

Fast Convergence Rate of Multiple Kernel Learning with Elastic-Net Regularization

† 鈴木 大慈 † 富岡 亮太 ‡ 杉山 将

† 東京大学大学院
情報理工学系研究科

‡ 東京工業大学大学院
情報理工学研究科

2011 年 3 月 29 日
IBISML 研究会@大阪

Outline

- 1 Introduction
 - MKL とその拡張
 - 本研究の概要
- 2 Mixed-Norm-Elasticnet-MKL
 - 準備
 - Mixed-Elasticnet-MKL の収束レート
- 3 Mini-max レート
- 4 Conclusion

Outline

- 1 Introduction
 - MKL とその拡張
 - 本研究の概要
- 2 Mixed-Norm-Elasticnet-MKL
 - 準備
 - Mixed-Elasticnet-MKL の収束レート
- 3 Mini-max レート
- 4 Conclusion

教師有りカーネル法

- カーネル関数 再生核ヒルベルト空間 (RKHS)

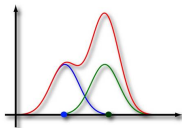
$$k(x, x') \Leftrightarrow \mathcal{H}_k$$

- 教師有り学習問題

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

- 表現定理

$$\exists \alpha_i \in \mathbb{R} \quad \text{s.t.} \quad \hat{f}(x) = \sum_{i=1}^n \alpha_i k(x_i, x)$$



カーネルの選択

カーネル法の良い点：データの構造をカーネルに詰め込める。

Challenge：どのようなカーネルを用いるか？

- ガウシアン, 多項式, カイ二乗, ...
- 沢山の特微量の候補

Multiple Kernel Learning：
凸最適化でカーネルを選択・統合

Multiple Kernel Learning

- Single Kernel Learning

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

- Multiple Kernel Learning (Lanckriet et al., 2004; Bach et al., 2004)

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

(\mathcal{H}_m : カーネル k_m に対応した RKHS)

- Group Lasso の無限次元への拡張
- スパースな解
- 表現定理により有限次元最適化で解ける (Sonnenburg et al., 2006; Rakotomamonjy et al., 2008; Suzuki & Tomioka, 2009)

様々な正則化

- L_1 -MKL (Lanckriet et al., 2004; Bach et al., 2004) : スパース

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

- L_2 -MKL : デンス

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

様々な正則化

- L_1 -MKL (Lanckriet et al., 2004; Bach et al., 2004) : スパース

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

- L_2 -MKL : デンス

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

- Elasticnet-MKL (Tomioka & Suzuki, 2009)

$$\min_{f_m \in \mathcal{H}_m} L \left(\sum_{m=1}^M f_m \right) + C_1 \sum_{m=1}^M \|f_m\|_{\mathcal{H}_m} + C_2 \sum_{m=1}^M \|f_m\|_{\mathcal{H}_m}^2$$

- Mixed-Norm-Elasticnet-MKL (Meier et al., 2009)

$$\min_{f_m \in \mathcal{H}_m} L \left(\sum_{m=1}^M f_m \right) + C_1 \sum_{m=1}^M \sqrt{\|f_m\|_n^2 + C_2 \|f_m\|_{\mathcal{H}_m}^2} + C_3 \sum_{m=1}^M \|f_m\|_{\mathcal{H}_m}^2$$

ただし, $\|f\|_n^2 := \frac{1}{n} \sum_{i=1}^n f(x_i)^2$.

様々な正則化

- L_1 -MKL (Lanckriet et al., 2004; Bach et al., 2004) : スパース

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

- L_2 -MKL : デンス

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

- Elasticnet-MKL (Tomioka & Suzuki, 2009)

$$\min_{f_m \in \mathcal{H}_m} L \left(\sum_{m=1}^M f_m \right) + C_1 \sum_{m=1}^M \|f_m\|_{\mathcal{H}_m} + C_2 \sum_{m=1}^M \|f_m\|_{\mathcal{H}_m}^2$$

- **Mixed-Norm-Elasticnet-MKL (Meier et al., 2009)** **本日のターゲット**

$$\min_{f_m \in \mathcal{H}_m} L \left(\sum_{m=1}^M f_m \right) + C_1 \sum_{m=1}^M \sqrt{\|f_m\|_n^2 + C_2 \|f_m\|_{\mathcal{H}_m}^2} + C_3 \sum_{m=1}^M \|f_m\|_{\mathcal{H}_m}^2$$

ただし, $\|f\|_n^2 := \frac{1}{n} \sum_{i=1}^n f(x_i)^2$.

本日のお題

Mixed-Norm-Elasticnet-MKL の汎化誤差を導出 .
既存のレートよりタイトなことを示す .

- これからは regression

$$L(f) = \frac{1}{n} \sum_{i=1}^n (f(x_i) - y_i)^2$$

を仮定 .

- 真の関数を

$$f^*(x) = \sum_{m=1}^M f_m^*(x) (= E[Y|x])$$

と書く .

既存の研究

$\|\hat{f} - f^*\|_{L_2}^2$ の収束レート, d は真の非ゼロ要素の数 $d = |\{m \mid \|f_m^*\|_{\mathcal{H}_m} \neq 0\}|$.

- L_1 -MKL (Koltchinskii & Yuan, 2008):

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

- Mixed-Norm-Elasticnet-MKL (Meier et al., 2009): mini-max ではない.

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

- Mixed-Norm- L_1 -MKL (Koltchinskii & Yuan, 2010): mini-max レート達成, 正則化項は $\sum_m (C_1 \|f_m\|_n + C_2 \|f_m\|_{\mathcal{H}_m})$

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

- Mini-max レート (Raskutti et al., 2009)

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

我々の結果（概要）

Mixed-Norm-Elasticnet-MKL の収束レート：

$$\|\hat{f} - f^*\|_{L_2}^2 = O_p \left(d^{\frac{1+q}{1+q+s}} n^{-\frac{1+q}{1+q+s}} R_2^{\frac{2s}{1+q+s}} + \frac{d \log(M)}{n} \right).$$

- 既存のレートよりタイト
- 真の関数 f^* の滑らかさ q を導入
- 真の関数 f^* の “ノルム” R_2 との関係を解明
- l_2 ボール上で mini-max 最適（既存のは l_∞ ボール上で最適）

既存の結果との関係

	滑らかさ (q)	最適性	収束レート
K&Y (2008)	$q = 1$?	$d^{\frac{1-s}{1+s}} n^{-\frac{1}{1+s}} + \frac{d \log(M)}{n}$
Meier et al. (2009)	$q = 0$	×	$d \left(\frac{\log(M)}{n} \right)^{\frac{1}{1+s}}$
K&Y (2010)	$q = 0$	l_∞ -ball	$dn^{-\frac{1}{1+s}} + \frac{d \log(M)}{n}$
IBIS2010	$0 \leq q \leq 1$	l_∞ -ball	$dn^{-\frac{1+q}{1+q+s}} + \frac{d \log(M)}{n}$
今回	$0 \leq q \leq 1$	l_2 -ball	$\left(\frac{d}{n}\right)^{\frac{1+q}{1+q+s}} R_2^{\frac{2s}{1+q+s}} + \frac{d \log(M)}{n}$

より速く, より一般的

Outline

- 1 Introduction
 - MKL とその拡張
 - 本研究の概要
- 2 Mixed-Norm-Elasticnet-MKL
 - 準備
 - Mixed-Elasticnet-MKL の収束レート
- 3 Mini-max レート
- 4 Conclusion

真がスパースであると仮定 .

$$l_0 := \{m \mid \|f_m^*\|_{\mathcal{H}_m} \neq 0\}$$

$$\|f_m^*\|_{\mathcal{H}_m} > 0 \quad (m \in l_0),$$

$$\|f_m^*\|_{\mathcal{H}_m} = 0 \quad (m \in l_0^c).$$

$d = |l_0|$ (真の非ゼロ要素の数) とおく .

Spectrum Condition (s)

$0 < s < 1$: モデルの複雑さを表わす.

Mercer の定理による分解 :

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

ただし, $\{\phi_{\ell, m}\}_{\ell=1}^{\infty}$ は $L_2(P)$ 内の ONS.

Spectrum Condition (s)

ある実数 $0 < s < 1$ が存在して,

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

s は RKHS の複雑さを表わす.

- s が大きいと複雑, s が小さいと単純

Proposition (Steinwart et al. (2009))

$$\mu_{\ell, m} \sim \ell^{-\frac{1}{s}} \Leftrightarrow N(B(\mathcal{H}_m), \epsilon, L_2(P)) \sim \epsilon^{-2s}$$

Convolution Condition (q)

$0 \leq q \leq 1$: 真 f^* の滑らかさを表わす .

$\Sigma_m : \mathcal{H}_m \rightarrow \mathcal{H}_m$ を $\langle f, \Sigma_m g \rangle_{\mathcal{H}_m} := E[f(X)g(X)]$ なるものと定義する .

Convolution Condition (q) (Caponnetto & de Vito, 2007)

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

$$f_m^* = \Sigma_m^{q/2} g_m^*$$

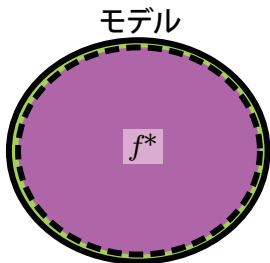
と表せる .

$k_m^{(q/2)}(x, x') := \sum_{\ell=1}^{\infty} \mu_{\ell, m}^{q/2} \phi_{\ell, m}(x) \phi_{\ell, m}(x')$ に対して ,

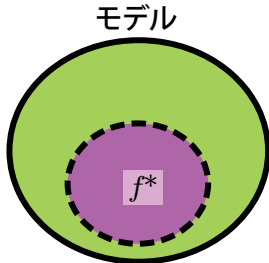
$$f_m^*(x) = \int k_m^{(q/2)}(x, x') g_m^*(x') dP(x'),$$

と書けることと同値 .

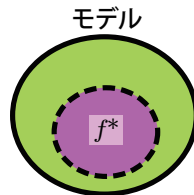
s と q の関係



(a) s 大, $q = 0$



(b) s 大, $q > 0$



(c) s 小, $q > 0$

Incoherence Condition

Incoherence Condition (Koltchinskii & Yuan, 2008; Meier et al., 2009)

ある定数 $0 < C$ が存在して,

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

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

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

I_0 の内側とも外側とも見分けがつく.

その他の条件

Basic Condition

- $E[Y|X] = f^*(X) = \sum_{m=1}^M f_m^*(X)$ であり, ノイズ $\epsilon := Y - f^*(X)$ は有界: $|\epsilon| \leq L$.
- $\sup_{X \in \mathcal{X}} |k_m(X, X)| \leq 1$ ($\forall m$).

∞ -norm Bound Condition

Spectrum Condition (s) と同時に次の不等式が満たされている:

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

Gaussian カーネルなど, Sobolev 空間に埋め込める空間はこれが成り立っている. Mendelson and Neeman (2010); Steinwart et al. (2009) で詳細な議論がされている.

我々の結果：Mixed-Norm-Elasticnet-MKL の収束レート

$$\min_{f_m \in \mathcal{H}_m} L \left(\sum_{m=1}^M f_m \right) + \lambda_1^{(n)} \sum_{m=1}^M \sqrt{\|f_m\|_n^2 + \lambda_2^{(n)} \|f_m\|_{\mathcal{H}_m}^2} + \lambda_3^{(n)} \sum_{m=1}^M \|f_m\|_{\mathcal{H}_m}^2.$$

Theorem

*Spectrum Condition (s), Convolution Condition (q), Incoherence Condition, Basic Condition, ∞ -norm Bound Condition*のもと，十分大きな n において，あるパラメータ $\lambda_1^{(n)}$, $\lambda_2^{(n)}$, $\lambda_3^{(n)}$ の値のもと，

$$\|\hat{f} - f^*\|_{L_2}^2 \leq C' \left(d^{\frac{1+q}{1+q+s}} n^{-\frac{1+q}{1+q+s}} R_{2,g^*}^{\frac{2s}{1+q+s}} + \frac{d \log(M)}{n} \right) \eta(t)^2,$$

が確率 $1 - e^{-\sqrt{nt}} - e^{-\sqrt{n}}$ ($\forall t \geq 1$) で成り立つ。

ただし $\eta(t) := \max(\sqrt{t}, t/\sqrt{n})$ であり， R_{2,g^*} を次のように定義する：

$$R_{2,g^*} := \left(\sum_{m=1}^M \|g_m^*\|_{\mathcal{H}_m}^2 \right)^{\frac{1}{2}}.$$

Bound の比較

$q = 0$ として具体的に比較

Koltchinskii and Yuan (2010) のレート : $dn^{-\frac{1}{1+s}} + \frac{d \log(M)}{n}$.

我々のレート : $d^{\frac{1+q}{1+q+s}} n^{-\frac{1+q}{1+q+s}} R_{2,g^*}^{\frac{2s}{1+q+s}} + \frac{d \log(M)}{n}$.

- ① $\|f_m^*\|_{\mathcal{H}_m} = 1$ ($m = 1, \dots, d$): 大きさ一様

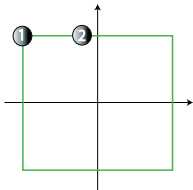
我々のレート : $dn^{-\frac{1}{1+s}} + \frac{d \log(M)}{n}$

Koltchinskii and Yuan (2010) と同じ .

- ② $\|f_m^*\|_{\mathcal{H}_m} = m^{-1}$ ($m = 1, \dots, d$): 大きさ急減衰

我々のレート : $d^{\frac{1}{1+s}} n^{-\frac{1}{1+s}} + \frac{d \log(M)}{n}$

Koltchinskii and Yuan (2010) より $d^{\frac{s}{1+s}}$ 倍だけ速い .



Outline

- 1 Introduction
 - MKL とその拡張
 - 本研究の概要
- 2 Mixed-Norm-Elasticnet-MKL
 - 準備
 - Mixed-Elasticnet-MKL の収束レート
- 3 Mini-max レート
- 4 Conclusion

Mini-max レート

Mini-max レート：どんな推定法も 超えられないレート。

$f_m^* = \sum_m^q g_m^*$ に注意する。

- ① $\left(\sum_{m=1}^M \|g_m^*\|_{\mathcal{H}_m}^2 \right)^{\frac{1}{2}} \leq R_2$ (g^* が半径 R_2 の ℓ_2 ボールに含まれる)

$$d^{\frac{1+q}{1+q+s}} n^{-\frac{1+q}{1+q+s}} R_2^{\frac{2s}{1+q+s}} + \frac{d \log(M/d)}{n}$$

我々のレートに一致。

- ② $\max_m \|g_m^*\|_{\mathcal{H}_m} \leq R_\infty$ (g^* が半径 R_∞ の ℓ_∞ ボールに含まれる)

$$dn^{-\frac{1+q}{1+q+s}} R_\infty^{\frac{2s}{1+q+s}} + \frac{d \log(M/d)}{n}$$

$q = 0, R_\infty = 1$ のとき, Koltchinskii and Yuan (2010) のレートに一致。

Outline

- 1 Introduction
 - MKL とその拡張
 - 本研究の概要
- 2 Mixed-Norm-Elasticnet-MKL
 - 準備
 - Mixed-Elasticnet-MKL の収束レート
- 3 Mini-max レート
- 4 Conclusion

Conclusion

- Mixed-Norm-Elasticnet-MKL の収束レートを導出 .
- 既存研究よりタイトなレートを導出 .
- 真 f^* の滑らかさ q を導入 .
- 導出されたレートは ℓ_2 ボール上の mini-max レートを達成 .

本研究のプレプリント (arXiv): <http://arxiv.org/abs/1103.0431>

slide: <http://www.simplex.t.u-tokyo.ac.jp/~s-taiji/data/IBISML2011.pdf>

- Bach, F., Lanckriet, G., & Jordan, M. (2004). Multiple kernel learning, conic duality, and the SMO algorithm. *the 21st International Conference on Machine Learning* (pp. 41–48).
- Caponnetto, A., & de Vito, E. (2007). Optimal rates for regularized least-squares algorithm. *Foundations of Computational Mathematics*, 7, 331–368.
- Koltchinskii, V., & Yuan, M. (2008). Sparse recovery in large ensembles of kernel machines. *Proceedings of the Annual Conference on Learning Theory* (pp. 229–238).
- Koltchinskii, V., & Yuan, M. (2010). Sparsity in multiple kernel learning. *The Annals of Statistics*, 38, 3660–3695.
- Lanckriet, G., Cristianini, N., Ghaoui, L. E., Bartlett, P., & Jordan, M. (2004). Learning the kernel matrix with semi-definite programming. *Journal of Machine Learning Research*, 5, 27–72.
- Meier, L., van de Geer, S., & Bühlmann, P. (2009). High-dimensional additive modeling. *The Annals of Statistics*, 37, 3779–3821.
- Mendelson, S., & Neeman, J. (2010). Regularization in kernel learning. *The Annals of Statistics*, 38, 526–565.
- Rakotomamonjy, A., Bach, F., Canu, S., & Y., G. (2008). SimpleMKL. *Journal of Machine Learning Research*, 9, 2491–2521.

- Raskutti, G., Wainwright, M., & Yu, B. (2009). Lower bounds on minimax rates for nonparametric regression with additive sparsity and smoothness. In *Advances in neural information processing systems 22*, 1563–1570. Cambridge, MA: MIT Press.
- Sonnenburg, S., Rätsch, G., Schäfer, C., & Schölkopf, B. (2006). Large scale multiple kernel learning. *Journal of Machine Learning Research*, 7, 1531–1565.
- Steinwart, I., Hush, D., & Scovel, C. (2009). Optimal rates for regularized least squares regression. *Proceedings of the Annual Conference on Learning Theory* (pp. 79–93).
- Suzuki, T., & Tomioka, R. (2009). SpicyMKL. arXiv:0909.5026.
- Tomioka, R., & Suzuki, T. (2009). Sparsity-accuracy trade-off in MKL. *NIPS 2009 Workshop:: Understanding Multiple Kernel Learning Methods*. Whistler. arXiv:1001.2615.