Doubly Debiased Lasso in Partially Linear Model Under Hidden Confounding

Chen Zhiyuan

November 2022

目录

1 Doubly Debiased Lasso

Doubly Debiased Lasso 是一个可以同时去除高维情况下 Lasso 的偏差（Lasso 没有无偏性）和 Hidden Confounding 带来的偏差的模型。它是由 Zijian Guo 等人提出的，这里简要介绍一下其方法论。

1.1 Hidden Confounding

混杂是指那些能同时影响自变量和因变量的因素，可能造成伪相关从而导致因果关系失真。模型中的混杂定义如下，其中 H 代表混杂：

$$Y_i = X_i'\beta + H_i'\phi + e_i; \quad X_i' = H_i'\psi + E_i'$$

为了推导的方便，混杂模型可以被改写为线性模型：

$$\begin{aligned} Y_i &= X_i'(\beta + b) + \epsilon_i; \quad X_i = \psi' H_i' + E_i' \\ \epsilon_i &= e_i + H_i'\phi - X_i'b; \quad b = \Sigma_X^{-1}\psi'\phi \end{aligned}$$

这个线性模型只含有因变量 Y 和自变量 X，但在估计参数时会出现一个误差 b，这个误差就是由混杂 H 带来的。

1.2 Spectral Transformations

去除混杂影响的核心手段是对原数据做一个谱变换。DDL 的提出者给出的模拟图像显示带混杂的矩阵 X 有一些奇异值明显大于不带混杂的矩阵的奇异值平均范围，所以构造的谱变换矩阵 $P^{(j)}$ 和 Q 作用在 X 上后要使 $P^{(j)X}$ 和 QX 的过大奇异值被压缩回正常范围。二者的构造自然是基于 SVD 的。

1.2.1 X_{-j} 的谱变换矩阵 $P^{(j)}$

$$\begin{aligned} X_{-j} &= U(X_{-j})\Lambda(X_{-j})V(X_{-j})' \quad by \quad SVD \\ P^{(j)} &= U(X_{-j})\Lambda(X_{-j})U(X_{-j})' \\ P^{(j)}X_{-j} &= U(S\Lambda)V' \end{aligned}$$

不妨将 Λ 对角线上的 m 个奇异值按降序排列。显然，只要合理定义 S 中的元素就能压缩 $S\Lambda$ 的奇异值，比如：

$$S_{l,l} = \begin{cases} \dfrac{\Lambda_{\rho_j m, \rho_j m}}{\Lambda_{l,l}} & if\ l \le \lfloor \rho_j m \rfloor \\ 1 & else \end{cases}$$

可以将前 $\rho_j m$ 个最大的奇异值全部压缩而后面较小的奇异值保持不变。压缩比例 ρ_j 越大，对混杂的修正越强（较大的奇异值），对原始数据的扭曲也越大（较小的奇异值），可能造成更大的误差。

1.2.2 X 的谱变换矩阵 Q

Q 的构造方式和 $P^{(j)}$ 基本一致，只是对象换成了 X：

$$Q = U(X)\Lambda(X)U(X)' \ \ with\ \ S_{l,l} = \begin{cases} \dfrac{\Lambda_{\rho m, \rho m}}{\Lambda_{l,l}} & if\ l \le \lfloor \rho m \rfloor \\ 1 & else \end{cases}$$

1.2.3 $\hat{\beta}^{init}$ 的求解

对谱变换过后的数据 QX 和 QY 使用 Lasso 估计即得到初始估计 $\hat{\beta}^{init}$:

$$\hat{\beta}^{init} = argmin\ \frac{1}{2n}||Q(y - X\beta)||_2^2 + \lambda \sum_{j=1}^{p} \frac{||QX_{.j}||_2}{\sqrt{n}} |\beta_j|$$
$$\lambda = A\sigma_e \sqrt{log\frac{p}{n}}, \quad A > \sqrt{2}$$

1.3 $\hat{\beta}_j$ 的求解

Doubly Debiased Lasso 每次只能估计 β 的一个固定的分量 $\hat{\beta}_j$，它通过拆分每次我们关心的分量所在的项和我们不关心的分量所在的项，并将误差项舍去来达到去偏的效果。渐进性质的证明结果在原文中都已给出。

从线性模型开始推导:

$$Y = X(\beta + b) + \epsilon$$
$$Y = X_j(\beta_j + b_j) + X_{-j}(\beta_{-j} + b_{-j}) + \epsilon \quad for \quad j \in 1, ..., p$$
$$Y - X_{-j}\hat{\beta}^{init}_{-j} = X_j(\beta_j + b_j) + X_{-j}(\beta_{-j} - \hat{\beta}^{init}_{-j}) + X_{-j}b_{-j} + \epsilon$$

至此可以看到两个来源的偏差: $X_{-j}(\beta_{-j} - \hat{\beta}^{init}_{-j})$ 源于初始估计 $\hat{\beta}^{init}$ 的 Lasso 有偏性，$X_{-j}b_{-j}$ 源于代表偏差影响的扰动向量 b。注意到两误差项都依赖于 X_{-j} 且目标 β_j 依赖于 X_j，将谱变换矩阵 $P^{(j)}$ 作用于上式得:

$$P^{(j)}(Y - X_{-j}\hat{\beta}^{init}_{-j}) = P^{(j)}X_j(\beta_j + b_j) + P^{(j)}X_{-j}(\beta_{-j} - \hat{\beta}^{init}_{-j}) + P^{(j)}X_{-j}b_{-j} + P^{(j)}\epsilon$$

并求 $P^{(j)}Z_j$ 作为 $P^{(j)}X_j$ 向 $P^{(j)}X_{-j}$ 投影的残差向量:

$$P^{(j)}Z_j = P^{(j)}X_j - P^{(j)}X_{-j}\gamma$$
$$\hat{\gamma} = argmin\ \frac{1}{2n}||P^{(j)}X_j - P^{(j)}X_{-j}\gamma||_2^2 + \lambda_j \sum_{l \neq j} \frac{||P^{(j)}X_l||_2}{\sqrt{n}} |\gamma_l|$$

于是在等式两边左乘上 $(P^{(j)}Z_j)'$ 使 $(P^{(j)}Z_j)'P^{(j)}X_j$ 被投影到 R^1，可以直接四则运算分离关心的真实值 β_j:

$$(P^{(j)}Z_j)'P^{(j)}(Y - X_{-j}\hat{\beta}^{init}_{-j}) = (P^{(j)}Z_j)'P^{(j)}X_j(\beta_j + b_j) + (P^{(j)}Z_j)'\ [P^{(j)}X_{-j}(\beta_{-j} - \hat{\beta}^{init}_{-j}) + P^{(j)}X_{-j}b_{-j} + P^{(j)}\epsilon]$$

$$\beta_j + b_j = \frac{(P^{(j)}Z_j)'P^{(j)}(Y - X_{-j}\hat{\beta}^{init}_{-j})}{(P^{(j)}Z_j)'P^{(j)}X_j} - \frac{(P^{(j)}Z_j)'P^{(j)}X_{-j}(\beta_{-j} - \hat{\beta}^{init}_{-j})}{(P^{(j)}Z_j)'P^{(j)}X_j} - \frac{(P^{(j)}Z_j)'P^{(j)}X_{-j}b_{-j}}{(P^{(j)}Z_j)'P^{(j)}X_j} - \frac{(P^{(j)}Z_j)'P^{(j)}\epsilon}{(P^{(j)}Z_j)'P^{(j)}X_j}$$

根据上文分离的两个误差项来源，提出 β 的估计 $\hat{\beta}_j$ 为:

$$\hat{\beta}_j = \frac{(P^{(j)}Z_j)'P^{(j)}(Y - X_{-j}\hat{\beta}^{init}_{-j})}{(P^{(j)}Z_j)'P^{(j)}X_j}$$
$$\hat{\beta}_j - \beta_j = \frac{(P^{(j)}Z_j)'P^{(j)}\epsilon}{(P^{(j)}Z_j)'P^{(j)}X_j} + \frac{(P^{(j)}Z_j)'P^{(j)}X_{-j}(\beta_{-j} - \hat{\beta}^{init}_{-j})}{(P^{(j)}Z_j)'P^{(j)}X_j} + \frac{(P^{(j)}Z_j)'P^{(j)}X_{-j}b_{-j}}{(P^{(j)}Z_j)'P^{(j)}X_j} + b_j$$

估计与真实值之差的第一项为方差；第二项中 $P^{(j)}Z_j$ 和 $P^{(j)}X_{-j}$ 是正交的（投影和残差）且 $\hat{\beta}^{init}$ 和 β_j 差距小；第三项中因为 $P^{(j)}$ 缩减了 X_{-j} 的奇异值，可以证明 $||\frac{1}{\sqrt{n}}P^{(j)}X_{-j}b_{-j}|| = O_p(\frac{1}{min(n,p)})$；最后一项 b_j 在 Dense Confounding 的假设下可以证明 $|b_j| \leq \frac{q\sqrt{logp}}{1+\lambda_q^2(\Psi)}$。所以这三项误差相对于方差都是可以忽略的，即估计 $\hat{\beta}_j$ 和真实值的差距仅为估计的方差，意味着它就是无偏的。

1.4 方差的估计和置信区间的构造

2 Partially Linear Models

部分线性模型的简单介绍

3 用 Doubly Debiased Lasso 估计带隐混杂的部分线性模型

在部分线性模型中加入隐混杂则得到：

$$y_i = X_i'\beta + H_i'\phi + g(T_i) + \epsilon_i$$
$$y_i - X_i'\beta + H_i'\phi = g(T_i) + \epsilon_i$$

不妨认为 β 和 ϕ 是固定的，则首先对 $g(T_i)$ 进行估计。$g(T_i)$ 可以在 t 处泰勒展开为：

$$g(T_i) = g(t) + g'(t)(T_i - t) =: a + b(T_i - t)a = g(t); b = g'(t)$$

于是 $g(T_i)$ 的估计转化为

4 参考文献