

論文解説：仮想的な0近傍法による高次バイアス補正

奥野彰文^{*1,2}, 下平英寿^{3,2}

¹ 統計数理研究所 統計基盤数理研究系 ² 理化学研究所 AIP センター

³ 京都大学大学院 情報学研究科

要旨

本稿は、NeurIPS2020 に採択された近傍法の高次バイアス補正に関する我々の論文 Okuno and Shimodaira (2020) に関する解説です。解説の平易さを優先するため、厳密な記述については当該論文をご参照ください。

キーワード: 近傍法, 高次バイアス補正

1 研究背景

(X, Y) を共変量と応答変数の組とし、その観測値を $(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)$ としましょう。この観測値を用いてクエリ x_* に付随するラベル y_* を予測する問題を考えます。ここで Y が2値 $\{0, 1\}$ をとると2値分類問題、 Y が実数値 \mathbb{R} をとると回帰問題になります。

Okuno and Shimodaira (2020) では分類問題を考えますが、数学的な扱いが難しくなるので、本稿では回帰問題のみを考えます。まずクエリ x_* からのユークリッド距離により、 $\|x_{(1)} - x_*\|_2 \leq \|x_{(2)} - x_*\|_2 \leq \dots \leq \|x_{(n)} - x_*\|_2$ となるよう入れ替えた添え字を $(1), (2), \dots, (n)$ とします。ユーザの指定するパラメータ $k \in \mathbb{N}$ と重み $w_1, w_2, \dots, w_k \geq 0$ を用いた

$$\hat{\eta}_{k,w}^{(kNN)}(X_*) = \sum_{i=1}^k w_i y_{(i)} \quad (1)$$

を特に重み付き k 近傍推定量と呼びます。ただし重みは $\sum_{i=1}^k w_i = 1$ を満たします。 $w_1 = w_2 = \dots = w_k = k^{-1}$ の場合を特に非重み付き k 近傍推定量と呼び、記号から重み w を省略します。 k 近傍推定量のイメージについては図1をご覧ください。

k 近傍法にはバイアスとバリエーションのトレードオフがあります。 k を大きく設定すると、より多くのデータのラベル平均を取るため、推定量のバリエーションが小さくなり、推定が安定します。しかしその分だけ遠くのデータまで参照することになるため、バイアスが大きくなります。逆に、 k を小さく設定するとクエリに近いデータのみを参照するためバイアスは小さくなりますが、平均を

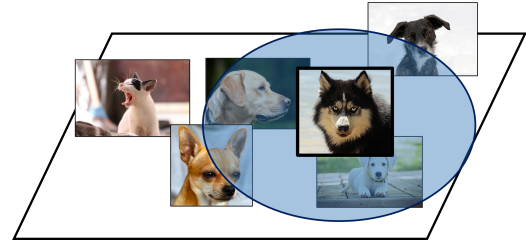


図1: 近傍法はあるクエリ (例えばここでは犬の画像) の近傍データ (他の犬の画像) を探し、付随するラベル $y_{(i)}$ の平均をとることでクエリのラベル y_* を予測する。

取るデータの数が減り、バリエーションが大きくなって推定の安定性が低下します。

ここで、数値実験を介して k 近傍推定量のバイアス-バリエーションのトレードオフを見てみましょう。図2は横軸を k とし、モンテカルロシミュレーションによって k 近傍推定量のバイアスとバリエーションを可視化したものです。青色が真値を表します。 k が大きくなると、黒実線で表される推定量の平均が青線から離れていき (つまりバイアスが大きくなります)、黒破線の幅が小さくなります (つまりバリエーションが小さくなります)。

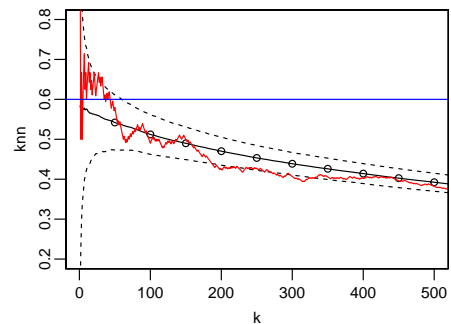


図2: 赤線: 非重み付き k 近傍推定量, 青線: 真値 $\eta(x_*) = \mathbb{E}[y_* | x_*]$, 黒の実線と破線はモンテカルロシミュレーションにおける k 近傍推定量の平均と標準偏差を表す。

2 本研究の貢献

2.1 仮想的な0近傍法

本研究では、バイアスとバリエーションのトレードオフを克服し、両方を小さくさせる、つまり「バリエーションを小さく保ちながらバイアスを減少させる」ために k を $k=0$

* 責任著者, okuno@ism.ac.jp

に外挿する

“仮想的な 0 近傍推定量”

を提案します. より具体的な手続きとしては, ユーザが指定する $1 \leq k_1 < k_2 < \dots < k_V \leq n$ を用いて,

- (1) 各 k_v について k 近傍推定量 $\hat{\eta}_{k_v}^{(k\text{NN})}$ を計算,
- (2) 各 k_v について距離 $r_{k_v} := \|x_{(k_v)} - x_*\|_2$ を計算,
- (3) 回帰関数 $f_\theta(r_{k_v})$ により, r_{k_v} を介して k_v 近傍推定量を予測します:

$$\hat{\theta} := \arg \min_{\theta \in \Theta} \sum_{v=1}^V \{\hat{\eta}_{k_v}^{(k\text{NN})} - f_\theta(r_{k_v})\}^2.$$

最後に, $r_0 = 0$ に外挿した $f_{\hat{\theta}}(0)$ を特に「マルチスケール k 近傍推定量」と呼び, これが仮想的な 0 近傍推定量に相当します.

理論的な点にも触れておきましょう. $\eta(x) = \mathbb{E}[y|x]$ が $\beta > 1$ での β -Hölder 条件を満たすとき, つまり $\eta(x)$ が非常になめらかであるとき, 回帰問題における L^2 損失を考えると, 通常近傍法の収束レートはよく知られた最適レート $n^{-2\beta/(d+2\beta)}$ より遅くなります. 一方で, $C = \max\{\gamma \in \mathbb{N} \cup \{0\} \mid \gamma < \beta/2\}$ とした多項式関数

$$f_\theta(r) = \theta_0 + \theta_1 r^2 + \theta_2 r^4 + \dots + \theta_C r^{2C}$$

を回帰関数として用いると, 仮想的な 0 近傍推定量は高次のバイアスを補正し, 最適レートを達成することが証明できます. さらに, プラグイン型の分類器 $\mathbb{1}(\hat{\eta}(X_*) \geq 1/2)$ を用いると, 分類問題における最適レートを達成します (Okuno and Shimodaira (2020) Theorem 2)^a.

2.2 実数重み付き近傍法との対応

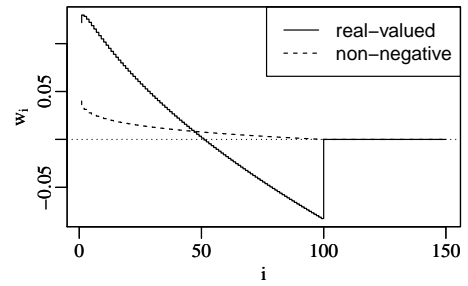
ここまで紹介した仮想的な 0 近傍法は, 実数の重みを付けた近傍法として解釈できます. 例えば, (1) で紹介した重み付き近傍法では, 通常, 非負の重み $w_1, w_2, \dots, w_n \geq 0$ を考えますが, 負の重みを許容する実数重み付き近傍法が Samworth (2012) などによって提案されています. Samworth (2012) は, 理論的に好ましい実数重み

$$w_i = \frac{1}{k} \left\{ 1 + \frac{d}{2} - \frac{d}{2k^{2/d}} (i^{1+2/d} - (i-1)^{1+2/d}) \right\}$$

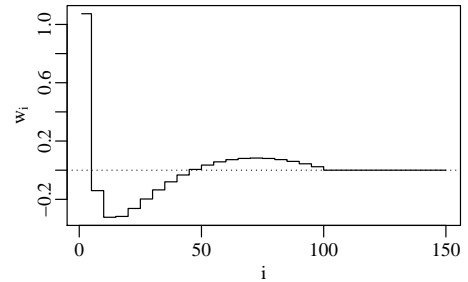
^a 回帰問題では $\eta(x) = \mathbb{E}[y|x]$ に Hölder 条件のみ仮定しますが, 分類問題では $\eta(x)$ が判別境界 $1/2$ と距離 t 以内になる確率が $C_0 t^\alpha$ で抑えられるという条件を追加で考えます. 回帰推定量の一般収束レートが r_n であれば, このとき分類問題の収束レートが r_n^α になることが Audibert and Tsybakov (2007) により示されています. 一方で通常の k 近傍法ではこれらの強い定理が利用できないので, Chaudhuri and Dasgupta (2014) では k 近傍法の分類問題での収束レートを議論しており, Okuno and Shimodaira (2020) ではさらに高次バイアスを補正したレートを導出しています.

($i = 1, 2, \dots, k$) を導出し, その実数重み付き近傍法が提案法と同じく最適レートを達成することを証明しています. Samworth (2012) の重みを $d = 10, k = 100$ で計算した例は, 図 3(a) の黒実線をご覧ください. 比較のため, 最適な非負重みも黒点線で示しています^b.

一方, 我々の提案法も, $k = 0$ 近傍への外挿を介して, 重み付き近傍法の実数重みを計算していると解釈できます. 提案法に対応する実数重みを $V = 20$ で計算した結果を, 図 3(b) に示しました. Samworth (2012) の重みが, (i が小さな一部領域を除き) $i \leq k$ で単調に減少するのに対して, 提案法に対応する重みは, i が増加するごとに一度減少し, 再度増加してから再び減少し, 滑らかに $w_{k+1} = 0$ へとつながっていくことがわかります.



(a) Samworth (2012)



(b) Okuno and Shimodaira (2020)

図 3: 実数重み付き近傍法とみなした場合の, 対応する実数重みの違い ($d = 10, k = 100, V = 20$).

3 まとめ

本研究では, k の値により近傍法がバイアス-バリエーションのトレードオフを持つことを示し, 異なる $k_1 < k_2, \dots < k_V$ で計算した k 近傍推定量の外挿を介して仮想的な 0 近傍推定量を定義しました. また, 提案した推定量が理論的に最適なレートを達成することを証明し, Samworth (2012) と異なる重みデザインによる実数重み付き近傍法に対応することも示しました.

^b 図は Okuno and Shimodaira (2020) Figure 2 の再利用であり, 本稿の著者が著作権を保持しています (<https://neurips.cc/FAQ/Copyright>)

謝辞

本稿のイラストには <https://pixabay.com/> から取得された画像が含まれており、これらは Creative Commons Zero (CC0) ライセンスの下で公開されています。

参考文献

- Audibert, J.-Y. and Tsybakov, A. B. (2007). Fast learning rates for plug-in classifiers. *Ann. Statist.*, 35(2):608–633.
- Chaudhuri, K. and Dasgupta, S. (2014). Rates of convergence for nearest neighbor classification. In *Advances in Neural Information Processing Systems*, volume 27. Curran Associates, Inc.
- Okuno, A. and Shimodaira, H. (2020). Extrapolation towards imaginary 0-nearest neighbour and its improved convergence rate. In *Advances in Neural Information Processing Systems*, volume 33, pages 21889–21899. Curran Associates, Inc.
- Samworth, R. J. (2012). Optimal weighted nearest neighbour classifiers. *The Annals of Statistics*, 40(5):2733–2763.