

# 高次元機械学習手法の統計的学習理論と計算理論

東京大学大学院情報理工学系研究科数理情報学専攻 鈴木 大慈

本講演では、複雑な構造を持つ高次元データを解析する機械学習手法の統計的学習理論およびその効率的計算法について紹介する。特に、構造を持つ高次元データ解析手法として、スパースカーネル加法モデルおよび低ランクテンソルモデル、構造的正則化学習法を主に扱う。

スパースカーネル加法モデルは、Lasso や Group Lasso を包含する広いクラスのモデルである。これは、各説明変数に非線形な関数を用いてそれらをすべての説明変数について足し合わせるカーネル加法モデルに、余分な説明変数を切り捨てるためにスパース性の構造を組み込んだモデルである。スパースカーネル加法モデルの推定には正則化を用いた推定法が広く用いられているが、それは高次元統計において基本的な手法である Lasso の非線形拡張とみなせる。本講演では、スパース性を誘導する標準的な  $L_1$  正則化以外にも様々な正則化を考察し、それらの汎化誤差を解析した結果を紹介する [2, 8]。理論によって、モデルの複雑さや真の関数の滑らかさによって  $L_1$  正則化以外の正則化が有利になりうることを示す。またベイズ推定を用いることで正則化学習の理論で仮定されていた相関への条件を外してもミニマックス最適レートを実現できる結果も紹介する [3]。

スパースカーネル加法モデルは各説明変数の関数について和をとるモデルと考えられるが、一方で和と積をとった関数も考えられる。そのようなモデルをテンソルモデルと呼ぶ。テンソルモデルは複数の変数間の関係を記述するため、その推定問題は高次元推定問題になる。そこで、真のテンソルが低ランクであると仮定してパラメータ推定することによって、次元の呪いを回避する方法が考えられる。本講演では、そのような推定手法としてベイズ推定や交互最適化法を考え、それらのミニマックス最適性を議論した結果について紹介する [6, 1, 7]。これらは、テンソルモデルを無限次元再生核ヒルベルトに拡張したものも含む。

また、データサイズの大きな高次元データ解析を高速に解く方法として、新しい確率的最適化手法を紹介する [4, 5]。これは、交互方向情報数法を確率的最適化に拡張した方法にあたり、広い範囲の構造的正則化学習に適用することができる。

- [1] H. Kanagawa, T. Suzuki, H. Kobayashi, N. Shimizu, and Y. Tagami. Gaussian process nonparametric tensor estimator and its minimax optimality. In *Proc. of the 29th International Conference on Machine Learning*, pages 1632–1641, 2016.
- [2] T. Suzuki. Unifying framework for fast learning rate of non-sparse multiple kernel learning. In *Advances in Neural Information Processing Systems 24*, pages 1575–1583, 2011.
- [3] T. Suzuki. PAC-bayesian bound for gaussian process regression and multiple kernel additive model. In *Proc. of the 25th Conference on Learning Theory*, volume 23, pages 8.1–8.20, 2012.
- [4] T. Suzuki. Dual averaging and proximal gradient descent for online alternating direction multiplier method. In *Proc. of the 30th International Conference on Machine Learning*, pages 392–400, 2013.
- [5] T. Suzuki. Stochastic dual coordinate ascent with alternating direction method of multipliers. In *Proc. of the 31th International Conference on Machine Learning*, pages 736–744, 2014.
- [6] T. Suzuki. Convergence rate of Bayesian tensor estimator and its minimax optimality. In *Proc. of the 32nd International Conference on Machine Learning*, pages 1273–1282, 2015.
- [7] T. Suzuki, H. Kanagawa, H. Kobayashi, N. Shimizu, and Y. Tagami. Minimax optimal alternating minimization for kernel nonparametric tensor learning. In *Advances In Neural Information Processing Systems 29*, pages 3783–3791, 2016.
- [8] T. Suzuki and M. Sugiyama. Fast learning rate of multiple kernel learning: trade-off between sparsity and smoothness. *The Annals of Statistics*, 41(3):1381–1405, 2013.