

Галкін О. А., к.ф.-м.н., м.н.с.

Дослідження процедур розпізнавання на основі розподільного розташування багатовимірних даних та функцій глибини

Київський національний університет імені
Тараса Шевченка, 03680, м. Київ, пр-т.
Академіка Глушкова 4д,
e-mail: galkin.o.a@gmail.com

O. A. Galkin, candidate of physical and mathematical
sciences, junior researcher

The research of recognition procedures based on the distributive location of multidimensional data and depth functions

Taras Shevchenko National University of Kyiv,
03680, Kyiv, Glushkova st., 4d,
e-mail: galkin.o.a@gmail.com

Стаття присвячена дослідженню непараметричних методів класифікації та їх асимптотичних властивостей на основі використання функцій напівпросторової та регресійної глибини, що використовуються для побудови лінійних та нелінійних розділових поверхонь серед конкуруючих множин даних. Запропоновано повністю незалежний від розподілу підхід, де для мінімізації коефіцієнтів помилкової класифікації застосовується розподільне розташування багатовимірних даних. Реалізовано оптимізаційний алгоритм для вирішення проблеми наявності декількох локальних мінімумів.

Ключові слова: функція глибини, розділова гіперплощина, байєсівський класифікатор.

Linear and quadratic classifiers are optimal Bayesian classifiers under the multivariate normal distribution based on simple parametric surfaces, and provide effective low-field class resolution. However, these methods are not reliable enough and sometimes are quite sensitive to extreme values and reductions in sample. In addition, these methods may not lead to correct classification if the data correspond to a certain distribution with heavy tails. This usually occurs when the assumption of multivariate normal distribution of data is violated. Most non-parametric methods do not use the information about the distribution of data and forms of dividing surface, but such flexible methods of statistical analysis for recognition problems sometimes tend to overfitting of training data and their performance can decrease significantly in the absence of large data sets. The article is devoted to research nonparametric classification methods and their asymptotic properties through the use of half-space and regression depth functions that are used to build linear and nonlinear separating surface between competing sets of data. A fully distribution free approach is proposed where distributive location of multidimensional data is used to minimize misclassification coefficients. An optimization algorithm was implemented for solving the problem of the presence of several local minima.

Key Words: depth function, separating hyperplane, Bayesian classifier.

Статтю представив д.ф.-м.н., проф. Анісімов А.В.

Вступ

Лінійні та квадратичні класифікатори на основі простих параметричних поверхонь є оптимальними байєсівськими класифікаторами в рамках багатовимірного нормального розподілу та забезпечують ефективне низькорозмірне поле класової роздільності. Однак, вказані класифікатори не є достатньо надійними та є досить чутливими до екстремальних значень та викидів, що наявні в вибірці даних. Це є наслідком того, що дані процедури потребують

оцінок проекційного вектора φ , постійної величини γ та симетричної матриці M , використовуючи моменти першого та другого порядку вибірки. Крім того, дані методи можуть призвести до некоректної класифікації, якщо дані відповідають певному розподілу з важкими хвостами. Як правило, це має місце, коли порушується припущення багатовимірного нормального розподілу даних. Більшість непараметричних методів не використовують інформацію щодо розподілу даних та форми

розділової поверхні, проте такі гнучкі методи статистичного аналізу для задач розпізнавання іноді мають тенденцію перенавчання на навчальних даних, а їх продуктивність може суттєво знизитись при відсутності великих множин даних.

У даній статті досліджуються деякі непараметричні методи класифікації на основі концепцій функцій напівпросторової та регресійної глибини, що використовуються для побудови лінійних та нелінійних розділових поверхонь серед конкуруючих множин даних. Досліджувані інструменти методів статистичного аналізу для задач розпізнавання є повністю незалежними від розподілу, оскільки ґрунтуються на функціях напівпросторової та регресійної глибини. Для мінімізації коефіцієнтів помилкової класифікації, дані класифікатори застосовують розподільне розташування багатовимірних даних. Крім того, вони є незалежними від будь-яких моделей для основних розподілів множин даних.

Виклад основного матеріалу

У багатовимірному просторі функція напівпросторової глибини визначає занурення певного елемента даних відносно багатовимірного розподілу, або заданих багатовимірних даних. Концепція регресійної глибини є поняттям глибини регресійної відповідності. Найпростішими формами розділових поверхонь є гіперплощини, що забезпечують лінійну роздільність між класами [1].

Далі у статті досліджуються процедури побудови вказаних глибинних інструментів лінійної класифікації на основі заданої навчальної вибірки з двома класами.

Напівпросторова глибина r -вимірних даних z відносно багатовимірного розподілу F знаходиться, як мінімальна ймовірність замкнутого напівпростору, що містить z , а саме:

$$\text{Нпр.Гл.}(z, F) = \inf_S P_F \{S\}, \quad (1)$$

де S є замкнутим напівпростором, а $z \in S$.

Емпіричний варіант даної функції глибини отриманий шляхом заміни F на емпіричну функцію розподілу F_m . Емпірична версія функції напівпросторової глибини, що є афінно-інваріантною, рівномірно збігається до функції глибини, коли F є неперервною функцією.

Розглянемо двокласову задачу з одномірними даними. Можна очікувати, що більшість різниць

$z_{1i} - z_{2j}$, де z_{1i} та z_{2j} належать до двох різних класів для $1 \leq i \leq m_1$ та $1 \leq j \leq m_2$, будуть мати однаковий знак у випадку, якщо класи є ефективно розділеними. Такий підхід може бути розвинутий до багатовимірних випадків, де можна отримати лінійну проекцію $\varphi'z$, для якої більшість різниць $\varphi'z_{1i} - \varphi'z_{2j}$ будуть мати однаковий знак. Це має місце лише у випадку, коли два класи можуть бути ефективно розділеними за допомогою лінійної роздільної функції [2].

Ми будемо оцінювати φ , мінімізуючи

$$\Phi_m(\varphi) = \frac{1}{m_1 m_2} \sum_{i=1}^{m_1} \sum_{j=1}^{m_2} \Lambda\{\varphi'(z_{1i} - z_{2j}) > 0\}, \quad (2)$$

де $m = (m_1, m_2)$ є вектором розмірів вибірки для двох класів. Задача максимізації може бути обмежена на множині $\{\varphi: \|\varphi\|=1\}$, що є задачею максимізації на скінченній вибірці. Зауважимо, що оцінена лінійна проекція є ортогональною до гіперплощини, що визначає напівпросторову глибину координат відносно даних, утворених різницями $z_{1i} - z_{2j}$ в r -вимірному просторі. Дана узагальнена незміщена статистика $\Phi_m(\varphi)$ є мірою лінійної роздільності між двома класами вздовж напрямку φ , а її максимальне значення серед різних можливих значень φ може розглядатися, як багатовимірна модель одномірної незміщеної статистики Манна-Уїтні.

Максимізуюча оцінка $\Phi_m(\varphi)$, позначена як $\bar{\varphi}_{\text{нпр}}$ може бути використана для побудови лінійного класифікатора $\bar{\varphi}'_{\text{нпр}}z + \gamma = 0$ для деякої відповідно вибраної постійної величини γ . Слід зауважити, що правило класифікації, а отже і відповідні ймовірності помилкової класифікації залежать від вибору цієї постійної величини.

Після знаходження оцінки $\bar{\varphi}_{\text{нпр}}$, $\bar{\gamma}_{\text{нпр}}$ може бути отримана шляхом мінімізації відносно γ середньої помилки помилкової класифікації $\Psi_{\text{нпр}}(\bar{\varphi}, \gamma)$, що задається виразом

$$\Psi_{\text{нпр}}(\bar{\varphi}, \gamma) = \frac{p_1}{m_1} \sum_{i=1}^{m_1} \Lambda\{\bar{\varphi}' z_{1i} + \gamma < 0\} + \frac{p_2}{m_2} \sum_{i=1}^{m_2} \Lambda\{\bar{\varphi}' z_{2i} + \gamma > 0\}, \quad (3)$$

де p_1 та p_2 є апіорними ймовірностями для двох класів.

Функція регресійної глибини визначає глибину занурення даних, що визначається за

допомогою вектора коефіцієнтів $\varphi_+ = (\varphi, \gamma) \in R^{r+1}$ в лінійній регресійній структурі. Маючи множину даних

$$\Delta_m = [\{z_i = (z_{i1}, z_{i2}, \dots, z_{ir}), y_i\}; i = 1, 2, \dots, m]$$

з m елементів даних, величина φ_+ буде незануренням до Δ_m тоді і тільки тоді, якщо існує така афінна гіперплощина S в z -просторі, що жоден z_i не належить до S , а остачі $y_i - \varphi'_+(z_i, 1)$ є всі додатні в одному відкритому напівпросторі (тобто, одна сторона S) в z -просторі та всі від'ємні в доповненні відкритого напівпростору (тобто, інша сторона S). Функція регресійної глибини занурення φ_+ визначається, як мінімальне число елементів даних, що повинні бути видалені для перетворення її в незанурення [3].

Такий підхід регресійної глибини був використаний для двокласових задач в контексті бінарної регресії для побудови лінійних класифікаторів [4]. Використовуючи мітки класу 0 або 1, як значення залежної змінної y та досліджуючи незанурення $\varphi_+ = (0, 0, \dots, 0, 0.7)$, величина φ_+ буде незануренням до Δ_m тоді і тільки тоді, якщо існує гіперплощина S в z -просторі, що повністю розділяє дані з двох класів. Як результат, регресійна глибина незанурення φ_+ може розглядатися, як мінімальне число помилково класифікованих даних, що може бути досягнуто за допомогою розділової гіперплощини S в z -просторі.

Оскільки в спеціалізованій літературі було досліджено лише задачу визначення розділової гіперплощини шляхом мінімізації загального числа помилково класифікованих даних [5], можна стверджувати, що лінійний класифікатор є емпірично оптимальним, коли два конкуруючих класи мають апріорні ймовірності, що пропорційні їх розмірам вибірки. На практиці необхідно коректно налаштувати ваги для різних даних та визначити виважену регресійну глибину занурення φ_+ , як мінімальне число ваг, що повинні бути видалені для перетворення її в незанурення. Тоді, виважена регресійна глибина в кінцевому підсумку стає середньою ймовірністю помилкової класифікації, що має наступний вигляд:

$$\Psi_m(\varphi, \gamma) = \frac{p_1}{m_1} \sum_{i=1}^{m_1} \Lambda\{\varphi' z_{1i} + \gamma < 0\} + \frac{p_2}{m_2} \sum_{i=1}^{m_2} \Lambda\{\varphi' z_{2i} + \gamma > 0\}. \quad (4)$$

Мінімізація $\Psi_m(\varphi, \gamma)$ відносно φ та γ дозволяє отримати оцінки $\bar{\varphi}_{регр}$ та $\bar{\gamma}_{регр}$, що визначають розділову гіперплощину для використання в задачах класифікації. Задача мінімізації може бути обмежена до $\{(\varphi, \gamma) : \|(\varphi, \gamma)\| = 1\}$. Також можна перевірити, що мінімізація $\Psi_m(\varphi, \gamma)$ виявляється оптимізаційною задачею на скінченній множині.

Крім того, було досліджено той факт, що оцінка максимальної ймовірності в задачі логістичної регресії існує лише тоді, коли є деякі збіги в коваріативному просторі між даними з двох класів, що відповідають значенням 0 та 1 залежної змінної. У випадках абсолютної роздільності не існує скінченної оцінки максимальної ймовірності для вектора регресійних коефіцієнтів φ_+ . Якщо дані з двох класів є повністю роздільними, можна побачити, що $(\bar{\varphi}_{регр}, \bar{\gamma}_{регр})$ є мінімізуючою оцінкою $\Psi_m(\varphi, \gamma)$ тоді і тільки тоді, коли $\bar{\varphi}_{регр}$ максимізує $\Phi_m(\varphi)$, а тому $\bar{\varphi}_{регр}$ є також $\bar{\varphi}_{нпр}$.

У випадку, коли межі класу є досить складними, лінійні класифікатори можуть бути недоцільними для використання. В такому разі, для розрізнення між конкуруючими класами необхідна залежність від нелінійних розділових поверхонь. Для побудови таких поверхонь, можна спроектувати дані $z_i = (z_{i1}, z_{i2}, \dots, z_{id})$ в простір більш високої розмірності для отримання нового вектора вимірюваних величин $t_i = (\omega_1(z_i), \omega_2(z_i), \dots, \omega_r(z_i))$ та виконати лінійну класифікацію в l -вимірному просторі.

Якщо ми проектуємо дані в простір квадратичних функцій, це може розглядатися, як лінійна класифікація з $l = r + \binom{r}{2}$ вимірюваними величинами, які в кінцевому підсумку призводять до квадратичного розділення в початковому r -вимірному просторі. Величини

$$\Phi_m(\varphi) = \frac{1}{m_1 m_2} \sum_{i=1}^{m_1} \sum_{j=1}^{m_2} \Lambda\{\varphi'(t_{1i} - t_{2j}) > 0\} \quad (5)$$

та

$$\Psi_m(\varphi, \gamma) = \frac{p_1}{m_1} \sum_{i=1}^{m_1} \Lambda\{\varphi' t_{1i} + \gamma < 0\} + \frac{p_2}{m_2} \sum_{i=1}^{m_2} \Lambda\{\varphi' t_{2i} + \gamma > 0\} \quad (6)$$

можуть бути оптимізовані для отримання відповідних оцінок φ та γ , які

використовуються для формування розділової поверхні в двокласовій задачі.

Традиційні лінійні та квадратичні методи статистичного аналізу для задач розпізнавання головним чином мотивовані багатовимірними нормальними розподілами. В двокласовій задачі лінійна розділова функція на основі моменту тісно пов'язана з відстанню Махаланобіса, що є досить чутливою до можливих викидів, присутніх в даних. З іншої сторони, незалежні від розподілу глибинні класифікатори є досить стійкими по відношенню до викидів [6].

Теорема 1. Припустимо, що $h_i(z) = c(z - \varepsilon_i)$, де $i = (1, 2)$ для деяких параметрів локалізації ε_i та загальної еліптично-симетричної функції щільності c , що задовольняє умову $c(kz) \geq c(z)$ для кожного z та $0 < k < 1$. У даному випадку, функції щільності h_1 та h_2 двох конкуруючих класів є еліптично-симетричними із загальною матрицею розсіювання Ξ .

Припускаючи, що апіорні ймовірності двох конкуруючих класів даних є рівними, ми можемо стверджувати, що:

(а) $\varphi = \Xi^{-1}(\varepsilon_1 - \varepsilon_2)$ є максимізуючою оцінкою $\Phi(\varphi)$, а також мінімальною оцінкою $\Psi(\varphi, \gamma)$ при належному виборі значення γ ;

(б) існує оптимальний лінійний байесівський класифікатор.

Доведення. (а) Характеристичними функціями розподілу є функції наступного вигляду:

$$O_{h_1}(\kappa) = e^{i\kappa\varepsilon_1} \alpha(\kappa\Xi\kappa) \quad (7)$$

та

$$O_{h_2}(\kappa) = e^{i\kappa\varepsilon_2} \alpha(\kappa\Xi\kappa) \quad (8)$$

оскільки розподіли мають загальну матрицю розсіювання Ξ та спільну еліптично-симетричну форму з параметрами локалізації ε_1 та ε_2 , а α є деякою загальною скалярною функцією.

Визначаючи

$$X = \varphi' \{ (Z_1 - Z_2) - (\varepsilon_1 - \varepsilon_2) \} / (\varphi\Xi\varphi)^{1/2}, \quad (9)$$

де $Z_1 \approx h_1$ та $Z_2 \approx h_2$, можна побачити, що характеристична функція X задається, як $O_X(\kappa) = \{ \alpha(\kappa^2) \}^2$. Оскільки розподіл X є симетричним відносно нуля та незалежний від таких параметрів множини, як ε та Ξ , $P\{\varphi'(Z_1 - Z_2) > 0\}$ може бути виражена, як

$$P\{\varphi'(Z_1 - Z_2) > 0\} = F_X \left(\left[\frac{\{\varphi'(\varepsilon_1 - \varepsilon_2)\}^2 / \varphi\Xi\varphi}{\varphi} \right]^{1/2} \right) \quad (9)$$

де F_X є кумулятивною функцією розподілу X . В результаті, $P\{\varphi'(Z_1 - Z_2) > 0\}$ матиме максимальне значення для деякого φ , якщо дане φ максимізує $\{\varphi'(\varepsilon_1 - \varepsilon_2)\}^2 / \varphi\Xi\varphi$. Це означає, що $\varphi = \Xi^{-1}(\varepsilon_1 - \varepsilon_2)$ є максимізуючою оцінкою $\Phi(\varphi)$.

(б) Функції щільності h_1 та h_2 можуть бути виражені, як

$$h_1(z) = V_r |\Xi|^{-1/2} \varpi \{ (z - \varepsilon_1)' \Xi^{-1} (z - \varepsilon_1) \} \quad (10)$$

та

$$h_2(z) = V_r |\Xi|^{-1/2} \varpi \{ (z - \varepsilon_2)' \Xi^{-1} (z - \varepsilon_2) \}, \quad (11)$$

що є результатом еліптичної симетрії з локалізаційним зсувом, де ε є монотонно спадною функцією на $[0, \infty)$, а V_r - постійною величиною, що залежить від розмірності r . Оптимальне байесівське правило класифікує дані до пергошо класу тоді і тільки тоді, якщо

$$\begin{aligned} h_1(z) \geq h_2(z) &\Leftrightarrow (z - \varepsilon_1)' \Xi^{-1} (z - \varepsilon_1) \leq \\ &\leq (z - \varepsilon_2)' \Xi^{-1} (z - \varepsilon_2) \Leftrightarrow \\ &\Leftrightarrow (\varepsilon_1 - \varepsilon_2)' \Xi^{-1} z \geq \frac{1}{2} \left[\varepsilon_1' \Xi^{-1} \varepsilon_1 - \varepsilon_2' \Xi^{-1} \varepsilon_2 \right]. \end{aligned} \quad (12)$$

Даний результат має місце у випадку рівних апіорних ймовірностей. Отже, оптимальний лінійний класифікатор є байесівським класифікатором, а $\varphi = \Xi^{-1}(\varepsilon_1 - \varepsilon_2)$ - мінімізуючою оцінкою $\Psi(\varphi, \gamma)$ при належному виборі значення γ .

Теорема 2. Нехай щільності h_1 та h_2 належать до класу з еліптично-симетричним багатовимірним нормальним розподілом, а також вони мають однакову форму, за винятком, можливо, їх параметрів локалізації та розкиду. Тоді, середній коефіцієнт помилкової класифікації квадратичного класифікатора, побудованого з використанням регресійної глибини збігається до оптимальної байесівської помилки, а квадратичний класифікатор збігається майже напевно до оптимального байесівського класифікатора, коли розмір навчальної вибірки прямує до нескінченності.

Доведення. У випадку, коли два конкуруючих розподіли множин даних є багатовимірними нормальними розподілами з параметрами локалізації та розсіювання (ε_1, Ξ_1) та (ε_2, Ξ_2) ,

$$\begin{aligned}
 p_1 h_1(z) &> p_2 h_2(z) \Leftrightarrow \\
 \Leftrightarrow p_1 |\Xi_1|^{-1/2} e^{-\frac{1}{2}(z-\varepsilon_1)' \Xi_1^{-1} (z-\varepsilon_1)} &> \\
 > p_2 |\Xi_2|^{-1/2} e^{-\frac{1}{2}(z-\varepsilon_2)' \Xi_2^{-1} (z-\varepsilon_2)} &\Leftrightarrow \\
 \Leftrightarrow (z-\varepsilon_2)' \Xi_2^{-1} (z-\varepsilon_2) - (z-\varepsilon_1)' \Xi_1^{-1} (z-\varepsilon_1) &> V,
 \end{aligned} \tag{13}$$

$$\text{де } V = 2 \log \left(\frac{p_2 |\Xi_1|^{1/2}}{p_1 |\Xi_2|^{1/2}} \right).$$

Оскільки достатньо показати, що оптимальний квадратичний класифікатор є унікальним байєсівським класифікатором за заданих умов, ми отримали, що оптимальне байєсівське правило є унікальним та квадратичним [7].

Ймовірнісна функція щільності $h(z)$ r -вимірною еліптично-симетричного багатомірною нормального розподілу задається, як

$$h(z) = V_r |\Xi|^{-1/2} \{1 + \mu^{-1} (z - \varepsilon)' \Xi^{-1} (z - \varepsilon)\}^{-U}, \tag{14}$$

де $\mu > 0$, $U > r/2$, $V_r = (p\mu)^{-r/2} I(U)/I(U - r/2)$, а ε та Ξ є параметрами локалізації та розсіювання, відповідно. Далі проведемо дослідження двох однакової форми еліптично-симетричних багатомірних нормальних розподілів за винятком їх параметрів локалізації та розсіювання.

Припустимо, що ε_i є параметром локалізації, а Ξ_i - матрицею розсіювання для i -ї ($i=1,2$) множини даних, а p_i є її апіорною ймовірністю.

Тоді, має місце наступна нерівність:

$$\begin{aligned}
 p_1 h_1(z) &> p_2 h_2(z) \Leftrightarrow \\
 \Leftrightarrow p_1 |\Xi_1|^{-1/2} \{1 + \mu^{-1} (z - \varepsilon_1)' \Xi_1^{-1} (z - \varepsilon_1)\}^{-U} &> \\
 > p_2 |\Xi_2|^{-1/2} \{1 + \mu^{-1} (z - \varepsilon_2)' \Xi_2^{-1} (z - \varepsilon_2)\}^{-U} & \\
 \Leftrightarrow \left\{ \frac{1 + \mu^{-1} (z - \varepsilon_1)' \Xi_1^{-1} (z - \varepsilon_1)}{1 + \mu^{-1} (z - \varepsilon_2)' \Xi_2^{-1} (z - \varepsilon_2)} \right\}^{-U} &> n \text{ для } n = \frac{p_2 |\Xi_2|^{-1/2}}{p_1 |\Xi_1|^{-1/2}} \tag{15} \\
 \Leftrightarrow \left\{ \frac{\mu + (z - \varepsilon_1)' \Xi_1^{-1} (z - \varepsilon_1)}{\mu + (z - \varepsilon_2)' \Xi_2^{-1} (z - \varepsilon_2)} \right\} < n_0 = n^{-1/U} & \\
 \Leftrightarrow (z - \varepsilon_1)' \Xi_1^{-1} (z - \varepsilon_1) - n_0 (z - \varepsilon_2)' \Xi_2^{-1} (z - \varepsilon_2) - (n_0 - 1)\mu < 0. &
 \end{aligned}$$

В результаті, оптимальне байєсівське правило має властивість унікальності та є квадратичним класифікатором, оскільки ліва частина цієї нерівності є квадратичною функцією від z .

Зауважимо, що ймовірнісна функція щільності r -вимірною еліптично-симетричного

багатомірною нормального розподілу задається, як

$$\begin{aligned}
 h(z) &= (p\mu)^{-r/2} I(U)/I(U - r/2) |\Xi|^{-1/2} \times \\
 &\times \{1 + \mu^{-1} (z - \varepsilon)' \Xi^{-1} (z - \varepsilon)\}^{-U}, \tag{16}
 \end{aligned}$$

де ε та Ξ є параметрами локалізації та розкиду, $\mu > 0$, а $U > r/2$. Коли $U = (\mu + r)/2$, а μ є цілим числом, відповідний розподіл є також багатомірним t -розподілом з функцією щільності μ . Як частковий випадок, має місце багатомірний розподіл Коші, коли $\mu = 1$.

Результати практичних експериментів

Максимізація $\Phi_m(\varphi)$ може бути зведена до φ

з $\|\varphi\| = 1$, а мінімізація $\Psi(\varphi, \gamma)$ - до (φ, γ) з $\|(\varphi, \gamma)\| = 1$. Враховуючи те, що порядок обчислювальної складності швидко зростає з розмірністю l , точна оптимізація $\Phi_m(\varphi)$ та $\Psi_m(\varphi, \gamma)$ не може бути застосованою за задачі великої розмірності, тому можна виконувати лише приблизну оптимізацію.

В практичних експериментах було застосовано підхід, де індикаторні функції, що фігурують у виразах для Φ_m та Ψ_m апроксимуються за допомогою відповідно визначених функцій згладжування. Використовуючи вказану апроксимацію, було використано похідні для визначення напрямку найшвидшого підйому, або спуску цільової функції, яку необхідно оптимізувати. Логістична функція $1/(1 + e^{-\kappa z})$ з великим додатнім κ є апроксимацією для індикаторної функції $\Lambda(z > 0)$.

Зрозуміло, що недостатньо велике значення κ призведе до неточної апроксимації. Однак, дуже велике значення κ призведе до достатньо точної апроксимації, однак числова оптимізація на основі найшвидшого підйому, або спуску буде чисельно нестійкою. Встановлено, що високий коефіцієнт чисельної стійкості в оптимізаційному алгоритмі може бути отриманий навіть для дуже великого значення t . Це стало можливим при умові, коли всі вимірювані величини були нормалізованими до процесу апроксимації. На основі практичних експериментів було встановлено, що якщо використовується значення $10 \leq \kappa \leq 15$ після процесу нормалізації вимірюваних величин, середні коефіцієнти помилкової класифікації для отриманих алгоритмів залишаються майже однаковими та достатньо малими. В результаті, були виведені

найбільш прийнятні значення, отримані в цьому діапазоні.

В двовимірному випадку для лінійних методів статистичного аналізу для задач розпізнавання, де точне визначення $\Phi_m(\varphi)$ та $\Psi_m(\varphi, \gamma)$ є досить простим, було порівняно ефективність точних та наближених глибинних методів класифікації, а також встановлено, що вони досягають практично однакових середніх коефіцієнтів помилкової класифікації.

Починаючи з різних випадкових початкових даних, наближені реалізації оптимізаційних алгоритмів виконувалися декілька разів для вирішення проблеми можливої наявності декількох локальних мінімумів.

Стосовно застосування класифікатора на основі напівпросторової глибини, необхідно було оцінити \mathcal{Y} з навчальної вибірки після оцінки φ . Це було реалізовано в результаті перерахування порядкових статистик спроектованих даних $\bar{\varphi}'_{np} t_{1i}$ та $\bar{\varphi}'_{np} t_{2j}$ ($1 \leq i \leq m_1$, $1 \leq j \leq m_2$) уздовж оціненого напрямку $\bar{\varphi}_{np}$. Було зроблено висновок, що обчислювальна складність при отриманні оцінки $\bar{\gamma}_{np}$ не збільшується з розмірністю l у результаті використання лінійних проєкцій.

Список використаних джерел

1. Mizera I. On depth and deep points: a calculus / I. Mizera // *The Annals of Statistics*. – 2002. – 30. – P. 1681-1736.
2. Mosler K. *Multivariate Dispersions, Central Regions and Depth* / K. Mosler. – New York: Springer-Verlag, 2002.
3. Serfling R. A depth function and a scale curve based on spatial quantiles / R. Serfling. – *Statistics and Data Analysis based on L_1 -Norm and Related Methods*. – 2002. – Birkhaeuser, Boston. – P. 25-38.
4. Vardi Y. The multivariate on L_1 -median and associated data depth / Y. Vardi, C.H. Zhang // *Proceedings of the National Academy of Sciences (USA)*. – 2000. – 97. – P. 1423-1426.
5. Zuo Y. Structural properties and convergence results for contours of sample statistical depth functions / Y. Zuo, R. Serfling // *The Annals of Statistics*. – 2000. – 28. – P. 483-499.
6. Godtliebsen F. Significance in scale space for bivariate density estimation / F. Godtliebsen, J.S. Marron, P. Chaudhuri // *Journal of Computational and Graphical Statistics*. – 2002. – 11. – P. 1-22.
7. Holmes C.C. A probabilistic nearest neighbor method for statistical pattern recognition / C.C. Holmes, N.M. Adams // *Journal of the Royal Statistical Society*. – 2002. – 64. – P. 295-306.

References

1. MIZERA, I. (2002) On depth and deep points: a calculus. *The Annals of Statistics*. 30. p. 1681-1736.
2. MOSLER, K. (2002) *Multivariate Dispersions, Central Regions and Depth*. New York: Springer-Verlag.
3. SERFLING, R. (2002) *A depth function and a scale curve based on spatial quantiles*. *Statistics and Data Analysis based on L_1 -Norm and Related Methods*. Birkhaeuser, Boston. p. 25-38.
4. VARDI, Y. & ZHANG, C.H. (2000) The multivariate on L_1 -median and associated data depth. *Proceedings of the National Academy of Sciences (USA)*. 97. p.1423-1426.
5. ZUO, Y., SERFLING R. (2000) Structural properties and convergence results for contours of sample statistical depth functions. *The Annals of Statistics*. 28. p. 483-499.
6. GODTLIEBSEN, F., MARRON J.S., CHAUDHURI P. (2002) Significance in scale space for bivariate density estimation. *Journal of Computational and Graphical Statistics*. 11. p. 1-22.
7. HOLMES, C.C., ADAMS N.M. (2002) Significance in scale space for bivariate density estimation. *Journal of the Royal Statistical Society*. 64. p. 295-306.

Надійшла до редколегії 17.04.15