ISS-SP-035

プライバシ保護機能を持つ ベータダイバージェンスを用いたロバスト逐次線形回帰

竹下　虎太朗† 　　福永　修一†
† 東京都立産業技術高等専門学校

1. はじめに

暗号プロトコルを用いたプライバシ保護データマイニング手法の 1 つとして線形回帰[1]がある. 線形回帰はデータに外れ値が含まれた場合に推定性能が劣化する. この問題に対して, 著者らはロバスト化に有効なベータダイバージェンスを用いたプライバシ保護機能を持つロバスト線形回帰[2]を提案した.

本研究では, 文献[2]の手法を時刻ごとにデータが与えられた際に逐次的にパラメータを推定するアルゴリズムに拡張する. そして提案手法の有効性を数値例を用いて示す.

2. プライバシ保護機能を持つベータダイバージェンスを用いたロバスト逐次線形回帰

パーティAとパーティBという2者がそれぞれ秘密データを保持しているとする. 各時刻tにおいて, パーティA の秘密データは説明変数x_tとする. また, パーティ B の秘密データは目的変数y_tとする. x_tとy_tの関係は以下の式で表されるとする.

$$y_t = \boldsymbol{\varphi}_t{}^{\mathrm{T}}\boldsymbol{\theta} + \varepsilon_t, \quad \varepsilon_t \sim \mathcal{N}(0, \sigma^2) \tag{1}$$

ここで$\boldsymbol{\varphi}_t = (x_t, 1)^{\mathrm{T}}$は基底関数, $\boldsymbol{\theta}$は未知パラメータである. この問題の目的は, パーティ A とパーティ B が秘密データを暗号化したまま, パラメータ$\boldsymbol{\theta}$を逐次的に推定することである. 暗号プロトコルには暗号化したまま加法と整数倍算が可能な加法準同型暗号を用いる.

ベータダイバージェンスを用いたロバストな逐次推定アルゴリズムは次式で与えられる.

$$\hat{\boldsymbol{\theta}}_t = \hat{\boldsymbol{\theta}}_{t-1} + \boldsymbol{K}_t\left(y_t - \boldsymbol{\varphi}_t{}^{\mathrm{T}}\hat{\boldsymbol{\theta}}_{t-1}\right) \tag{2}$$

$$\boldsymbol{K}_t = \frac{w_t \boldsymbol{P}_{t-1}\boldsymbol{\varphi}_t}{(1 + w_t \boldsymbol{\varphi}_t{}^{\mathrm{T}}\boldsymbol{P}_{t-1}\boldsymbol{\varphi}_t)} \tag{3}$$

$$\boldsymbol{P}_t = \boldsymbol{P}_{t-1} - \frac{w_t \boldsymbol{P}_{t-1}\boldsymbol{\varphi}_t\boldsymbol{\varphi}_t{}^{\mathrm{T}}\boldsymbol{P}_{t-1}}{(1 + w_t \boldsymbol{\varphi}_t{}^{\mathrm{T}}\boldsymbol{P}_{t-1}\boldsymbol{\varphi}_t)} \tag{4}$$

$$w_t = \exp\left(\frac{-\beta\left(y_t - \boldsymbol{\varphi}_t{}^{\mathrm{T}}\hat{\boldsymbol{\theta}}_{t-1}\right)^2}{2\sigma^2}\right) \tag{5}$$

秘密データを暗号化した状態で, 式(2)-(5)のアルゴリズムによる逐次推定を行う.

本研究では, パーティAは暗号化と復号, パーティBは暗号化の処理を行うことができるものとする. パーティAはx_tを暗号化しパーティBに送信する. パーティBは暗号化した状態で秘密データを含む計算を行いパーティAに送信する. パーティAは重み関数を計算する. パーティAは$\boldsymbol{P}_{t-1}$と$\boldsymbol{K}_t$を計算し暗号化する. 暗号化したx_tとともにパーティ B に送信する. パーティ B は暗号化した状態で$\hat{\boldsymbol{\theta}}_t$の更新を行う.

3. 数値例

(1)式において, $\boldsymbol{\theta} = (\theta_1, \theta_2)^{\mathrm{T}} = (8.1,\ 1.7)^{\mathrm{T}}$とする. データ数$N = 100$とし, 入力データ$x_t$は区間$[0, 5]$の一様分布から生成した値とする. 雑音は標準正規分布から生成した値とする. ただし雑音のうち, $t = 5, 10$は, 外れ値として 5 の値をとる.

提案手法と通常の逐次線形回帰による, 推定値の推移の比較を図 1 に示す. 緑・茶の破線は$\theta_{t,1}, \theta_{t,2}$の真値, 青・赤の実線は提案手法による$\theta_{t,1}, \theta_{t,2}$の推定値, 橙・紫の一点鎖線は通常の逐次線形回帰による$\theta_{t,1}, \theta_{t,2}$の推定値である. 図より, 通常の線形回帰に比べ, 提案手法は外れ値の影響を受けていないことがわかる.

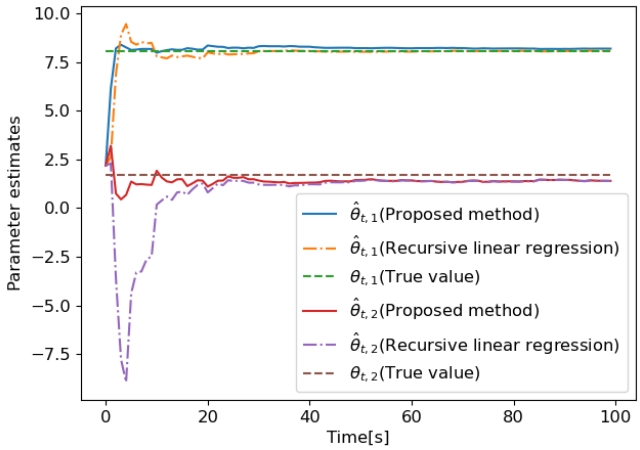

図1. 推定値の推移の比較

4. おわりに

本研究ではロバスト化に有効なベータダイバージェンスを用いたプライバシ保護機能を持つロバスト逐次線形回帰を提案した. 数値例より, 提案手法は外れ値の影響を受けず推定が行えることがわかった.

最後に本研究は JSPS 科研費 19K04448 の助成を受けたものであり, ここに感謝の意を表します.

参考文献

[1] R. Hall, S. E. Fienberg and Y. Nardi. Secure multiple linear regression based on homomorphic encryption. *Journal of Official Statistics*, 2011.

[2] 竹下, 福永, 田中, 黄. プライバシ保護機能を持つベータダイバージェンスを用いたロバスト線形回帰. 第 22 回情報論的学習理論ワークショップ, 2019.