Kullback-Leibler 情報量に基づく Alternating Direction Method of Multipliers による 精度行列のスパース推定

九州大学 大学院数理学府 上田 勇祐
九州大学 マス・フォア・インダストリ研究所 西井 龍映
九州大学 大学院数理学府 江田 智尊

1 はじめに

制約付き凸最適化問題に対するアルゴリズム **ADMM**(Alternating Direction Method of Multipliers, [1]) は、目的関数を逐次的に最小化する手法であり、理論的収束性が保証されている。さて、多次元正規分布の精度行列において、成分が 0 となる行・列に対応する変数同士は条件付き独立となる。そのため、変数間の関係性を理解する上で精度行列のスパース推定は重要な課題である。本研究では精度行列のスパース推定のために ADMM を修正した手法を提案し、従来法と比較する。

2 精度行列の ADMM によるスパース推定

$N_p(\mathbf{0}, X^{-1})$ からの無作為標本 $\mathbf{a}_1, \dots, \mathbf{a}_n$ の負の対数尤度関数に L_1 罰則を加えた目的関数の最小化による精度行列 X のスパース推定を考える。

$$\text{minimize}_{X>0} \left[\text{tr}(SX) - \log |X| + \lambda \|X\|_1 \right] \quad \left(S := \frac{1}{n} \sum_{i=1}^n \mathbf{a}_i \mathbf{a}_i^T, \lambda > 0 \right)$$

ADMM は X に近い正定値行列 Z を導入し、次の目的関数の逐次的最小化により X を推定する。

$$L_\rho(X, Z, Y) = \text{tr}(SX) - \log |X| + \lambda \|Z\|_1 + \text{tr}\{Y(X - Z)\} + \frac{\rho}{2} \|X - Z\|_F^2 \quad (1)$$

ただし行列 Y は Lagrange 未定乗数、 $\rho > 0$ はトレードオフパラメータ、 $\|\cdot\|_F$ は Frobenius ノルムを表す。この手法は X に近い行列 Z を用意し、 X には尤度最大化の役割を、 Z には L_1 ノルム最小化の役割を担わせることで、それぞれの更新を単純化する。さらに $\|X - Z\|_F^2$ をペナルティとして加えることによって X と Z を近い行列に保とうというものである。

3 提案手法

式 (1) では X と Z の間の距離を Frobenius ノルムで測っているが、本研究ではこれを Kullback-Leibler (KL) 情報量に改める。ただし、目的関数の凸性や推定手順の単純さの観点から、2 つの正定値行列 Z_1, Z_2 を用意し、 X と Z_1 の距離を KL 情報量で、 Z_1 と Z_2 の距離を Frobenius ノルムで測る次の目的関数を提案する。

$$\begin{aligned} L_{\rho, \alpha}(X, Z_1, Z_2, Y_1, Y_2) &= \text{tr}(SX) - \log |X| + \lambda \|Z_2\|_1 + \text{tr}\{Y_1(X - Z_1)\} + \text{tr}\{Y_2(Z_1 - Z_2)\} \\ &+ \rho \{ \alpha \log |X Z_1^{-1}| + \text{tr}(X^{-1} Z_1) \} + (1 - \alpha) \|Z_1 - Z_2\|_F^2 \end{aligned} \quad (2)$$

行列 Y_1, Y_2 は Lagrange 未定乗数、 $\rho > 0$ 、 $0 < \alpha < 1$ ($\rho\alpha \leq 1$) はトレードオフパラメータを表す。

式 (2) の最小化手順は以下の通りである。ただし、 k は反復回数を表す。

$$\begin{aligned} X^{k+1} &= \text{argmin}_{X>0} \left[\text{tr}(SX) - \log |X| + \text{tr}\{Y_1^k(X - Z_1^k)\} + \rho\alpha \{ \log |X(Z_1^k)^{-1}| + \text{tr}(X^{-1} Z_1^k) \} \right] \\ Z_1^{k+1} &= \text{argmin}_{Z_1>0} \left[\text{tr}\{Y_1^k(X^{k+1} - Z_1)\} + \text{tr}\{Y_2^k(Z_1 - Z_2^k)\} \right. \\ &\quad \left. + \rho \{ \alpha \log |X^{k+1} Z_1^{-1}| + \text{tr}\{(X^{k+1})^{-1} Z_1\} \} + (1 - \alpha) \|Z_1 - Z_2^k\|_F^2 \right] \\ Z_2^{k+1} &= \text{argmin}_{Z_2>0} \left[\lambda \|Z_2\|_1 + \text{tr}\{Y_2^k(Z_1^{k+1} - Z_2)\} + \rho(1 - \alpha) \|Z_1^{k+1} - Z_2\|_F^2 \right] \\ Y_1^{k+1} &= Y_1^k + \rho\alpha \{ (Z_1^{k+1})^{-1} - (X^{k+1})^{-1} \}, \quad Y_2^{k+1} = Y_2^k + 2\rho(1 - \alpha)(Z_1^{k+1} - Z_2^{k+1}) \end{aligned}$$

Elastic Net 型の罰則でも同様の最適化が可能である。本手法に基づく数値実験の結果は当日報告する。

参考文献

- [1] Boyd et al. (2010). “Distributed optimization and statistical learning via the Alternating Direction Method of Multipliers,” *Foundations and Trend in Machine Learning* **3** (1), 1-122.