

低ランク分解を用いたノンパラメトリックテンソル回帰

東京大学 今泉允聡
情報学研究所 林浩平

1 テンソル回帰問題

本研究では多次元配列（テンソル）を共変量とする回帰問題を扱う。
 K 次のテンソル X は

$$X \in \mathbb{R}^{I_1 \times \dots \times I_K}$$

の様に表現される。ただし I_k は次数 k の次元である。例として、行列は $K=2$ のテンソルで、fMRI による画像データは $K=4$ (x 軸・y 軸・z 軸・時間) のテンソルとして表現される。

各観測値 $i=1, \dots, n$ ごとに実数 Y_i と共変量 X_i が与えられているとし、回帰問題

$$Y_i = f(X_i) + \epsilon_i$$

を考える。ここで ϵ_i は誤差項である。ここでの目的は、回帰関数 f を推定し、また新しい入力 X に対して Y を予測することである。

関連研究として、[1] は線形回帰モデルを、[2] はノンパラメトリック回帰モデルとガウス過程事前分布を持ちいたその推定方法を提案している。

2 提案手法

本研究は、ノンパラメトリックなモデルに基づく回帰関数を提案する。ただしテンソルは一般に非常に多くの要素 ($\prod_k I_k$ 個) の要素を含むため、通常のノンパラメトリック回帰モデルは次元の呪いに直面し、推定の収束レートが非常に遅くなる。それを踏まえて、本研究はテンソルの CP(CANDECOMP/PARAFAC) 分解を応用し、分解した要素毎にノンパラ回帰を行うモデルを提案する。

提案モデルは以下の様になる。

$$f(X_i) = \sum_{m=1}^M \sum_{r=1}^R \lambda_{r,i} \prod_{k=1}^K f_m^{(k)}(x_{r,i}^{(k)}).$$

ここで、 $M \in \mathbb{N}$ はモデルの複雑度を表すパラメータ、 $R \in \mathbb{N}$ はテンソル X_i の階数である。また $\lambda_r \in \mathbb{R}$ と $x_r^{(k)} \in \mathbb{R}^{I_k}$ は、テンソル X に CP 分解を行って得られる要素である。このモデルは、[1] による線形モデルに CP 分解を行い、要素毎にノンパラメトリックな関数を考えることで得られる。

本研究のモデルは、比較的次元な要素 $x_r^{(k)}$ 毎にノンパラメトリックな関数を考えているため、推定の収束レートを改善する。またパラメータ M を調整することで、多くの関数を表現することが可能である。

文献

[1] H. Zhou, L. Li, and H. Zhu, “Tensor regression with applications in neuroimaging data analysis,” *Journal of the American Statistical Association*, vol. 108, no. 502, pp. 540–552, 2013.

[2] Q. Zhao, G. Zhou, L. Zhang, and A. Cichocki, “Tensor-variate gaussian processes regression and its application to video surveillance,” in *Acoustics, Speech and Signal Processing, 2014 IEEE International Conference*, pp. 1265–1269, IEEE, 2014.