

劣モジュラ性を用いた データ生成過程の学習

大阪大学 産業科学研究所
/ 日本科学技術機構 (さきがけ)

河原 吉伸



Email: kawahara@ar.sanken.osaka-u.ac.jp
web: <http://www.ar.sanken.osaka-u.ac.jp/~kawahara/>

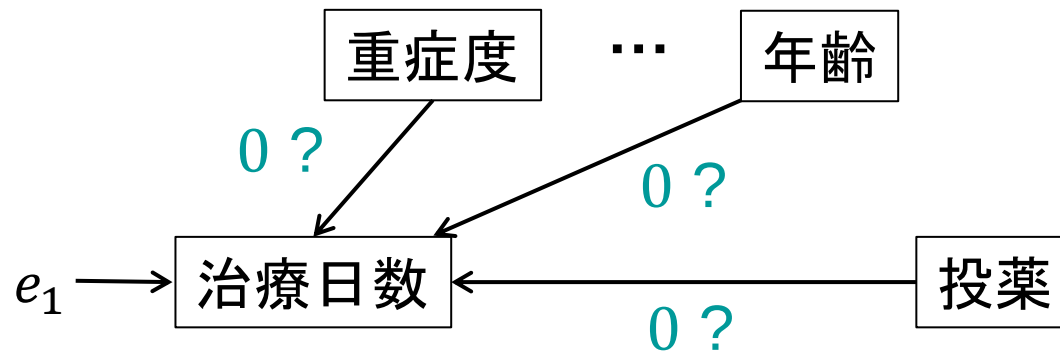
発表の概要

- ▶ 機械学習における組合せ的計算によるアプローチの重要性と難しい点, 劣モジュラ性利用の動機
- ▶ 劣モジュラ性の導入的説明と特徴選択への利用
- ▶ 劣モジュラカット法を中心として, 劣モジュラ性を用いた大域的解法の紹介

特徴選択(1)

例) 治療日数を決定する親変数の中から,

- 1. 治療日数を予測するのに有用な変数を選ぶ
- 2. 治療日数の要因を特定する



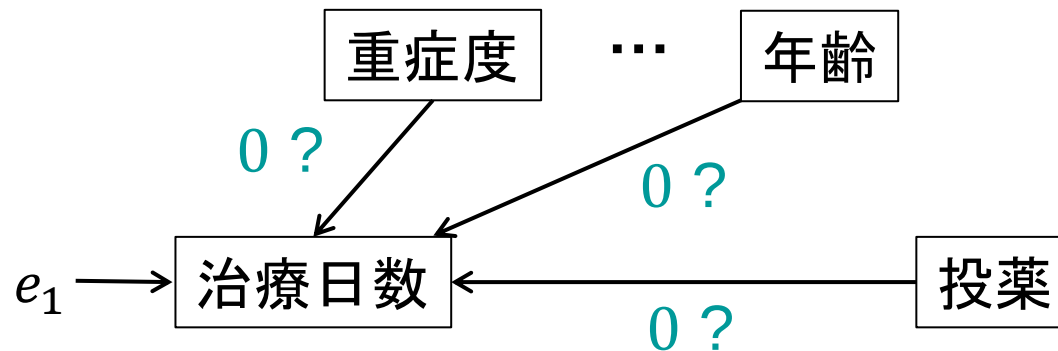
$$\begin{array}{ccccccc} \text{治療日数} & = & b_{11} \times \text{投薬} & + & b_{12} \times \text{重症度} & + & b_{13} \times \text{年齢} & + \cdots + e_1 \\ \hline (y) & & (x_1) & & (x_2) & & (x_3) & \end{array}$$

特徴選択(1)

例) 治療日数を決定する親変数の中から,

- 1. 治療日数を予測するのに有用な変数を選ぶ
- 2. 治療日数の要因を特定する

データ生成過程の学習



$$\begin{array}{ccccccc}
 \text{治療日数} & = & b_{11} \times \text{投薬} & + & b_{12} \times \text{重症度} & + & b_{13} \times \text{年齢} & + \cdots + e_1 \\
 (y) & & (x_1) & & (x_2) & & (x_3) &
 \end{array}$$

特徴選択(2)

- ▶ d 次元の入力 $x = \{x^1, x^2, \dots, x^p\}$ が与えられた時に, その部分集合を選ぶ

(目的1) 予測モデルの獲得(汎化誤差最小化)

⇒ 一般に, 精度と計算効率のバランスが重要

(目的2) データを生成する機構の同定

⇒ **大域的な最適性**を持った解のみが有意

⇒ 全探索的なアプローチが必要(組合せ最適化)

⇒ **組合せ爆発の発生**(サイズが大きくなると現実的時間内での計算が困難)

特徴選択(2)

- ▶ d 次元の入力 $x = \{x^1, x^2, \dots, x^p\}$ が与えられた時に, その部分集合を選ぶ

(目的1) 予測モデルの獲得(汎化誤差最小化)

⇒ 一般に, 精度と計算効率のバランスが重要

(目的2) データを生成する機構の同定

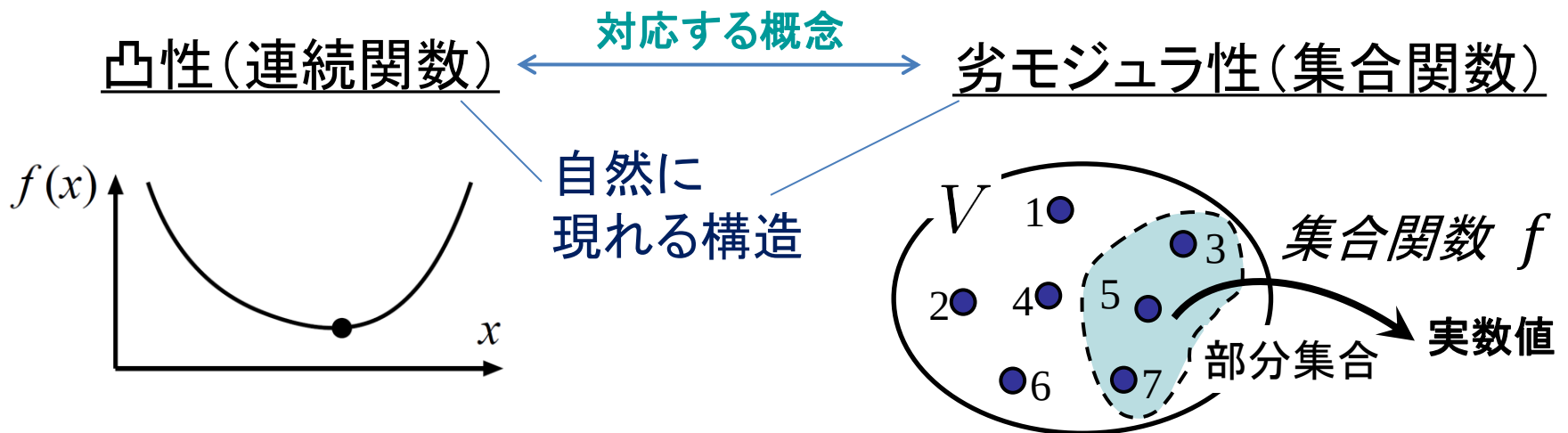
⇒ 大域的な最適性を持った解のみが有意

⇒ 全探索的なアプローチが必要(組合せ最適化)

⇒ 組合せ爆発の発生(サイズが大きくなると現実的時間内での計算が困難)

劣モジュラ性の利用

- ▶ 近年、劣モジュラ性(凸性に対応する離散構造)が、組合せ的計算において有用であるということが知られてきた
- ⇒ 組合せ的なデータ解析の飛躍的進歩の寄与への期待



機械学習では、サポートベクターマシンなどで、凸性が有効に利用されてきた

$$f(S) + f(T) \leq f(S \cap T) + f(S \cup T)$$

$(S, T \subseteq V)$

機械学習問題における利用例

▶ クラスタリング:

- Q-Clustering (Narashimhan, Jojic & Bilmes (NIPS'06))
- MAC Clustering (Nagano, Kawahara & Iwata (NIPS'10))
- バランス・クラスタリング (Kawahara, Nagano & Okamoto (PatRec(in press)))

▶ 特徴選択:

- Dictionary学習 (Krause & Cevher (ICML'10))
- 正則化関数としての劣モジュラ関数 (Bach (NIPS'10))

▶ 能動学習:

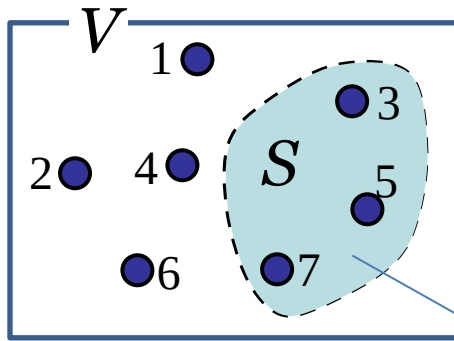
- 画像分類への応用 (Hoi et al. (ICML'06))

▶ その他の工学的問題:

- センサ配置 (Krause, Singh & Guestrin (JMLR))
- グラフ・マイニング (Thoma et al. (SDM'09))
- ネットワークの構造推定 (Chechotka & Guestrin (NIPS'08), Steer et al. (ICML'09), Gomez Rodriguez et al. (KDD'10))

(補) 集合関数

- ▶ 集合関数 $f : 2^V \rightarrow \mathbb{R}$ ($2^V = \{S : S \subseteq V\}$)
 - 有限集合 $V = \{1, \dots, n\}$ の部分集合 $S \subseteq V$ から実数への関数



$$V = \{1, 2, \dots, 7\}$$

$$S = \{3, 5, 7\}$$

こういった一つ一つの部分集合が,
 f により実数値に対応づけられる

劣モジュラ性とは (1)

- ▶ 連続関数における凸性に対応する集合関数の構造:

連続関数における凸性



集合関数における劣モジュラ性

共通する部分が多い

- 最小化に関する効率的なアルゴリズム構築が可能:
 - Queyranneのアルゴリズム(Queyranne,1998)や最小ノルム点アルゴリズム(Fujishige,1980)など.
- いくつかの凸関数への連続緩和が知られている:
 - Lovász拡張(Lovász,1983)や, (Nemhauser et al.,1978), (Feige,STOC'06) など.

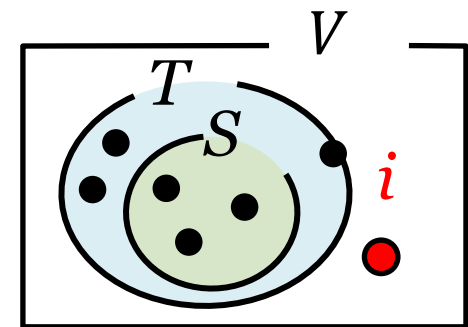
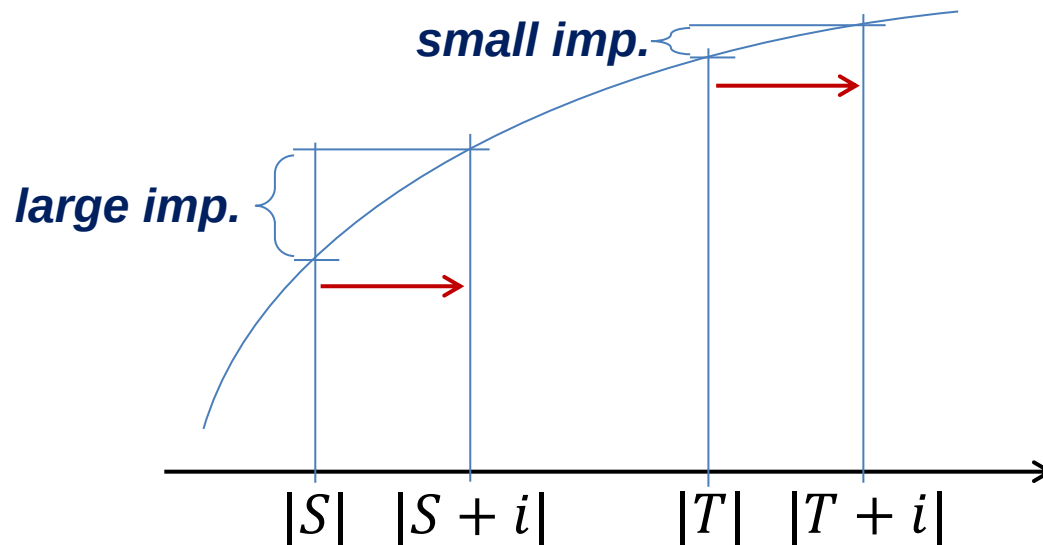
劣モジュラ性とは (2)

- ▶ 劣モジュラ関数: 次式を満たす集合関数 f ($S \subseteq T \subseteq V, i \in V \setminus T$)

$$f(S + i) - f(S) \geq f(T + i) - f(T)$$

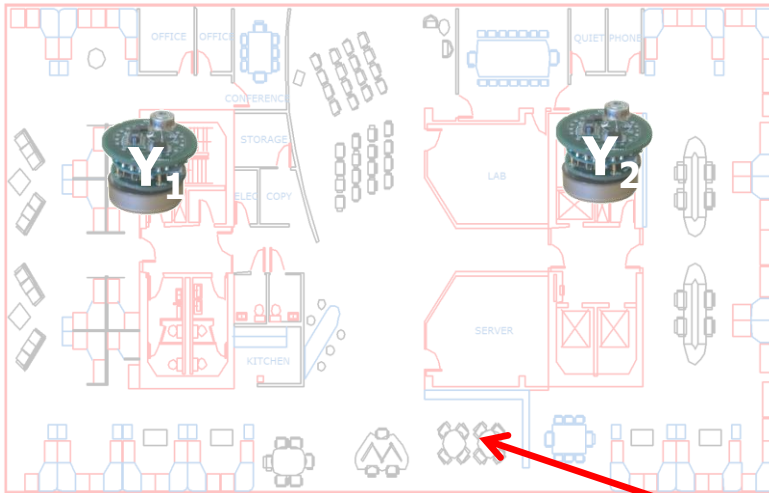


Diminishing Returns (規模の経済性, 限界効用逓減の法則など)

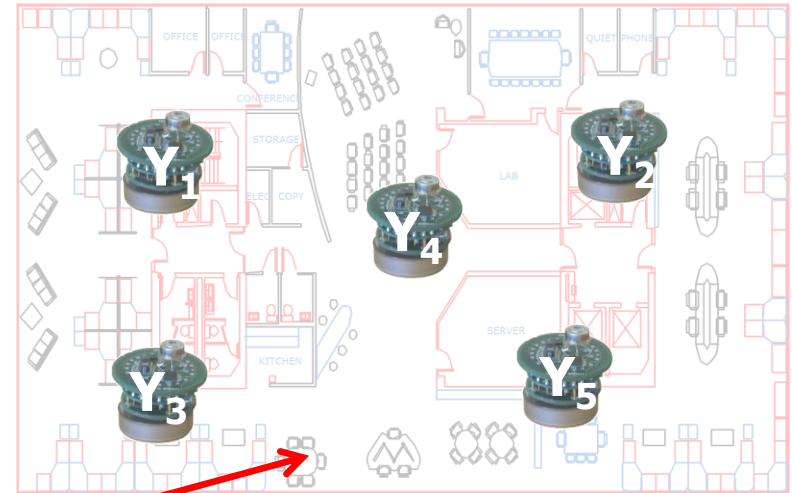


劣モジュラ性とは (3)

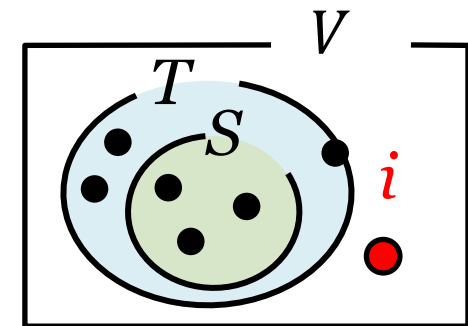
Location $S = \{Y_1, Y_2\}$



Location $S = \{Y_1, Y_2, Y_3, Y_4, Y_5\}$



$$\underbrace{f(S + i) - f(S)}_{\text{'large' imp.}} \geq \underbrace{f(T + i) - f(T)}_{\text{'small' imp.}}$$



劣モジユラ関数の例

- ▶ 組合せ最適化をはじめとして、情報理論やグラフ理論、ゲーム理論、経済学など、多くの数理工学や社会科学の分野において頻繁に現れる.
 - 情報理論
 - 同時エントロピー, (対称)相互情報量, 情報利得関数など
 - グラフ理論
 - カット関数, ネットワークのカット容量, マトロイドのランクなど
 - その他
 - 経済学における効用関数(優モジユラ), Convex game, 被覆関数, 正定値対称行列の行列式, 自乗重相関係数(負の自乗誤差)など

凸性との関係 (Lovász拡張) (1)

Lovász拡張: 集合関数 f の連続関数への拡張の一つ

(定義) 任意の実ベクトル $p \in \mathbb{R}^n$ の要素の値を大きい順に

$$\hat{p}_1 > \hat{p}_2 > \dots > \hat{p}_m$$

と並べたとき, 任意の集合関数 f に対して, Lovász拡張 $\hat{f}$ は次のように定義される ($U_k = \{i \in V : p_i \geq \hat{p}_k\}$):

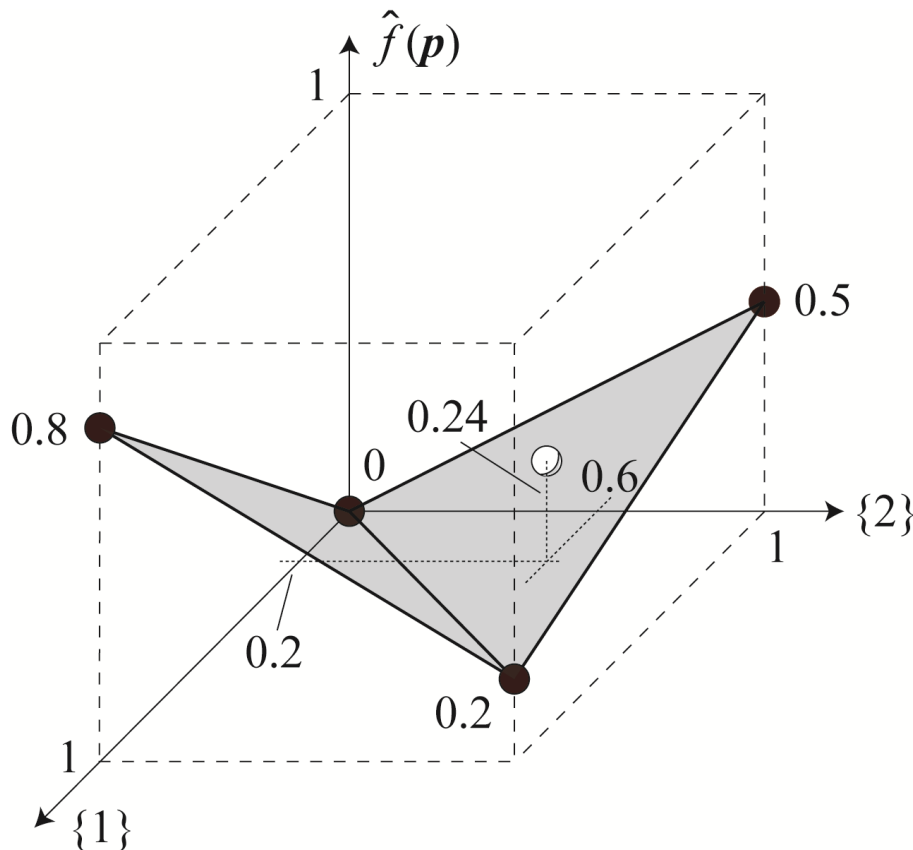
$$\hat{f}(p) = \sum_{k=1}^{m-1} (\hat{p}_k - \hat{p}_{k+1}) f(U_k) + \hat{p}_m f(U_m)$$

$$\begin{array}{ccc} f: 2^V \mapsto \mathbb{R} & \xrightarrow{\text{Lovasz拡張}} & \hat{f}: \mathbb{R}^n \mapsto \mathbb{R} \\ \text{(集合関数)} & & \text{(連続関数)} \end{array}$$

凸性との関係 (Lovász拡張) (2)

Lovász拡張: 集合関数 f の連続関数への拡張

例) $|V| = 2$. $f(0) = 0, f(V) = 0.2, f(\{1\}) = 0.8, f(\{2\}) = 0.5$



$p = (0.2, 0.6)$ における $\hat{f}$

1. p の要素を降順に並べる.

$$p_1 = 0.6 > p_2 = 0.2$$

$$(U_1 = \{2\}, U_2 = (\{1, 2\}))$$

2. 定義に従って,

$$\begin{aligned} \hat{f}(p) &= (0.6 - 0.2) \times f(\{2\}) \\ &\quad + 0.2 \times f(V) = 0.24 \end{aligned}$$

凸性との関係 (Lovász拡張) (3)

Lovász拡張: 集合関数 f の連続関数への拡張

(定義) 任意の実ベクトル $p \in \mathbb{R}^n$ の要素の値を大きい順に

$$\hat{p}_1 > \hat{p}_2 > \dots > \hat{p}_m$$

と並べたとき, 任意の集合関数 f に対して, Lovász拡張 $\hat{f}$ は次のように定義される ($U_k = \{i \in V : p_i \geq p_k\}$):

$$\hat{f}(p) = \sum_{k=1}^{m-1} (\hat{p}_k - \hat{p}_{k+1}) f(U_k) + \hat{p}_m f(U_m)$$

定理 (Lovász, 1983):

A set-function f is submodular if and only if its Lovász extension $\hat{f}$ is convex.

正則化学習による特徴選択(1)

- ▶ 正則化損失関数最小化としての特徴選択:

$$\min_{\mathbf{w} \in \mathbb{R}^p} \underbrace{\sum_{i=1}^n l(y_i, f(x_i))}_{\text{損失関数}} + \underbrace{\lambda \cdot \Omega(\mathbf{w})}_{\text{正則化項}}$$

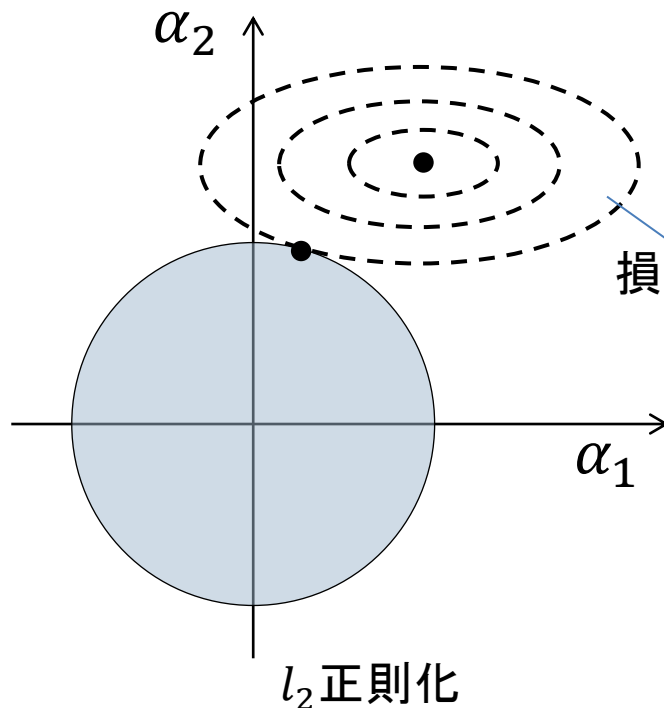
$$\left[\Leftrightarrow \min_{\mathbf{w} \in \mathbb{R}^p} \sum_{i=1}^n l(y_i, f(x_i)) \quad \text{s.t. } \Omega(\mathbf{w}) \leq k \right]$$

パラメータへの拘束を加える事で、モデルパラメータを $\mathbf{w}$ をスパースにする(多くの次元が0)

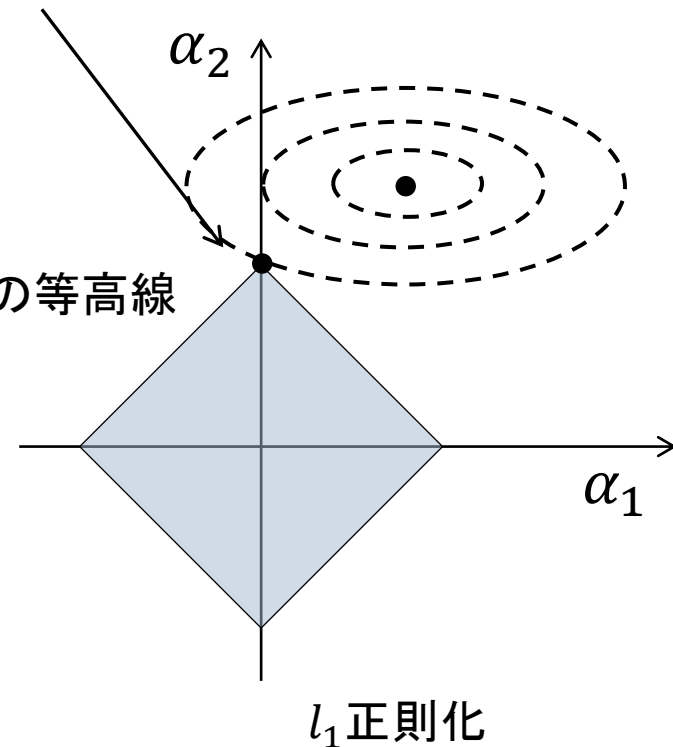
正則化学習による特徴選択(2)

- ▶ 正則化項の選択により, スパースな解を得る事ができる.

解が軸に乗りやすいため, スパースな解が得られやすい



損失関数値の等高線



特徴選択における劣モジュラ性の利用

- ▶ 正則化損失関数最小化としての特徴選択:

$$\min_{\mathbf{w} \in \mathbb{R}^p} \underbrace{\sum_{i=1}^n l(y_i, f(x_i))}_{\text{損失関数}} + \underbrace{\lambda \cdot \Omega(\mathbf{w})}_{\text{正則化項}}$$

$$\left[\Leftrightarrow \min_{\mathbf{w} \in \mathbb{R}^p} \sum_{i=1}^n l(y_i, f(x_i)) \quad \text{s.t.} \quad \Omega(\mathbf{w}) \leq k \right]$$

1. 損失関数における劣モジュラ性

2. 正則化における劣モジュラ性 (Bach, NIPS'10)

損失関数における劣モジュラ性(1)

- ▶ 特徴選択における損失関数には, 集合関数としてみた場合, 優モジュラ性(負の劣モジュラ性)を有するものが多い:

$$\min_{\mathbf{w} \in \mathbb{R}^p} \sum_{i=1}^n l(y_i, f(x_i)) + \lambda \cdot \Omega(\mathbf{w})$$

正則化項により, スパースな解を得る事で特徴の部分集合を選択



直接的に, 集合関数最適化として特徴の部分集合を選択

$$\max_{S \subseteq V} \underline{g(S)} \quad \text{s.t. } \Omega(S) \leq k \quad (\text{劣モジュラ最大化})$$

劣モジュラ関数.

例えば, 自乗重相関係数(負の自乗誤差)や相互情報量

損失関数における劣モジュラ性(2)

例) { 損失関数: 自乗誤差 $l(y_i, f(\mathbf{x}_i)) = |y_i - f(\mathbf{x}_i)|^2$
 モデル: 線型モデル $f(\mathbf{x}_i) = \mathbf{w} \cdot \mathbf{x}_i^T$
 正則化項: l_0 ノルム $\Omega(\mathbf{w}) = |\mathbf{w}|^0$

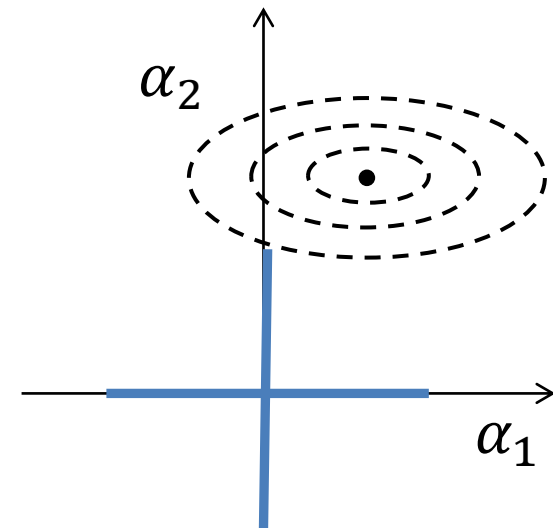


$V := \{1, \dots, p\}$ (パラメータ $\mathbf{w} \in \mathbb{R}^p$ のインデックス)

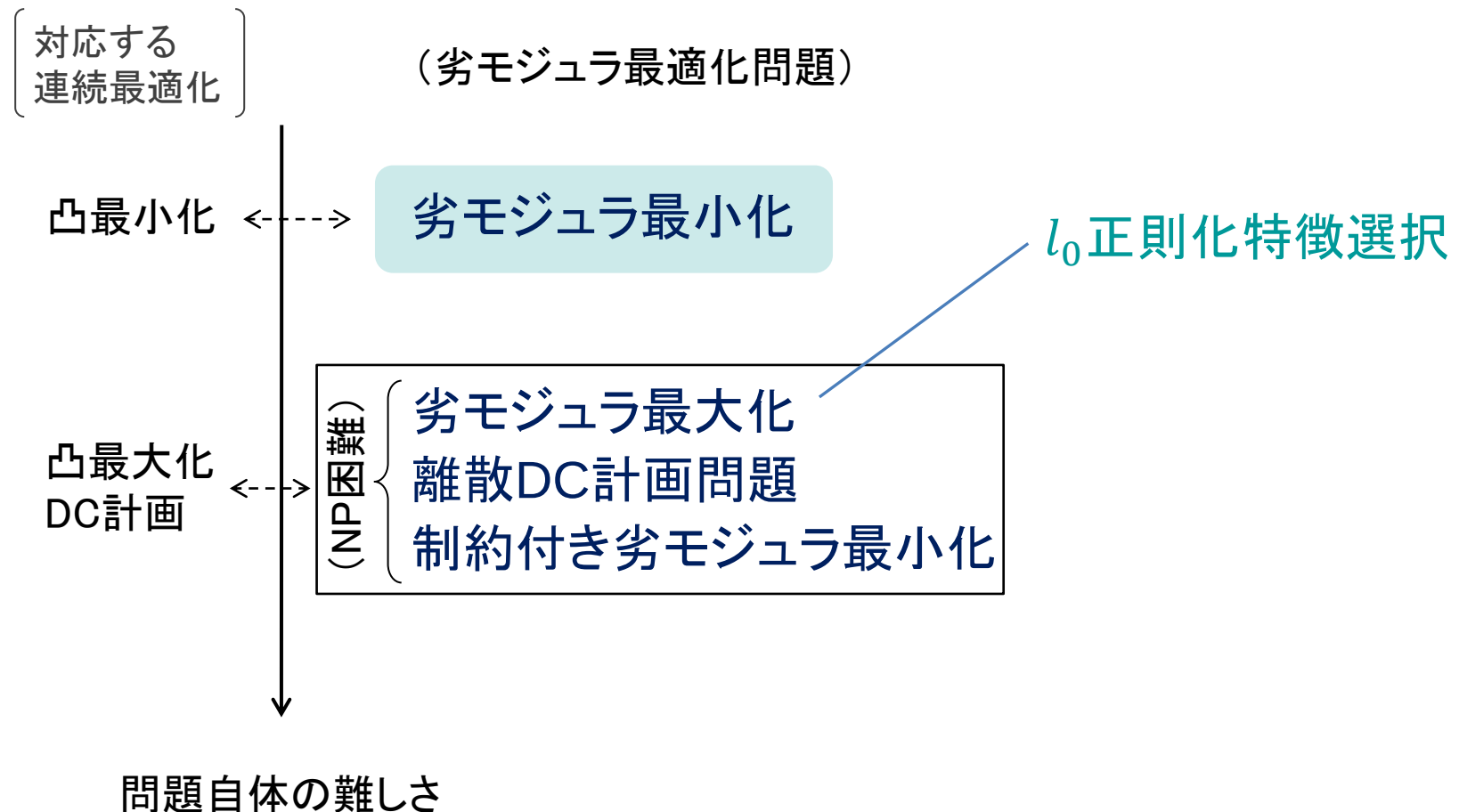
$g(S) := -\sum_i \left| y_i - \sum_{j \in S (\subseteq V)} w^j x_i^j \right|^2$ (劣モジュラ (Das&Kempe'08))

$$\Rightarrow \boxed{\max_{S \subseteq V} g(S) \quad \text{s.t. } |S| \leq k}$$

(サイズ制約下の劣モジュラ関数最大化)



劣モジュラ最適化における位置づけ



劣モジュラ最大化の厳密解法

▶ 劣モジュラ最大化:

$$\max_{S \subseteq V} f(S) \quad \text{s.t. } |S| \leq k$$

- NP困難(従って, (強)多項式時間で計算可能な厳密アルゴリズムの実現は絶望的)
- f が単調増加関数の場合, 貪欲法で $(1 - 1/e)$ -近似が得られる事が知られている (Nemhauser et al., 1978)

⇒ 厳密解法(大域的最適化アルゴリズム)

- 制約生成アルゴリズム (Nemhauser et al., 1978)
- 劣モジュラカット法 (Kawahara et al., NIPS'09)

劣モジュラ・カット(動機)

(問題点) 実行可能解が $n := |V|$ と共に指数関数的に増加



$$(\text{解の数}) = \binom{n}{k} + \binom{n}{k-1} + \cdots + \binom{n}{1} \quad (k : \text{cardinality})$$

(疑問) 最適性を保証しつつ, 実行可能解を効率的に減らす事はできないか?



(解決案) Lovasz拡張を通した凸性と劣モジュラ性の関係を用いて, 連続の大域的最適化の手法を適用する

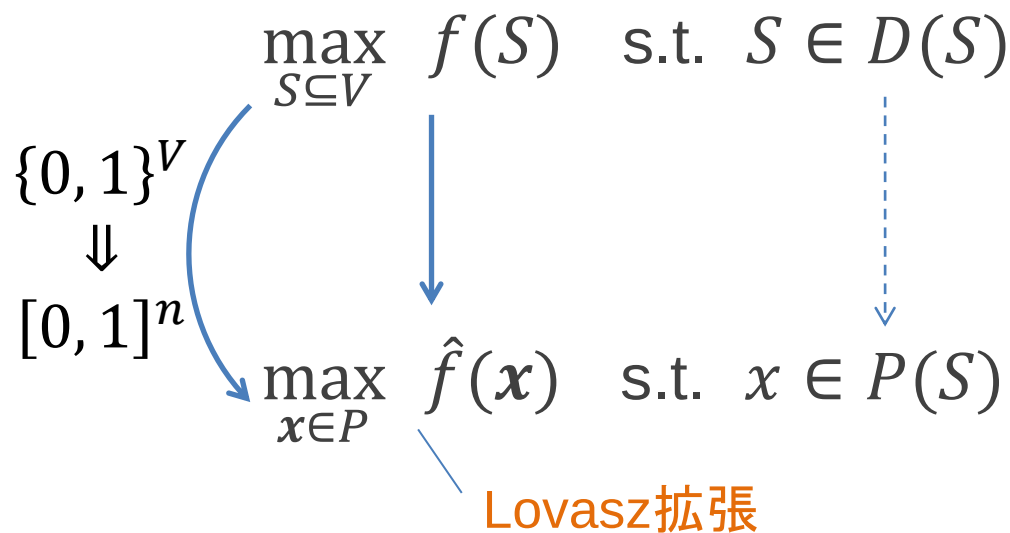
劣モジュラ・カット(1)

- ▶ 現在, 最良値 γ と探索域 $D(S)$ を知っているとする
- ▶ 劣モジュラ最大解の連続への緩和を考える:

問題 ①

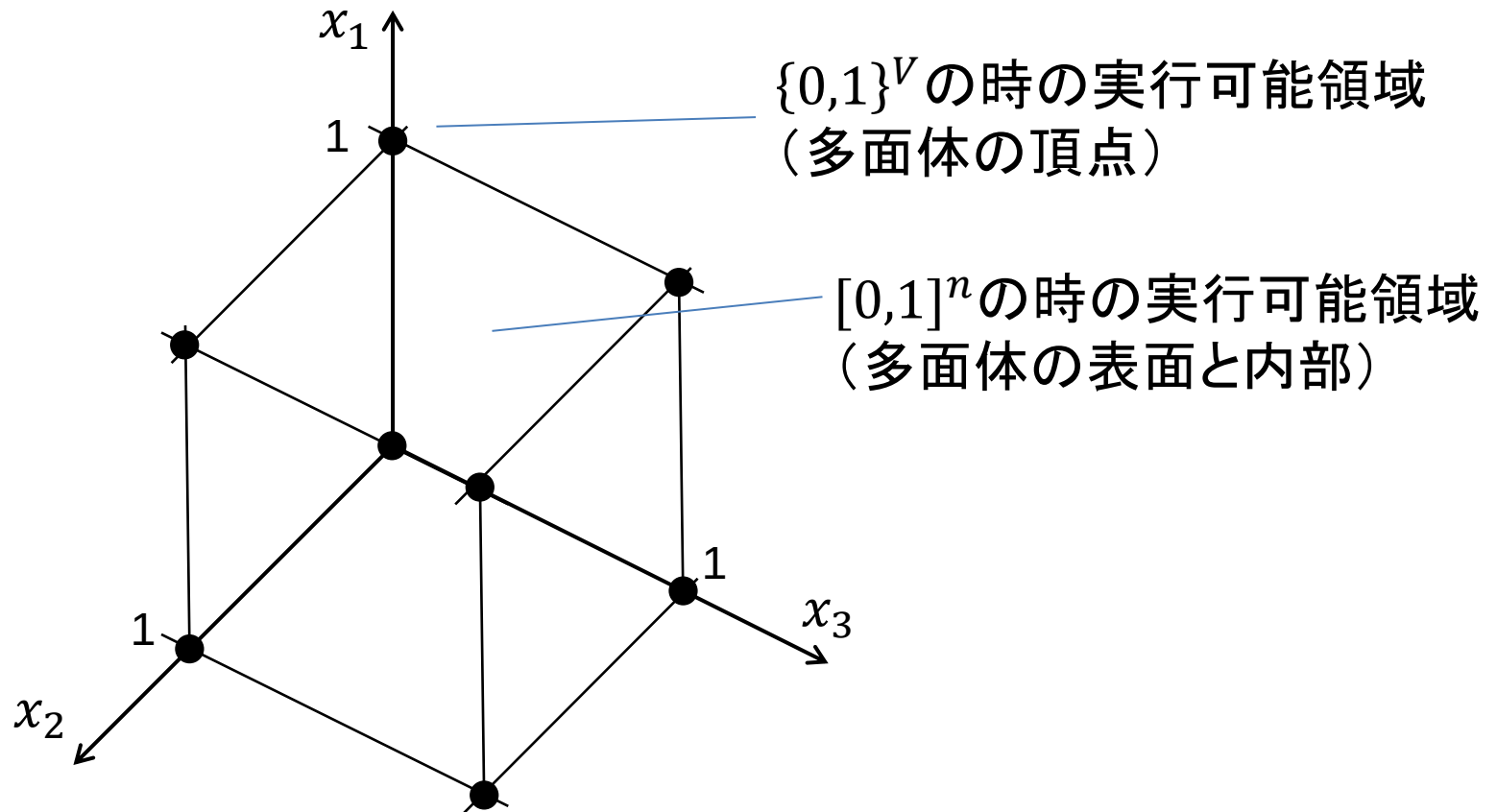


問題 ②



劣モジュラ・カット(1)

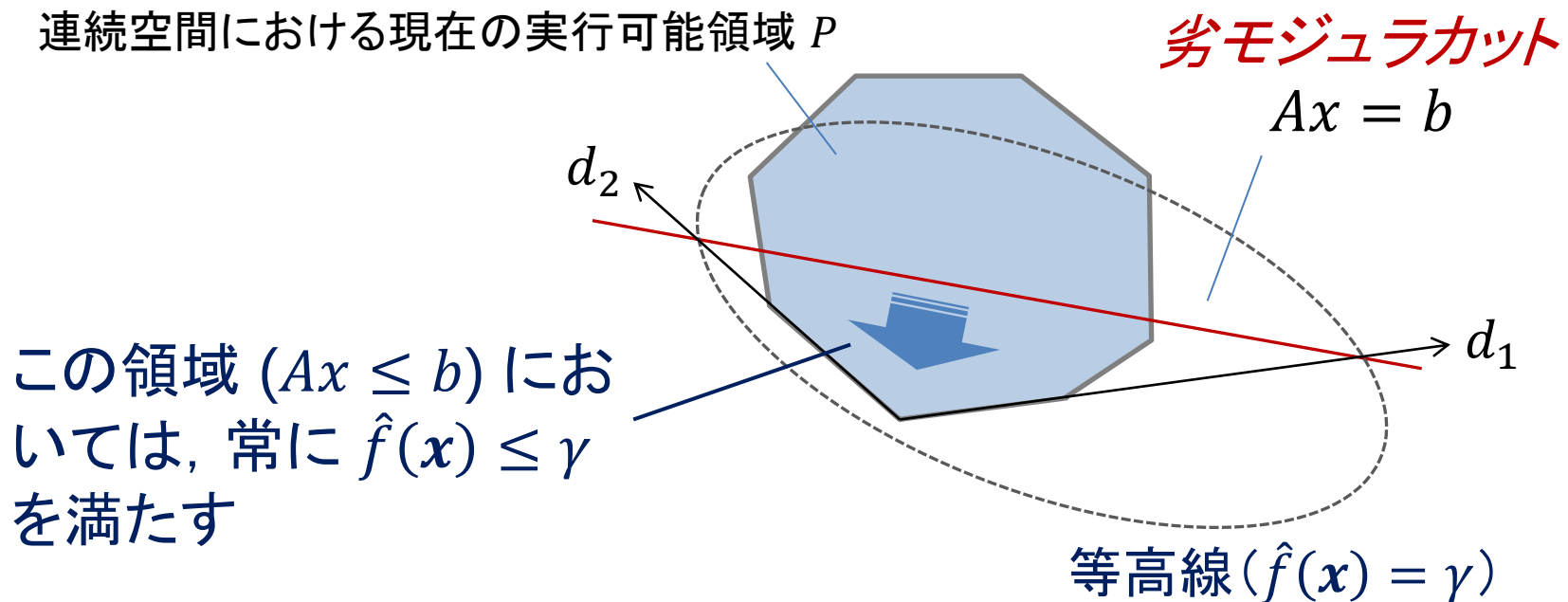
- ▶ 現在, 最良値 γ と探索域 $D(S)$ を知っているとする
- ▶ 劣モジュラ最大解の連続への緩和を考える:



劣モジュラ・カット(3)

- ▶ 現在, 最良値 γ と探索域 $D(S)$ を知っているとする

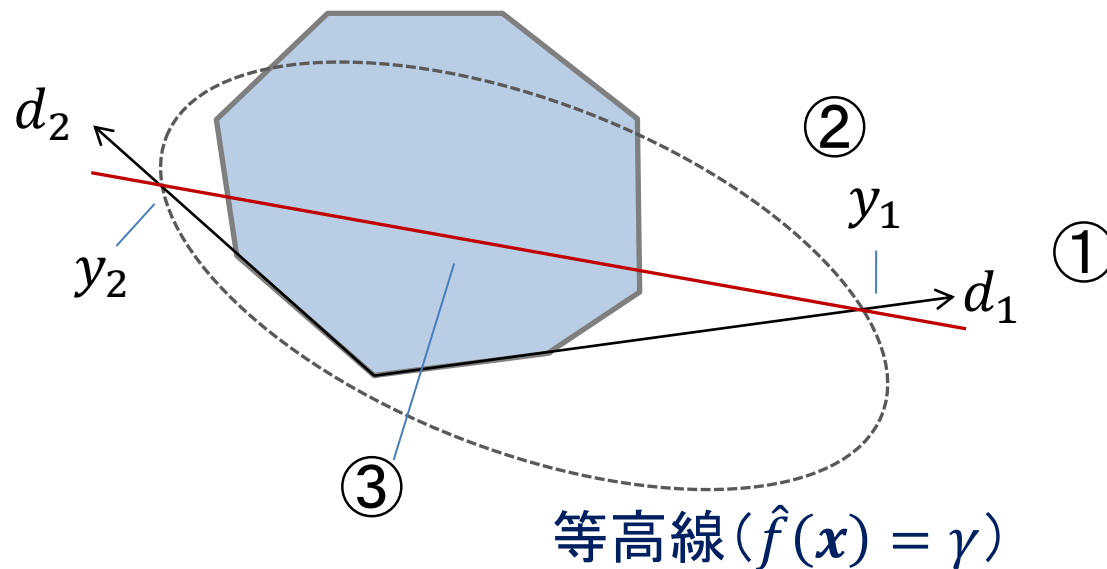
劣モジュラ・カット: Lovász拡張を通した凸性より, $\hat{f}(x) \leq \gamma$ for $Ax \leq b$ となる線形制約 (valid cut)



劣モジュラ・カットの構成

▶ 基本的な手順:

1. 一次独立な方向 $d_1, \dots, d_n$ を決める
2. 各方向 d_i と等高線との交点 y_i (γ -拡張) を計算する
3. $y_1, \dots, y_n$ を通る超平面を求める



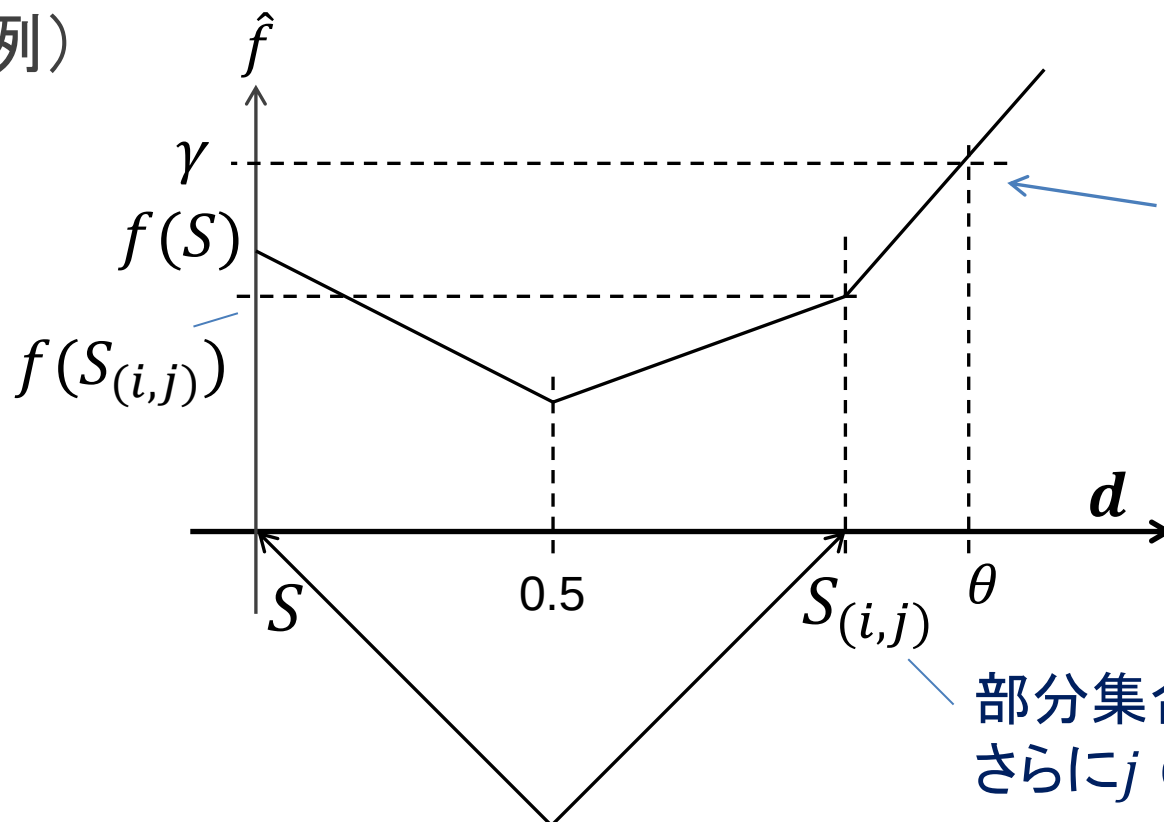
劣モジュラ・カットの構成 (γ 拡張の計算)

- ▶ γ -拡張: 方向 d に対して, 次式を満たす点

$$y = x + \theta \cdot d$$

⇒ Lovász 拡張に関しては, Closed に計算可能

例)

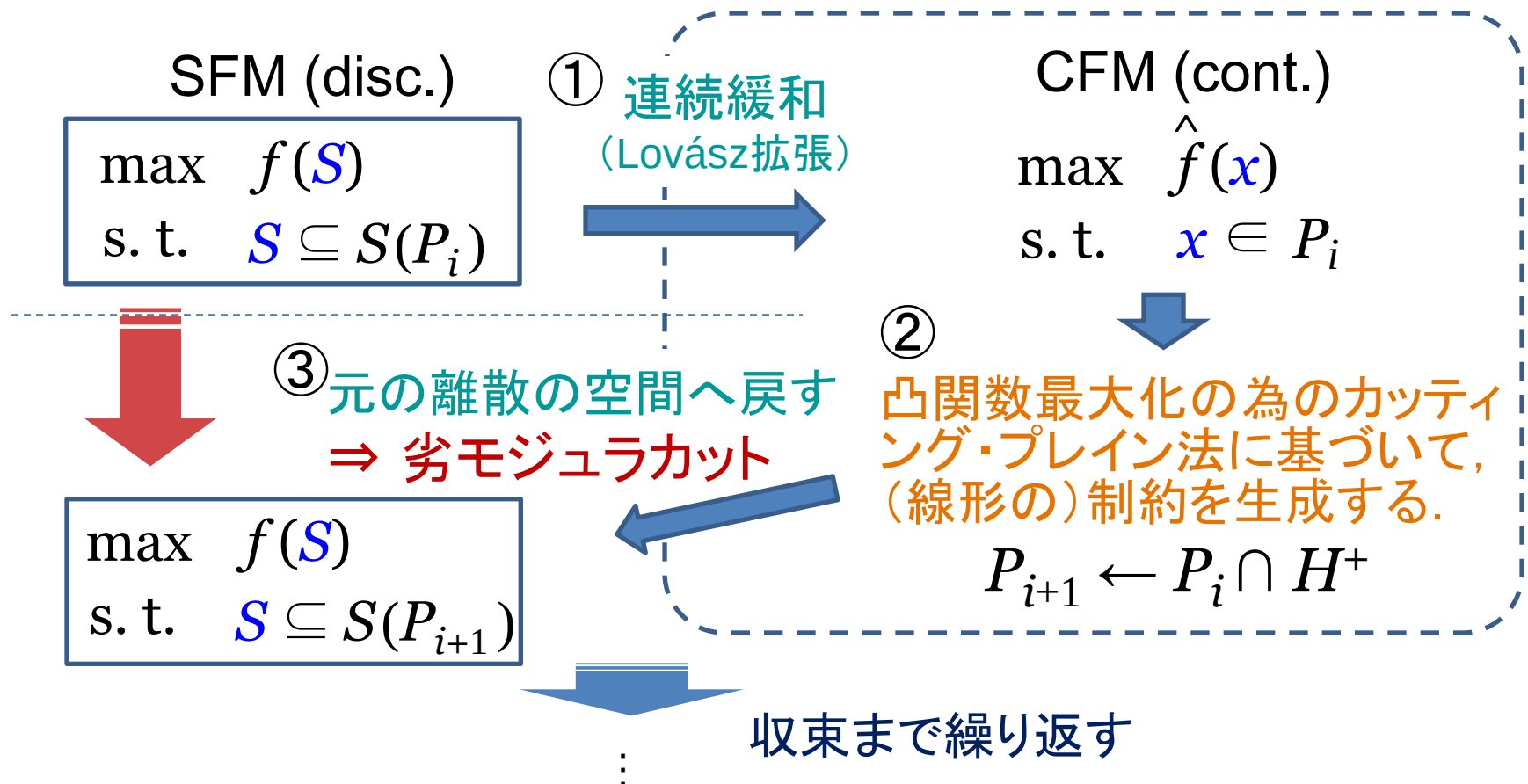


単に(傾きが既知の)
直線上で値が γ となる
点を探せばいい

部分集合 S から $i \in S$ を取り除き,
さらに $j \in V \setminus S$ を足した集合

Cutting-Plane法への適用

- 反復的に、劣モジュラカットを生成し、実行可能領域を絞っていく。最終的に大域解が得られる。



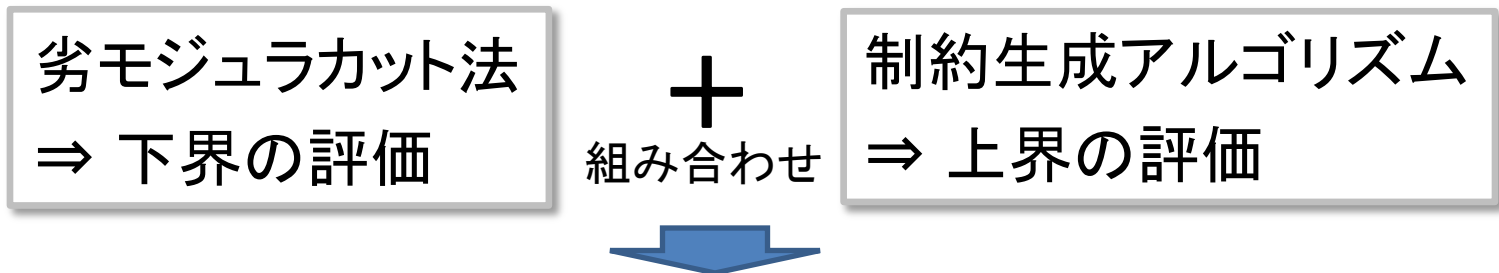
ε -近似解法の導入

- ▶ 劣モジュラカット法で厳密解を求めるには、それなりの計算時間がかかってしまう.

⇒「指定した $\varepsilon > 0$ に対し ε -近似解*を効率よく求めたい」



- ▶ 制約生成アルゴリズムを上界評価に利用:



現在の解の精度保証が可能

*) ここでは、 ε -近似解とは最適解からの誤差が ε 以内の解をいう

制約生成アルゴリズム

(Nemhauser&Wolsey,1981)

(元の問題)

$$\begin{array}{ll} \max_{S} & f(S) \\ \text{s. t.} & |S| \leq k \\ & S \subseteq V \end{array}$$

最適値 η^*



$$\begin{array}{ll} \max_{\eta, S} & \eta \\ \text{s. t.} & \eta \leq f(S) \\ & |S| \leq k \\ & S \subseteq V \end{array}$$

制約生成アルゴリズム:

- 小さい $Q \subseteq 2^V$ からスタート
- 各反復: $Q \subseteq 2^V$ を集合族として大きくしていく

等価



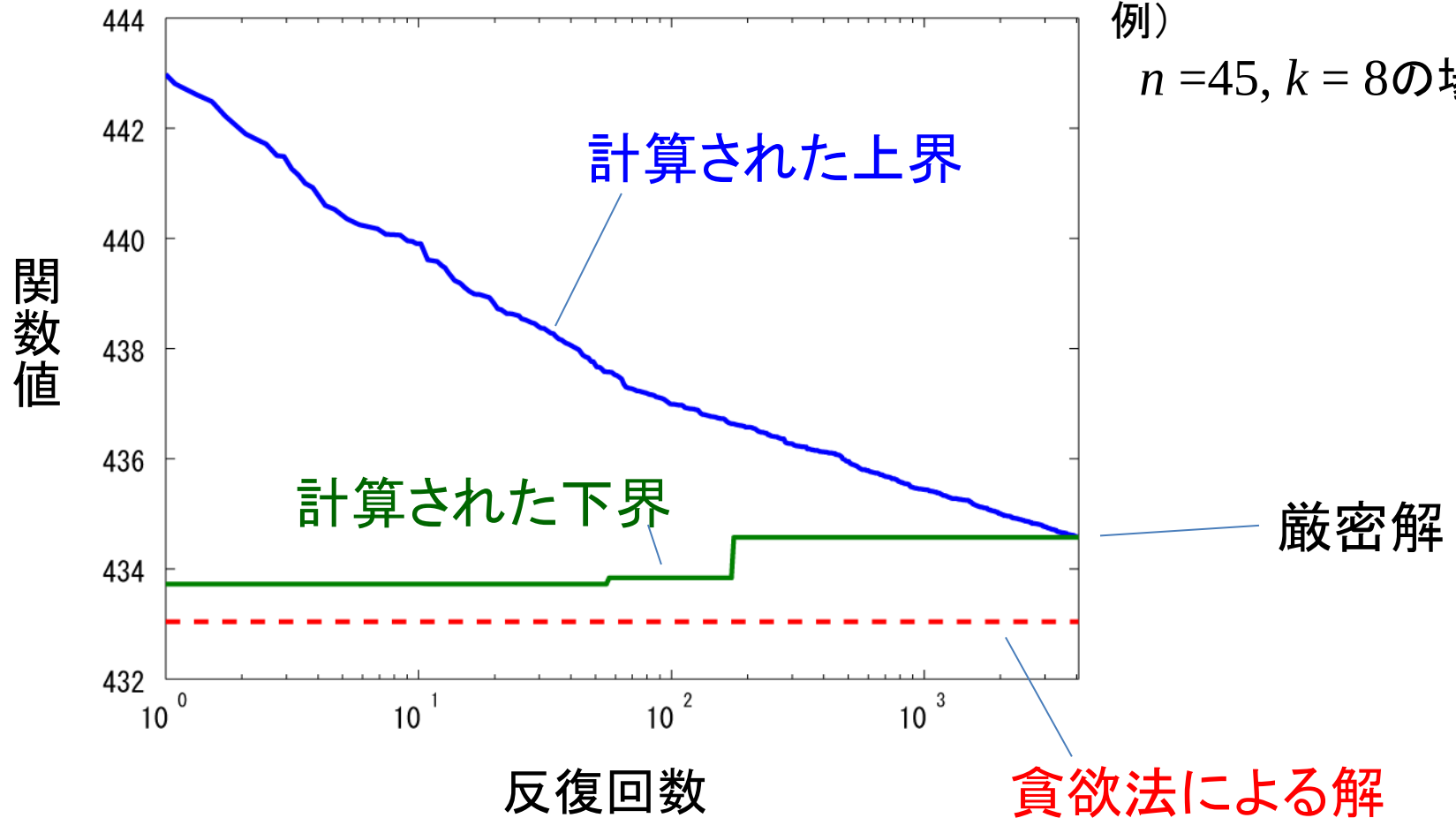
$$\begin{array}{ll} \max_{\eta, x} & \eta \\ \text{s. t.} & \eta \leq f(T) + \sum_{j \in V-T} \rho_j(T) x_j, \quad \forall T \in 2^V \\ & \sum_{j \in N} x_j \leq k \\ & x_j \in \{0,1\}, \quad \forall j \in V \end{array}$$

$$\rho_j(S) = f(S \cup \{j\}) - f(S)$$

人工データへの適用例(収束の様子)

例)

$n = 45, k = 8$ の場合



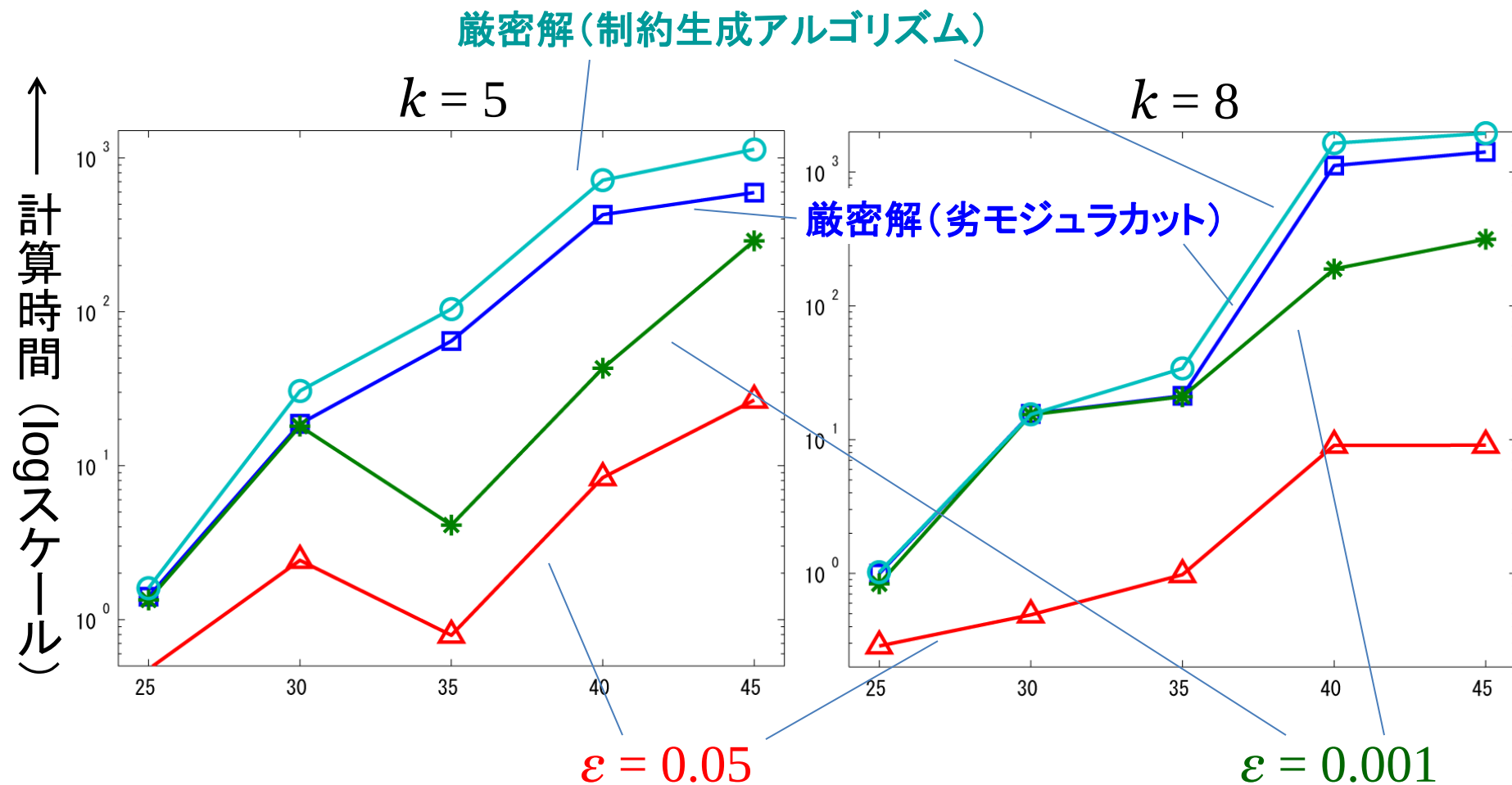
人工データへの適用例(説明)

▶ *K*-配置問題: 次の劣モジュラ関数の最大化

$$f(S) = \sum_{i=1}^m \max_{j \in S} c_{ij} \quad (C = \{c_{ij}\} \text{は非負行列})$$

- 複数の n に対して, ランダムに非負行列 C を生成し, 劣モジュラカット法, 及び, 従来手法(制約生成アルゴリズム)を適用.
- 劣モジュラカット法に関しては, 厳密解に加え, $\varepsilon = 0.05$, 0.001 の場合の近似解も計算.

人工データへの適用例(計算時間)



実データへの適用例(文書分類)(1)

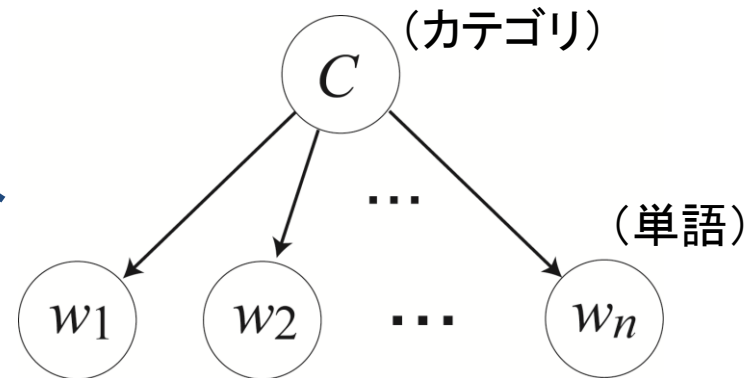
- ▶ 文書分類における重要単語の選択に対して, 劣モジュラカット法, 及び, 貪欲法を適用(Reuters-21587を利用)

(選択の基準) 単語を与えた事による情報利得関数

$$I(C;w) = H(C) - H(C|w)$$

⇒ 単語の生成プロセスが, 単純ベイズ仮定の下に行われるとすると, $I(C;w)$ は単語選択に関して劣モジュラ性を有する.

- ▶ トップ5つのカテゴリに関して, 両アルゴリズムを用いて2値分類として単語を選択して評価.



実データへの適用例(文書分類)(2)

(カテゴリ)		(選ばれた単語)	(利得)	(精度)
'earn'	Greedy	vs, dividend, qtr, split, writedown, payout, rev, frederick, nonperform, startup	0.615	0.834
	SubmoCut	profit, split, vs, qtr, earn, dividend, payabl, auditor, restat, rev	0.653	0.816
'acq'	Greedy	cyclop, twa, unsolicit, purol, chemlawn, tfb, ferruzzi, bor, spc, chalmer	0.187	0.788
	SubmoCut	cyclop, twa, unsolicit, purol, chemlawn, tfb, ferruzzi, bor, spc, chalmer	0.187	0.788
'money-fx'	Greedy	bank, dollar, sai, pct, currenc, blah, treasuri, new, two, prior	0.331	0.902
	SubmoCut	dlr, year, currenc, pct, bank, rate, record, sai, blah, dollar	0.335	0.924
'grain'	Greedy	wheat, grain, corn, maiz, rice, barlei, year, april, three, period	0.279	0.951
	SubmoCut	dlr, tonn, year, wheat, corn, washington, grain, agricultur, ct, offici	0.290	0.937
'crude'	Greedy	oil, dlr, sai, barrel, total, unit, two, today, y, year	0.269	0.952
	SubmoCut	mln, april, oil, pct, compani, report, today, state, blah, crude	0.271	0.953

発表のまとめ

- ▶ 特徴選択を例に，データ生成過程学習における厳密解探索法の動機と問題点，その解決における劣モジューラ性を持つ重要性について述べた．
- ▶ 特徴選択における損失関数および正則化項に現れる劣モジューラ性とその性質について述べた．
- ▶ 特徴選択などの多くの機械学習が含まれる劣モジューラ最大化のための厳密解法を紹介した．特に，劣モジューラカット法において，劣モジューラ性がアルゴリズム構築に有用である事について述べた．

謝辞

- ▶ 本発表に関連して、次の方々に感謝致します：
 - 永野清仁氏(東大), 津田宏治氏(AIST), Jeff A. Bilmes氏(Univ. of Washington), 岡本吉夫氏(JAIST), 岩田覚氏(京大), Stefanie Jegelka氏(MPI), Karsten M. Borgwardt氏(MPI), 鷲尾隆氏(阪大)
- ▶ 明日(11/4), 本発表に(少し?)関連したポスター発表があります.
「最小平均費用クラスタリング」(P4-30)
永野清仁(東大), 河原吉伸(阪大), 岩田覚(京大)

また本発表で紹介した研究の一部は、次のプログラムの下に実施されたものです：

- JSTさきがけ「知の創生と情報社会」, ERATO「湊離散構造処理系プロジェクト」, GCOE「計算世界観の深化と展開」, 科学研究費補助金(若手B・22700147)