Е. В. Б у р н а е в (Москва, ИППИ РАН). **Оценка точности в задаче аппроксимации многомерной зависимости.**

Пусть $D_N = \{x_i, y_i\}_{i=1}^N$ — выборка, порожденная моделью $y(x) = f(x) + \varepsilon(x)$, $x \in \mathbf{R}^p$, $y \in \mathbf{R}^1$, где $\varepsilon(x) \sim N(0, \sigma^2(x))$ — нормально распределенный шум. Задача состоит в том, чтобы по выборке D_N оценить $f(x)$ и $\sigma(x)$. Опишем алгоритм бустинга, позволяющий это сделать.

Пусть $E(D_N, f, \sigma) = \sum_{i=1}^N [(y_i - f(x_i))^2/2\sigma^2(x_i) + \log\sqrt{2\pi}\sigma(x_i)]$ — функция ошибок; b^* — номер последней итерации бустинга, на которой произошло уменьшение значения ошибки; $y^{(b)}(x)$ — процесс (являющийся модификацией процесса $y(x)$), порождающий выборку $D_N^{(b)}$ (являющуюся модификацией исходной выборки D_N), которая используется во время b-й итерации бустинга.

Во время инициализации алгоритма бустинга: подсчитываются начальные оценки $\widehat{f}^{(0)}(x) = \sum_{i=1}^N y_i/N$ и $\widehat{\sigma}^{(0)}(x) = \sum_{i=1}^N (y_i - \widehat{f}^{(0)}(x_i))^2/(N-1)$ функций $f(x)$ и $\sigma(x)$ соответственно, а также начальное значение ошибки $E^{(0)} = E(D_N, \widehat{f}^{(0)}, \widehat{\sigma}^{(0)})$; задаются начальные значения $b = 0$, $b^* = 0$, $E^* = E^{(0)}$, $D_N^{(0)} = D_N$, $y^{(0)}(x) = y(x)$, $\widehat{f}_0(x) = \widehat{f}^{(0)}(x)$, $\widehat{\sigma}_0(x) = \widehat{\sigma}^{(0)}(x)$, $f_0(x) = f(x)$.

Условие остановки алгоритма: пусть b — текущий номер итерации алгоритма, тогда дальнейшее выполнение итераций прекращается, если одновременно выполнено $b - b^* = L_1$ и $E^{(b)} \geqslant E^*$, где параметр L_1 задается пользователем.

Во время b-й итерации бустинга выполняются следующие шаги.

Шаг 1. Положим $D_N^{(b)} = \{x_i \in D_N,\ y_i = y^{(b)}(x_i)\}_{i=1}^N$, где $y^{(b)}(x) = (y^{(b-1)}(x) - \widehat{f}_{b-1}(x))/\widehat{\sigma}_{b-1}(x)$. Очевидно, $y^{(b)}(x) \sim N(f_b(x), \sigma_b^2(x))$, где $f_b(x) = (f_{b-1}(x) - \widehat{f}_{b-1}(x))/\widehat{\sigma}_{b-1}(x)$ и $\sigma_b^2(x) = \sigma_{b-1}^2(x)/\widehat{\sigma}_{b-1}^2(x)$.

Шаг 2. В качестве параметрических моделей для $f_b(x)$ и $\sigma_b(x)$ будем использовать разложения по адаптивному базису $\widehat{f}_b(x) = \sum_{k=1}^{p_{f,b}} c_{k,1}^{f,b}\varphi^f(\gamma_k^{f,b}, x)$, $\widehat{\sigma}_b(x) = \exp\{\sum_{k=1}^{p_{\sigma,b}} c_k^{\sigma,b}\varphi^\sigma(\gamma_k^{\sigma,b}, \omega(x)\} + \sum_{k=1}^{p_{f,b}} c_{k,2}^{f,b}\varphi^f(\gamma_k^{f,b}, x))$, где $\varphi^f(\gamma_k^{f,b}, x)$ и $\varphi^\sigma(\gamma_k^{\sigma,b}, \omega)$ — адаптивные базисные функции, например, $\varphi^f(\gamma_k^{f,b}, x) = 1/(1 + \exp\{\sum_{i=1}^p \gamma_{k,i}^{f,b} x^i\})$ (сигмоидная функция); $\omega(x)$ — какой-либо входной вектор, содержащий информацию о локальном поведении функции $\widehat{f}_b(x)$, например, $\omega(x) = (x^1, \ldots, x^p, \|\nabla\widehat{f}_b(x)\|)$. Для оценки параметров функций $\widehat{f}_b(x)$ и $\widehat{\sigma}_b(x)$ по выборке $D_N^{(b)}$ воспользуемся методом максимизации (с помощью градиентного спуска) функции правдоподобия нормального распределения.

Шаг 3. Оценки функций $f(x)$ и $\sigma(x)$, полученные на b-м шаге бустинга, имеют вид $\widehat{f}^{(b)}(x) = \sum_{i=1}^b \widehat{f}_i(x)\prod_{j=1}^i \widehat{\sigma}_{j-1}(x)$ и $\widehat{\sigma}^{(b)}(x) = \prod_{j=0}^b \widehat{\sigma}_j(x)$ соответственно. Оценим ошибку по формуле $E^{(b)} = E(D_N, \widehat{f}^b, \widehat{\sigma}^{(b)})$.

Шаг 4. Проверим, выполняется ли условие $E^{(b)} < E^*$. Если это условие не выполняется, то значения b^* и E^* обновляются, т.е. $b^* = b$ и $E^* = E^{(b)}$, в противном случае обновления значений b^* и E^* не происходит.

После b-й итерации алгоритма бустинга проверяется, выполнено ли условие остановки. Если это условие не выполнено, то $b \leftarrow b + 1$, и выполняется следующая итерация алгоритма.

Результатом алгоритма будут оценки $\widehat{f}(x) = \sum_{i=1}^{b^*} \widehat{f}_i(x)\prod_{j=1}^i \widehat{\sigma}_{j-1}(x)$ и $\widehat{\sigma}(x) = \prod_{j=0}^{b^*} \widehat{\sigma}_j(x)$ функций $f(x)$ и $\sigma(x)$ соответственно. Функция $\widehat{f}(x)$ является аппроксимацией функции $f(x)$, а функция $\widehat{\sigma}(x)$ по сути определяет точность этой аппроксимации и может использоваться для построения доверительного интервала прогноза $\widehat{f}(x)$ значения $f(x)$.

Работа выполнена при финансовой поддержке РФФИ, проекты № 09–01–00584а, № 09–07–00180а.