Sparse inverse covariance estimation with the graphical lasso

李信来

2023 年 7 月 5 日

上海交通大學
SHANGHAI JIAO TONG UNIVERSITY

核心问题

假设我们有 N 个维度为 p 的多元正态观测值，其均值为 ，协方差矩阵为 。按照 Banerjee 等人（2007）的方法，令 $\Theta = \Sigma^{-1}$，并且令 S 为经验协方差矩阵，问题是在非负定矩阵 Θ 上最大化受惩罚的对数似然

$$\arg\max_{\Theta \in \mathbb{R}^{N \times N}} (\log\det\Theta - \mathrm{tr}(S\Theta) - \rho\|\Theta\|_1) \tag{1}$$

问题的转化

设 W 为 Σ 的估计值。Banerjee 等人（2007）表明问题（1）是凸的，他们证明可以通过以分块坐标下降的方式在 W 的每一行和相应列上进行优化来解决这个问题。将 W 和 S 进行分块划分：

$$W = \begin{pmatrix} W_{11} & w_{12} \\ w_{12}^T & w_{22} \end{pmatrix}, S = \begin{pmatrix} S_{11} & s_{12} \\ s_{12}^T & s_{22} \end{pmatrix} \tag{2}$$

他们证明 w_{12} 的解满足

$$w_{12} = \arg\min_y \{y^T W_{11}^{-1} y : \|y - s_{12}\|_\infty \le \rho\} \tag{3}$$

利用凸对偶性，Banerjee 等人（2007）继续展示了解决（3）与解决下面的对偶问题等价：

$$\min_\beta \{\frac{1}{2}\|W_{11}^{1/2}\beta - b\|^2 + \rho\|\beta\|_1\} \tag{4}$$

其中 $b = W_{11}^{1/2} s_{12}$ 如果 β 是 (4) 的解，那么 $w_{12} = W_{11}\beta$ 是 (3) 的解。

Graphical Lasso Algorithm

上海交通大学 SHANGHAI JIAO TONG UNIVERSITY

1. 以 $W = S + \rho I$ 为初始值。在接下来的步骤中，W 的对角线保持不变。
2. 对于每个 $j = 1, 2, ..., p$ $1, 2, ..., p$ ...，解决套索问题（4），该问题的输入是内积 W_{11} 和 s_{12} 。这将给出一个长度为 $p - 1$ 的向量解 $\hat{\beta}$。使用 $w_{12} = W_{11}\hat{\beta}$ 来填充相应的行和列。
3. 继续进行直到收敛。

步骤 2

令 $V = W_{11}$ 和 $u = s_{12}$，那么更新形式如下：

$$\hat{\beta}_j \leftarrow \frac{S(u_j - \sum_{k \neq j} V_{kj}\hat{\beta}_k \ , \ \rho)}{V_{jj}}, \tag{5}$$

对于 $j = 1, 2, ..., p \ 1, 2, ..., p$ ... 利用软阈值操作符 S 进行更新，其中 S 的定义如下：

$$S(x, t) = \text{sign}(x)(|x| - t)_+. \tag{6}$$

循环遍历预测变量直到收敛。在实现中，当 W 的平均绝对变化小于 $t \cdot \text{ave}|S^{-\text{diag}}|$ 时，过程停止，其中 $S^{-\text{diag}}$ 是经验协方差矩阵 S 的非对角元素，t 是一个固定的阈值，默认设置为 0.001。

从 W 到 Θ

上海交通大学 SHANGHAI JIAO TONG UNIVERSITY

由 $W\Theta = I$ 和（2）可得

$$
\begin{aligned}
W_{11}\theta_{12} + w_{12}\theta_{22} &= 0, \\
w_{12}^T\theta_{12} + w_{22}\theta_{22} &= 1, f
\end{aligned}
$$

结合 $\hat{\beta} = W_{11}^{-1}w_{12}$，可得

$$
\begin{aligned}
\hat{\theta}_{22} &= 1/(w_{22} - w_{12}^T\hat{\beta}) \\
\hat{\theta}_{12} &= -\hat{\beta}\hat{\theta}_{22}
\end{aligned}
\tag{7}
$$

因此可以将每个问题的所有系数 β 存储在一个 $p \times p$ 矩阵 $\hat{B}$ 中，并在收敛后计算 $\hat{\Theta}$。

谢谢