Національна академія наук України
Інститут кібернетики ім. В.М. Глушкова

Препринт 2009 – 1

МЕТОДИ НЕГЛАДКОЇ ОПТИМІЗАЦІЇ У СПЕЦІАЛЬНИХ ЗАДАЧАХ КЛАСИФІКАЦІЇ

Київ – 2009

УДК 519.8

Автори
П.І. Стецюк, О.А. Березовський, М.Г. Журбенко, Д.О. Кропотов

Методи негладкої оптимізації у спеціальних задачах класифікації / П.І. Стецюк, О.А. Березовський, М.Г. Журбенко, Д.О. Кропотов. – Київ, 2009. – 28 с. – (Препр. / НАН України. Ін-т кібернетики ім. В.М.Глушкова; 2009–1).

Розглянуто математичні моделі задачі класифікації, задач побудови лінійних і квадратичних дискримінантних функцій та задачі кодування монохроматичних зображень. Математичні моделі представлені у вигляді дискретних, негладких та багатоекстремальних задач оптимізації. Вони орієнтовані при їх розв’язуванні на використання субградієнтних методів мінімізації недиференційовних функцій. Моделі та розроблені методи можуть бути застосовані при розв'язуванні окремих задач розпізнавання й аналізу даних.

Для спеціалістів в області математичного програмування та застосувань чисельних методів оптимізації.

Іл. 1, Табл. 1. Бібліогр.: с. 26–27 (18 найменувань).

Рецензент чл.-кор. НАН України А.М. Гупал

Затверджено до друку науковою радою НАН України з проблеми “Кібернетика”

ВСТУП

Відомий спеціаліст у теорії розпізнавання образів М.І. Шлезінгер зазначав [1]: «...вряд ли в какой-то другой области различные разделы прикладной математики приходят в столь тесное соприкосновение, как в распознавании. Поэтому распознавание образов может служить испытательным полигоном…в прикладной информатике». Враховуючи, що на цьому «полігоні» виникає цілий ряд задач, які можуть бути представлені у вигляді задач недиференційовної оптимізації (наприклад, задача Андерсона [1]), природною виглядає зацікавленість у розробці відповідних чисельних алгоритмів для їхнього розв'язання на основі досягнень в області негладкої оптимізації.

На сьогодні розроблені ефективні субградієнтні та ε-субградієнтні методи мінімізації негладких опуклих функцій і накопичено значний досвід їх використання в різних областях застосувань. Значне місце серед цих методів займають субградієнтні методи з розтягом простору [2, 3, 4], до яких належать і *r*-алгоритми, що виявилися конкурентоздатними з найбільш вдалими реалізаціями методів спряжених напрямків та методів квазіньютонівського типу. Важливу роль методи недиференційовної оптимізації відіграють при розробці алгоритмів для вирішення широкого класу складних дискретних та багатоекстремальних оптимізаційних задач на основі методу гілок та границь. Використання структурних особливостей таких задач може істотно сприяти підвищенню ефективності алгоритмів, що розробляються для розв'язування "оціночних" задач для методу гілок та границь. Далі розглянемо декілька таких задач, які можуть бути використані при аналізі даних у задачах розпізнавання об'єктів зі складними зв'язками*.

У першому розділі розглянута задача класифікації, що відіграє значну роль при розбитті множини об'єктів на задану кількість класів. Вона сформу-

*Робота виконана в рамках спільного конкурсу НАН України та Російського фонду фундаментальних досліджень 2008 р. при фінансовій підтримці НАН України (Постанова Президії НАН України від 02.04.2008 №104 , проект №4).

льована у вигляді задачі цілочислового лінійного програмування, яка має блочну структуру матриці обмежень. Роль негладкої оптимізації відведено обчисленню ефективних нижніх оцінок для цільової функції, які враховують структурні особливості задачі при використанні схеми декомпозиції по різних групах обмежень. Для знаходження нижніх оцінок рекомендується використовувати субградієнтні та ε-субградієнтні методи мінімізації негладких опуклих функцій.

У другому розділі сформульовані математичні моделі для побудови лінійних та квадратичних дискримінантних функцій (розділяючих гіперповерхонь для двох скінченних підмножин точок) у вигляді задач квадратичного програмування та негладкого опуклого програмування з обмеженнями, що використовують негладкі матричні функції (максимальне або мінімальне власне число симетричної матриці). Для лінійних дискримінантних функцій розглянуто задачу побудови розділяючої гіперплощини з максимальним зазором та задачу робастного розділення. Для квадратичних дискримінантних функцій розглянуто задачі розділення довільною опуклою або увігнутою квадратичними функціями. Для спеціальної задачі розділення m $(m > 2)$ множин точок паралельними гіперплощинами сформульована математична модель у вигляді задачі негладкого опуклого програмування.

У третьому розділі розглянута задача представлення монохроматичних зображень у вигляді неопуклої квадратичної моделі. На ній продемонстровано техніку двоїстих лагранжевих оцінок Шора [3, 5] з використанням функціонально надлишкових обмежень для отримання та уточнення нижньої оцінки цільової функції в багатоекстремальних квадратичних задачах. Доведено, що знаходження оцінки Шора зводиться до задачі безумовної опуклої недиференційовної оптимізації. Показано, що розширення базової моделі за допомогою функціонально надлишкових обмежень дозволяє отримати більш точні нижні оцінки.

1. МАТЕМАТИЧНА МОДЕЛЬ ЗАДАЧІ КЛАСИФІКАЦІЇ ТА ЇЇ ЛАГРАНЖЕВІ ОЦІНКИ

Математична модель задачі класифікації наведена у вигляді спеціалізованої задачі цілочислового лінійного програмування. Побудовані дві координуючі задачі недиференційовної оптимізації (які відповідають схемам декомпозиції за різними групами обмежень) для обчислення нижніх лагранжевих оцінок оптимального значення цільової функції. Для знаходження цих оцінок рекомендується використовувати субградієнтні та ε-субградієнтні методи мінімізації негладких опуклих функцій.

1.1. Математична модель. Задачу класифікації розглянемо згідно з роботою [6]. Зміст задачі полягає в наступному.

Нехай задано множину $J = \{1,2,\dots,N\}$ об'єктів та множину $I = \{1,2,\dots,M\}$ класів ($M > 2$). Для кожної пари i та j ($i \in I$, $j \in J$) поставлено у відповідність число $d_{ij} \geq 0$, яке прийнято називати індексом неузгодженості між класом i та об'єктом j. Необхідно кожний з об'єктів із множини J зарахувати до одного з p класів із множини I ($p < M$) так, щоб мінімізувати критерій їхньої взаємної неузгодженості.

Задача класифікації формулюється у вигляді такої задачі цілочислового лінійного програмування:

$$f^* = \min_x \sum_{i \in I} \sum_{j \in J} d_{ij} x_{ij} , \tag{1.1}$$

$$x_{ij} \leq y_i \ \ \forall i \in I , \ \forall j \in J , \tag{1.2}$$

$$\sum_{i \in I} x_{ij} = 1 \ \ \forall j \in J , \tag{1.3}$$

$$\sum_{i \in I} y_i = p , \tag{1.4}$$

$$y_i = 0 \vee 1 \ \ \forall i \in I , \tag{1.5}$$

$$x_{ij} = 0 \vee 1 \ \ \forall i \in I , \forall j \in J . \tag{1.6}$$

Інтерпретація змісту обмежень задачі враховує цілочисельність змінних y (1.5) та x (1.6). Оптимальні значення змінних y_i визначають множину класів $I^* \subset I$. Обмеження (1.4) задає кількість класів, які необхідно вибрати, та означає, що $\left|I^*\right| = p$. Коли клас i належить до p вибраних класів, то $y_i = 1$. Оптимальне значення змінної $x_{ij} = 1$ визначає той об'єкт j, який належить до класу $i \in I^*$. Це забезпечують обмеження (1.2) та (1.3). Обмеження (1.2) означають, що об'єкт j може належати тільки до того класу i, для якого $y_i = 1$. Обмеження (1.3) відображають той факт, що кожний об'єкт має бути пов'язаний тільки з одним класом із I^*.

Задача (1.1)–(1.6) в загальному випадку є *NP*-повною. Ефективність її розв'язування методом гілок та границь визначається алгоритмом знаходження нижніх оцінок для цільової функції (1.1). Опишемо метод обчислення нижніх оцінок на основі негладкої двоїстої задачі, отриманої за допомогою лагранжевої релаксації частини обмежень початкової задачі. Для розв'язування двоїстої задачі можна використовувати ефективні алгоритми субградієнтного типу.

Розглянемо два способи отримання негладких оціночних задач для оптимального значення f^* цільової функції в задачі (1.1)–(1.6). Способи відрізняються тим, що для лагранжевої релаксації будуть вибрані різні групи обмежень із задачі (1.1)–(1.6).

1.2. Оцінка ψ_1^*. Перший спосіб пов'язаний із застосуванням схеми декомпозиції по групі обмежень (1.3). Якщо кожному обмеженню типу $\sum_{i \in I} x_{ij} = 1$ відповідає множник Лагранжа λ_j, то значення $\psi_1(\lambda)$ двоїстої функції у точці λ визначається оптимальним розв'язком задачі

$$\psi_1(\lambda) = \min_x \left(-\sum_{j \in J} \lambda_j + \sum_{j \in J} \sum_{i \in I} (d_{ij} + \lambda_j) \right) \qquad (1.7)$$

при обмеженнях

$$x_{ij} \le y_i \ \ \forall i \in I\,,\ \forall j \in J\,, \tag{1.8}$$

$$\sum_{i \in I} y_i = p\,, \tag{1.9}$$

$$y_i = 0 \vee 1 \ \ \forall i \in I\,, \tag{1.10}$$

$$x_{ij} = 0 \vee 1 \ \ \forall i \in I\,, \forall j \in J\,. \tag{1.11}$$

Задача (1.7)–(1.11) розв'язується наступним чином. Припустимо, що відомі ті p змінних y_i, які в оптимальному розв'язку дорівнюють одиниці. Нехай $I_1 \subset I$ – множина індексів цих змінних. Тоді оптимальне значення задачі (1.7)–(1.11) дорівнює

$$-\sum_{j \in J} \lambda_j + \sum_{i \in I_1} \sum_{j \in J} \min\{d_{ij} + \lambda_j ;0\}\,,$$

а оптимальний розв'язок, який йому відповідає, –

$$\tilde{x}_{ij}(\lambda) = \begin{cases} 1, \text{якщо}\ i \in I_1 \wedge d_{ij} + \lambda_j \le 0, \\ 0 - \text{у протилежному випадку.} \end{cases}$$

Таким чином, щоб визначити оптимальну множину I_1, достатньо для кожного $i \in I$ обчислити величину $S(\lambda) = \sum_{j \in J} \min\{d_{ij} + \lambda_j ;0\}$. Тоді I_1 виявляється підмножиною індексів I, які відповідають p найменшим значенням величини $S(\lambda)$. Розв'язок задачі (1.7)–(1.11) визначає також і суперградієнт функції ψ_1 в точці λ. Його j-а компонента визначається за формулою

$$g_j(\lambda) = \sum_{j \in J} \tilde{x}_{ij}(\lambda) - 1.$$

Найкраща нижня оцінка для цільової функції (1.1) буде відповідати оптимальному значенню двоїстої функції

$$\psi_1^* = \max_{\lambda} \psi_1(\lambda)\,. \tag{1.12}$$

Задача (1.12) є задачею максимізації негладкої увігнутої функції. Для її розв'язування можна використовувати методи негладкої оптимізації, наприклад методи субградієнтного типу з регулюванням кроку як в [7] або *r*-алгоритми [2, 3, 4]. Далі наведені псевдокоди для цих двох методів, аналогічно тому, як вони були застосовані при знаходженні подібних оцінок для задач розміщення [8].

"Простий" метод

Procedure Dual_S;
begin
$0 < q < 1$ – *параметр*;
$\psi^* = -\infty$;
$k = 0$; $\pi_0 = 2$;
$\lambda_j^0 = \min_{i \in I} d_{ij}$, $j = 1,...,m$;
repeat
Обчислити $\psi(\lambda^k)$, g_k;
$\psi^* = \max\{\psi^*; \psi(\lambda^k)\}$;
$h_k = \pi_k (f^{up} - \psi(\lambda^k)) / \| g_k \|^2$;
$\lambda^{k+1} = \lambda^k + h_k g_k$;
$\pi_{k+1} = \pi_k \cdot q$;
$k = k + 1$;
until
$\|\lambda^k - \lambda^{k-1}\| \le \varepsilon_\lambda$ або
$k > k_{\max}$ або
$\psi(\lambda^k) \ge f^{up}$;
end;

r-алгоритм

Procedure Dual_r;
begin
$2 < \alpha < 4$ – *параметр*;
$\psi^* = -\infty$;
$k = 1$; $B_0 = E$; $g_0 = 0$;
$\lambda_j^0 = \min_{i \in I_0} d_{ij}$, $j = 1,...,m$;
repeat
Обчислити $\psi(\lambda^k)$, g_k;
$\psi^* = \max\{\psi^*; \psi(\lambda^k)\}$;
$\xi_k = \dfrac{B_{k-1}^T (g_k - g_{k-1})}{\left\| B_{k-1}^T (g_k - g_{k-1}) \right\|}$;
$B_k = B_{k-1} + (1/\alpha - 1) B_{k-1} \xi_k \xi_k^T$;
$p_k = B_k B_k^T g_k / \left\| B_k^T g_k \right\|$;
$\lambda^{k+1} = \lambda^k + h_k p_k$;
де h_k таке, що $(p_k, g_{k+1}) \le 0$;
$k = k + 1$;
until
$\|\lambda^k - \lambda^{k-1}\| \le \varepsilon_\lambda$ або
$k > k_{\max}$ або
$\psi(\lambda^k) \ge f^{up}$;
end;

При описанні псевдокодів значення суперградієнта в точці λ^k позначено g_k, а значення цільової функції для найкращого згенерованого допустимого розв’язку задачі (1.1)–(1.6) (верхня оцінка для f^*) позначено f^{up}.

1.3. Оцінка ψ_2^*. Другий спосіб отримання нижньої оцінки для задачі (1.1)–(1.6) пов'язаний з лагранжевим послабленням обмежень (1.2), тобто з декомпозицією задачі по обмеженнях (1.2). З кожним обмеженням типу $x_{ij} - y_i \le 0$ зв'яжемо двоїсту змінну $u_{ij} \ge 0$. Для набору заданих значень u_{ij} двоїсту функцію запишем у вигляді

$$\psi_2(u) = \min_x \left(-\sum_{i\in I} y_i \sum_{j\in J} u_{ij} + \sum_{i\in I}\sum_{j\in J} (d_{ij} + u_{ij}) x_{ij} \right) \tag{1.13}$$

при обмеженнях

$$\sum_{i\in I} x_{ij} = 1 \;\; \forall j \in J, \tag{1.14}$$

$$\sum_{i\in I} y_i = p, \tag{1.15}$$

$$y_i = 0 \vee 1 \;\; \forall i \in I, \tag{1.16}$$

$$x_{ij} = 0 \vee 1 \;\; \forall i \in I, \forall j \in J. \tag{1.17}$$

Оптимальний розв'язок $(\tilde{y}(u), \tilde{x}(u))$ задачі (1.13)–(1.17) визначається таким шляхом:

1) покладемо рівними одиниці ті компоненти y_i вектора $\tilde{y}(u)$, які відповідають p найменшим значенням величин $u_i = \sum_{j\in J} u_{ij}$, а решту – рівними нулю;

2) для кожного $j \in J$ покладемо $\tilde{x}(u) = 0$ при $i \ne i_0$ та $\tilde{x}(u) = 1$ при $i = i_0$, де i_0 визначається за формулою $d_{i_0 j} + u_{i_0 j} = \min_{i\in I}\{d_{ij} + u_{ij}\}$.

Вектор, (i, j)-а складова якого дорівнює $\tilde{x}(u) - \tilde{y}(u)$, є суперградієнтом двоїстої функції ψ_2 в точці u.

Найкраща нижня оцінка для задачі (1.1)–(1.6) згідно з другим способом буде відповідати оптимальному значенню двоїстої функції

$$\psi_2^* = \max_{u\ge 0} \psi_2(u). \tag{1.18}$$

Задача (1.18) – задача максимізації негладкої увігнутої функції з обмеженнями на невід'ємність множників Лагранжа. Для її розв'язування доцільно використовувати методи негладкої оптимізації.

1.4. Про оцінки ψ_1^* та ψ_2^*. Обидва способи лагранжевої релаксації обмежень у задачі (1.1)–(1.6) дають однакові значення нижніх оцінок для оптимального значення f^*. Обидві оцінки дорівнюють оптимальному значенню цільової функції (1.1) в задачі лінійного програмування, яка отримана із задачі (1.1)–(1.6) релаксацією обмежень (1.5) та (1.6) на цілочисельність змінних.

Однак кожний із цих двох способів має як свої переваги, так і свої недоліки. Так, наприклад, перевагою другого способу є те, що у випадку існування багатьох близько розташованих об'єктів (з однаковими d_{ij}) він дозволяє у процесі оптимізації двоїстої функції ψ_2 знаходити хороші допустимі цілочислові розв'язки задачі (1.1)–(1.6). Для цього на кожній ітерації субградієнтного процесу необхідно вибирати такі індекси класів, які відповідають p змінним $\tilde{y}(u)$, що дорівнюють 1, та визначати розбиття об'єктів, прив'язуючи кожний об'єкт до найближчого класу щодо вихідних індексів неузгодженості d_{ij}.

Недолік другого способу полягає в тому, що задача (1.18) для оцінки ψ_2^* має розмірність $N \times M$, яка набагато більша за N – розмірність задачі (1.12) для оцінки ψ_1^*. Отже при розв'язуванні задач для декількох сотень об'єктів за допомогою першого способу доцільно орієнтуватися на використання *r*-алгоритму [2, 3] як ефективного засобу мінімізації негладких опуклих функцій. При застосуванні другого способу доцільно орієнтуватися на використання субградієнтних методів з регулюванням кроку по типу "збіжного ряду" [7] або спеціалізованих алгоритмів на основі ε-субградієнтних методів негладкої оптимізації [9], адаптованих для розв'язування задач даного класу великої розмірності.

2. МАТЕМАТИЧНІ МОДЕЛІ ДЛЯ ПОБУДОВИ ЛІНІЙНИХ ТА КВАДРАТИЧНИХ ДИСКРИМІНАНТНИХ ФУНКЦІЙ

Сформульовані математичні моделі для побудови лінійних та квадратичних дискримінантних функцій, що визначають розділяючі гіперповерхні для двох скінченних підмножин точок. Для лінійних дискримінантних функцій розглянуто задачу побудови розділяючої гіперплощини з максимальним зазором і задачу робастного розділення. Математичні моделі для них сформульовані у вигляді задачі квадратичного програмування та задач негладкого опуклого програмування. Для квадратичних дискримінантних функцій розглянуто задачі розділення довільною, опуклою або увігнутою квадратичними функціями. Математичні моделі для них сформульовані у вигляді задач опуклого програмування з обмеженнями, що використовують негладкі матричні функції, які відповідають максимальному або мінімальному власному числу симетричної матриці. Для спеціальної задачі розділення m ($m > 2$) множин точок паралельними гіперплощинами сформульована математична модель у вигляді задачі негладкого опуклого програмування.

2.1. Лінійні дискримінантні функції. Розглянемо задачі побудови лінійної дискримінантної функції (лінійного класифікатора), тобто гіперплощини, що розділяє дві скінченні підмножини точок в n-вимірному евклідовому просторі. Головною перевагою лінійного класифікатора є його простота та обчислювальна ефективність.

Нехай X та Y – дві скінченні множини векторів ознак в евклідовому просторі E^n, що належать до класу Ω_X та Ω_Y відповідно. $L(z)$ – лінійна функція: $L(z) = (b, z) + c$, де $b^T = (b_1, b_2, \ldots, b_n)$ – ваговий вектор, c – поріг. Довільна гіперплощина в E^n задається множиною точок $z \in E^n$, які задовольняють рівності $L(z)=0$. Лінійна дискримінантна функція розділяє множини X та Y гіперплощиною $L(z)=0$, якщо $L(x)>0$ для будь-якого з векторів x із множини X та $L(y)<0$ для будь-якого з векторів y із множини Y. Задача полягає в тому, щоб знайти гіперплощину, розділяючу вектори із множини X та вектори із множини Y, або встановити неможливість лінійної роздільності множин.

У випадку, коли розділяюча гіперплощина існує, виникає додаткова задача відокремлення множин з максимальною маржею – щоб відстань від гіперплощини до найближчої точки була максимальною. Така гіперплощина називається гіперплощиною з максимальним зазором (maximum margin hyperplane), відповідно лінійний класифікатор називається класифікатором з максимальним зазором (maximum margin classifier). Задача знаходження такого класифікатора еквівалентна знаходженню максимальної ширини для смуги (задається двома паралельними гіперплощинами), що розділяє дві задані множини.

2.1.1. Лінійний класифікатор з максимальним зазором. Нехай задано дві скінченні множини точок $X=\{x_i\}_{i=\overline{1,m_1}}$ та $Y=\{y_i\}_{i=\overline{1,m_2}}$ в евклідовому просторі E^n. Задача полягає у такому виборі для гіперплощини параметрів b та c, щоб максимізувати відстань між паралельними їй гіперплощинами, що відокремлюють множини X та Y. Паралельні гіперплощини можуть бути описані рівняннями $(b,z)+c=1$ та $(b,z)+c=-1$. Якщо множини X та Y лінійно роздільні, то умова, щоб між цими паралельними гіперплощинами не було жодної точки, еквівалентна наступним обмеженням: $(b,x_i)+c\ge 1$, $i=\overline{1,m_1}$, $(b,y_i)+c\le -1$, $i=\overline{1,m_2}$.

Відстань між обома паралельними гіперплощинами дорівнює $2/\|b\|$, тому вона буде тим більшою, чим меншим є $\|b\|$. Таким чином, математична модель для лінійного класифікатора з максимальним зазором представляється наступною задачею квадратичного програмування [10]:

$$f^* = \min_{b,c} \|b\|^2, \tag{2.1}$$

$$(b, x_i) + c \ge 1,\ i = \overline{1, m_1}, \tag{2.2}$$

$$-(b, y_i) - c \ge 1,\ i = \overline{1, m_2}\ . \tag{2.3}$$

У випадку, коли система лінійних обмежень (2.2)–(2.3) несумісна, це означає факт нероздільності множин X та Y. У протилежному випадку множини X та Y лінійно розділяються. Оптимальний розв'язок $(b^*; c^*)$ для задачі (2.1)–(2.3) визначає лінійний класифікатор з максимальним зазором.

Математичну модель лінійного класифікатора з максимальним зазором можна представити також у вигляді такої задачі опуклого програмування з негладкою цільовою функцією:

$$f^* = f(b^*, c^*) = \min_{b,c} \max\left\{-(b, x_i) - c,\ i = \overline{1, m_1}; (b, y_j) + c,\ j = \overline{1, m_2}\right\}, \tag{2.4}$$

$$\sum_{i=1}^{n} b_i^2 \le 1. \tag{2.5}$$

Ця задача випливає із моделі "оптимального разделения конечных множеств точек" (див. [1, с. 183, формула (1.45)]) .

Оптимальне значення цільової функції (2.4) є недодатнім і характеризує лінійну роздільність множин X та Y наступним чином.

Якщо $f^* < 0$, тоді розв'язок задачі (2.4)–(2.5) визначає лінійний класифікатор з максимальним зазором. При цьому f^* дорівнює взятому зі знаком "мінус" значенню відстані від оптимальної гіперплощини до опуклих оболонок множин X та Y. Знайдена гіперплощина є серединою смуги максимальної ширини $d = -2f^*$, що відокремлює множину X від множини Y.

Якщо $f^* = 0$, це означає, що множини X та Y лінійно нероздільні. При цьому можливі два варіанти розв'язку задачі (2.4)–(2.5):

1) єдиним розв'язком є нульовий вектор ($b_i^* = 0$, $i = \overline{1,n}$; $c^* = 0$). Тоді опуклі оболонки множин X та Y "перекриваються";

2) існують розв'язки, відмінні від ($b_i^* = 0$, $i = \overline{1,n}$; $c^* = 0$). Це відповідає випадку, коли смуга стає виродженою (має нульову ширину), тобто знайдена гіперплощина обов'язково містить тільки граничні точки опуклих оболонок множин X та Y.

Задачі (2.1)–(2.3) та (2.4)–(2.5) можуть бути розв'язані за допомогою алгоритмів мінімізації недиференційовних опуклих функцій з використанням методу негладких штрафних функцій. Кількість змінних для обох задач визначається розміром простору ознак плюс одиниця. При досить малих розмірах простору ознак (до десяти) ефективними для розв’язування цих задач будуть методи еліпсоїдів [2, 3]. При розмірах простору ознак від декількох десятків до декількох сотень при розв’язуванні цих класів задач досить ефективними будуть *r*-алгоритми.

Вищенаведені математичні моделі відповідають таким задачам лінійного розділення, для яких суттєвим фактором є вимога до існування (або встановлення відсутності) строгого розділення двох множин. Однак задачі оптимального (за деяким критерієм) лінійного розділення часто ставляться і для тих випадків, коли строгого розділення множин не існує. Розглянемо побудову лінійної дискримінантної функції для однієї з таких задач.

2.1.2. Задача робастного розділення. В роботі [11] запропонована математична модель задачі лінійного розділення двох множин точок, для якої розділяюча множини X та Y гіперплощина визначається в два етапи.

На першому етапі для гіперплощини визначаються параметри $(b;c)$ шляхом розв'язання такої задачі мінімізації:

$$F^* = \min_{b,c} \frac{1}{m_1} \sum_{i=1}^{m_1} \max\{0; -(b, x_i) - c + 1\} + \frac{1}{m_2} \sum_{i=1}^{m_2} \max\{0; (b, y_i) + c + 1\}. \quad (2.6)$$

На другому етапі проводиться процедура корекції розділяючої гіперплощини. Ця процедура полягає в розв'язуванні такої задачі мінімізації:

$$\min_{c_{low} \le c \le c_{up}} \left(\frac{1}{m_1} \sum_{i=1}^{m_1} \max\{0; -(b, x_i) - c\} + \frac{1}{m_2} \sum_{i=1}^{m_2} \max\{0; (b, y_i) + c\} \right), \qquad (2.7)$$

де $c_{low} = \min\limits_{i=\overline{1,m_1}} (b, x_i)$, $c_{up} = \max\limits_{i=\overline{1,m_2}} (b, y_i)$.

У задачі (2.7) вектор b фіксований, і його значення визначено на першому етапі. Задача (2.7) є задачею мінімізації одновимірної кусочно-лінійної опуклої функції.

Основна чисельна трудомісткість побудови розділяючої гіперплощини визначається першим етапом – розв'язуванням задачі (2.6).

У роботі [11] досліджуються різні властивості запропонованої моделі побудови розділяючої гіперплощини. Показано, що модель відображає основні інтуїтивні уявлення розділення. Так, наприклад, якщо опуклі оболонки множин точок X та Y не перетинаються, то оптимальна гіперплощина, що відповідає розв'язку задачі (2.6), строго розділяє ці оболонки (при цьому оптимальне значення цільового функціоналу (2.6) $F^* = 0$).

Розв'язування задачі (2.6) в роботі [11] здійснюється шляхом попереднього зведення її до еквівалентної їй задачі лінійного програмування ($t \in E^{m_1}$, $s \in E^{m_2}$):

$$\min_{b,c,t,s} \frac{1}{m_1} \sum_{i=1}^{m_1} t_i + \frac{1}{m_2} \sum_{i=1}^{m_2} s_i \, , \qquad (2.8)$$

$$(b, x_i) + c + t_i \ge 1, \; i = \overline{1, m_1}, \qquad (2.9)$$

$$-(b, y_i) - c + s_i \ge 1, \; i = \overline{1, m_2}, \qquad (2.10)$$

$$t, s \ge 0. \qquad (2.11)$$

Для задач, що включають велику кількість точок при невеликих розмірностях евклідового простору їх визначення ($m_1, m_2 >> n$), зведення задачі (2.6) до задачі лінійного програмування (2.8)–(2.11) не ефективне. У цьому випадку остання має велику кількість як змінних (змінні t і s), так і обмежень. Але саме такими є більшість практичних задач (це видно і з наведених в [11] прикладів). Для розв'язування такого роду задач доцільно використовувати ефективні методи негладкої оптимізації, а саме *r*-алгоритми безпосередньо для розв'язання задачі (2.6) [12]. Для цього задача (2.6) має корисні властивості: вона є задачею безумовної мінімізації опуклої функції з невеликою кількістю змінних.

2.2. Квадратичні дискримінантні функції. Розглянемо задачу розділення двох множин точок квадратичною поверхнею $(Az,z)+(b,z)+c=0$, де параметрами поверхні виступають симетрична матриця $A=\{A_{ij}\}_{i,j=1}^{n}$, вектор $b\in E^n$ та скаляр c. Квадратичну поверхню, яка розділяє множини X та Y, будемо називати квадратичною дискримінантною функцією (квадратичним класифікатором).

Математичну модель для квадратичного класифікатора сформулюємо як задачу квадратичного програмування

$$f^* = \min_{A,b,c}\left(\sum_{i=1}^{n}\sum_{j=1}^{n}A_{ij}^2+\sum_{i=1}^{n}b_i^2\right), \tag{2.12}$$

$$(Ax_i,x_i)+(b,x_i)+c\ge 1,\ i=\overline{1,m_1}, \tag{2.13}$$

$$-(Ay_i,y_i)-(b,y_i)-c\ge 1,\ i=\overline{1,m_2}. \tag{2.14}$$

Аналогічно випадку лінійного класифікатора, для задачі (2.12)–(2.14) мають місце такі твердження. Якщо система обмежень (2.13)–(2.14) несумісна, то множини X та Y є квадратично нероздільними. Інакше розв'язок задачі (2.12)–(2.14) задає параметри оптимальної за критерієм (2.12) квадратичної поверхні, яка строго розділяє множини X та Y.

Вибір критерію (2.12) для квадратичного класифікатора пояснюється такими міркуваннями. Розглянемо розширений (щодо E^n) евклідів простір розмірністю $n+n^2$ – E^{n+n^2}. У відповідність точкам $x \in E^n$ поставимо точки $(x; xx^T) \in E^{n+n^2}$. Множинам X та Y будуть відповідати їх образи X' та Y', вектори для яких мають розмірність $n+n^2$ і визначаються за вказаним правилом. Задача (2.12)–(2.14) відповідає побудові лінійного класифікатора з максимальним зазором для розділу множин X' та Y' у розширеному евклідовому просторі E^{n+n^2}.

По аналогії із задачею (2.4)–(2.5) математичну модель квадратичного класифікатора (2.12)–(2.14) можна сформулювати у вигляді негладкої задачі опуклого програмування:

$$f^* = \min_{A,b,c} \max \left\{ -(Ax_i, x_i) - (b, x_i) - c\ , i = \overline{1, m_1}; (Ay_j, y_j) + (b, y_i) + c,\ j = \overline{1, m_2} \right\}, \quad (2.15)$$

$$\sum_{i=1}^{n} \sum_{j=1}^{n} A_{ij}^2 + \sum_{i=1}^{n} b_i^2 \le 1. \quad (2.16)$$

Оптимальне значення цільової функції (2.15) є недодатнім та характеризує квадратичну роздільність множин наступним чином. Якщо $f^* < 0$, тоді розв'язок задачі (2.15)–(2.16) визначає такий квадратичний класифікатор, який для множин X' та Y' у розширеному евклідовому просторі відповідає лінійному класифікатору з максимальним зазором відповідно до моделі (2.4)–(2.5). Якщо $f^* = 0$, то це означає, що множини X та Y квадратично нероздільні. При цьому можливо встановити (при ненульових оптимальних параметрах) таку гіперплощину в розширеному евклідовому просторі, яка проходить через перетин границь опуклих оболонок множин X' та Y'.

На квадратичну дискримінантну функцію можуть накладатися додаткові умови, наприклад, вимоги до її опуклості або ввігнутості. У цих випадках до постановок задач добавляються обмеження на максимальне або мінімальне власні числа симетричної матриці *A*. Так, для знаходження опуклої квадратичної дискримінантної функції до обмежень задачі (2.12)–(2.15) та задачі (2.16)–(2.17) слід додати обмеження

$$\lambda_{\min}(A) \geq 0, \tag{2.17}$$

а для знаходження увігнутої квадратичної дискримінантної функції –

$$\lambda_{\max}(A) \leq 0. \tag{2.18}$$

Обмеження (2.17) та (2.18) є опуклими. Для існування опуклої квадратичної дискримінантної функції достатньо впевнитися у сумісності системи обмежень (2.13), (2.14), (2.17), а для існування ввігнутої квадратичної дискримінантної функції – у сумісності системи обмежень (2.13), (2.14), (2.18).

Математичні моделі квадратичного класифікатора мають форму негладких задач опуклого програмування (негладкі функції, матричні обмеження). Розмірність цих задач дорівнює $N = (n+1)(n+2)/2$, де n – розмірність простору ознак. Для їх ефективного розв'язання можна використати методи негладкої оптимізації субградієнтного типу з перетворенням простору, наприклад, *r*-алгоритми.

2.3. Спеціальна лінійна дискримінантна функція. Розглянемо задачу побудови спеціального лінійного класифікатора, коли кількість класів $m > 2$, і потрібно розділити їх паралельними гіперплощинами (порядок розташування класів заданий). Серед усіх можливих варіантів необхідно знайти такий варіант, для якого найменший із зазорів, що розділяє пари множин, є максимальним.

Нехай задано m множин точок $\{x_i\}_{i \in I_j}$, $j = \overline{1,m}$, у просторі R^n. Паралельні гіперплощини визначаються рівняннями $(b,z) + c_k = 0$, $k = \overline{1,m-1}$ (вектор $b \in E^n$ для всіх $(m-1)$ гіперплощин однаковий, але вільні члени $C = \{c_k\}_{k=\overline{1,m-1}}$ є різними). Тоді математичну модель для лінійного класифікатора можна записати у вигляді такої задачі опуклого недиференційовного програмування:

$$f^* = \min_{b,C} \underline{\max_{k=1,m-1}} \left\{ \max\{(b, x_i) + c_k, i \in I_k; -(b, x_i) - c_k, i \in I_{k+1}\} \right\}, \quad (2.19)$$

$$\sum_{i=1}^{n} b_i^2 \le 1. \quad (2.20)$$

У цільовій функції (2.19) враховано, що замість порівняння відхилень k-ї гіперплощини від точок усіх множин достатньо розглянути лише ті множини, що до неї прилягають: I_k та I_{k+1} (вважається заданим). Критерієм оптимальності виступає умова $\max_{b,C} \underline{\min_{k=1,m-1}} \{d_k\}$, де d_k – відстань від k-ї гіперплощини до найближчої точки.

Аналіз розв'язків задачі (2.19)–(2.20) ідентичний тому, який мав місце для задачі знаходження лінійного класифікатора з максимальним зазором (2.4)–(2.5). У випадку $m = 2$ задача (2.19)–(2.20) в точності відповідає задачі (2.4)–(2.5).

3. КВАДРАТИЧНА МОДЕЛЬ ДЛЯ МОНОХРОМАТИЧНОГО ЗОБРАЖЕННЯ, ЛАГРАНЖЕВІ ОЦІНКИ ТА ЇХ УТОЧНЕННЯ

Побудовано неопуклу квадратичну модель для задачі кодування монохроматичних зображень. Для отримання нижньої оцінки цільової функції застосовано техніку двоїстих лагранжевих оцінок Шора. Доведено, що знаходження оцінки Шора зводиться до задачі безумовної опуклої недиференційовної оптимизації. Показано, що розширення базової моделі за допомогою функціонально надлишкових обмежень дозволяє отримати більш точні нижні оцінки.

3.1. Постановка задачі. Нехай монохроматичне (чорно-біле) зображення на площині визначається чорними та білими клітинами сітки розміром $n \times m$. Зрозуміло, що найпростіший спосіб його кодування (задання) – це поставити у відповідність кожній клітині притаманний їй колір. При цьому досягається точна ідентифікація об'єкта, що, однак, потребує значних ресурсів пам'яті ($n \times m$ параметрів). Поступившись певною мірою точністю задання зображения, можна суттєво зменшити кількість параметрів кодування, і, відповідно, об'єм необхідної пам'яті. Це можна зробити, наприклад, таким чином. За параметри кодування приймемо n рядків і m стовпців сітки, точніше їх кольори (чорний або білий). Правило "відтворення" зображення (побудови за цими параметрами зображення, яке буде наближенням до заданого) приймемо таким: клітина буде білою тільки у випадку, коли вона належить білому рядку і білому стовпцю, у протилежному випадку клітина має чорний колір. Задача полягає в знаходженні таких кольорів рядків та стовців сітки, які за наведеним правилом найкращим чином наближають задане монохроматичне зображення. Іншими словами, за допомогою $(n+m)$ параметрів (невідомі кольори стовпців та рядків) слід найкращим чином закодувати задане монохроматичне зображення.

Побудуємо математичну модель для даної задачі кодування. Нехай зображення G, що кодується, задається матрицею $W = \{w_{ij}\}_{i=\overline{1,n},\, j=\overline{1,m}}$, де w_{ij} характеризує колір клітини сітки, в якій перетинаються i-й рядок та j-й стовпець; x_i, $i=\overline{1,n}$, та y_j, $j=\overline{1,m}$ – параметри кодування, які визначають колір i-го рядка й j-го стовпця відповідно. Будемо вважати, що значення w_{ij}, x_i та y_j, $i=\overline{1,n}$, $j=\overline{1,m}$, дорівнюють 1, якщо колір білий, і 0, якщо колір чорний. У цих термінах правило побудови зображення $\widetilde{G}$ за параметрами x та y ("відтворення" зображення G), яке характеризується матрицею $\widetilde{W} = \{\widetilde{w}_{ij}\}_{i=\overline{1,n},\, j=\overline{1,m}}$, приймає вигляд

$$\widetilde{w}_{ij} = x_i y_j,\ i=\overline{1,n},\ j=\overline{1,m},$$

а критерій якості кодування зображення визначається величиною

$$\sum_{i=1}^{n}\sum_{j=1}^{m}(\widetilde{w}_{ij}-w_{ij})^2 = \sum_{i=1}^{n}\sum_{j=1}^{m}(x_i y_j - w_{ij})^2 = \sum_{i=1}^{n}\sum_{j=1}^{m}(1-2w_{ij})x_i y_j + \sum_{i=1}^{n}\sum_{j=1}^{m} w_{ij},$$

яка задає суму неспівпадаючих за кольором клітин у зображеннях G и $\widetilde{G}$.

Тоді задача знаходження оптимального коду зображення може бути записана у вигляді оптимізаційної квадратичної задачі:

$$f^* = \min_{x,y} \sum_{i=1}^{n} \sum_{j=1}^{m} (1-2w_{ij}) x_i y_j + \sum_{i=1}^{n} \sum_{j=1}^{m} w_{ij} \tag{3.1}$$

при обмеженнях

$$x_i^2 = x_i, \ i = \overline{1,n}, \tag{3.2}$$

$$y_j^2 = y_j, \ j = \overline{1,m}. \tag{3.3}$$

У загальному випадку задача (3.1)–(3.3) є NP-повною, що викликає певні обчислювальні труднощі при її розв'язуванні. В зв'язку з цим пропонується розглядати оцінки для оптимального значення цільової функції f^*, які можна отримати шляхом розв'язання опуклих матричних задач.

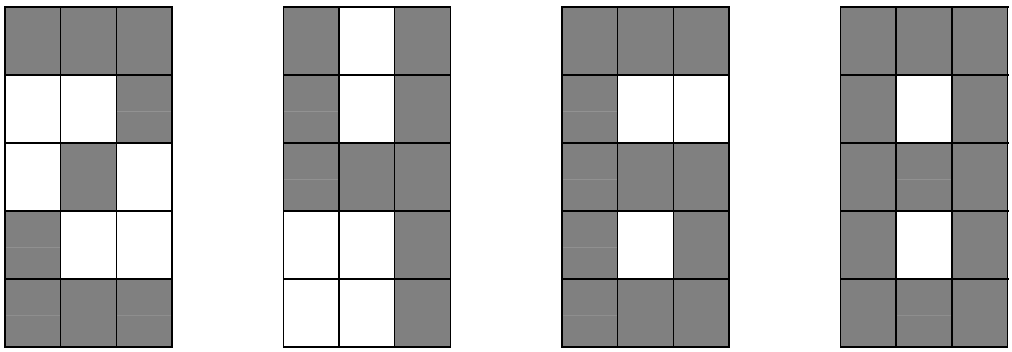
3.2. Нижня оцінка ψ^*. З метою визначення нижньої оцінки для f^* у неопуклій задачі (3.1)–(3.3) застосуємо техніку двоїстих лагранжевих оцінок [3, 5]. Нижня оцінка ψ^* може бути знайдена за допомогою субградієнтних методів негладкої оптимізації як розв'язок такої задачі:

$$f^* \geq \psi^* = \sup_{u \in D} \inf_{z} L(z,u) + \sum_{i=1}^{n} \sum_{j=1}^{m} w_{ij}, \tag{3.4}$$

де $z = (x_1, \ldots, x_n, y_1, \ldots, y_m)^T \in R^{n+m}$; $u \in R^{n+m}$ – вектор множників Лагранжа; $L(z,u) = \sum_{i=1}^{n} \sum_{j=1}^{m} (1-2w_{ij}) z_i z_{n+j} + \sum_{i=1}^{n+m} u_i (z_i^2 - z_i) = z^T K(u) z - \sum_{i=1}^{n+m} u_i z_i$ – функція Лагранжа задачі (3.1)–(3.3); $D = \{u : K(u) \in S_{n+m}^+\}$ – область значень множників Лагранжа, при яких квадратична матриця функції Лагранжа $K(u)$ належить до класу симетричних додатньо визначених матриць S_{n+m}^+, що еквівалентно умові: мінімальне власне число матриці $K(u)$ є додатнім: $\lambda_{\min}(K(u)) > 0$.

Верхню оцінку $f(\tilde{z})$ для f^* можна обчислити шляхом побудови допустимих розв'язків $\tilde{z}$ для задачі (3.1)–(3.3) різними способами. Один зі способів – це прийняти за значення $\tilde{z}_i$, $i = \overline{1,(n+m)}$, відповідні компоненти вектора $z^* = \arg\sup\limits_{u \in D} \inf\limits_{z \in R^{n+m}} L(z,u)$, "округлені" до найближчих кінців відрізка [0;1]. Інший спосіб побудови допустимих розв'язків $\tilde{z}$ буде описано далі.

Наведемо кілька тестових прикладів використання описаного методу (на рисунку показано первісні зображення, в таблиці – результати обчислень).



1 2 3 4

РИСУНОК

ТАБЛИЦЯ

№, п/п	ψ^*	f^*	$f(x,y)$	Допустимий розв'язок
1	3,00	3	3	x=(0,1,1,1,0), y=(1,1,1)
2	1,60	2	2	x=(0,0,0,1,1), y=(1,1,0)
3	0,63	1	1	x=(0,1,0,1,0), y=(0,1,0)
4	-0,02	0	0	x=(0,1,0,1,0), y=(0,1,0)

У наведених тестових прикладах нижні оцінки ψ^* знайдено за допомогою програми DSQTPR [13], яка реалізує одну з модифікацій r-алгоритму [2], що належить до субградієнтних методів з розтягом простору.

3.3. Безумовна задача для оцінки ψ^*. Знаходження оцінки ψ^* (3.4) можна звести до задачі безумовної опуклої мінімізації [13].

Представимо задачу (3.1)–(3.3) у вигляді однорідної квадратичної задачі. Для цього перепишемо задачу (3.1)–(3.3) в бінарних змінних, зробивши заміни $x_i = (1+v_i)/2$, $i=\overline{1,n}$, та $y_j = (1+v_{n+j})/2$, $j=\overline{1,m}$:

$$f^* = \min_v \sum_{i=1}^{n}\sum_{j=1}^{m}\frac{(1-2w_{ij})}{4}(v_i v_{n+j} + v_i + v_{n+j}) + \sum_{i=1}^{n}\sum_{j=1}^{m}\frac{(1+2w_{ij})}{4}$$

при обмеженнях

$$v_i^2 = 1,\ i=\overline{1,(n+m)}.$$

Якщо домножити всі лінійні члени на додаткову бінарну змінну v_{n+m+1} та ввести позначення $c = \sum_{i=1}^{n}\sum_{j=1}^{m}\frac{(1+2w_{ij})}{4}$, то отримаємо

$$f^* = \min_v \sum_{i=1}^{n}\sum_{j=1}^{m}\frac{(1-2w_{ij})}{4}(v_i v_{n+j} + v_i v_{n+m+1} + v_{n+j} v_{n+m+1}) + c \qquad (3.5)$$

при обмеженнях

$$v_i^2 = 1,\ i=\overline{1,(n+m+1)}. \qquad (3.6)$$

У роботі [14] доведено, що лагранжева двоїста оцінка при таких операціях не змінюється, тобто оцінка для задачі (3.5)–(3.6)

$$\psi^* = \sup_{\lambda_{\min}(A(u))>0} \inf_{\nu} \left(v^T A v + \sum_{i=1}^{n+m+1} u_i (v_i^2 - 1) \right) + c, \qquad (3.7)$$

дорівнює оцінці (4) для задачі (3.1)–(3.3). Тут A – матриця квадратичної форми (3.5), а $A(u) = A + diag(u)$ – матриця квадратичної форми (3.7).

Оскільки внутрішня задача мінімізації однорідної квадратичної форми за умови $\lambda_{\min}(A(u)) > 0$ має тривіальний розв'язок $\nu = \mathbf{0}$, де $\mathbf{0}$ – нульовий вектор, то задача (3.7) набуде вигляду задачі опуклого програмування

$$\psi^* = \sup_{\lambda_{\min}(A(u))>0} \left(-\sum_{i=1}^{n+m+1} u_i \right) + c. \qquad (3.8)$$

Для спрощення подальших викладок розглянемо задачу

$$\tilde{\psi}^* = \inf_{\lambda_{\min}(A(u))>0} \left(\sum_{i=1}^{n+m+1} u_i \right). \qquad (3.9)$$

(Зрозуміло, що $\psi^* = -\tilde{\psi}^* + c$). Для врахування обмеження цієї задачі використаємо функцію негладкого штрафу у вигляді функції максимуму. Для цього скористаємось результатом із [15].

Розглянемо задачу оптимізації

$$\min_{y \in M} f_0(y), \qquad (3.10)$$

$$f_i(y) \le 0,\ i = \overline{1,k},\ y \in M, \qquad (3.11)$$

де $f_i(y)$, $i = \overline{0,k}$ – неперервні функції, $M \subseteq R^n$ – деяка множина. Введемо сімейство задач, що залежить від вектора параметрів $z \in R^k$:

$$V(z) = \inf\left\{ f_0(y) : f_i(y) \le z_i; i = \overline{1,k}, y \in M \right\},$$

(зазначимо, що $V(\mathbf{0})$, де $\mathbf{0}$ – k-вимірний нульовий вектор, відповідає початковій задачі (3.10)–(3.11)).

Якщо використати для врахування обмежень (3.11) негладку штрафну функцію у вигляді функції максимуму

$$\Phi_N(y) = f_0(y) + N \max\left\{0, f_1(y), \ldots, f_k(y)\right\}, \qquad (3.12)$$

то справедлива

Теорема 1 [15]. Нехай $\inf\limits_{t>0} \dfrac{V(te) - V(0)}{t} = -L > -\infty$, де e – k-вимірний вектор, всі компоненти якого рівні одиниці. Якщо $N > L$, тоді точки мінімуму задачі $V(0)$ і задачі $\inf\limits_{y \in M} \Phi_N(y)$, де $\Phi_N(y)$ визначається формулою (3.12), збігаються.

Застосуємо наведену теорему до задачі (3.9). У даному випадку

$$V(te) = \inf\left\{ \sum_{i=1}^{n+m+1} u_i : -\lambda_{\min}\left(A + diag(u)\right) \le t \right\} =$$

$$= \inf\left\{ \sum_{i=1}^{n+m+1} u_i : -\lambda_{\min}\left(A + diag(u) + tI\right) \le 0 \right\} =$$

$$= \inf\left\{ \sum_{i=1}^{n+m+1} (\tilde{u}_i - t) : \lambda_{\max}\left(A + diag(\tilde{u}\right) \le 0 \right\} = V(0) - \sum_{i=1}^{n+m+1} t = V(0) - (n+m+1)t.$$

Тоді $\inf\limits_{t>0}\dfrac{V(t)-V(0)}{t}=\inf\limits_{t>0}\dfrac{V(0)-(n+m+1)t-V(0)}{t}=-(n+m+1)=-L>-\infty$.

Таким чином, згідно з теоремою 1, при $N>n+m+1$

$$\tilde{\psi}^* = \inf_{\lambda_{\min}(A(u))>0}\left(\sum_{i=1}^{n+m+1}u_i\right)=\inf_u\left(\sum_{i=1}^{n+m+1}u_i+N\max\{0,-\lambda_{\min}(A+diag(u))\}\right).$$

Тобто справедлива

Теорема 2 [16]. При $N>n+m+1$ нижня оцінка для задачі (3.1)–(3.3), що визначається еквівалентними задачами (3.4) та (3.7),

$$\psi^* = -\min_u\left(\sum_{i=1}^{n+m+1}u_i+N\max\{0;\lambda_{\max}(-A-diag(u))\}\right)+\frac{1}{4}\sum_{i=1}^{n}\sum_{j=1}^{m}(1+2w_{ij}). \qquad (3.13)$$

Зазначимо, що задача (3.13) дає лише оцінку для f^*. Для знаходження допустимого розв'язку $\tilde{v}$, а разом з тим і верхньої оцінки $f(\tilde{v})$ задачі (3.5)–(3.6), можна застосовувати відому евристичну процедуру:

нехай u^* – розв'язок задачі (3.13), $\xi^*=(\xi_1^*,\xi_2^*,\ldots,\xi_{n+m+1}^*)$ – власний вектор матриці $A(u^*)$, який відповідає мінімальному власному числу $\lambda_{\min}(A(u^*))$;

а) впорядкуємо компоненти вектора ξ^* у незростаючому порядку – $\xi^*_{\alpha(1)}\ge\xi^*_{\alpha(2)}\ge\ldots\ge\xi^*_{\alpha(n+m+1)})$;

б) для різних натуральних k, $1\le k\le n+m$, будемо розбивати усі змінні задачі на дві групи: $\{v_{\alpha(1)},\ldots,v_{\alpha(k)}\}$ та $\{v_{\alpha(k+1)},\ldots,v_{\alpha(n+m+1)}\}$. Змінним тієї групи, куди зараховується v^*_{n+m+1}, присвоюємо значення 1, іншим -1. Вибираємо той варіант допустимого розв'язку $\tilde{v}$, за якого значення $\tilde{v}^T A\tilde{v}$ мінімальне.

3.4. Функціонально надлишкові обмеження. Використання техніки лагранжевих двоїстих оцінок дозволяє покращувати оцінку ψ^* (3.4) шляхом розширення квадратичної постановки (3.1)–(3.3) за рахунок функціонально надлишкових обмежень [3, 5]. Один з способів побудови таких обмежень (запропонований у [5]) полягає у перемноженні нерівностей $z_i \ge 0$ та $z_j \le 1$, які є наслідком рівнянь (3.2)–(3.3) (нагадаємо, що $z = (x_1, \dots, x_n, y_1, \dots, y_m)^T$:

$$\begin{cases} z_i z_j \ge 0, \\ z_i(1 - z_j) \ge 0, \\ (1 - z_i) z_j \ge 0, \\ (1 - z_i)(1 - z_j) \ge 0, \end{cases} \quad i = \overline{1, (n+m-1)}, \; j = \overline{(i+1), (n+m)}. \tag{3.14}$$

Зрозуміло, що лагранжева двоїста оцінка ψ_r^*, побудована для розширеної задачі (3.1)–(3.3), (3.14) за тими ж правилами, що й ψ^* для задачі (3.1)–(3.3), буде не гіршою за оцінку (3.4). Застосувавши ті ж самі міркування, як і в попередньому випадку (тобто провівши заміну булевих змінних на бінарні і домноження лінійних частин на бінарну змінну), задачу (3.1)–(3.3), (3.14) можна звести до однорідної квадратичної задачі:

$$f^* = \min_{v} \sum_{i=1}^{n} \sum_{j=1}^{m} \frac{(1 - 2w_{ij})}{4} (v_i v_{n+j} + v_i v_{n+m+1} + v_{n+j} v_{n+m+1}) + c \tag{3.15}$$

при обмеженнях

$$v_i^2 = 1, \; i = \overline{1, (n+m+1)}. \tag{3.16}$$

$$\begin{cases} -v_i v_j - v_i v_{n+m+1} - v_j v_{n+m+1} - 1 \le 0, \\ v_i v_j - v_i v_{n+m+1} + v_j v_{n+m+1} - 1 \le 0, \\ v_i v_j + v_i v_{n+m+1} - v_j v_{n+m+1} - 1 \le 0, \\ -v_i v_j + v_i v_{n+m+1} + v_j v_{n+m+1} - 1 \le 0, \end{cases}, \; i = \overline{1, (n+m-1)}, \; j = \overline{(i+1), (n+m)}. \tag{3.17}$$

Для лагранжевої двоїстої оцінки ψ_r^* для задачі (3.15)–(3.17) (або, що те саме, для задачі (3.1)–(3.3), (3.14)) аналогічно теоремі 2 доводиться

Теорема 3. При $N > n + m + 1$

$$\psi_r^* = -\min_{\substack{u \\ t \ge 0}} \left(\sum_{i=1}^{n+m+1} u_i + \sum_{i=1}^{l} t_i + N \max\{0; \lambda_{\max}(-B(t) - diag(u))\} \right) +$$

$$+ \frac{1}{4} \sum_{i=1}^{n} \sum_{j=1}^{m} (1 + 2w_{ij}). \tag{3.18}$$

У виразі (3.18) застосовані такі позначення: $B(t)+diag(u)$ – квадратична матриця функції Лагранжа задачі (3.15)–(3.17); $u \in R^{n+m+1}$ – вектор множників Лагранжа для обмежень (3.16); $t \in R^{l}$ – вектор множників Лагранжа для обмежень (3.17), де $l = 2(n+m)(n+m-1)$.

Стосовно ефективності використання функціонально надлишкових обмежень зазначимо, що вже при розширенні задачі (3.1)–(3.3) тільки обмеженнями вигляду $z_i z_j \geq 0$ з (3.14) маємо покращення нижньої оцінки на тестових прикладах (див. рисунок): для другого вона дорівнює 1.66, для третього – 0.83, для четвертого – 0.00 (точна). Якщо ж врахувати всі обмеження (3.14), то оцінки ψ_r^* стають точними для всіх наведених тестових прикладів.

ВИСНОВКИ

Для узагальнених задач класифікації, побудови лінійних та нелінійних дискримінантних функцій, представлення монохроматичних зображень сформульовані математичні моделі у вигляді матричних задач негладкої оптимізації, спеціалізованих задач цілочисельного лінійного програмування та неопуклих квадратичних задач з булевими змінними. Для їх розв'язання застосовані схеми декомпозиції, техніка двоїстих лагранжевих оцінок Н.З. Шора та використані ефективні методи негладкої оптимізації.

Програмні реалізації розглянутих методів можуть бути включені в пакети програм для розв’язування задач розпізнавання та класифікації. Наприклад, в універсальну програмну систему “RECOGNITION” інтелектуального аналізу даних, розпізнавання та прогнозування [17, 18] поряд з методом опорного вектора може бути включений метод робастного розділення.

Список літератури

1. *Шлезингер М., Главач В.* Десять лекций по статистическому и структурному распознаванию. – Киев: Наук. думка, 2004. – 545 с.
2. *Шор Н.З.* Методы минимизации недифференцируемых функций и их приложения. – Киев: Наук. думка, 1979. – 199 с.
3. *Shor N.Z.* Nondifferentiable Optimization and Polynomial Problems. – Dordrecht, Kluwer, 1998. – 394 p.
4. *Шор Н.З., Журбенко Н.Г., Лиховид А.П., Стецюк П.И.* Развитие алгоритмов недифференцируемой оптимизации и их приложения // Кибернетика и системный анализ. – 2003. – № 4. – С. 80–94.
5. *Шор Н.З., Стеценко С.И.* Квадратичные экстремальные задачи и недифференцируемая оптимизация. – Киев: Наук. думка, 1989. – 208 с.
6. *Мину М.* Математическое программирование. – М.: Наука,1990.– 488 с.
7. *Christofides N., Beasley J.E.* A tree search algorithm for the p-median problem // European Journal of Operational Research. – 1986. – N 10 – P.196–204.
8. *Кузьменко В.Н., Гольденгорин Б.И., Тсо М., Стецюк П.И.* Сравнение двух субградиентных методов при нахождении оценок для задач размещения // Теория оптимальных решений. – Киев: Ин-т кибернетики им. В.М. Глушкова НАН Украины, 2004. – С. 111–119.
9. *Журбенко Н.Г.* Об одном ε-субградиентном алгоритме минимизации // Теория оптимальных решений. – Киев: Ин-т кибернетики им. В.М. Глушкова НАН Украины, 2006. – С. 111–118.
10. Support vector machine. – Wikipedia, the free encyclopedia. – Last modified 16.16.2008. – http://en.wikipedia.org/wiki/Support_vector_machine.
11. *Bennett K.P., Mangasarian O.L.* Robust Linear Programming Discrimination of Two Linearly Inseparable Sets // Optimization Methods and Software. – 1996. –N 5. – P. 23–34.

12. *Журбенко Н.Г., Саимбетов Д.Х.* К численному решению одного класса задач робастного разделения двух множеств // Методы исследования экстремальных задач. – Киев: Ин-т кибернетики им. В.М. Глушкова НАН Украины, 1994. – С. 52–55.
13. *Shor N.Z., Stetsyuk P.I.* Dual Solution of Quadratic-Type Problems by *r*-algorithm (subroutine DSQTPr) // Abstracts of Second International Workshop "Recent Advances in Non-Differentiable Optimization" (October, 1-4, 2001, Kyiv, Ukraine). – P. 36.
14. *Березовский О.А., Стецюк П.И.* Об одном способе нахождения двойственных квадратичных оценок Шора // Кибернетика и системный анализ. – 2008. – № 6. – С. 89–99.
15. *Пшеничный Б.Н.* Метод линеаризации. – М.: Наука, 1983. – 136 с.
16. *Березовский О.А.* Двойственная квадратичная оценка Шора. Задача представления монохроматического изображения // Праці IV міжнар. школи-семінара "Теорія прийняття рішення". – Ужгород: УжНУ, 2008. – С. 13–14.
17. *Разработка* универсальной программной системы интеллектуального анализа данных, распознавания и прогноза / Ю.И. Журавлев, В.В. Рязанов,..., Д.А. Кропотов и др. // Докл. 11 Всерос. конф. "Математические методы распознавания образов" (ММРО-11). – М.: Регион-Холдинг, 2003. – С. 311–313.
18. *RECOGNITION*: A Universal Software System for Recognition, Data Mining, and Forecasting / Yu.I. Zhuravlev, V.V. Ryazanov, D.A. Kropotov at all // Pattern Recognition and Image Analysis. – 2005. – 15, N 2. – P. 476–479.

Наукове видання

СТЕЦЮК Петро Іванович
БЕРЕЗОВСЬКИЙ Олег Анатолійович
ЖУРБЕНКО Микола Григорович
КРОПОТОВ Дмитро Олександрович

МЕТОДИ НЕГЛАДКОЇ ОПТИМІЗАЦІЇ У СПЕЦІАЛЬНИХ ЗАДАЧАХ КЛАСИФІКАЦІЇ

Редактор Г.Ф. Трохимець

Підп. до друку 21.04.09. Формат 60×84/16. Бум. офс. Офс. друк. Ум. друк. арк. 1,63.
Ум. фарб.-відб. 1,75. Обл.-вид. арк. 1,88. Тираж 80 прим. Зам. 41.

Редакційно-видавничий відділ з поліграфічною дільницею
Інституту кібернетики ім. В.М. Глушкова НАН України
03680, МСП, Київ-187, проспект Академіка Глушкова, 40