УДК 519.7

О. А. Галкін
Київський національний університет ім. Тараса Шевченка

ПРОЦЕС РОЗПІЗНАВАННЯ НА ОСНОВІ ЯДЕРНИХ ОЦІНОК ЩІЛЬНОСТІ З ВИКОРИСТАННЯМ МНОЖИНИ ОПТИМАЛЬНИХ МІР ЗГЛАДЖУВАННЯ

Запропоновано комплексний метод для розв'язання задач розпізнавання образів, що грунтується на використанні множини оптимальних мір згладжування для кожної оцінки щільності класу. Для доведення ефективності використання різних конкуруючих доменів застосовано міру типу *p*-начення з апостеріорною ймовірністю. Для забезпечення гнучкості у проблемі залежності смуг пропускання від характерних спостережень вдавалися до залежних від даних скоригованих вагових функцій.

Предложен комплексный метод для решения задач распознавания образов, основанный на использовании множества оптимальных мер сглаживания для каждой оценки плотности класса. Для доказательства эффективности применения различных конкурирующих доменов использовали меру типа *p*-значения с апостериорной вероятностью. Для обеспечения гибкости в проблеме зависимости полос пропускания от характерных наблюдений использовали зависимые от данных скорректированные весовые функции.

In the article there have been proposed a complex method for solving problems of pattern recognition based on the use of multiple measures of optimal smoothing for each class density estimation. Measure of p-values type is used with the posterior probability for proving the effectiveness of the use of various competing domains. Depending on the data corrected weighting functions are used for providing flexibility in the problem of dependence of bandwidths on specific observations.

Ключові слова: міра згладжування, ядерна оцінка щільності, смуга пропускання.

Вступ. Ефективність класифікатора на основі ядерних оцінок щільності вирішальною мірою залежить від значень параметрів смуги пропускання. Існуючі методики вибору смуги пропускання, метою яких є мінімізація середньоквадратичних інтегрованих помилок для оцінки щільності не є достатньо ефективні для задач розпізнавання образів, а отже, можуть призводити до низьких показників помилкової класифікації. З іншого боку, відомий метод *V*-кратної перехресної перевірки та інші подібні методи вибору смуги пропускання в непараметричних задачах класифікації є не достатньо ефективні через кусково-постійну природу оцінених імовірнісних функцій помилкової класифікації з нескінченною множиною мінімумів.

Більшість таких методів на основі перехресної перевірки потребують значних обчислень, коли є декілька конкуруючих класів. Двома іншими важливими чинниками використання ядерних оцінок щільності є такі:

- вибір смуг пропускання повинен залежати від характерного спостереження, що необхідно класифікувати. Крім того, слід враховувати залежність від щільності доменів; у разі класифікації характерного спостереження треба оцінювати переваги того чи іншого домену для різних смуг пропускання в оцінках щільності відповідно до різних конкуруючих доменів;
- у багатокласовій задачі розпізнавання замість використання однієї смуги пропускання для кожної оцінки щільності домену більш ефективне є використання різних смуг пропускання для оцінки щільності класу у разі, коли її порівнювати з оцінками щільності для різних конкуруючих класів в класифікації характерного спостереження.

Замість вибору однієї оптимальної смуги пропускання для кожної оцін-ки щільності класу будемо розглядати сімейство оцінок щільності $\left\{ \hat{f}_{lS_l} : S_l \in \mathrm{H}_l \right\}$ для кожного комплексу в широкому діапазоні смуг пропускання. Припускаємо, що за спільного розгляду різних рівнів згладжування можна отримати більше інформації, необхідної для класифікації, ніж отриманої за допомогою методів на основі однієї оптимальної смуги пропускання для кожної оцінки щільності класу.

Постановка задачі. В методах класифікації на основі ядерних оцінок щільності використовуатимемо ядерні оцінки невідомих щільностей популяцій $f_l(x)$ $(l = 1, 2, ..., L)$. Такі оцінки підключаються в байєсівське правило для побудови такого класифікатора:

$$\mathrm{d}_K(x) = \arg\max_l p_l \hat{f}_{lS_l}(x),$$

де p_l – апріорні ймовірності, а $\hat{f}_{jS_j}(x)$ – ядерні оцінки щільності відповідних класів. Якщо $x_{l1}, x_{l2}, ..., x_{ln_l}$ є *d*-вимірними спостереженнями в навчальній вибірці з *l*-го домену $(l = 1, 2, ..., L)$, то ядерні оцінки щільності $\hat{f}_{lS_l}(x)$ *j*-го домену будемо задавати як

$$\hat{f}_{lS_l}(x) = n_l^{-1} S_l^{-d} \sum_{k=1}^{n_l} K\left\{ S_l^{-1}(x_{lk} - x) \right\},$$

де $K(\cdot)$ – d-вимірна функція щільності, а $S_l > 0$ – параметр смуги пропускання [1].

Припустимо, що $x_{l1}, x_{l2}, ..., x_{ln_l}$ є навчальна вибірка спостережень l-го класу, де $1 \le l \le L$. Для класифікації спостереження x в один з L класів спочатку необхідно отримати оцінки щільності $\hat{f}_{lS_l}(x)$ в точці x для всіх $l = 1, 2, ..., L$. Перед обчисленням оцінки щільності класу необхідно нормалізувати дані в класі, застосовуючи оцінку дисперсійної матриці класу для того, щоб зробити дані більш сферичними за природою і тим самим зробити використання загальної смуги пропускання S_l для всіх координатних змінних більш виправданим. Оцінку щільності для вихідних векторів даних можна отримати з нормалізованих векторів даних, скористовшись простою формулою перетворення для ймовірнісної функції щільності, коли випадкові вектори проходять лінійне перетворення. Для даної пари конкуруючих класів, наприклад класу-1 та класу-2, та фіксованої пари смуг пропускання S_1 та S_2 для двох оцінок щільності класу існує порядок між функціями $p_1 \hat{f}_{1S_1}(x)$ та $p_2 \hat{f}_{2S_2}(x)$, що визначає, який один з двох класів є більш ймовірний. Введемо тепер деякі міри для підтвердження доказів на користь одного чи іншого класу [2].

Метод розв'язування. У двокласовій задачі для заданого спостереження x та заданої пари смуг пропускання S_1 та S_2 апостеріорну ймовірність використання першого домену задамо як

$$\wp_{S_1,S_2}(1 \mid x) = \frac{p_1 \hat{f}_{1S_1}(x)}{p_1 \hat{f}_{1S_1}(x) + p_2 \hat{f}_{2S_2}(x)}.$$

Можна застосувати широкий спектр значень S_1 та S_2 для обчислення апостеріорних ймовірностей. У двокласовому ядерному дискримінантному аналізі спостереженя x класифікується в домен 1, якщо $p_1 \hat{f}_{1S_1}(x) > p_2 \hat{f}_{2S_2}(x)$. Для заданого спостереження x розглянемо таку ймовірність:

$$P_{S_1,S_2}(x) = P\left\{ p_1 \hat{f}_{1S}(x) > p_2 \hat{f}_{2S}(x) \mid x \right\}.$$

Зрозуміло, що великі та малі значення цієї ймовірності дають розв'язки на користь першого та другого доменів відповідно. Для фіксованих S_1 та S_2, оскільки оцінки щільності є середніми значеннями незалежних однаково розподілених випадкових величин, а оцінки щільності для різних популяцій грунтуються на незалежних множинах доменів, ми можемо використовувати нормальне наближення для оцінки вказаної ймовірності з великим ступенем точності навіть для дуже великих розмірів навчальної вибірки. Використовуючи таке нормальне наближення з оціненими середніми значеннями та дисперсіями, ми отримуємо таку ймовірність:

$$P_{S_1,S_2}(x) = \Phi\left(\frac{p_1 E\left[\hat{f}_{1S_1}(x) \mid x \right] - p_2 E\left[\hat{f}_{2S_2}(x) \mid x \right]}{\sqrt{p_1^2 Var\left[\hat{f}_{1S_1}(x) \mid x \right] + p_2^2 Var\left[\hat{f}_{2S_2}(x) \mid x \right]}} \right) =$$
$$= \Phi\left(\left\{ p_1 \hat{f}_{1S_1}(x) - p_2 \hat{f}_{2S_2}(x) \right\} / \sqrt{p_1^2 \hat{s}_{1S_1}^2(x) + p_2^2 \hat{s}_{2S_1}^2(x)} \right),$$

де Φ – стандартна функція нормального розподілу; n_1 та n_2 – розміри навчальної вибірки для двох класів, а $\hat{s}_{lS_l}^2(x)$ – оцінена дисперсія $\hat{f}_{lS_l}(x)$ $(l = 1, 2)$, отримана з навчальної вибірки з використанням вибіркової дисперсії від $S_l^{-d} K\left\{ S_l^{-1}(x_{l1} - x) \right\}$, $S_l^{-d} K\left\{ S_l^{-1}(x_{l2} - x) \right\}, \ldots, S_l^{-d} K\left\{ S_l^{-1}(x_{ln_l} - x) \right\}$.

Альтернативну інтерпретацію нормального наближення $P_{S_1,S_2}(x)$ можна задати наступним чином. Для заданого спостереження x та пари смуг пропускання S_1 та S_2 скористаємося парою гіпотез H_0: $p_1 E\left\{ \hat{f}_{1S}(x) \right\} \ge p_2 E\left\{ \hat{f}_{2S}(x) \right\}$ та $H_A : p_1 E\left\{ \hat{f}_{1S}(x) \right\} < p_2 E\left\{ \hat{f}_{2S}(x) \right\}$. Якщо взяти навчальну вибірку для тестування даних гіпотез з використанням ядерних оцінок щільності і розглядати їх як статистики, тобто вибіркові середні значення використовувати в двовибіркових задачах, то нормальне наближення можна прийняти як одностороннє p-значення, пов'язане з задачею тестування. Це і є суттєвою підтавою назвати його дискримінантною мірою типу p-значення [3].

Теорема 1. Припустимо, що $E\left[K^2 \left\{ S_i^{-1}(x - x_{i1}) \right\} \mid x \right] < \infty$ для $i = 1, 2$ для заданого спостереження x. Далі визначимо $\mu_{iS_i}(x) = E\left\{ \hat{f}_{iS_i}(x) \right\}$ для $i = 1, 2$. Якщо $n_1 / N \to \lambda (0 < \lambda < 1)$ за $N = n_1 + n_2 \to \infty$, ми маємо:

а) $\left| P_{S_1,S_2}(1 \mid x) - \dfrac{p_1 \mu_{1S_1}(x)}{p_1 \mu_{1S_1}(x) + p_2 \mu_{2S_2}(x)} \right| = O_P(N^{-1/2})$;

б) $\left| P_{S_1,S_2}(x) - I\left\{ p_1 \mu_{1S_1}(x) > p_2 \mu_{2S_2}(x) \right\} \right| = O_P(N^{-1/2} e^{-CN})$ для деякого $C > 0$, де $I\{\cdot\}$ означає характеристичну функцію.

Доведення. (а) Для того щоб спростити вирази, визначимо $T_i = p_i \hat{f}_{iS_i}(x)$ для $i = 1,2$. Оскільки T_i є середнє значення незалежних та однаково розподілених випадкових величин, з центральної граничної теореми виходить, що за допустимої умови моменту для великих розмірів вибірки T_i прямує до нормального розподілу із середнім значенням $\tau_i = p_i \varepsilon_{iS_i}(x)$ та дисперсією $v_i = p_i^2 s_{iS_i}^2(x)$, що має порядок $O(n_i^{-1})$.

Далі, визначимо $\Psi(T_1, T_2) = T_1/(T_1 + T_2)$. У даному випадку T_1 та T_2 є додатньо визначені та незалежні випадкові величини. Крім того, функція Ψ є неперервно диференційовною в T_1 та T_2. Асимптотичний розклад Тейлора приводить до того, що

$$\frac{\left\{\Psi(T_1,T_2) - \Psi(\tau_1,\tau_2)\right\}}{\nu} \xrightarrow{L} \textit{Нормаль}(0,1), \text{ де } \nu = \left\{\sum_{i=1}^{2} v_i (\partial\Psi/\partial T_i)^2_{T_1=\tau_1, T_2=\tau_2}\right\}^{1/2}.$$

Оскільки $n_1/N \to \lambda$, а $n_2/N \to 1-\lambda (0 < \lambda < 1)$ за $N \to \infty$, ми маємо, що $\left|\Psi(T_1,T_2) - \Psi(\tau_1,\tau_2)\right| = O_P(N^{-1/2})$.

(б) Припустимо, що $\tau_1 > \tau_2$, тобто $I\left\{p_1\varepsilon_{1S_1}(x) > p_2\varepsilon_{2S_2}(x)\right\} = 1$. Відтак для деяких фіксованих S_1, S_2 та x з частини (а) даної теореми виходить, що

$$\frac{1}{\sqrt{v_1 + v_2}}\left[(T_1 - T_2) - (\tau_1 - \tau_2)\right] \xrightarrow{L} \textit{Нормаль}(0,1), \text{ за } N \to \infty.$$

Тепер визначимо $Z_{S_1,S_2}(x) = \frac{1}{\sqrt{v_1+v_2}}(T_1 - T_2)$. З порядку v_1 та v_2 бачимо, що $Z_{S_1,S_2}(x) = O_P(N^{1/2})$, а $\frac{1}{\sqrt{N}} Z_{S_1,S_2}(x)$ – збіжний за ймовірністю до постійного значення *С*. Тому для $x > 0$, а також з урахуванням того, що $\frac{1}{x}\phi(x) < 1 - \Phi(x) < \left(\frac{1}{x} - \frac{1}{x^3}\right)\phi(x),$ де $\phi(\cdot)$ та $\Phi(\cdot)$ означають імовірністну та кумулятивну функції щільності стандартного нормального розподілу відповідно, ми отримуємо, що $1 - P_{S_1,S_2}(x) = 1 - \Phi(Z_{S_1,S_2}(x)) = O_p(N^{-1/2}e^{-CN})$. Теорему доведено.

Дана теорема дає уявлення про асимптотичну поведінку дискримінантних мір. Для будь-якого заданого *х* та пари смуг пропускання (S_1, S_2) оцінена апостеріорна імовірність $P_{S_1,S_2}(x)$ збігається до $p_1\mu_{1S_1}(x)/\left[p_1\mu_{1S_1}(x) + p_2\mu_{2S_2}(x)\right]$ зі швидкістю $O(N^{-1/2})$, однак залежно від $\mu_{1S_1}(x), \mu_{2S_2}(x)$ та апостеріорних імовірностей $P_{S_1,S_2}(1|x)$ вона експоненціально збігається до 0 або 1. Наприклад, якщо $p_1\mu_{1S_1}(x) < p_2\mu_{2S_2}(x)$, $P_{S_1,S_2}(1|x)$ матиме швидкість збіжності $\sqrt{N}$ до значення, меншого 0.5, однак відповідні міри типу *р*-значення експоненціально збігаються до нуля набагато швидше. Тому для заданих (S_1, S_2) за зростання об'єму навчальної вибірки $P_{S_1,S_2}(x)$ завжди даватиме більш переконливі докази, ніж $P_{S_1,S_2}(1|x)$, за чи проти домену 1.

Для того щоб досягти рішення відносно класифікації спостереження *х*, важливо знати, які значення пари смуг пропускання (S_1, S_2) найшвидше приведуть до статистично більш достовірного результату класифікації. Для деякого фіксованого вибору (S_1, S_2) розглянемо середню ймовірність помилкової класифікації для двокласової задачі, заданої таким чином:

$$\Xi(S_1,S_2) = p_1 \int_{x\in\Re^c_{S_1,S_2}} f_1(x)dx + p_2 \int_{x\in\Re_{S_1,S_2}} f_2(x)dx,$$

де $\Re_{S_1,S_2}$ є множина всіх *х*, класифікованих до домену 1, а $\Re^c_{S_1,S_2}$ – множина доповнення. Існують декілька методів на основі перехресної перевірки для оцінки показника помилкової класифікації $\Xi(S_1,S_2)$ для класифікатора з використанням навчальної вибірки. Дані підходи потребують свого роду емпіричної пропорції помилково класифікованих спостережень, що призводить до оцінок, які є кусково-постійні в природі, у той час коли реальна функція може бути хорошою гладкою функцією. Для різних варіантів смуг пропускання було використано гладку та більш точну оцінку середньої ймовірності помилкової класифікації для класифікаторів на основі ядерних оцінок щільності [4; 5].

Природний спосіб об'єднання результатів, отриманих на різних рівнях згладжування для досягнення остаточного рішення, полягає у формуванні відповідних середньозважених апостеріорних ймовірностей, обчислених для різних варіантів вибору (S_1, S_2). Для цього необхідно використовувати відповідну вагову функцію. Зрозуміло, що вагова функція має набувати більш великих значень для даної пари смуг пропускання, що приводить до менших показників помилкової класифікації. Існує багато відповідних вагових функцій, що задовольняють дану властивість. Беггінг, бустінг та дуговий класифікатор є досить відомі методи для об’єднання результатів різних класифікаторів для підвищення їх продуктивності. Дані методи також припускають різні ваги різним класифікаторам на основі їх ймовірностей помилкової класифікації та комбінують результати, використовуючи ці ваги [6–8].

Далі визначимо $\Xi_0 = \min_{S_1,S_2} \hat{\Xi}(S_1,S_2)$ та розглянемо вагові функції $w(S_1,S_2)$, які є спадними функціями від $\hat{\Xi}(S_1,S_2)$, що еквівалентно $\hat{\Xi}(S_1,S_2)-\Xi_0$. Крім того, $w(S_1,S_2)$ повинні зникати кожного разу, коли відповідне значення $\hat{\Xi}(S_1,S_2)$ перевищує будь-яку з двох апріорних ймовірностей, оскільки продуктивність класифікатора буде більш низькою у порівнянні з тривіальним класифікатором, що класифікує всі спостереження до класу, що має найвищу апріорну ймовірність. Зауважимо, що для того щоб класифікувати спостереження *x*, варто включати у вагах відповідну міру $P_{S_1,S_2}(x)$ типу *p*-значення. Крім того, варто більше покладатися на ті пари смуг пропускання, що приводять до більш обгрунтованих доказів для одного або двох класів та відповідно налаштовувати значення вагової функції. Дані налаштовані ваги будуть не лише залежати від загально оцінених ймовірностей помилкової класифікації, а й від класифікованого характерного спостереження. У всіх числових розрахунках використано таку налаштовану вагову функцію:

$$w_x(S_1,S_2) = w(S_1,S_2)\left|P_{S1,S2}(x)-0.5\right|,$$

де

$$w(S_1,S_2)=\begin{cases} \exp\left\{-\dfrac{1}{2}\dfrac{(\hat{\Xi}(S_1,S_2)-\Xi_0)^2}{\Xi_0(1-\Xi_0)/N}\right\}, \text{ якщо } \dfrac{\hat{\Xi}(S_1,S_2)-\Xi_0}{\left[\Xi_0(1-\Xi_0)/N\right]^{1/2}} \le \tau \\ \qquad\qquad\qquad\qquad \text{та } \hat{\Xi}(S_1,S_2) < \min\left\{p_1,p_2\right\} \\ 0 \qquad\qquad\qquad\qquad \text{в іншому випадку.} \end{cases}$$

Для $N = n_1 + n_2$, Ξ_0 та $\Xi_0(1-\Xi_0)/N$ можна розглядати оцінки для середнього значення та дисперсії емпіричного показника помилкової класифікації найбільш продуктивного класифікатора на основі оцінок щільності ядра, коли такий класифікатор використовують для класифікації N незалежних спостережень. Постійна величина τ визначає максимальне число відхилень від мінімального оціненого показника помилкової класифікації в стандартизованій шкалі, за якою вагова схема ігнорує пару смуги пропускання (S_1,S_2), поставивши на них нульову вагу. Зрозуміло, що $\tau=0$ відповідає ситуації прикріплення всіх ваг лише на пару смуги пропускання (S_1,S_2), для якої $\hat{\Xi}(S_1,S_2)=\Xi_0$. Зазначимо також, що вибір вагової функції гауссова типу визначає, що для практичних цілей немає потреби розглядати значення τ, що є більше трьох. Даний вибір скоригованої вагової функції є дещо суб'єктивний, тому можна використовувати інші відповідні функції для тих же цілей. Однак нашим емпіричним досвідом є те, що кінцевий результат не є дуже чутливий до вагової процедури, якщо використовувати прийнятну вагову функцію [9–11].

Для деякого фіксованого S_l $\hat{f}_{lS_l}(x)$ є середнє значення незалежних та однаково розподілених випадкових величин. Коли розмір вибірки n_l збільшується, її дисперсія прагне до нуля і збігається за імовірністю до $\mu_{lS_l}(x) = E\left\{\hat{f}_{lS_l}(x)\right\} = K_{S_l} * f_l(x)$, що є згорткою функції щільності f_l з ядром K та смугою пропускання S_l. У двокласовій задачі порядок $p_1\mu_{1S_1}(x)$ та $p_2\mu_{2S_2}(x)$ визначає асимптотичне правило рішення. Крім того, якщо розподіли комплексу задовольняють моделі зсуву локалізації (тобто $f_l(x) = g(x-\mu_l)$ для деякої загальної щільності g та параметрів локалізації μ_l), а однакову смугу пропускання S використовують для обох комплексів, $\mu_{lS}(x)$ $(l=1,2)$ зберігає порядок фактичних щільностей для всіх значень S. Більше того, у випадку рівних апріоних ймовірностей для всіх позитивних значень $S = S_1 = S_2$ відповідна ймовірність помилкової класифікації асимптотично стає оптимальним байєсовим ризиком. Для великих значень S, оскільки дисперсія прагне до нуля досить швидко, дана збіжність є значно швидша. Наступна теорема дає уявлення щодо поведінки класифікатора на основі ядерних оцінок щільності для великих розмірів вибірок та великих смуг пропускання, коли щільності комплексу не обов’язково мають задовольняти будь-яку умову симетрії.

Теорема 2. Припустимо, що f_1 та f_2 є такі, що $\int \|x\|^6 f_i(x)dx < \infty$, а ядро K є *d*-вимірна функція щільності з модою в 0 та обмеженою третьою похідною. Визначимо постійну величину $C_\pi = p_2 / p_1$ та припустимо, що S_1, S_2 варіюються таким чином, що $S_2 / S_1 = C_S$ та є постійна. Оскільки $S_1 \to \infty$, $\Xi(S_1,S_2)$ має таку асимптотичну поведінку.

(а) Коли $C_\pi > C_S^d$, за $S_1,S_2 \to \infty$, $\Xi(S_1,S_2) \to p_1$.

(б) Коли $C_\pi < C_S^d$, за $S_1,S_2 \to \infty$, $\Xi(S_1,S_2) \to p_2$.

(в) Коли $C_\pi = C_S^d$, за $S_1, S_2 \to \infty$, $\Xi(S_1, S_2)$ прямує до ймовірності помилкової класифікації квадратичного правила класифікації, що задається так:

$$\begin{aligned} \mathrm{d}_Q(x) &= 1, \quad \text{якщо } C_S^2 E_{f_1}\left\{(x-X)'\nabla^2 K(0)(x-X)\right\} > \\ &> E_{f_2}\left\{(x-X)'\nabla^2 K(0)(x-X)\right\} \\ &= 2, \quad \text{в іншому випадку.} \end{aligned}$$

Доведення. Спочатку зауважимо, що $\Xi(S_1,S_2) = p_1 E_{f_1}\{I(p_1 \hat f_{1S_1} < p_2 \hat f_{2S_2})\} + p_2 E_{f_2}\{I(p_1 \hat f_{1S_1} > p_2 \hat f_{2S_2})\}$. Із визначення $\hat f_{lS_l}(x)$ $(l = 1, 2)$, можна побачити, що

$$E_{f_l}\{\hat f_{S_l}(x)\} = S_l^{-d} E_{f_l}[K\{(x-X)/S_l\}] \text{ та } Var_{f_l}\{\hat f_{S_l}(x)\} = n_l^{-1} S_l^{-2d} Var_{f_l}[K\{(x-X)/S_l\}].$$

Використовуючи розклад Тейлора в 0, $K\{(x-X)/S_l\}$ можна виразити як

$$\begin{aligned} K\{(x-X)/S_l\} &= K(0) + (1/2S_l^2)\{(x-X)'\}\nabla^2 K(0)(x-X) + \\ &+ (1/6S_l^3)\sum_{i,k,m} Y_{i,k,m}, (\text{оскільки } \nabla K(0) = 0), \end{aligned}$$

де $Y_{i,k,m} = (x_i - X_i)(x_k - X_k)(x_m - X_m)\dfrac{\partial^3 K(t)}{\partial t_i \partial t_k \partial t_m}|_{t=\xi}$, для деякого проміжного вектора ξ між 0 та $(x-X)/S_l$.

Тому, використовуючи той факт, що функція K обмежена третіми похідними, а $\int \|x\|^6 f(x)dx < \infty$, ми отримуємо

$$E_{f_l}\{\hat f_{lS_l}(s)\} = S_l^{-d}\left[K(0) + (1/2S_l^2)E_{f_l}\{(x-X)'\nabla^2 K(0)(x-X)\} + O(S_l^{-3})\right] \text{та}$$

$$Var_{f_l}\{\hat f_{lS_l}(s)\} = (4 n_l S_l^{2d+4})^{-1}\left[Var_{f_l}\{(x-X)'\nabla^2 K(0)(x-X)\} + O(S_l^{-3})\right].$$

Оскільки дисперсія ядерних оцінок щільності асимптотично прагне до нуля, для будь-якого спостереження *x* та будь-якої пари смуг пропускання S_1, S_2 відповідний класифікатор класифікує спостереження *x* до домену 1 тоді і тільки тоді, коли

$$\begin{aligned} & p_1 E_{f_1}\{\hat f_{1S_1}(x)\} > p_2 E_{f_2}\{\hat f_{2S_2}(x)\} \\ & \Leftrightarrow p_1 S_1^{-d}\left[K(0) + (1/2S_1^2)E_{f_1}\{(x-X)'\nabla^2 K(0)(x-X)\} + O(S_1^{-3})\right] > \\ & > p_2 S_2^{-d}\left[K(0) + \left[K(0) + (1/2S_2^2)E_{f_2}\{(x-X)'\nabla^2 K(0)(x-X)\} + O(S_2^{-3})\right]\right] \\ & \Leftrightarrow C_\pi C_S^{-d}\left[K(0) + (1/2S_1^2)E_{f_1}\{(x-X)'\nabla^2 K(0)(x-X)\} + O(S_1^{-3})\right] > \\ & > \left[K(0) + (1/2S_2^2)E_{f_2}\{(x-X)'\nabla^2 K(0)(x-X)\} + O(S_2^{-3})\right]. \end{aligned}$$

(а) Коли $C_\pi < C_S^d$, для великих значень S_1 та $S_2 = C_S S_1$, верхня нерівність є справедлива, яким б не було значення спостереження *x*.

(б) Таким же чином, коли $C_\pi > C_S^d$, для кожного спостереження *x* результуючий класифікатор асимптотично завжди класифікує його до домену 2.

(в) Коли $C_\pi = C_S^d$, для великих значень S_1 та S_2 легко перевірити, що вищенаведена нерівність є справедлива тоді і тільки тоді, коли $C_S^2 E_{f_1}\{(x-X)'\nabla^2 K(0)(x-X)\} > E_{f_2}\{(x-X)'\nabla^2 K(0)(x-X)\}$. Теорему доведено.

Зауважимо, що коли $C_\pi = C_S = 1$ (тобто $p_1 = p_2$), квадратичний класифікатор стає лінійним класифікатором, що задається таким чином:

$$\mathrm{d}_L(x) = \arg\min_i \left[x'\nabla^2 K(0) E_{f_i}(X) - \frac{1}{2} E_{f_i}\left\{X'\nabla^2 K(0) X\right\}\right].$$

Враховуючи припущення щодо локалізаційного зсуву та сферичної симетрії, а також що ядерна функція K є також сферична (відзначимо, що $\nabla^2 K(0)$ є від'ємно визначена функція), даний лінійний класифікатор можна виразити у спрощеній формі, а саме

$$\mathrm{d}_l(x) = \arg\max_i \left\{x'\mu_i - \frac{1}{2}\mu_i'\mu_i\right\},$$

де μ_i – параметр локалізації для *i*-го комплексу ($i = 1, 2$). Варто відзначити, що описаний лінійний класифікатор є оптимальний байєсовий класифікатор. Тому у даному випадку ймовірність помилкової класифікації Ξ асимптотично збігається до оптимального байєсового ризику.

Висновки. Запропонований комплексний метод для розв'язання задач розпізнавання образів, що грунтується на використанні множини оптимальних мір згладжування для кожної оцінки щільності класу є більш інформатичний метод у порівнянні зі звичайним ядерним дискримінантним аналізом. Водночас з результатами класифікації було отримано змогу мати уявлення про ефективність дискримінантної міри та пов'язаних з нею статистичних невизначенностей. В методі на основі єдиної оптимальної смуги пропускання дана смуга пропускання не залежить від характерного спостереження, що підлягає класифікації. Використання залежної від даних скоригованої вагової функції в запропонованому методі забезпечує таку гнучкість.

У багатокласових задачах розпізнавання часто є обчислювально складно отримати оптимальні смуги пропускання у разі мінімізації оціненої ймовірності помилкової класифікації $\hat{\Xi}(S_1, S_2, ..., S_L)$. Метод попарної класифікації, застосовуваний у даному дослідженні, не лише значно знижує обчислювальне навантаження, але й забезпечує гнучкість у використанні різних смуг пропускання для класу, де один клас порівнюється з різними конкуруючими класами. Оскільки відповідний вибір смуги пропускання для оцінки щільності домену в задачі класифікації може залежати як від її конкуруючих оцінок щільності домену, так і від відповідних апріорних ймовірностей, даний метод попарного порівняння є досить ефективний.

Бібліографічні посилання

1. **Yang, C.** Improved fast gauss transform and efficient kernel density estimation [Text] / C. Yang, R. Duraiswami, N.A. Gumerov, L. Davis. – ICCV. – 2003.
2. **Shi, Q.** Hash kernels for structured data [Text] / Q. Shi, J. Petterson, G. Dror, J. Langford, A. Smola, S. Vishwanathan // Journal of Machine Learning Research. – 2009. – V.10. – P. 2615–2637.
3. **Shawe-Taylor, J.** Kernel Methods for Pattern Analysis [Text] / J. Shawe-Taylor, N. Cristianini. – Cambridge University Press, 2004.
4. **Scholkopf, B.** Learning with Kernels: Support Vector Machines, Regularization, Optimization, and Beyond [Text] / B. Scholkopf, A. J. Smola, MIT Press, 2002.
5. **Raykar, V.C.** Fast computation of kernel estimators [Text] / V. C. Raykar, R. Duraiswami, L. H. Zhao // J. of Computational and Graphical Statistics. – 2010. – 19(1). – P. 205–220.
6. **Raykar, V.C.** Fast optimal bandwidth selection for kernel density estimation [Text] / V. C. Raykar, R. Duraiswami. – In SDM, 2006.
7. **Chen, Y.** Super-samples from kernel hearding [Text] / Y. Chen, M. Welling, A. Smola // In Conference on Uncertainty in Artificial Intellegence. – 2010.
8. **Raykar, V.C.** Very fast optimal bandwidth selection for univariate kernel density estimation [Text] / V.C. Raykar, R. Duraiswami. Technical Report CS-TR-4774, University of Maryland, CollegePark, 2005.
9. **Girolami, M.** Probability density estimation from optimally condensed data samples [Text] / M. Girolami and C. He // IEEE Transactions on Pattern Analysis and Machine Intelligence. – 2003. – V.25, №10. – P. 1253–1264.
10. **Hall, P.** Cross-Validation and the Estimation of Conditional Probability Densities [Text] / P. Hall, J. Racine, Q. Li // Journal of the American Statistical Association. – 2004. – V. 99. – P. 1015–1026.
11. **Hansen, B.E.** Uniform Convergence Rates for Kernel Estimation with Dependent Data [Text] / B.E. Hansen, Econometrics Theory, 2008.

Надійшла до редколегії 19.01.2015