НЕКОТОРЫЕ ПРИМЕНЕНИЯ НЕГЛАДКОГО АНАЛИЗА В МАТЕМАТИЧЕСКОЙ ДИАГНОСТИКЕ*

Г. Ш. Тамасян

g.tamasyan@spbu.ru

18 мая 2022 г.

1°. Введение. В докладе пойдет речь о некоторых результатах в математической диагностике полученных В. Ф. Демьяновым и его учениками [1–3].

Положим $\mathrm{I} = 1 : s$, $\mathrm{J} = 1 : m$. Рассмотрим два конечных множества

$$\mathcal{A} = \{a_i \in \mathbb{R}^n \mid i \in \mathrm{I}\}, \qquad \mathcal{B} = \{b_j \in \mathbb{R}^n \mid j \in \mathrm{J}\}.$$

Пусть $d \in \mathcal{A} \cup \mathcal{B}$. Задача идентификации заключается в том, чтобы как можно лучше разделить множества $\mathcal{A}$ и $\mathcal{B}$ согласно определенному правилу, по которому выбранный элемент d относится к тому или иному множеству.

Обозначим через $\mathcal{L}$ — семейство классификаторов (функций). Идентификация точек множеств $\mathcal{A}$ и $\mathcal{B}$ проводится с помощью функции $\ell \in \mathcal{L}$ следующим образом: будем считать, что точка d, принадлежащая $\mathcal{A} \cup \mathcal{B}$,

- *правильно* идентифицирована, если
$$\Big(d \in \mathcal{A} \ \text{ и } \ \ell(d) > 0\Big) \quad \text{или} \quad \Big(d \in \mathcal{B} \ \text{ и } \ \ell(d) < 0\Big);$$
- *существенно неидентифицируемая*, если $\ell(d) = 0$;
- *неправильно* идентифицирована, если
$$\Big(d \in \mathcal{A}, \text{ но } \ell(d) \leqslant 0\Big) \quad \text{или} \quad \Big(d \in \mathcal{B}, \text{ но } \ell(d) \geqslant 0\Big).$$

Теперь мы можем немного уточнить постановку задачи. Необходимо найти такой классификатор ℓ^* из семейства $\mathcal{L}$, чтобы количество неправильно идентифицированных точек было наименьшим.

*Семинар по оптимизации, машинному обучению и искусственному интеллекту «O&ML» `http://www.apmath.spbu.ru/oml/`

2°. Семейства классификаторов. Приведем примеры двух семейств классификаторов. В качестве первого выберем множество

$$\mathcal{L} = \big\{ \ell(x,g) = \langle w, x\rangle + \beta \mid g \in \Omega \big\},$$

где

$$\Omega = \big\{ g = (w,\beta) \in \mathbb{R}^n \times \mathbb{R} \mid \|w\| = 1 \big\},$$

$\|\cdot\|$ — евклидова норма.

Таким образом, идентификация проводится с помощью гиперплоскости

$$H(g) = \{ x \in \mathbb{R}^n \mid \ell(x,g) = 0,\ g \in \Omega \}\,.$$

Величина $|\ell(x,g)|$ является расстоянием от точки x до гиперплоскости $H(g)$. Классификатор ℓ идентифицирует точку $d \in \mathcal{A} \cup \mathcal{B}$ как точку множества $\mathcal{A}$, если $\ell(d,g) > 0$, и как точку множества $\mathcal{B}$, если $\ell(d,g) < 0$.

В качестве другого семейства классификаторов рассмотрим множество

$$\Psi = \big\{ \psi(x,\lambda) \mid \lambda \in \Lambda \big\},$$
$$\Lambda = \big\{ \lambda = (z_1, z_2, z_3, z_4) \in \mathbb{R}^{4n} \mid z_k \in \mathbb{R}^n,\ k = 1:4 \big\},$$

где $x \in \mathbb{R}^n$,

$$\psi(x,\lambda) = \min\big\{ \|x - z_3\|, \|x - z_4\| \big\} - \min\big\{ \|x - z_1\|, \|x - z_2\| \big\}.$$

Классификатор ψ идентифицирует точку $d \in \mathcal{A} \cup \mathcal{B}$ как точку множества $\mathcal{A}$, если $\psi(d,\lambda) > 0$, т. е. она находится ближе всего к одной из точек z_1 или z_2, и как точку множества $\mathcal{B}$, если $\psi(d,\lambda) < 0$, т. е. точка d ближе всего к одной из точек z_3 или z_4.

Далее будет идти речь только о семействе классификаторов $\mathcal{L}$.

3°. Постановка задачи. С помощью классификатора $\ell \in \mathcal{L}$, которому взаимнооднозначно сопоставляется вектор g из Ω, можно осуществить разбиение индексных множеств I и J, а именно $\mathrm{I} = \mathrm{I}_-(g) \cup \mathrm{I}_0(g) \cup \mathrm{I}_+(g)$ и $\mathrm{J} = \mathrm{J}_-(g) \cup \mathrm{J}_0(g) \cup \mathrm{J}_+(g)$, где

$$\begin{array}{ll}
\mathrm{I}_-(g) = \{ i \in \mathrm{I} \mid \ell(a_i, g) < 0 \}\,, & \mathrm{J}_-(g) = \{ j \in \mathrm{J} \mid \ell(b_j, g) > 0 \}\,, \\
\mathrm{I}_0(g) = \{ i \in \mathrm{I} \mid \ell(a_i, g) = 0 \}\,, & \mathrm{J}_0(g) = \{ j \in \mathrm{J} \mid \ell(b_j, g) = 0 \}\,, \\
\mathrm{I}_+(g) = \{ i \in \mathrm{I} \mid \ell(a_i, g) > 0 \}\,, & \mathrm{J}_+(g) = \{ j \in \mathrm{J} \mid \ell(b_j, g) < 0 \}\,.
\end{array}$$

Здесь $\mathrm{I}_-(g)$ и $\mathrm{J}_-(g)$ — индексные множества, состоящие из номеров элементов множеств $\mathcal{A}$ и $\mathcal{B}$ соответственно, которые *ошибочно* идентифицированы; $\mathrm{I}_0(g)$ и $\mathrm{J}_0(g)$ — неидентифицируемые; $\mathrm{I}_+(g)$ и $\mathrm{J}_+(g)$ — правильно идентифицированы.

Качество классификатора $\ell \in \mathcal{L}$ можно измерить, например, количеством ошибочно или неправильно идентифицированных точек. Для этого рассмотрим следующие критерии (*натуральные* критериальные функционалы):

$$K_1(g) := |\mathrm{I}_-(g)| + |\mathrm{J}_-(g)|,$$
$$K_2(g) := |\mathrm{I}_-(g)| + |\mathrm{J}_-(g)| + |\mathrm{I}_0(g)| + |\mathrm{J}_0(g)|,$$
$$K_3(g) := \frac{1}{s}\Big(|\mathrm{I}_-(g)| + |\mathrm{I}_0(g)|\Big) + \frac{1}{m}\Big(|\mathrm{J}_-(g)| + |\mathrm{J}_0(g)|\Big),$$
$$K_4(g) := \max\Big\{\frac{1}{s}|\mathrm{I}_-(g)|,\ \frac{1}{m}|\mathrm{J}_-(g)|\Big\},$$

где $|\mathcal{C}|$ — число элементов множества $\mathcal{C}$.

Таким образом, задача идентификации может быть формализована так: *найти такое g^* из Ω (или $\ell^* \in \mathcal{L}$), что*

$$K(g^*) = \min_{g\in\Omega} K(g).$$

Здесь K — один из вышеперечисленных критериев.

Поясним постановку задачи идентификации в случае выбора натурального функционала K_2. Требуется найти такое положение гиперплоскости $H(g)$, чтобы количество неправильно идентифицируемых точек было как можно меньше.

Очевидно, что натуральные функционалы разрывные, поэтому их исследование затруднительно. Указанная проблема преодолевается построением аппроксимации натурального функционала, называемого в дальнейшем "суррогатным" функционалом, для которого применимы как гладкие, так и негладкие методы оптимизации.

4°. Суррогатные функционалы. Выберем $\varepsilon > 0$ и рассмотрим следующие функции:

$$F_1(g) = \sum_{i\in\mathrm{I}} [-\ell(a_i,g)]_+ + \sum_{j\in\mathrm{J}} [\ell(b_j,g)]_+\,,$$
$$F_2(g) = \sum_{i\in\mathrm{I}} \Big([-\ell(a_i,g)]_+\Big)^2 + \sum_{j\in\mathrm{J}} \Big([\ell(b_j,g)]_+\Big)^2,$$
$$F_{3,\varepsilon}(g) = \sum_{i\in\mathrm{I}} \left[\frac{-\ell(a_i,g)}{|\ell(a_i,g)|+\varepsilon}\right]_+ + \sum_{j\in\mathrm{J}} \left[\frac{\ell(b_j,g)}{|\ell(b_j,g)|+\varepsilon}\right]_+,$$
$$F_{4,\varepsilon}(g) = \sum_{i\in\mathrm{I}} \left[\frac{-\ell(a_i,g)+\varepsilon}{|\ell(a_i,g)|+\varepsilon}\right]_+ + \sum_{j\in\mathrm{J}} \left[\frac{\ell(b_j,g)+\varepsilon}{|\ell(b_j,g)|+\varepsilon}\right]_+.$$

Функции F_1 и F_2 имеют следующую "геометрическую" интерпретацию. Значение функции F_1 при конкретном g соответствует сумме расстояний от ошибочно идентифицируемых точек до гиперплоскости $H(g)$, а для функции F_2 — сумме квадратов расстояний.

З а м е ч а н и е. При $g \in \Omega$ справедливо

$$F_{3,\varepsilon}(g) \underset{\varepsilon\downarrow 0}{\longrightarrow} |\mathrm{I}_-(g)| + |\mathrm{J}_-(g)| = K_1(g).$$

А также для достаточно малых $\varepsilon > 0$ имеем

$$F_{4,\varepsilon}(g) = |\mathrm{I}_-(g)| + |\mathrm{I}_0(g)| + |\mathrm{J}_-(g)| + |\mathrm{J}_0(g)| = K_2(g).$$

5°. Условие оптимальности для функционала F_1 на множестве Ω. Рассмотрим задачу минимизации функционала F_1 при ограничении $\|w\| = 1$. Эту задачу условной оптимизации сведем к задаче безусловной минимизации с помощью точных штрафных функций (см. [4]), т. е.

$$\Phi_\lambda(g) := F_1(g) + \lambda\varphi(g) \longrightarrow \min_{g\in\mathbb{R}^{n+1}},$$

где $\varphi(g) = \Big|\|w\| - 1\Big|$, λ — фиксированная положительная величина превосходящая точный штрафной параметр λ^*.

Ясно, что функции F_1 и φ являются субдифференцируемыми (см. [5, 6]). При этом их субдифференциалы равны

$$\partial F_1(g) = \sum_{i\in \mathrm{I}_-(g)} \begin{pmatrix} -a_i \\ -1 \end{pmatrix} + \sum_{i\in \mathrm{I}_0(g)} \operatorname{conv}\left\{\mathbf{0}_{n+1}, \begin{pmatrix} -a_i \\ -1 \end{pmatrix}\right\} +$$
$$+ \sum_{j\in \mathrm{J}_-(g)} \begin{pmatrix} b_j \\ 1 \end{pmatrix} + \sum_{j\in \mathrm{J}_0(g)} \operatorname{conv}\left\{\mathbf{0}_{n+1}, \begin{pmatrix} b_j \\ 1 \end{pmatrix}\right\},$$

$$\partial\varphi(g) = \begin{cases} \begin{pmatrix} w/\|w\| \\ 0 \end{pmatrix}, & \text{если } \|w\| > 1, \\ \begin{pmatrix} -w/\|w\| \\ 0 \end{pmatrix}, & \text{если } \|w\| < 1, \\ \operatorname{conv}\left\{\begin{pmatrix} -w \\ 0 \end{pmatrix}, \begin{pmatrix} w \\ 0 \end{pmatrix}\right\}, & \text{если } \|w\| = 1. \end{cases}$$

Здесь $\operatorname{conv}\mathcal{C}$ — выпуклая оболочка множества $\mathcal{C}$.

Запишем общий вид субградиента v для функции $\varphi(g)$ при $g \in \Omega$. В этом случае $v \in \operatorname{conv}\left\{\begin{pmatrix} -w \\ 0 \end{pmatrix}, \begin{pmatrix} w \\ 0 \end{pmatrix}\right\}$, что эквивалентно

$$v(\gamma) = (1-\gamma)\begin{pmatrix} -w \\ 0 \end{pmatrix} + \gamma\begin{pmatrix} w \\ 0 \end{pmatrix} = (2\gamma - 1)\begin{pmatrix} w \\ 0 \end{pmatrix} \text{ при } \gamma \in [0, 1].$$

Аналогично поступая, выпишем общий вид субградиента v для функции $F_1(g)$ при $g \in \Omega$. Положим $\mu_i \in [0,1]$, $i \in \mathrm{I}_0(g)$, $\nu_j \in [0,1]$, $j \in \mathrm{J}_0(g)$. Имеем

$$v(\mu,\nu) = \sum_{i \in \mathrm{I}_-(g)} \begin{pmatrix} -a_i \\ -1 \end{pmatrix} + \sum_{i \in \mathrm{I}_0(g)} \mu_i \begin{pmatrix} -a_i \\ -1 \end{pmatrix} + \sum_{j \in \mathrm{J}_-(g)} \begin{pmatrix} b_j \\ 1 \end{pmatrix} + \sum_{j \in \mathrm{J}_0(g)} \nu_j \begin{pmatrix} b_j \\ 1 \end{pmatrix}.$$

Таким образом, каждый субградиент v функции $\Phi_\lambda(g)$ при $g \in \Omega$ может быть представлен так:

$$v(\mu,\nu,\widetilde{\lambda}) = \sum_{i \in \mathrm{I}_-(g)} \begin{pmatrix} -a_i \\ -1 \end{pmatrix} + \sum_{i \in \mathrm{I}_0(g)} \mu_i \begin{pmatrix} -a_i \\ -1 \end{pmatrix} + \sum_{j \in \mathrm{J}_-(g)} \begin{pmatrix} b_j \\ 1 \end{pmatrix} + \sum_{j \in \mathrm{J}_0(g)} \nu_j \begin{pmatrix} b_j \\ 1 \end{pmatrix} + \widetilde{\lambda} \begin{pmatrix} w \\ 0 \end{pmatrix}, \tag{1}$$

где $\widetilde{\lambda} \in [-\lambda, \lambda]$, $\mu_i \in [0,1]$, $i \in \mathrm{I}_0(g)$, $\nu_j \in [0,1]$, $j \in \mathrm{J}_0(g)$.

Известно, что для того чтобы точка $g^* \in \mathbb{R}^{n+1}$ являлась точкой минимума функционала Φ_λ на $\mathbb{R}^{n+1}$, необходимо выполнение условия

$$\mathbf{0}_{n+1} \in \partial \Phi_\lambda(g^*).$$

В силу представления (1), перепишем необходимое условие минимума функционала Φ_λ покоординатно.

УТВЕРЖДЕНИЕ 1. *Для того чтобы точка $g^* \in \mathbb{R}^{n+1}$ доставляла минимальное значение функционалу $F_1(g)$ на множестве Ω, необходимо, чтобы нашлись $\mu_i^* \in [0,1]$, $i \in \mathrm{I}_0(g^*)$, $\nu_j^* \in [0,1]$, $j \in \mathrm{J}_0(g^*)$, $\widetilde{\lambda}^* \in [-\lambda, \lambda]$, при которых выполнялись следующие соотношения:*

$$-\sum_{i \in \mathrm{I}_-(g^*)} a_i - \sum_{i \in \mathrm{I}_0(g^*)} \mu_i^* a_i + \sum_{j \in \mathrm{J}_-(g^*)} b_j + \sum_{j \in \mathrm{J}_0(g^*)} \nu_j^* b_j + \widetilde{\lambda}^* w^* = \mathbf{0}_n, \tag{2}$$

$$-\left|\mathrm{I}_-(g^*)\right| - \sum_{i \in \mathrm{I}_0(g^*)} \mu_i^* + \left|\mathrm{J}_-(g^*)\right| + \sum_{j \in \mathrm{J}_0(g^*)} \nu_j^* = 0. \tag{3}$$

СЛЕДСТВИЕ 1. *Пусть $\mathrm{I}_0(g^*) = \varnothing$, $\mathrm{J}_0(g^*) = \varnothing$. Тогда из (2) и (3) следует, что*

$$\widetilde{\lambda}^* w^* = \sum_{i \in \mathrm{I}_-(g^*)} a_i - \sum_{j \in \mathrm{J}_-(g^*)} b_j =: r,$$

$$\left|\mathrm{I}_-(g^*)\right| = \left|\mathrm{J}_-(g^*)\right|.$$

То есть вектор нормали w^* коллинеарен вектору r, а количество ошибочно классифицированных элементов множеств $\mathcal{A}$ и $\mathcal{B}$ совпадают.

6°. Условие оптимальности для функционала $F_{3,\varepsilon}$ на множестве Ω. Осуществив анализ задачи минимизации суррогатного функционала $F_{3\varepsilon}$ на множестве Ω аналогично тому, как было сделано в предыдущем пункте, получим следующее утверждение (см. [3]).

УТВЕРЖДЕНИЕ 2. *Для того чтобы точка $g^* \in \mathbb{R}^{n+1}$ доставляла минимальное значение функционалу $F_{3\varepsilon}(g)$ на множестве Ω, необходимо, чтобы нашлись $\mu_i^* \in [0,1]$, $i \in \mathrm{I}_0(g^*)$, $\nu_j^* \in [0,1]$, $j \in \mathrm{J}_0(g^*)$, $\widetilde{\lambda}^* \in [-\lambda, \lambda]$, при которых выполнялись следующие соотношения:*

$$\widetilde{\lambda}^* w^* = \sum_{i \in \mathrm{I}_-(g^*)} \frac{\varepsilon a_i}{(\varepsilon - \ell(a_i, g^*))^2} + \sum_{i \in \mathrm{I}_0(g^*)} \mu_i^* \frac{a_i}{\varepsilon} - \sum_{j \in \mathrm{J}_-(g^*)} \frac{\varepsilon b_j}{(\ell(b_j, g^*) + \varepsilon)^2} - \sum_{j \in \mathrm{J}_0(g^*)} \nu_j^* \frac{b_j}{\varepsilon},$$

$$\varepsilon \sum_{i \in \mathrm{I}_-(g^*)} \frac{1}{(\varepsilon - \ell(a_i, g^*))^2} + \frac{1}{\varepsilon} \sum_{i \in \mathrm{I}_0(g^*)} \mu_i^* = \varepsilon \sum_{j \in \mathrm{J}_-(g^*)} \frac{1}{(\ell(b_j, g^*) + \varepsilon)^2} + \frac{1}{\varepsilon} \sum_{j \in \mathrm{J}_0(g^*)} \nu_j^*. \quad (4)$$

СЛЕДСТВИЕ 2. *Пусть $\mathrm{I}_0(g^*) = \varnothing$, $\mathrm{J}_0(g^*) = \varnothing$. Тогда из (4) следует, что сумма квадратов обратных расстояний (с точностью ε) ошибочно идентифицированных точек множеств $\mathcal{A}$ и $\mathcal{B}$ совпадают.*

Работа выполнена в Институте проблем машиноведения РАН при финансовой поддержке Российского научного фонда (проект № 20-71-10032).

ЛИТЕРАТУРА

1. Demyanov V. F. *Mathematical diagnostics via nonsmooth analysis* // Optimization Methods and Software. 2005. Vol. 20. No 2–3. P. 191–212. DOI 10.1080/10556780512331318236.

2. Ананьев К. И., Демьянова В. В., Демьянов В. Ф., Кокорина А. В., Свистун С. Я., Стегалин И. С. *Оптимизационные методы в задачах диагностики* // Вестн. С.-Петербург. ун-та. Сер. 10. Прикл. матем. Информ. Проц. упр. 2011. № 3. С. 3–12.

3. Григорьева К. В. *Суррогатные функционалы в задачах диагностики* // Вестник Санкт-Петербургского университета. Прикладная математика. Информатика. Процессы управления. 2009. № 1. С. 33–43.

4. Долгополик М. В. *Обзор по точным штрафным функциям* // Семинар «CNSA & NDO». Избранные доклады. 2 октября 2014 г. (`http://apmath.spbu.ru/cnsa/reps14.shtml#1002`)

5. Демьянов В. Ф., Малозёмов В. Н. *Введение в минимакс.* М.: Наука, 1972. 368 с.

6. Демьянов В. Ф., Рубинов А. М. *Основы негладкого анализа и квазидифференциальное исчисление.* М.: Наука, 1990. 431 с.