Отчет по конкурсу “Молодая математика России” за 2022 год

Пучкин Никита Андреевич

1 Результаты исследований

Рассмотрена задача оценки гладкого многообразия по неточным наблюдениям $Y_1, \dots, Y_n \in \mathbb{R}^D$, сгенерированным независимо соглано модели

$$Y = X + \varepsilon, \tag{1}$$

где $X \in \mathbb{R}^D$ – точка на гладком многообразии $\mathcal{M}^*$ размерности $d \ll D$, $\varepsilon \in \mathbb{R}^D$ – случайный многомерный шум с нулевым средним. Доказана новая минимаксная нижняя оценка на точность восстановления многообразия.

Были сформулированы следующие предположения. Во-первых, предполагается, что истинное многообразие $\mathcal{M}^*$ принадлежит классу $\mathscr{M}^d_\varkappa$ компактных, линейно связных многообразий без края, которые содержатся в шаре $\mathcal{B}(0, R)$, имеют рич не менее $\varkappa$ и размерность d:

$$\begin{aligned}\mathcal{M}^* \in \mathscr{M}^d_\varkappa = \big\{\mathcal{M} \subset \mathbb{R}^D : \mathcal{M} &- \text{компактное линейно-связное} \\ &\text{многообразие без края}, \mathcal{M} \subseteq \mathcal{B}(0, R), \\ &\operatorname{reach}(\mathcal{M}) \geqslant \varkappa, \dim(\mathcal{M}) = d < D\big\}.\end{aligned} \tag{A1}$$

Рич многообразия $\mathcal{M}$ определяется как супремум таких $r > 0$, что любая точка в $\mathcal{M} + \mathcal{B}(0, r)$ имеет единственную проекцию на $\mathcal{M}$. Здесь и далее $A + B = \{a + b : a \in A, b \in B\}$ – сумма Минковского множеств A и B.

Во-вторых, плотность $p(x)$ векторов $X_1, \dots, X_n$ (по отношению к d-мерной мере Хаусдорфа на $\mathcal{M}^*$) должна удовлетворять следующему условию:

$$\exists\, p_1 \geqslant p_0 > 0 : \forall x \in \mathcal{M}^* \quad p_0 \leqslant p(x) \leqslant p_1. \tag{A2}$$

Помимо условий (A1) и (A2) на регулярность многообразия $\mathcal{M}^*$ и на распределение X, необходимо наложить условия и на распределение шума ε. Мы предполагаем, что условное распределение $(\varepsilon \,|\, X)$ имеет следующие свойства:

$$\begin{aligned}&\mathbb{E}(\varepsilon \,|\, X) = 0, \quad \text{почти-наверное}, \\ &\|\varepsilon\| \leqslant M < \varkappa, \quad (\varepsilon \,|\, X) \in \mathcal{T}^\perp_X \mathcal{M}^* \quad \mathbb{P}(\cdot \,|\, X)\text{-почти-наверное для всех } X \in \mathcal{M}^*,\end{aligned} \tag{A3}$$

где $\mathcal{T}^\perp_X \mathcal{M}^*$ – ортогональное к $\mathcal{M}^*$ подпространство в точке X. Другими словами, рассмотрен случай ограниченного ортогонального шума. Модель, удовлетворяющая условиям (A1)–(A3), часто используется в литературе (см., например, [3, 1, 2]).

Была доказана следующая минимаксная нижняя оценка на точность восстановления $\mathcal{M}^*$.

Теорема 1.1 *Пусть выборка* $\mathbb{Y}_n = \{Y_1, \dots, Y_n\}$ *сгенерирована из модели* (1)*, где* $\mathcal{M}^* \in \mathscr{M}^d_{\varkappa}$*, плотность* $p(x)$ *случайного вектора* X *удовлетворяет* (A2) *(с достаточно большим* p_1 *и малым* p_0*), а для шума* ε *выполнено условие* (A3)*. Тогда, какова бы ни была оценка* $\widehat{\mathcal{M}}$*, расстояние Хаусдорфа* $d_H(\widehat{\mathcal{M}}, \mathcal{M}^*)$ *между ней и истинным многообразием* $\mathcal{M}^*$ *удовлетворяет неравенству:*

$$\sup_{\mathcal{M}^* \in \mathscr{M}^d_{\varkappa}} \mathbb{E}_{\mathcal{M}^*} d_H(\widehat{\mathcal{M}}, \mathcal{M}^*) \geqslant \frac{C}{\varkappa} \left[M^2 \vee \left(\frac{M^2 \varkappa^2 \log n}{n} \right)^{\frac{2}{d+4}} \right],$$

где C *– некоторая абсолютная константа.*

Теорема 1.1 существенно улучшает известные прежде в литературе результаты [3, 2] и дополняет минимаксную нижнюю оценку, полученную в [4]. По результатам работы готовится препринт, который будет подан для публикации в ведущий журнал по математической статистике.

2 Опубликованные и поданные в печать работы

В отчетном году было опубликовано и подано в печать три работы.

- N. Puchkin and V. Spokoiny. Structure-adaptive Manifold Estimation. *Journal of Machine Learning Research*, volume 23(40), pages 1–62, 2022.
- N. Puchkin and N. Zhivotovskiy, “Exponential Savings in Agnostic Active Learning Through Abstention”, *IEEE Transactions on Information Theory*, volume 68(7), pages 4651–4665, 2022.
- N. Puchkin and V. Ulyanov, “Inference via Randomized Test Statistics”, *Annales de l'Institut Henri Poincaré (в печати).*

3 Участие в конференциях и летних школах

В 2022 году выступил с докладом на четырех конференциях и летних школах.

- Конференция “Ломоносовские чтения 2022”, секция вычислительной математики и кибернетики, 14–22 апреля 2022 г.
- Летняя школа “Обучение, понимание и оптимизация в моделях искусственного интеллекта”, Пушкин, Санкт-Петербург, 20–26 июня 2022 г.
- Летняя школа “Math for Machine Learning”, Бен-Герир, Марокко, 25–30 июля 2022 г.
- Конференция “Fall into ML 2022”, Москва, 1–3 ноября 2022 г.

4 Работа в научных центрах

Являюсь младшим научным сотрудником Международной лаборатории стохастических алгоритмов и анализа многомерных данных НИУ ВШЭ и Института проблем передачи информации им. А. А. Харкевича РАН.

5 Педагогическая деятельность

В 2022 году было принято участие в пяти курсах в качестве лектора, семинариста или учебного ассистента.

- Онлайн-методы машинного обучения, МФТИ, весенний семестр, лектор.
- Статистическая теория машинного обучения, МФТИ, осенний семестр, лектор.
- Advanced statistical methods, НИУ ВШЭ и Сколтех, весенний семестр, семинарист.
- Random matrix theory, НИУ ВШЭ, осенний семестр, семинарист.
- Введение в теорию случайных процессов, НИУ ВШЭ, весенний семестр, учебный ассистент.

Список литературы

[1] E. Aamari and C. Levrard. Stability and minimax optimality of tangential Delaunay complexes for manifold reconstruction. *Discrete Comput. Geom.*, 59(4):923–971, 2018.

[2] E. Aamari and C. Levrard. Nonasymptotic rates for manifold, tangent space and curvature estimation. *Ann. Statist.*, 47(1):177–204, 2019.

[3] C. R. Genovese, M. Perone-Pacifico, I. Verdinelli, and L. Wasserman. Minimax manifold estimation. *J. Mach. Learn. Res.*, 13:1263–1291, 2012.

[4] A. K. H. Kim and H. H. Zhou. Tight minimax rates for manifold estimation under Hausdorff loss. *Electron. J. Stat.*, 9(1):1562–1582, 2015.