ОБОЗРЕНИЕ
ПРИКЛАДНОЙ И ПРОМЫШЛЕННОЙ
Том 19 МАТЕМАТИКИ Выпуск 3
2012

А. В. Б е р н ш т е й н (Москва, ИСА РАН). **Локальная обобщающая способность в задаче восстановления по данным нелинейного многообразия.**

Задача снижения размерности формулируется как задача восстановления многообразия: по выборке $\mathbf{X}_n = \{X_1, X_2, \ldots, Xn\}$ из неизвестного q-мерного многообразия данных $\mathbf{X} = \{X = f(b) \in \mathbf{R}^p:\ b \in \mathbf{B} \subset \mathbf{R}^q\}$ в $\mathbf{R}^p$, покрытого одной координатной системой (картой), необходимо построить такие отображение вложения h многообразия $\mathbf{X}$ в q-мерное множество $\mathbf{Y} = h(X) \subset \mathbf{R}^q$ и отображение восстановления g множества $\mathbf{Y}$ в $\mathbf{R}^p$, что пара $\theta = (h, g)$ обеспечивает выполнение приближенных равенств

$$r_\theta(X) \equiv g(h(X)) \approx X \text{ для всех } X \in \mathbf{X}. \tag{1}$$

Ошибки восстановления $\delta_\theta(X) = \|X - r_\theta(X)\|$ могут быть вычислены для точек $X \in \mathbf{X}_n$ выборки, а для новых (out-of-sample) точек $X \in \mathbf{X}\backslash\mathbf{X}_n$ величина $\delta_\theta(X)$ характеризует обобщающую способность процедуры θ в точке X.

Рассмотрим эмпирическое многообразие $\mathbf{X}_\theta = \{X = g(y) \in \mathbf{R}^p:\ y \in \mathbf{Y} \subset \mathbf{R}^q\}$ в $\mathbf{R}^p$ и касательные подпространства $T(X) \equiv X \oplus L(X)$ и $T_\theta(r_\theta(X)) \equiv r_\theta(X) \oplus L_\theta(r_\theta(X))$ к многообразиям $\mathbf{X}$ и $\mathbf{X}_\theta$ в точках $X \in \mathbf{X}$ и $r_\theta(X) \in \mathbf{X}_\theta$ соответственно, где q-мерные линейные подпространства L и L_θ — точки многообразия Грассмана $\mathrm{Grass}(p, q)$ [1], состоящего из q-мерных линейных подпространств в $\mathbf{R}^p$. Для точек $L, L' \in \mathrm{Grass}(p, q)$ величина $d_{P,2}(L, L') = \|P_L - P_{L'}\|_2$ является метрикой (проекционной 2-нормой) [2] на многообразии Грассмана, где P_L и $P_{L'}$ — операторы проектирования на линейные подпространства L и L'.

Если многообразие данных $\mathbf{X}$ лежит в трубке $\mathrm{Tube}(\mathbf{X}_\theta)$ эмпирического многообразия, состоящей из точек в $\mathbf{R}^p$, проекции которых на многообразие $\mathbf{X}_\theta$ единственны, то можно рассмотреть новое решение $\theta(g) = (h_g, g)$, в котором $h_g(X) = \arg\min\{\|X - g(y)\|,\ y \in \mathbf{Y}\}$ для $X \in \mathrm{Tube}(\mathbf{X}_\theta)$ есть функция проектирования на эмпирическое многообразие $\mathbf{X}_\theta$. По определению, при $\theta(g) = (h_g, g)$ с $r_{\theta(g)}(X) = g(h_g(X))$ для решения имеет место неравенство

$$\|X - r_{\theta(g)}(X)\| \leqslant \|X - r_\theta(X)\| \text{ для } X \in \mathbf{X}. \tag{2}$$

Пусть $X_0 \in \mathbf{X}$ и $\delta_\theta(X_0, \varepsilon) = \max\{\delta_\theta(X):\ X \in U_\varepsilon(X_0)\}$ есть максимальная ошибка восстановления в ε-окрестности $U_\varepsilon(X_0) = \{X \in \mathbf{X}:\ \|X - X_0\| \leqslant \varepsilon\}$ точки X_0, характеризующая локальную обобщающую способность процедуры θ в окрестности точки X_0. Следующая теорема устанавливает нижнюю границу для локальной обобщающей способности процедуры θ в произвольной точке $X_0 \in \mathbf{X}$.

Теорема. *Если многообразие данных $\mathbf{X}$ лежит в трубке Tube($\mathbf{X}_\theta$) эмпирического многообразия $\mathbf{X}_\theta$, а отображения h и g — гладкие и имеют ранг q, то при $\varepsilon \to 0$ имеет место неравенство*

$$\delta^2_{\theta(g)}(X_0, \varepsilon) \geqslant \delta^2_{\theta(g)}(X_0) + \varepsilon^2\, d^2_{P,2}(L(X_0), L_{\theta(g)}(r_{\theta(g)}(X_0))) + o(\varepsilon^2). \tag{3}$$

Здесь и далее символ $o(\cdot)$ в векторном случае понимается в покомпонентном смысле. При $\delta_{\theta(g)}(X_0) = 0$ неравенство (3) превращается в равенство.

С учетом (2), теорема определяет оценку снизу для величины $\delta_\theta(X_0, \varepsilon)$, а при $X \in U_\varepsilon(X_0)$ и $\varepsilon \to 0$ имеют место асимптотические неравенства

$$\|(X - X_0) - (r_{\theta(g)}(X) - r_{\theta(g)}(X_0))\| \leqslant \|X - X_0\| \, d_{P,2}(L(X_0), L_{\theta(g)}(r_\theta(X_0)) + o\,(|X - X_0\|);$$

$$\|X - X_0\|\sqrt{1 - d^2_{P,2}(L(X_0), L_{\theta(g)}(r_{\theta(g)}(X_0)))} \leqslant \|r_{\theta(g)}(X) - r_{\theta(g)}(X_0)\| \\ +o\,(\|X - X_0\|) \leqslant \|X - X_0\|.$$

Левая часть в первом неравенстве характеризует, насколько отображение $r_\theta(g)$ сохраняет локальную структуру многообразия данных $\mathbf{X}$, а второе неравенство характеризует степень локальной неизометричности этого отображения.

Следовательно,чем больше расстояния между линейными подпространствами $L(X)$ и $L_{\theta(g)}(X)$, тем ниже локальная обобщающая способность построенного решения, хуже сохраняется локальная структура многообразия данных и обеспечиваются свойства локальной изометричности. Поэтому от процедуры снижения размерности $\theta = (h, g)$ естественно требовать, чтобы она обеспечивала не только близость (1) между точками $X \in \mathbf{X}$ и их образами $r_{\theta(g)} \in \mathbf{X}_\theta$, но и тангенциальную близость между линейными пространствами L и $L_{\theta(g)}$ в выбранной метрике на многообразии Грассмана Grass(p, q).

СПИСОК ЛИТЕРАТУРЫ

1. *Lee J. M.* Introduction to smooth manifolds. New-York: Springer-Verlag, 2003.
2. *Hamm J.*, *Lee D. D.* Grassmann Discriminant Analysis: a Unifying View on Subspace-Based Learning. — In: Proceedings of the 25th International Conference on Machine Learning (ICML 2008), July 2006.