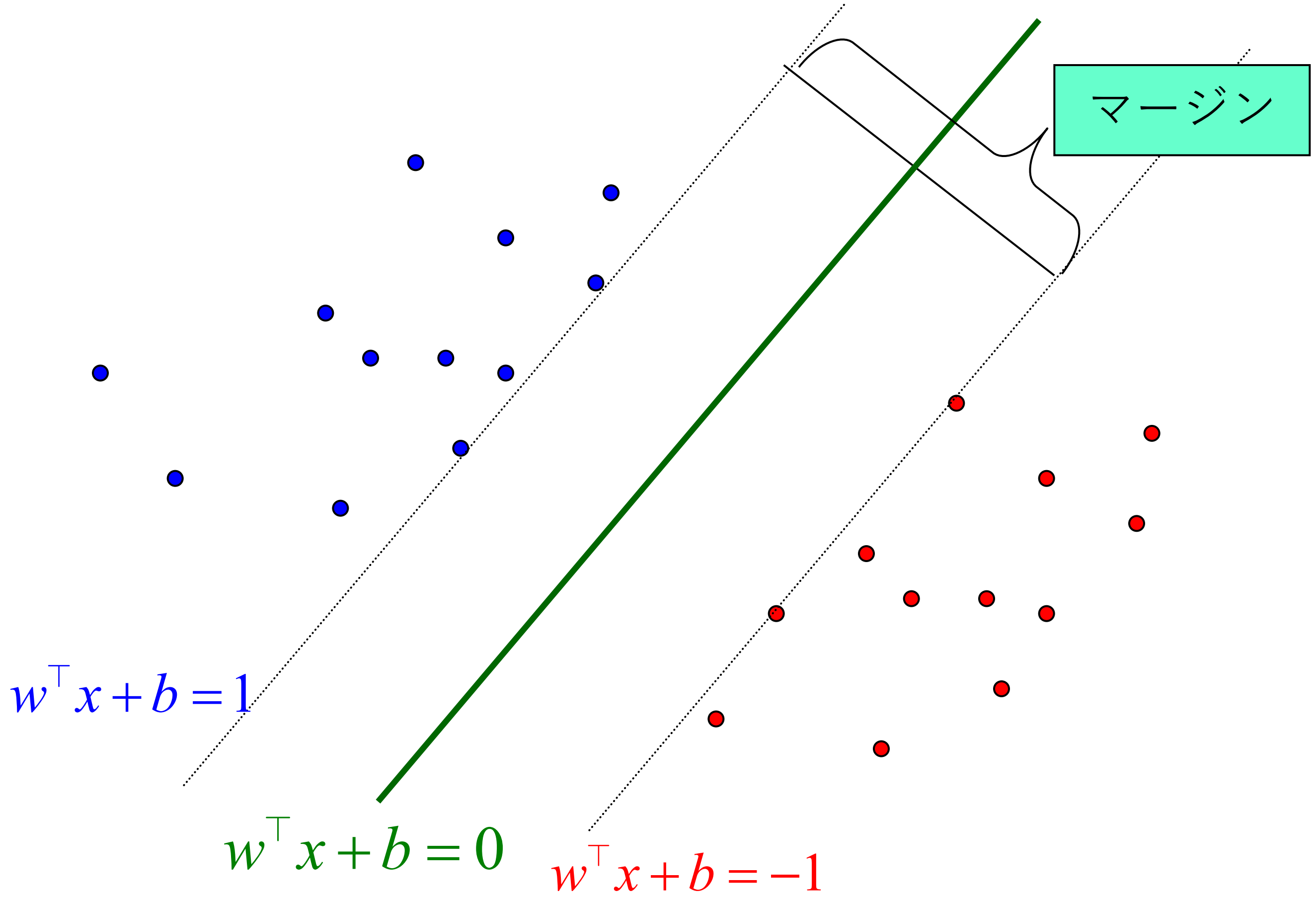
$$w^\top x^i + b \geq 1$$

負のデータに対しては

$$w^\top x^i + b \leq -1$$

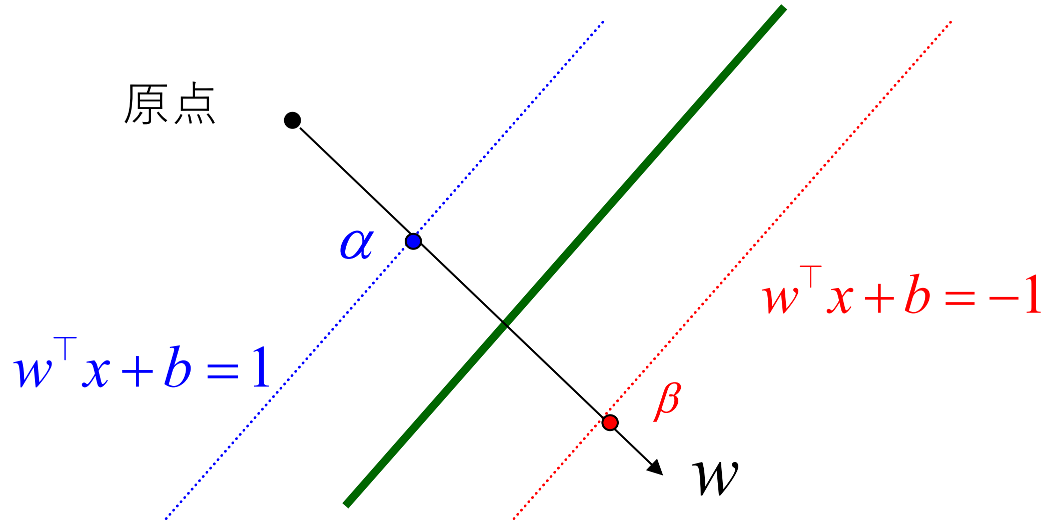
$$y^i(w^\top x^i + b) \geq 1,\ i = 1, \ldots, m$$

SVMの目的：マージンの最大化



マージンの最大化

超平面 $w^\top x+b=0$ は，ベクトル w と直交している．



原点から $w^\top x+b=1$ 上に一番近い点を α,

$w^\top x+b=-1$ 上に一番近い点を β とする．

マージンの計算

α と β は，tw と表せる．

α は，$w^\top x + b = 1$ 上にあるから，

$$w^\top (tw) + b = 1 \Rightarrow t = \frac{1-b}{\|w\|^2} \Rightarrow \alpha = \frac{(1-b)w}{\|w\|^2}$$

同様にして，$\beta = \dfrac{(-1-b)w}{\|w\|^2}$

よって，マージン $= \|\alpha - \beta\| = \dfrac{2}{\|w\|}$

線形SVM

マージン $\frac{2}{\|w\|}$ の最大化 ⇔ $\|w\|^2$ の最小化

$$\min \ \frac{1}{2}\|w\|^2$$

$$\text{s.t.} \quad y_i(w^\top x_i + b) \geq 1 \ \ i = 1, \ldots, m$$

⇒ 凸２次計画問題

ソフトマージン

線形関数で分離できないとき，実行可能解が存在しない．

⬇

$$\min \ \frac{1}{2}\|w\|^2 + C\sum_{i=1}^{m} v_i$$

$$\text{s.t.} \quad y_i(w^\top x_i + b) + v_i \geq 1 \ \ i = 1, \ldots, m$$

$$v_i \geq 0 \ \ i = 1, \ldots, m$$

KKT条件

$$w - \sum_{i=1}^{m} \alpha_i y_i x^i = 0$$

$$\sum_{i=1}^{m} \alpha_i y_i = 0$$

$$y_i(w^\top x^i + b + v_i) \geq 0,\ \alpha_i \geq 0,\ y_i(w^\top x^i + b + v_i)\alpha_i = 0$$

$$v_i \geq 0,\ \beta_i \geq 0,\ v_i \beta_i = 0$$

ソフトマージンSVMの双対問題

ラグランジュ関数

$$
\begin{aligned}
L(w,b,v,\alpha,\beta) &:= \frac{1}{2}\|w\|^2 + C\sum_{i=1}^{m} v_i + \sum_{i=1}^{m} \alpha_i (1 - y_i w^\top x^i - b y_i - v_i) - \sum_{i=1}^{m} \beta_i v_i \\
&= \frac{1}{2}\|w\|^2 - \sum_{i=1}^{m} \alpha_i y_i w^\top x^i - \sum_{i=1}^{m} \alpha_i y_i b + \sum_{i=1}^{m} (C - \alpha_i - \beta_i) v_i + \sum_{i=1}^{m} \alpha_i
\end{aligned}
$$

双対問題の目的関数

$$L(w,b,v,\alpha,\beta)=\frac{1}{2}\|w\|^2-\sum_{i=1}^{m}\alpha_i y_i w^\top x^i-\sum_{i=1}^{m}\alpha_i y_i b+\sum_{i=1}^{m}(C-\alpha_i-\beta_i)v_i+\sum_{i=1}^{m}\alpha_i$$

よって,

$$\omega(\alpha,\beta)=\begin{cases}-\infty & \text{if } \sum\alpha_i y_i\neq 0 \text{ or } C-\alpha_i-\beta_i\neq 0 \text{ for some } i\\ \min_w \frac{1}{2}\|w\|^2-\sum\alpha_i y_i w^\top x^i & \text{otherwise}\end{cases}$$

ここで, $\min_w \frac{1}{2}\|w\|^2-\sum\alpha_i y_i w^\top x^i$ の最小解は $\bar{w}=\sum_{i=1}^{m}\alpha_i y_i x^i$

双対問題

$$\max \; -\frac{1}{2}\alpha^\top K\alpha + \sum_{i=1}^{m}\alpha_i$$

$$\text{s.t.} \quad \sum_{i=1}^{m}\alpha_i y_i = 0$$

$$C - \alpha_i - \beta_i = 0 \quad i = 1,\ldots,m$$

$$\alpha_i \geq 0,\; \beta_i \geq 0 \quad i = 1,\ldots,m$$

ただし $K_{ij} = y_i y_j (x^i)^\top x^j$

$$\min \; \frac{1}{2}\alpha^\top K\alpha - \sum_{i=1}^{m}\alpha_i$$

$$\text{s.t.} \quad \sum_{i=1}^{m}\alpha_i y_i = 0$$

$$0 \leq \alpha_i \leq C \quad i = 1,\ldots,m$$

双対問題の解を用いた主問題の解表現

双対問題の解を α^* とすると，$w^* = \sum_{i=1}^{m} \alpha_i^* y_i x^i$

$0 < \alpha_i^* < C$ となる i に対して

$$y_i((w^*)^\top x^i + b^*) = 1$$

となることから，b^* が求まる．

識別関数

識別関数は

$$
\begin{aligned}
f(\color{red}{x}) &= (w^*)^\top \color{red}{x} + b^* \\
&= \sum_{i=1}^{m} \alpha_i^* y_i (x^i)^\top \color{red}{x} + b^*
\end{aligned}
$$

非線形SVM

ある関数 ϕ を用いて，識別関数は

$$f(x) = w^\top \phi(x) + b$$

と表されているとする．

双対問題と識別関数

$$\min \ \frac{1}{2}\alpha^\top K\alpha - \sum_{i=1}^{m}\alpha_i$$

$$\text{s.t.} \ \sum_{i=1}^{m}\alpha_i y_i = 0$$

$$0 \le \alpha_i \le C \quad i = 1, \ldots, m$$

ただし

$$K_{ij} = y_i y_j \phi(x^i)^\top \phi(x^j)$$

$$f(x) = (w^*)^\top \phi(x) + b^*$$

$$= \sum_{i=1}^{m}\alpha_i^* y_i \phi(x^i)^\top \phi(x) + b^*$$

カーネルトリック

ここで,

$$K(x,z) = \phi(x)^{\top}\phi(y)$$

とすると，双対問題と識別関数は ϕ を用いずに関数 $K(x,z)$ で表すことができる.

$K(x,z) = \phi(x)^{\top}\phi(y)$ となるような ϕ が存在する関数をカーネル関数という.

カーネル関数を決めて，識別関数を求めることをカーネルトリックという.

カーネルの例

ガウスカーネル

$$K(x,z) = \exp\left(-\frac{\|x-z\|^2}{\sigma}\right)$$

多項式カーネル

$$K(x,z) = (x^\top z + c)^p$$

サポートベクター回帰

$$\min_{w\in R^n, b\in R} \ \frac{1}{2}\|w\|^2 + C\sum_{t=1}^{T} \max\left\{0, |w^\top x^t + b - y_t| - \varepsilon\right\}$$

Fenchel双対問題

$$\min \ \frac{1}{2}\lambda K^\top \lambda - \sum_{t=1}^{T} b_t \lambda_t + \varepsilon\|\lambda\|_1$$

$$\text{s.t.} \quad \sum \lambda_t = 1$$

$$-C \le \lambda_t \le C \quad (t = 1,\ldots,T)$$

$$\lambda_t = \alpha_t - \beta_t,\ \ \alpha_t \ge 0,\ \ \beta_t \ge 0$$

$$\Rightarrow |\lambda_t| = \alpha_t + \beta_t$$

教科書によく出ているラグランジュ双対問題

$$\min \ \frac{1}{2}(\alpha-\beta)K^\top(\alpha-\beta) - \sum_{t=1}^{T} b_t(\alpha_t - \beta_t) + \varepsilon\sum_{t=1}^{T}(\alpha_t + \beta_t)$$

$$\text{s.t.} \quad \sum(\alpha_t - \beta_t) = 1$$

$$0 \le \alpha_t, \beta_t \le C \quad (t = 1,\ldots,T)$$

紹介すること

1. 最適値関数と感度解析
2. サポートベクターマシン
3. **ロバスト最適化**
   2023年夏季学校(後藤先生)
4. Performance Estimation Problem

ロバスト最適化

データが不確実なとき，最悪の場合を考えての最適化

ロバスト最適化

不確実性集合

$$\begin{aligned} &\min \quad f(x) \\ &\text{s.t.} \quad g(x;u) \le 0 \quad \forall u \in U \end{aligned}$$

不確実なデータ

$\Longleftrightarrow$

$$\begin{aligned} &\min \quad f(x) \\ &\text{s.t.} \quad \max_{u \in U} g(x;u) \le 0 \end{aligned}$$

半無限計画(semi-infinite programing)か，
制約条件中に最適化問題が含まれる問題になる．

ロバスト最適化がうまくいく理屈

制約条件中の問題

$$\max \; g(x;u)$$
$$\text{s.t.} \quad u \in U$$

ロバスト最適化

$$\min \; f(x)$$
$$\text{s.t.} \quad \max_{u \in U} \; g(x;u) \le 0$$

強双対性が成り立てば等価

双対問題

$$\min \; \omega(x;\lambda)$$
$$\text{s.t.} \quad \lambda \in \Omega$$

$$\min \; f(x)$$
$$\text{s.t.} \quad \omega(x;\lambda) \le 0$$
$$\lambda \in \Omega$$

$$\lambda \in \Lambda, \; w(x;\lambda) \le 0 \;\Rightarrow\; \max_{u \in U} g(x,u) \le w(x,\lambda) \le 0$$

弱双対

応用先によっては弱双対性でもよい？

- 理論が好きな研究者は等価性(強双対性)にこだわる.
- 双対ギャップが小さければ，十分じゃない？

SDP緩和とか使えるのでは？

紹介すること

1. 最適値関数と感度解析
2. サポートベクターマシン
3. ロバスト最適化
4. **Performance Estimation Problem**

Performance Estimation Problem

とある**アルゴリズム**の とある**問題のクラス内での最悪の計算量**の見積もりを
とある**最適化問題の最適値(or 上界値)**として与える.

とある最適化問題 ＝ Performance Estimation Problem

キーワード：ロバスト最適化，SDP緩和，弱双対性

Drori, Y., Teboulle, M. Performance of first-order methods for smooth convex minimization: **a novel approach**. *Math. Program.* **145**, 451–482 (2014).

Glineur, F., Performance estimation of optimization methods: A guided tour, Workshop on Nonsmooth Optimization and Applications (NOPTA 2024), University of Antwerp, April 8, 2024. https://perso.uclouvain.be/francois.glineur/files/talks/NOPTA2024.pdf

PEPの定式化例

とある問題のクラス： $F^{1,L}$

制約なしの凸最適化問題で，解集合 X^* は空でない．

目的関数 f は微分可能で勾配 ∇f は係数 L のリプシッツ連続

とあるアルゴリズム：

最急降下法 $x^{k+1} = x^k - \frac{1}{L}\nabla f(x^k)$

ざっくりと定式化すると

$$
\begin{aligned}
\max\quad & f(x^N) - f(x^*) \\
\text{s.t.}\quad & f \in F^{1,L},\ x^* \in X^* \\
& x^{k+1} = x^k - \frac{1}{L}\nabla f(x^k),\ k = 1,\ldots,N-1 \\
& \|x^0 - x^*\|^2 \le R
\end{aligned}
$$

$f \in F^{1,L}$ をどう表現する？

PEPの中に出てくるのは関数値 $f(x^k)$ と勾配ベクトル $\nabla f(x^k)$ だけ.

$f^k = f(x^k),\ g^k = \nabla f(x^k)$と考えて, $\{(x^k, f^k, g^k)\}, k = 1, \ldots, N-1$ を決定変数とする.

$f \in F^{1,L}$ を $\{(x^k, f^k, g^k)\}, k = 1, \ldots, N-1$ で表現できる？

$f \in F^{1,L}$ であれば $(x^k, f^k, g^k) = (x^k, f(x^k), \nabla f(x^k)), k = 1, \ldots, N-1$ は次の不等式をみたす.

$$f^i \geq f^j + \langle g^j, x^i - x^j \rangle + \frac{1}{2L} \|g^i - g^j\|^2 \quad \forall i, j$$

逆に $\{(x^k, f^k, g^k)\}, k = 1, \ldots, N-1$ がこの不等式を満たすとき,

$(x^k, f(x^k), \nabla f(x^k)) = (x^k, f^k, g^k), k = 1, \ldots, N-1$ となる関数 $f \in F^{1,L}$ が存在する.

$x^* \in X^*$ をどう表現する？

$x^* = 0,\ f(x^*) = 0$ としても計算量に影響しない.

以上よりPEPは以下の非凸な２次制約２次最適化問題として表せる.

$$
\begin{aligned}
\max\ & f^N - f^* \\
\text{s.t.}\ \ & x^* = 0,\ f^* = 0,\ g^* = 0 \\
& f^i \ge f^j + \left\langle g^j, x^i - x^j \right\rangle + \frac{1}{2L}\left\| g^i - g^j \right\|^2 \quad \forall i \ne j,\ i = 0\ldots, N-1,\ j = 0,\ldots, N-1 \\
& f^* \ge f^j + \left\langle g^j, x^* - x^j \right\rangle + \frac{1}{2L}\left\| g^* - g^j \right\|^2 \quad j = 0,\ldots, N-1 \\
& f^i \ge f^* + \left\langle g^*, x^i - x^* \right\rangle + \frac{1}{2L}\left\| g^i - g^* \right\|^2 \quad i = 0,\ldots, N-1 \\
& \left\| x^{k+1} - \left(x^k - \frac{1}{L} g^k \right) \right\|^2 = 0,\ k = 0,\ldots, N-1 \\
& \left\| x^0 - x^* \right\|^2 \le R
\end{aligned}
$$

PEPのSDP緩和

PEPではベクトル x^k, g^k は出てこないで,
内積の形 $\langle x^i, x^j \rangle, \langle g^i, g^j \rangle, \langle x^i, g^j \rangle$ で出てくる.
例えば, $N=1$ のときは

$$G = \begin{pmatrix} \langle x^0, x^0 \rangle & \langle x^0, x^1 \rangle & \langle x^0, g^0 \rangle & \langle x^0, g^1 \rangle \\ \langle x^1, x^0 \rangle & \langle x^1, x^1 \rangle & \langle x^1, g^0 \rangle & \langle x^1, g^1 \rangle \\ \langle g^0, x^0 \rangle & \langle g^0, x^1 \rangle & \langle g^0, g^0 \rangle & \langle g^0, g^1 \rangle \\ \langle g^1, x^0 \rangle & \langle g^1, x^1 \rangle & \langle g^1, g^0 \rangle & \langle g^1, g^1 \rangle \end{pmatrix}$$

そこで x^k, g^k の内積を G_{ij} に置き換える.
G は必ず半正定値対象行列となるため, $G \succeq 0$ を加える.
このように変換した問題は線形SDPとなり, 元のPEPの緩和問題になっている.

SDP緩和問題の双対問題の上界値

SDP緩和問題を解いても数値が求まるだけで，一般的なバウンド：

例：
$$f(x^N) - f^* \le \frac{LR^2}{2(2N+1)}$$

は出てこない．

そこでSDP緩和問題の双対問題を考え，

その**上界値**(**実行可能解の目的関数値**)によって，きれいなバウンドを出す．

$$\mathrm{opt(PEP)} \le \mathrm{opt}(\text{SDP緩和問題}) \le \mathrm{opt}(\text{SDP緩和問題の双対問題}) \le \mathrm{opt}(\text{その双対問題の上界値})$$

さらにそのバウンドとなる具体例を構成できれば最適なバウンドといえる．

ロバスト最適化として

PEPは とある問題のクラスの中で**最悪**の計算量を与える問題である.

本当は, **最高**のアルゴリズムを求めたい.

最適なステップ幅を求める問題はロバスト最適化として定式化できる.

$$
\begin{aligned}
\min_{t_0,\cdots,t_{N-1}} \quad \max \quad & f(x^N) - f(x^*) \\
\text{s.t.} \quad & f \in F^{1,L},\, x^* \in X^* \\
& x^{k+1} = x^k - t_k \nabla f(x^k),\, k = 0,\ldots,N-1 \\
& \left\| x^0 - x^* \right\|^2 \le R
\end{aligned}
$$