IBIS2010

スパース正則化学習の学習性能，特にスパース性と汎化誤差の関係について

鈴木　大慈

東京大学

情報理工学系研究科

数理情報学専攻

2010年11月4日

冨岡亮太（東京大学），杉山将（東京工業大学）との共同研究

スパース性と汎化誤差の関係
どのような正則化が好ましい？

Multiple Kernel Learning (MKL)

↓

Elasticnet MKL

Lp-norm MKL

汎化誤差を理論的に解析

教師有りカーネル法

カーネル関数

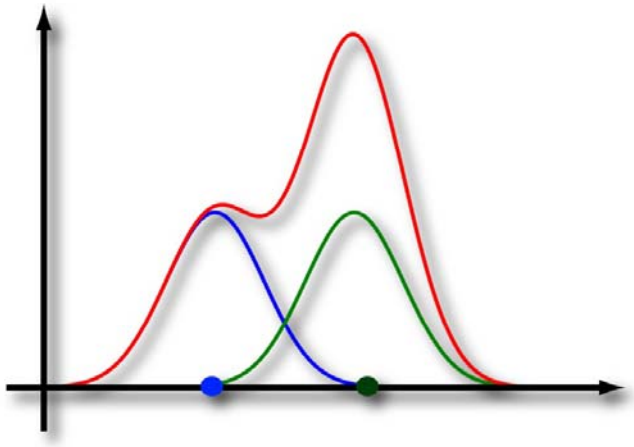
$$k(x, x')$$

（$\mathcal{H}_k$: 再生核ヒルベルト空間）

$$\min_{f \in \mathcal{H}_k} \frac{1}{n} \sum_{i=1}^{n} \ell(y_i, f(x_i)) + C \|f\|^2_{\mathcal{H}_k}$$

回帰, 判別: SVM, SVR, ….

$$f(x) = \sum_{i=1}^{N} \alpha_i k(x_i, x)$$



カーネル関数の例

- ガウシアン，多項式, カイ二乗, ….
  - パラメータ:ガウス幅, 多項式の次数，…

- 特徴量
  - Computer Vision:色, 勾配, sift (sift, hsvsift, huesift, scaling of sift), Geometric Blur, 画像領域の切り出し, . . .

⇒ **MKL: カーネルを選択して統合**

MKL: Multiple Kernel Learning

[Lanckriet et al. 2004]

$\{k_m\}_{m=1}^M$：M個のカーネル関数

$$\min_{f_m \in \mathcal{H}_m} \frac{1}{n}\sum_{i=1}^{n} \ell\left(y_i, \sum_{m=1}^{M} f_m(x_i)\right) + C\sum_{m=1}^{M} \|f_m\|_{\mathcal{H}_m}$$

$\mathcal{H}_m$：カーネル関数k_mに付随したRKHS

L1正則化：スパース

- Gourp Lassoの無限次元への拡張

[Bach, Lanchriet, Jordan:ICML2004]

カーネル重みとの関係

f : given

$$\min_{\{f_m\}_m} \left\{ \left(\sum_{m=1}^{M} \|f_m\|_{\mathcal{H}_m}^{p} \right)^{\frac{1}{p}} \;\middle|\; f = \sum_{m=1}^{M} f_m, f_m \in \mathcal{H}_m \right\}$$

Youngの不等式

$$= \min_{\{\lambda_m\}_m} \left\{ \|f\|_{\mathcal{H}_k} \;\middle|\; k = \sum_{m=1}^{M} \lambda_m k_m, \;\; \sum_{m=1}^{M} \lambda_m^{\frac{p}{2-p}} = 1, \;\; \lambda_m \geq 0 \right\}$$

[Micchelli & Pontil: JMLR2005]

$$\min_{f_m \in \mathcal{H}_m} \;\; \frac{1}{n} \sum_{i=1}^{n} \ell \left(y_i, \sum_{m=1}^{M} f_m(x_i) \right) + C \sum_{m=1}^{M} \|f_m\|_{\mathcal{H}_m}$$

$$\min_{\substack{k=\sum_{m=1}^{M} \lambda_m k_m, \\ \|\lambda\|_{\ell_1}=1, \lambda_m \geq 0}} \left\{ \min_{f \in \mathcal{H}_k} \;\; \frac{1}{n} \sum_{i=1}^{n} \ell \left(y_i, f(x_i) \right) + C \|f\|_{\mathcal{H}_k} \right\}$$

kはk_mらの凸結合

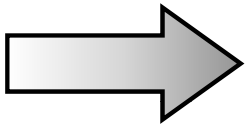
目的関数をカーネル関数の凸結合の中で最小化

カーネル重み: L_2

L_1 (MKL)

$$\min_{f_m \in \mathcal{H}_m} \quad L\left(\sum_{m=1}^{M} f_m\right) + C \sum_{m=1}^{M} \|f_m\|_{\mathcal{H}_m}$$

スパース

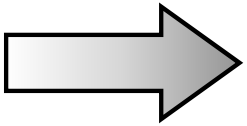


$$\min_{k=\sum_{m=1}^{M} \lambda_m k_m,\ \|\lambda\|_{\ell_1}=1, \lambda_m \geq 0} \left\{ \min_{f \in \mathcal{H}_k} \quad L(f) + C\|f\|_{\mathcal{H}_k} \right\}$$

L_2 (Uniform)

$$\min_{f_m \in \mathcal{H}_m} \quad L\left(\sum_{m=1}^{M} f_m\right) + C \sum_{m=1}^{M} \|f_m\|^2_{\mathcal{H}_m}$$

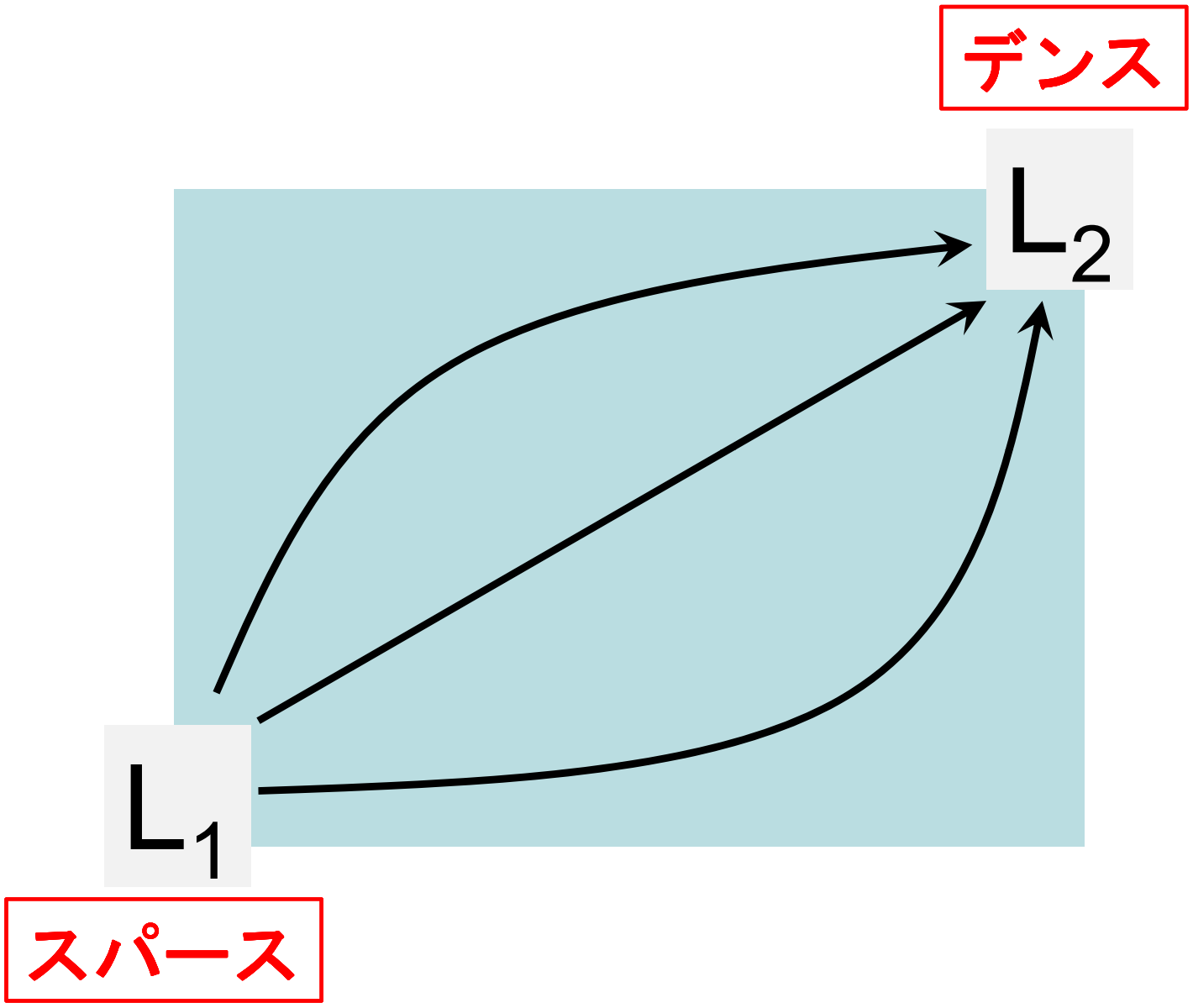
デンス



$$\min_{f \in \mathcal{H}_k} \quad L(f) + C\|f\|^2_{\mathcal{H}_k}$$

結構良い性能

$k = \sum_{m=1}^{M} k_m$: 単なる一様重みでの重ね合わせ


デンス
L_2
L_1
スパース

L_1とL_2の橋渡し

Elasticnet MKL

[Shawe-Taylor: NIPS workshop 2008,
Tomioka & Suzuki: NIPS workshop 2009]
cf. elastic-net: [Zou & Hastie: JRSS, 2005]

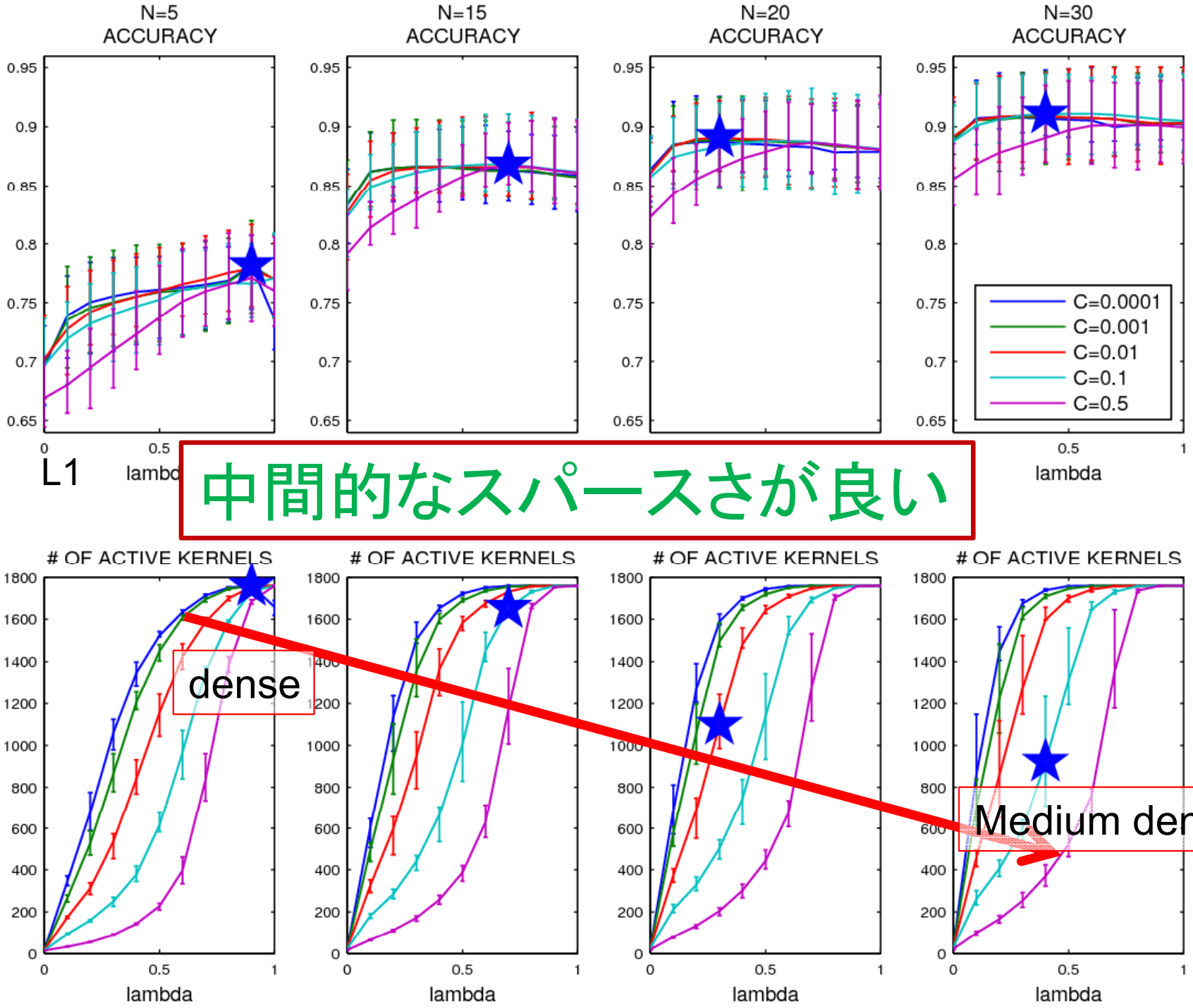
$$\min_{f_m \in \mathcal{H}_m} \frac{1}{n}\sum_{i=1}^{n} \ell\left(y_i, \sum_{m=1}^{M} f_m(x_i)\right) + C_1 \sum_{m=1}^{M} \|f_m\|_{\mathcal{H}_m} + C_2 \sum_{m=1}^{M} \|f_m\|_{\mathcal{H}_m}^2$$

Lp-norm MKL (1≦p≦2)

[Marius et al.: NIPS2009]

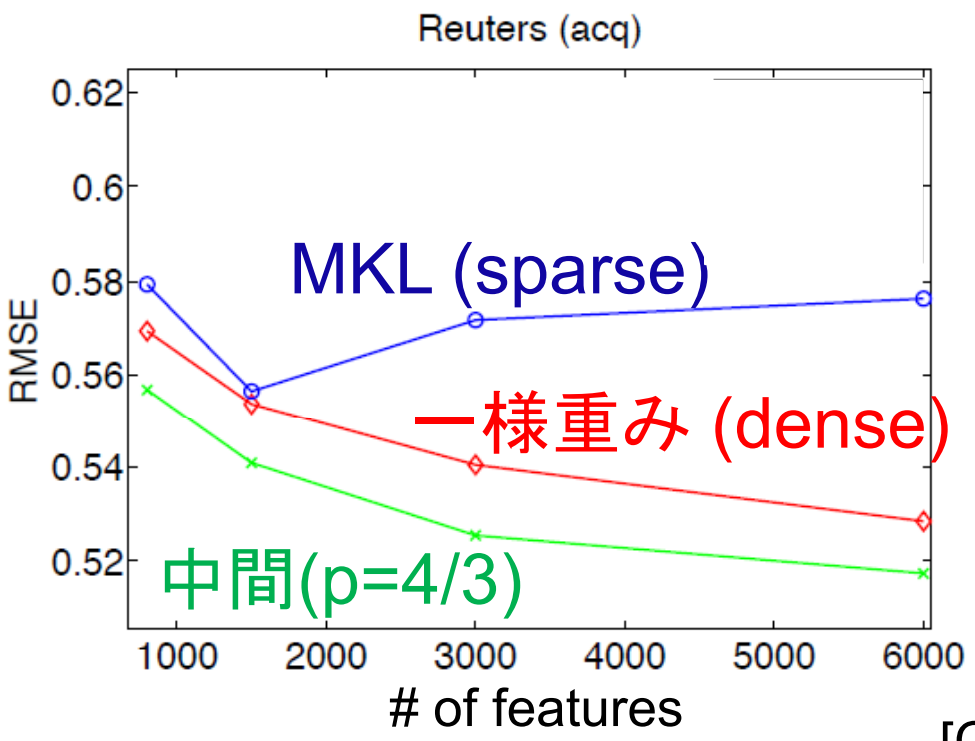
$$\min_{f_m \in \mathcal{H}_m} \frac{1}{n}\sum_{i=1}^{n} \ell\left(y_i, \sum_{m=1}^{M} f_m(x_i)\right) + C \sum_{m=1}^{M} \|f_m\|_{\mathcal{H}_m}^p$$

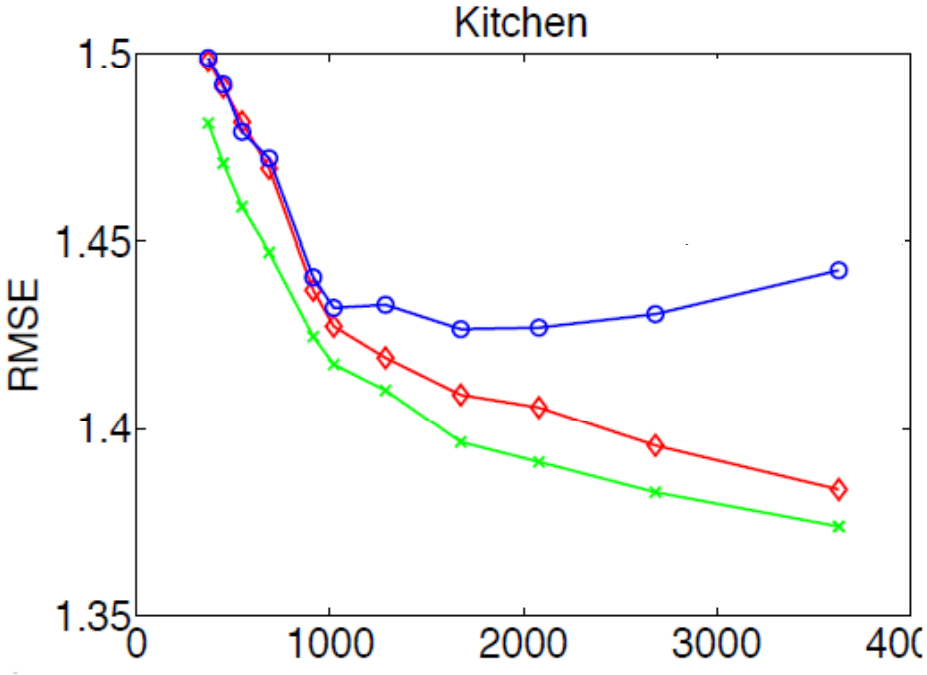
Elasticnet MKL: caltech 101 dataset



[Tomioka & Suzuki: NIPS 2009 Workshop]

Lp-norm MKL





[Cortes, Mohri, and Rostamizadeh: UAI 2009]

ここまでのまとめ

- L_1(MKL)とL_2(一様重み)の中間的なスパースさ
  → elasticnet/Lp-norm MKL
  → 実験的に性能○

実は,

- 計算量も少なくてすむ(後述)

- なぜ，性能が良いのか？
- どのような条件のとき，中間的スパースさが良いのか？

以後，主にElasticnet MKLを扱う.

概要

- 導入
- 効率的計算法
- 漸近的汎化誤差の解析
  - Elasticnet MKLの収束レート
    - 真がスパースな状況
    - 真がスパースでない状況
  - Lp-norm MKLの収束レート

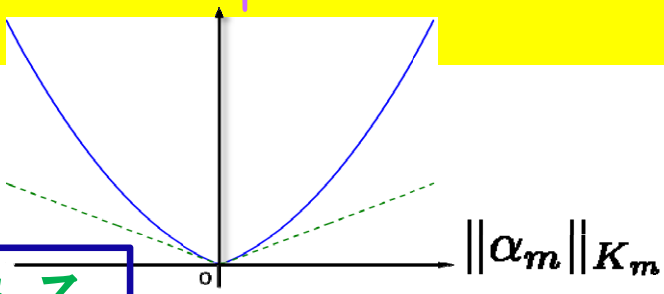
効率的計算法

双対問題

表現定理: $f_m(x_i) = (K_m \alpha_m)_i$

$$(K_m)_{i,j} := k_m(x_i, x_j) \quad \|\alpha_m\|_{K_m} := \sqrt{\alpha_m^T K_m \alpha_m}$$

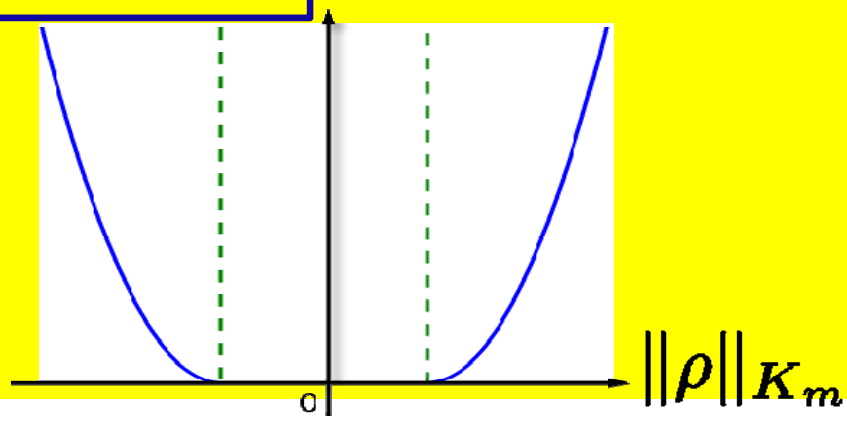
$$\min_{\alpha_m \in \mathbb{R}^n} \quad \phi_\ell \left(\sum_{m=1}^{M} K_m \alpha_m \right) + \sum_{m=1}^{M} \left(C_1 \|\alpha_m\|_{K_m} + C_2 \|\alpha_m\|_{K_m}^2 \right)$$



Fenchel双対

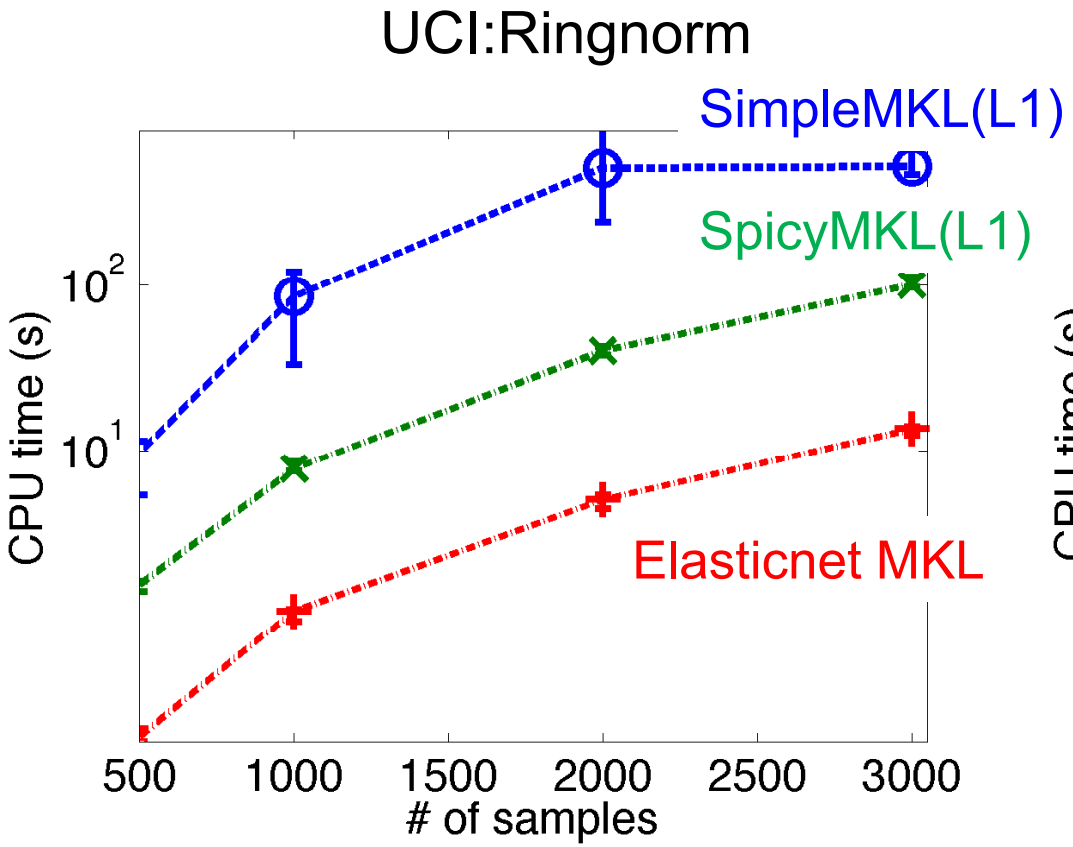
降下法（Newton法など）が使える

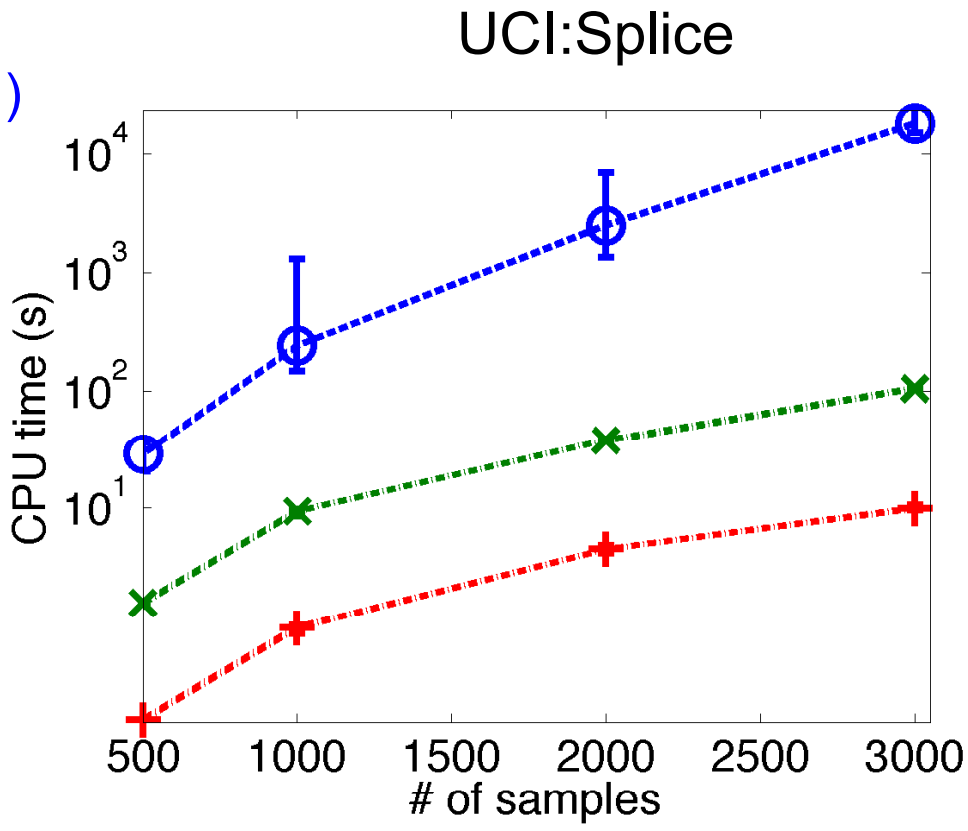
$$\max_{\rho \in \mathbb{R}^n} \quad -\phi_\ell^* (-\rho) - \sum_{m=1}^{M}$$



なめらか！

数値実験


UCI:Ringnorm
SimpleMKL(L1)
SpicyMKL(L1)
Elasticnet MKL
CPU time (s)
10^2
10^1
500
1000
1500
2000
2500
3000
# of samples


UCI:Splice
CPU time (s)
10^4
10^3
10^2
10^1
500
1000
1500
2000
2500
3000
# of samples

概要

- 導入
- 効率的計算法
- 漸近的汎化誤差の解析
  - Elasticnet MKLの収束レート
    - 真がスパースな状況
    - 真がスパースでない状況
  - Lp-norm MKLの収束レート

漸近的汎化誤差の解析

これからは二乗ロス（回帰）を想定：$\ell(y, f) = (y - f)^2$

スパース学習の収束レート

- Lasso & Dantzig Selector
  - Candes & Tao: AS2007 (Dantzig selector)
  - Bunea, Tsybakov & Wegkamp: AS2007 (Lasso)
  - Meinshausen & Yu: AS2009 (Lasso)
  - Bickel, Ritov & Tsybakov: AS2009 (Dantzig&Lasso)

$$\|\hat{\beta} - \beta^*\|_{\ell_2}^2 = O_p\left(\frac{d\log(M)}{n}\right)$$

  - Raskutti, Wainwright & Yu: arXiv:0910.2042, 2009.

mini-max レート $\min_{\hat{\beta}} \max_{\beta^*} \mathrm{E}\|\hat{\beta} - \beta^*\|_{\ell_2}^2 \geq C\left(\frac{d\log(M/d)}{n}\right)$

MKLに関する既存の結果

- L_1-MKL
  - Koltchinskii & Yuan: COLT2008

$$O_p\left(d^{\frac{1-s}{1+s}}n^{-\frac{1}{1+s}}+\frac{d\log(M)}{n}\right)$$

タイトではない

- Minimax-レート
  - Raskutti, Wainwright & Yu: NIPS2009

$$O_p\left(dn^{-\frac{1}{1+s}}+\frac{d\log(M/d)}{n}\right)$$

- Elasticnet型正則化
  - Meier, van de Geer & Bühlmann: AS2009

• Sobolev空間

$$O_p\left(d\left(\frac{\log(M)}{n}\right)^{\frac{1}{1+s}}\right)$$

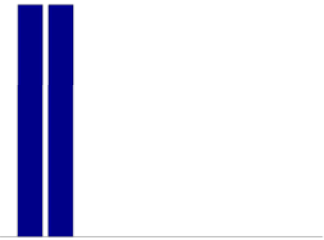
$$f^* = \sum_{m=1}^{M} f_m^* \quad : \text{真の関数}$$

$$I_0 := \{m \mid \|f_m^*\|_{\mathcal{H}_m} \neq 0\}$$

$$d := |I_0|$$

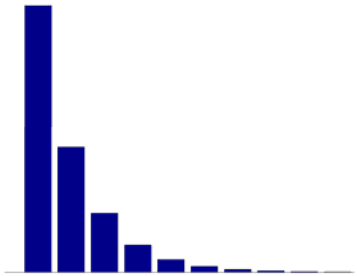
- 収束レートの導出：二つの状況設定
  - 真が厳密にスパース

$$\|f_m^*\|_{\mathcal{H}_m} > 0 \quad (m \in I_0),$$
$$\|f_m^*\|_{\mathcal{H}_m} = 0 \quad (m \in I_0^c)$$



  - 真が大体スパース

$$\|f_m^*\|_{\mathcal{H}_m} \leq C m^{-\beta}$$

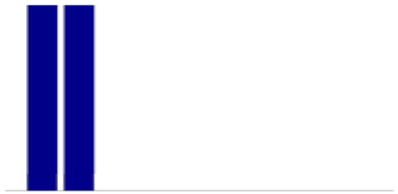


dやMはサンプル数に応じて増えてもよい.

スパースな状況

$$I_0 = \{m \mid \|f_m^*\|_{\mathcal{H}_m} > 0\}$$

$$\|f_m^*\|_{\mathcal{H}_m} > 0 \quad (m \in I_0),$$
$$\|f_m^*\|_{\mathcal{H}_m} = 0 \quad (m \in I_0^c)$$



- RKHSの複雑さの仮定（s, スペクトルの条件）

s :RKHSの複雑さを表わす定数
s が大きい → 複雑
s が小さい → 単純

Mercerの定理

$$k_m(x,x') = \sum_{\ell=1}^{\infty} \mu_{\ell,m}\phi_{\ell,m}(x)\phi_{\ell,m}(x')$$

$\{\phi_{\ell,m}\}_{\ell=1}^{\infty}$: ONS in $L_2(P_X)$

ある実数 s が存在して

$$\mu_{\ell,m} \leq C\ell^{-\frac{1}{s}} \qquad (\forall \ell, m)$$

covering numberとも関係あり，[Steinwart, Hush & Scovel: COLT2009]

- **真の関数の積分表現の仮定**

真はある関数と積分核とのたたみこみで，十分なめらか．

ある $g_m^* \in \mathcal{H}_m$ が存在して，

$$f_m^*(x) = \int_{\mathcal{X}} k_m^{(1/2)}(x, x') g_m^*(x') \mathrm{d}P(x') \qquad (\forall m = 1, \ldots, M),$$

ただし $k_m^{(1/2)}(x, x') = \sum_{\ell=1}^{\infty} \mu_{\ell,m}^{1/2} \phi_{\ell,m}(x) \phi_{\ell,m}(x')$.

$$\Sigma_{m,m} : \mathcal{H}_m \to \mathcal{H}_m$$

$$\langle f, \Sigma_{m,m} g \rangle := \mathrm{E}_X[f(X)g(X)]$$

$$f_m^* = \Sigma_{m,m}^{1/2} g_m^*$$

[Koltchinskii & Yuan: COLT2008, Bach: JMLR2008]

$$\min_{f_m \in \mathcal{H}_m} \frac{1}{n}\sum_{i=1}^{n} \ell\left(y_i, \sum_{m=1}^{M} f_m(x_i)\right) + C_1 \sum_{m=1}^{M} \|f_m\|_{\mathcal{H}_m} + C_2 \sum_{m=1}^{M} \|f_m\|_{\mathcal{H}_m}^2$$

定理：Elasticnet MKLの収束レート

前述の仮定と次の二つの条件が満たされているとする：

- $\|f_m^*\|_{\mathcal{H}_m} \leq R \ (\forall m \in I_0)$
- あるC¤ ĸが存在して $\|g_m^*\|_{\mathcal{H}_m} / \|f_m^*\|_{\mathcal{H}_m} < C \ (\forall m \in I_0)$

ちょっと強い仮定

すると，$C_1 = C_2 = K \max\left\{n^{-\frac{1}{2+s}}, \frac{F\log(Mn)}{\sqrt{n}}\right\}$ のとき，

$$\|\hat{f} - f^*\|_{L_2(P)}^2 = O_p\left(dn^{-\frac{2}{2+s}} + \frac{d\log(M)}{n}\right).$$

mini-max レートを達成

$$\min_{f_m \in \mathcal{H}_m} \frac{1}{n}\sum_{i=1}^{n} \ell\left(y_i, \sum_{m=1}^{M} f_m(x_i)\right) + C_1 \sum_{m=1}^{M} \|f_m\|_{\mathcal{H}_m}$$

定理：L_1-MKLの収束レート

より強い仮定が必要

前述の仮定と次の二つの条件が満たされているとする：

- $\sum_{m=1}^{M} \|f_m^*\|_{\mathcal{H}_m} \leq R$ ←dによらない：L_1では真が大きくなれない
- あるC¤ κが存在して $\|g_m^*\|_{\mathcal{H}_m} / \|f_m^*\|_{\mathcal{H}_m} < C \ (\forall m \in I_0)$

すると，$C_1 = K \max\left\{n^{-\frac{1}{2+s}}, \frac{F\log(Mn)}{\sqrt{n}}\right\}$のとき，

$$\|\hat{f} - f^*\|_{L_2(P)}^2 = O_p\left(dn^{-\frac{2}{2+s}} + \frac{\log(M)}{n}\right).$$

これはKoltchinskii & Yuan (COLT2008)による結果を改善:

$$O_p\left(d^{\frac{1-s}{1+s}} n^{-\frac{1}{1+s}} + \frac{d\log(M)}{n}\right)$$

その他の仮定

- Incoherence条件

[Koltchinskii & Yuan:COLT2008,
Meier, van de Geer & Bühlmann:AS2009]

ある定数 $Hs< C$ が存在して

$$0 < C^{-1} < \kappa(I_0)(1 - \rho^2(I_0))$$

$$\kappa(I) := \sup \left\{ \kappa \geq 0 \mid \kappa \leq \frac{\| \sum_{m \in I} f_m \|^2_{L_2(P)}}{\sum_{m \in I} \|f_m\|^2_{L_2(P)}}, \ \forall f_m \in \mathcal{H}_m \ (m \in I) \right\}$$

$$\rho(I) := \sup \left\{ \frac{\langle f_I, g_{I^c} \rangle_{L_2(P)}}{\|f_I\|_{L_2(P)} \|g_{I^c}\|_{L_2(P)}} \mid f_I \in \mathcal{H}_I, g_{I^c} \in \mathcal{H}_{I^c}, f_I \neq 0, g_{I^c} \neq 0 \right\}$$

非ゼロ成分はその他の成分で代用できず，かつ解の一意性が保証されている．

- カーネルは有界

$$\sup_x |k_m(x,x)| \leq 1$$

$$f_m(x) \leq \langle f_m, \phi_m(x) \rangle_{\mathcal{H}_m} \leq \|f_m\|_{\mathcal{H}_m} \sqrt{\langle \phi_m(x), \phi_m(x) \rangle_{\mathcal{H}_m}}$$
$$\leq \|f_m\|_{\mathcal{H}_m} \sqrt{k_m(x,x)} \leq \|f_m\|_{\mathcal{H}_m}$$

実は,

- $f_m^*(x) = \int_{\mathcal{X}} k_m^{(1/2)}(x, x') g_m^*(x') \mathrm{d}P(x')$
- $\frac{\|g_m^*\|_{\mathcal{H}_m}}{\|f_m^*\|_{\mathcal{H}_m}} < C$

といった強い条件は取り除ける.

正則化項の変換

Elasticnet MKL

$$\min_{f_m \in \mathcal{H}_m} \frac{1}{n}\sum_{i=1}^{n} \ell\left(y_i, \sum_{m=1}^{M} f_m(x_i)\right) + C_1 \sum_{m=1}^{M} \|f_m\|_{\mathcal{H}_m} + C_2 \sum_{m=1}^{M} \|f_m\|_{\mathcal{H}_m}^2$$

cf. Meier, van de Geer & Bühlmann: AS2009.

$$\min_{f_m \in \mathcal{H}_m} \frac{1}{n}\sum_{i=1}^{n} \ell\left(y_i, \sum_{m=1}^{M} f_m(x_i)\right) + C_1 \sum_{m=1}^{M} \sqrt{\|f_m\|_n^2 + C_2\|f_m\|_{\mathcal{H}_m}^2} + C_3 \sum_{m=1}^{M} \|f_m\|_{\mathcal{H}_m}^2$$

$$\|f_m\|_n^2 := \frac{1}{n}\sum_{i=1}^{n} f_m(x_i)^2$$

- より弱い条件で速い収束レートを実現
- 最適化は難しい

- 真の関数の積分表現の仮定 (q)

q: 真の滑らかさを表わす定数（大きいほどなめらか）

先の仮定はq=1を仮定していることに相当

$$\Sigma_{m,m} : \mathcal{H}_m \to \mathcal{H}_m$$
$$\langle f, \Sigma_{m,m} g \rangle := \mathrm{E}_X[f(X)g(X)]$$

ある実数 $0 \le q \le 1$ と $g_m^* \in \mathcal{H}_m$ が存在して

$$f_m^* = \Sigma_{m,m}^{q/2} g_m^*$$

$$f_m^*(x) = \int_{\mathcal{X}} k_m^{(q/2)}(x, x') g_m^*(x') \mathrm{d}P(x') \qquad (\forall m = 1, \ldots, M),$$

ただし $k_m^{(q/2)}(x, x') = \sum_{\ell=1}^{\infty} \mu_{\ell,m}^{q/2} \phi_{\ell,m}(x) \phi_{\ell,m}(x')$.

[Caponnetto & de Vito: FCM2007]では単一カーネルの場合を考察.

定理

$$\min_{f_m \in \mathcal{H}_m} \frac{1}{n}\sum_{i=1}^{n} \ell\left(y_i, \sum_{m=1}^{M} f_m(x_i)\right) + C_1 \sum_{m=1}^{M} \sqrt{\|f_m\|_n^2 + C_2\|f_m\|_{\mathcal{H}_m}^2} + C_3 \sum_{m=1}^{M} \|f_m\|_{\mathcal{H}_m}^2$$

前述の仮定と次の条件が満たされているとする:

- $\|g_m^*\|_{\mathcal{H}_m} \leq R \ (\forall m \in I_0)$

すると $\lambda_n = K \max\left\{ n^{-\frac{1}{1+q+s}}, \left(\frac{\log(M)}{n}\right)^{\frac{1}{1+q}} \right\}$ に対し,

$C_1 = \lambda_n^{\frac{1+q}{2}}, C_2 = C_3 = \lambda_n$ の時,

$$\|\hat{f} - f^*\|_{L_2(P)}^2 = O_p\left(dn^{-\frac{1+q}{1+q+s}} + \frac{d\log(M)}{n}\right)$$

非ゼロ要素がd個の場合のmini-maxレート

既存のバウンドとの比較

我々の結果: $\|\hat{f} - f^*\|^2_{L_2(P)} = O_p\left(dn^{-\frac{1+q}{1+q+s}} + \frac{d\log(M)}{n}\right)$

- Meier, van de Geer & Bühlmann: AS2009

  qI Hの時,

$$\|\hat{f} - f\|^2_{L_2(P)} = O_p\left(d\left(\frac{\log(M)}{n}\right)^{\frac{1}{1+s}}\right).$$

- 我々のバウンドは彼らのバウンドを改善
- qI Hを$0 \leq q \leq 1$へ拡張
- 一般のRKHSで議論

有限次元(Lasso)との比較

- Lassoはmini-maxレートを達成
- L_1-MKLは難しそう

  理由：無限次元ではL_2ノルムとRKHSノルムが同値でない.

$$c_1\|f_m\|_{\mathcal{H}_m} \not\leq \|f_m\|_{L_2(P)} \leq c_2\|f_m\|_{\mathcal{H}_m}$$

その他の仮定

- 無限大ノルムのバウンド

ある定数 $C \geq 1$ が存在して，$\forall f_m \in \mathcal{H}_m, \forall m$ に対し，

$$\|f_m\|_\infty \leq C \|f_m\|_{L_2(P)}^{1-s} \|f_m\|_{\mathcal{H}_m}^{s}$$

Sobolev空間や Gaussian RKHSはこの条件を満たす.

cf. [Mendelson and Neeman, 2008; Steinwart, Hush and Scovel, COLT2009]

- Incoherence条件

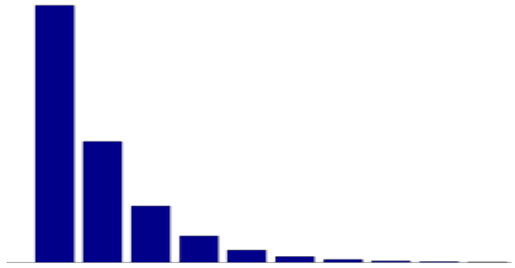
$$0 < C^{-1} < \kappa(I_0)(1 - \rho^2(I_0))$$

- カーネルは有界

$$\sup_x |k_m(x, x)| \leq 1$$

大体スパースな状況

$$\|f_m^*\|_{\mathcal{H}_m} \leq Cm^{-\beta} \quad (m = 1, \ldots, M)$$



$$\min_{f_m \in \mathcal{H}_m} \frac{1}{n} \sum_{i=1}^{n} \ell \left(y_i, \sum_{m=1}^{M} f_m(x_i) \right) + C_1 \sum_{m=1}^{M} \|f_m\|_{\mathcal{H}_m} + C_2 \sum_{m=1}^{M} \|f_m\|_{\mathcal{H}_m}^2$$

もとの正則化を考える

- 一般化Incoherence条件

b: 各RKHS間の相関の強さを表す定数

（小さいほど低い相関）

ある定数 $\mathrm{Hs} < C$ が存在して

$$|I|^{-b} < C\kappa(I)(1-\rho^2(I)) \quad (\forall I \subset \{1,\ldots,M\})$$

デザイン行列のIにおける最小個有値

$$\kappa(I) := \sup\left\{\kappa \geq 0 \mid \kappa \leq \frac{\|\sum_{m\in I} f_m\|^2_{L_2(P)}}{\sum_{m\in I}\|f_m\|^2_{L_2(P)}},\ \forall f_m \in \mathcal{H}_m\ (m\in I)\right\}$$

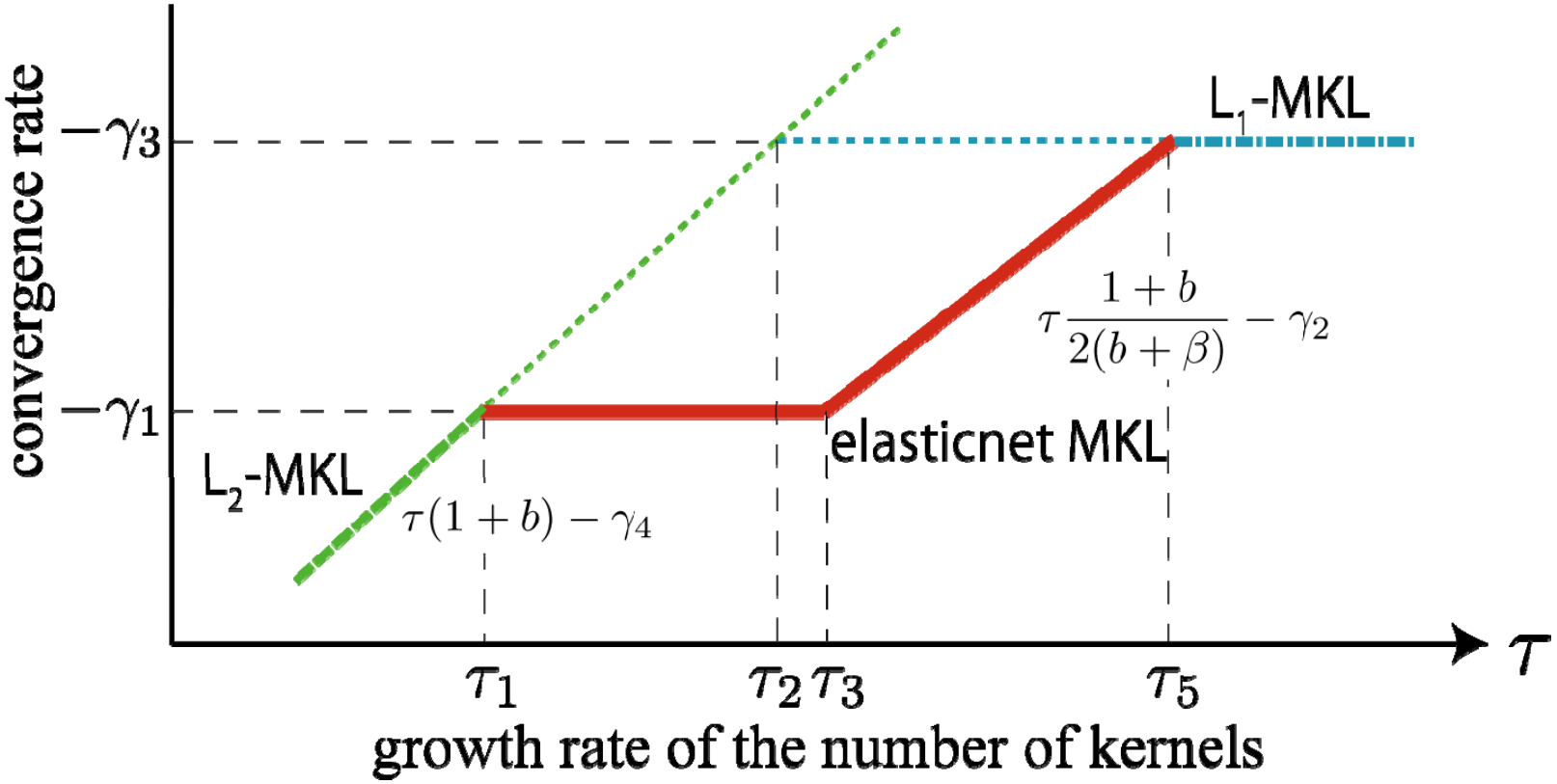
$$\rho(I) := \sup\left\{\frac{\langle f_I, g_{I^c}\rangle_{L_2(P)}}{\|f_I\|_{L_2(P)}\|g_{I^c}\|_{L_2(P)}} \mid f_I \in \mathcal{H}_I, g_{I^c} \in \mathcal{H}_{I^c}, f_I \neq 0, g_{I^c} \neq 0\right\}$$

I と I^cの相関

qI fを仮定し，カーネルの数はべき乗で増えるとする:

$$M = \lceil n^{\tau} \rceil.$$

$2\beta(s-1)+bs>1$の時，各手法の収束レートは
下の図のような関係になる:



$2\beta(s-1)+bs>1$: β が小さく (非スパース) で b が大 (相関が強い).

Lp-norm MKL

Lp-norm MKLの収束レート

$$\min_{f_m \in \mathcal{H}_m} \frac{1}{n} \sum_{i=1}^{n} \ell \left(y_i, \sum_{m=1}^{M} f_m(x_i) \right) + C_1 \sum_{m=1}^{M} \|f_m\|_{\mathcal{H}_m}^p$$

$$\|\hat{f} - f^*\|_{L_2(P)}^2 = O_p \left(n^{-\frac{1}{1+s}} M^{1-\frac{2s}{p(1+s)}} \left(\sum_{m=1}^{M} \|f_m^*\|_{\mathcal{H}_m}^p \right)^{\frac{2s}{p(1+s)}} \right)$$

pが大で大
pが小で小

pが大で小
pが小で大

トレードオフ

真のスパースさと推定量のスパースさの兼ね合い

まとめ

- L1とL2の中間的なスパースさ
  - 実験的に性能〇
  - 計算量も少なくて済む
- 理論的に性能の良さを考察
  - Elasticnet MKLは弱い条件で速い収束
  - 正則化項を変えることによりさらに条件を緩和
  - RKHS間の相関が強く，あまりスパースでないときにElasticnet MKLは良い性能
  - Lp-norm MKLは真のスパースさに応じて最適なノルムが変化