

УДК 519.7

Галкін О. А., к.ф.-м.н., м.н.с.

**Вплив варіацій смуги пропускання на поведінку показника помилкової класифікації ядерного класифікатора**

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

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

**Effect of bandwidth variations on the behaviour of misclassification rate of kernel classifier**

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

*Стаття присвячена проблемі виявлення відмінностей між використанням оптимальної смуги пропускання, що мінімізує середньо-квадратичні інтегровані помилки ядерної оцінки щільності розподілу та оптимальної смуги пропускання, що мінімізує середню ймовірність помилкової класифікації, коли ядерні оцінки щільності розподілу використовуються для дискримінантного аналізу.*

*Ключові слова: смуга пропускання, ядерна оцінка, класифікація.*

*The article is devoted a problem of detecting differences between using the optimal bandwidth that minimizes the mean-integrated square error of kernel density estimate and the optimal bandwidth that minimizes the average probability of misclassification when kernel estimates of distribution used for discriminant analysis. Attention is paid to the behavior of the average probability of misclassification for different choices of the bandwidth parameter depending on whether a priori probabilities for different populations are equal or not. Basic work of experts assignments, and also demonstrated that when all a priori probabilities are equal, depending on the nature of density functions and kernel, large bandwidths can also lead to nearly optimal misclassification rates. The results presented in the article were obtained taking into account the fact that with increasing sample size one should gradually reduce bandwidth to zero to achieve greater performance of kernel density estimates of distribution.*

*Key Words: bandwidth, kernel estimate, classification.*

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

**Вступ**

Ядерні оцінки щільності розподілу є одним з найбільш відомих методів побудови непараметричних оцінок щільності комплексів. Використання ядерних оцінок щільності в дискримінантному аналізі є досить поширеним в науковій літературі та в багатьох програмних продуктах. Якщо  $x_{i1}, x_{i2}, \dots, x_{in_i}$  є  $d$ -вимірними спостереженнями в навчальній вибірці з  $l$ -м комплексом, ядерна оцінка щільності  $f_l(x)$  обчислюється як

$$\hat{f}_{lS}(x) = n_l^{-1} S^{-d} \sum_{k=1}^{n_l} K \{ S^{-1} (x_{ik} - x) \},$$

де функція ядра  $K(\cdot)$  є функцією щільності в  $d$ -вимірному просторі, а  $S > 0$  - відповідним згладжувальним параметром, відомим як смуга пропускання. Правило класифікації, що ґрунтується на даних ядерних оцінках щільності розподілу може бути задане як

$$d_K(x) = \arg \max_l \pi_l \hat{f}_{lS}(x).$$

Для звичайної задачі оцінки щільності оптимальна пропускна здатність береться як така, що мінімізує середньо-квадратичні інтегровані помилки ядерної оцінки щільності розподілу [1], а саме

$$E_{\text{int}} = E \left[ \int \{ \hat{f}_{IS}(x) - f_I(x) \}^2 dx \right].$$

Оскільки ефективність непараметричного класифікатора залежить від відповідних оцінок щільності класу, вибір згладжувального параметра відіграє важливу роль в задачах класифікації. Постає питання: наскільки відповідним є середній показник помилкової класифікації, коли смуга пропускання, що мінімізує середньо-квадратичні інтегровані помилки в задачах оцінки щільності використовується для задач класифікації?

#### Виклад основного матеріалу

У спробі дослідити дане питання почнемо з розгляду простої двокласової проблеми з рівними апріорними ймовірностями, де класи є багатовимірними нормальними з однаковою дисперсійною матрицею  $\Sigma = I$ , але різними середніми векторами  $\mu_1$  та  $\mu_2$ . У даному випадку смуга пропускання, що мінімізує середньо-квадратичні інтегровані помилки є однаковою для обох класів, якщо вона має однакову кількість даних з двох класів навчальної вибірки. Крім того, якщо використовується нормальне ядро, можна обчислити смугу пропускання, що аналітично мінімізує середньо-квадратичні інтегровані помилки для нормально розподілених даних. Для даної проблеми усереднена ймовірність помилкової класифікації  $\Xi$  також може бути оцінена та виведена у якості функції від смуги пропускання  $S$ . Оскільки ядерна оцінка щільності розподілу є середнім значенням незалежних та однаково розподілених випадкових величин, можна використовувати нормальне наближення для її розподілу. Середнє значення та дисперсія даного нормального наближення мають відповідні аналітичні вирази, коли розподіли даних та ядро є нормальними. Було зроблено спробу оцінити  $\Xi(S)$  для даного значення  $S$  по двом різним процедурам, одна з використанням нормального наближення (описана вище), а інша по багатомасштабній симуляції Монте-Карло. Помітної різниці не спостерігалось на побудованих значеннях  $\Xi(S)$  для двох різних підходів – розмір вибірки ( $n_1 = n_2 = n = 50$ ) був достатньо вдалим для дуже високого ступеня точності в нормальному наближенні для розподілу ядерних оцінок щільності.

На рис. 1-4 значення  $\Xi(S)$  були отримані для різних варіацій  $S$  та для різних розмірностей ( $d=1, 2, 4, 6$ ), де було вибрано  $\mu_1 = (0, 0, \dots, 0)$  та  $\mu_2 = (2, 0, \dots, 0)$ , а розмір вибірки дорівнював 50 для обох класів. Дані рисунки показують чітку різницю між оптимальною смугою пропускання для звичайної задачі оцінки щільності та для задачі класифікації.

Для різних розмірностей оптимальна смуга пропускання для задач класифікації (смуга пропускання призводить до найнижчої ймовірності помилкової класифікації) позначена як '\*', а ширина смуги, що мінімізує середньо-квадратичні інтегровані помилки позначена як 'o'. Дана різниця між двома смугами пропускання стає більшою при зростанні розмірності  $d$ . Для розмірності  $d=6$  найкраща ширина смуги для задачі класифікації зменшує середній показник помилкової класифікації майже на 34%, у порівнянні з частотою помилок, що відповідає оптимальній смузі пропускання, яка мінімізує середньо-квадратичні інтегровані помилки в задачі оцінки щільності.

Досить цікавою на рис. 1-4 є поведінка  $\Xi(S)$  для великих значень  $S$ . Відомо, що для задачі оцінки щільності середньо-квадратичні інтегровані помилки виявляються великими для дуже малих значень смуги пропускання (через велику дисперсію) так само, як і для дуже великих значень смуги пропускання (через великий зсув) [2]. Однак, у всіх випадках на рис. 1-4  $\Xi(S)$  стає має гладким після досягнення його мінімального значення. На відміну від того, що відбувається у випадку звичайної оцінки щільності, великі значення ширини смуг не є поганим вибором для розглянутих задач класифікації. Змінюючи  $\mu_1$ ,  $\mu_2$  та  $\Sigma$ , ми отримуємо різні графіки для  $\Xi(S)$ , але основна модель залишається незмінною.

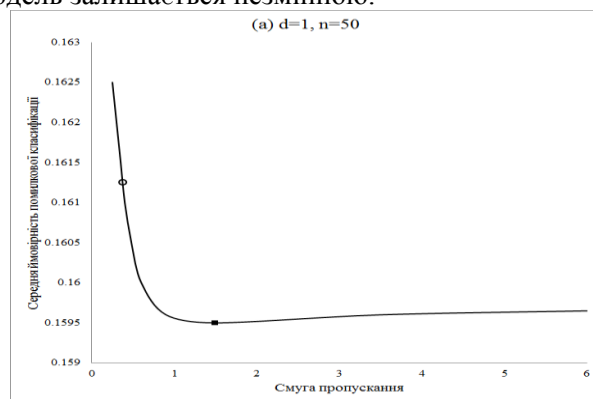


Рис. 1 Істинні  $\Xi$ -функції та оптимальні смуги пропускання (1 експеримент)

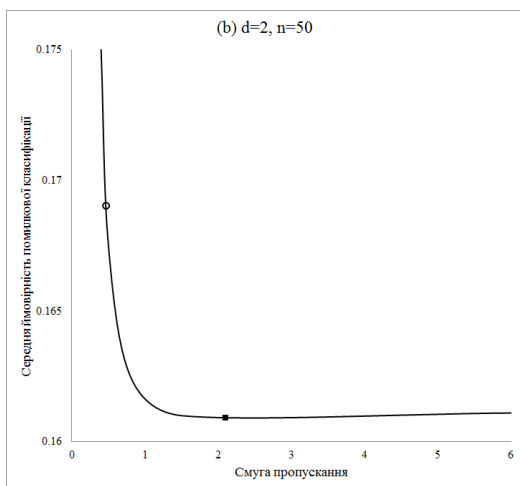


Рис. 2 Істинні  $\Xi$ -функції та оптимальні смуги пропускання (2 експеримент)

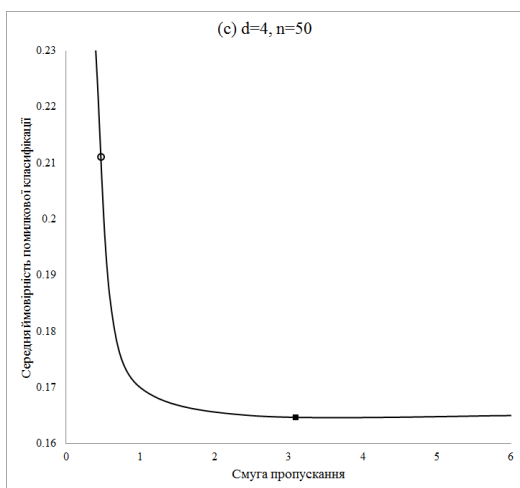


Рис. 3 Істинні  $\Xi$ -функції та оптимальні смуги пропускання (3 експеримент)

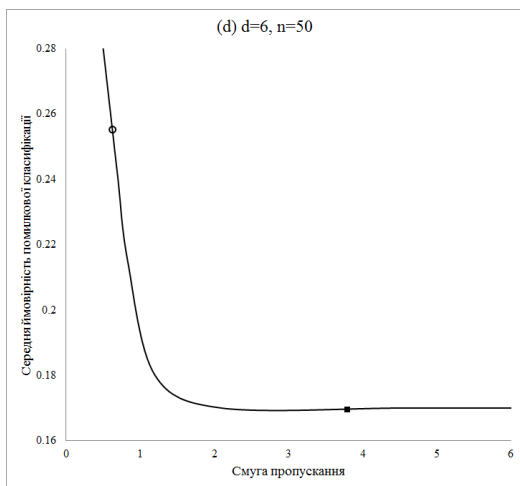


Рис. 4 Істинні  $\Xi$ -функції та оптимальні смуги пропускання (4 експеримент)

Існують також інші методи вибору оптимальної смуги для задач класифікації, що базуються на технології перехресної перевірки [3-4]. Наприклад,  $Z$ -кратний метод перехресної перевірки розділяє навчальну вибірку на  $Z$  частин, розміри яких є майже рівними. Зазвичай, стратифікована випадкова вибірка використовується для формування складок, де спостереження, що належать до різних класів використовуються як різні формації. Беручи одну складку в той момент, коли і тестову вибірку, вона використовує різні ширини смуг для класифікації своїх членів на основі навчальної вибірки, сформованої всіма спостереженнями, що належать до інших  $Z-1$  складок. Дана процедура повторюється по  $Z$  складках, а загальна частка помилкової класифікації використовується для оцінки  $\Xi(S)$ . Ширина смуги  $S$ , для якої оціночне значення  $\Xi(S)$  є мінімальним вважається оптимальною шириною смуги. Коли ми маємо усі  $N$  спостережень, пропускна або  $N$ -кратна перехресна перевірка може розглядатися як спеціальний випадок цієї процедури, де кожна складка складається з одного спостереження [5]. Оскільки спостережувані частки помилкової класифікації використовуються для оцінки  $\Xi(S)$ , оцінки представляють собою ступінчасті функції замість того, щоб бути гладкими кривими, коли істинне  $\Xi(S)$  є прийнятною гладкою функцією. Отже, замість єдиного унікального мінімуму дана процедура часто призводить до інтервалу, або об'єднання декількох інтервалів, як можливих варіантів вибору для гладкого параметра, з якого важко вибрати єдине оптимальне значення.

Рис. 5-8 чітко демонструють обмеження даної технології на основі перехресної перевірки. Тут було розглянуто таку ж проблему, як і на рис. 1-4 та згенеровано вибірки з тих же нормальних комплексів. Істинні та оцінені (по пропускній та 10-кратній перехресній перевірках) середні ймовірності помилок класифікації виведені одночасно на рис. 4-8. Оцінені криві не лише мають поведінку ступінчастих функцій, але і в деяких випадках пропускають відповідні місця розташування оптимальних смуг пропускання з великим запасом.

Для досить великих смуг пропускання середньо-квадратичної інтегрованої помилки ядерна оцінка щільності розподілу є досить великою, що є результатом значного зсуву. Встановлено, що значення показника помилкової класифікації  $\Xi(S)$  спочатку досягає свого мінімуму, а потім стає майже гладким для

широкого діапазону великих значень смуги пропускання  $S$ . Спочатку пояснимо таку явну аномальну поведінку  $\Xi(S)$  на прикладах. Припустимо, що існує  $n$  спостережень в навчальній вибірці з кожного комплексу, а загальна смуга пропускання використовується для різних оцінок щільності комплексів, що обґрунтовані для таких випадків, як наприклад, моделі локалізації міграції комплексів.

Величини  $E\{\hat{f}_{IS}(x)\}$  та  $\hat{f}_{IS}(x)$  можуть розглядатися відповідно як теоретичні та емпіричні масштабно-просторові функції від  $l$ -го комплексу для різних варіацій згладжувального параметра  $S$  [6-7]. Теоретичні масштабно-просторові функції  $E\{\hat{f}_{IS}(x)\}$  є згортками істинних щільностей  $f_i(x)$  з ядром  $K$  та зі смугою пропускання  $S$ .

Відомо, що із зростанням розміру вибірки дисперсія ядерної оцінки щільності розподілу, що є середнім значенням множини незалежних та однаково розподілених випадкових величин зменшується і як наслідок, для деякої фіксованої смуги пропускання  $S$  розподіл  $\hat{f}_{IS}(x)$  має тенденцію бути майже дегенеративним в  $E\{\hat{f}_{IS}(x)\}$ , коли розмір вибірки є досить великим.

Для фіксованого значення  $S$ , коли апостеріорні ймовірності для різних комплексів є рівними, а також коли розмір вибірки  $n$  прямує до нескінченності, класифікатор на основі ядерної оцінки щільності розподілу має тенденцію класифікувати спостереження до класу, що має найбільше значення для теоретичної функції масштабованого простору. Це виконується, коли апостеріорні ймовірності для різних комплексів є рівними, а розмір вибірки  $n$  прямує до нескінченності при фіксованому значенні  $S$ . Те ж саме має місце і для випадку згортки, коли функції  $f$  та  $K$  є сферично-симетричними та строго спадними функціями відстані від їх центрів симетрії. Задовольняючи умови моделі локалізації зсуву, функції масштабованого простору зберігають порядок серед початкових щільностей для всіх значень  $S$ .

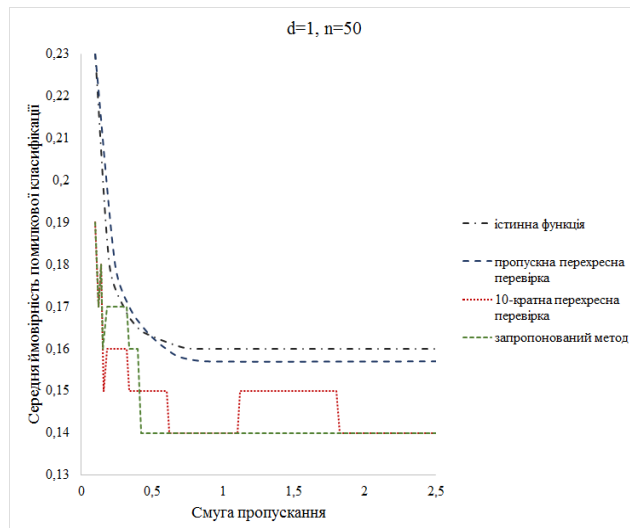


Рис. 5 Середні ймовірності помилкової класифікації (1 експеримент)

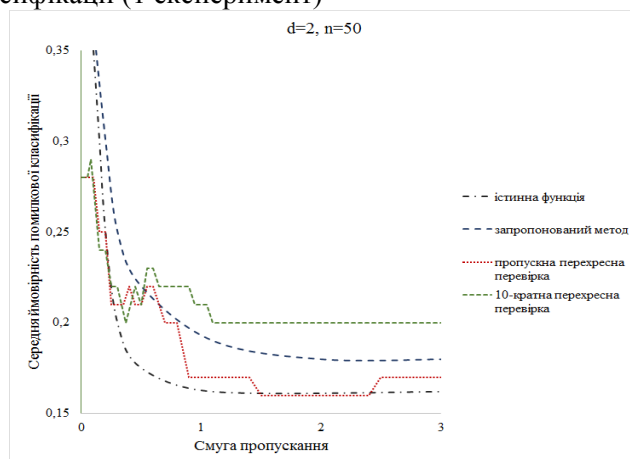


Рис. 6 Середні ймовірності помилкової класифікації (2 експеримент)

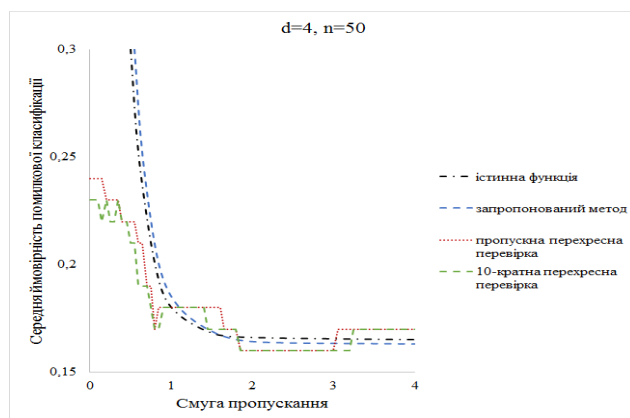


Рис. 7 Середні ймовірності помилкової класифікації (3 експеримент)

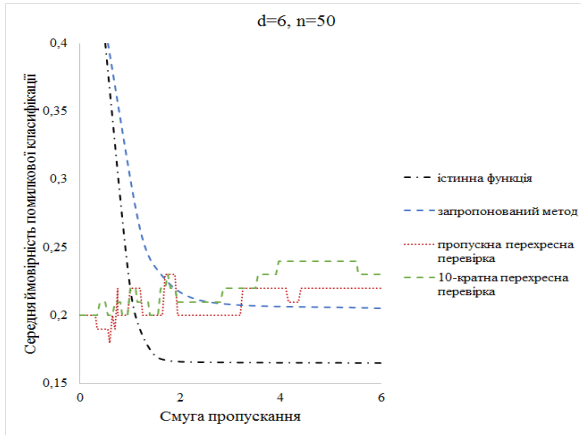


Рис. 8 Середні ймовірності помилкової класифікації (4 експеримент)

На попередньо представлених рисунках продемонстровано випадки рівних апостеріорних ймовірностей.

**Теорема 1.** Припустимо, що  $f_1, f_2, \dots, f_L$  та  $K$  є сферично-симетричними щільностями, а функції  $f_l$  задовольняють умовам моделі локалізації зсуву, тобто  $f_l(x) = g(x - \mu_l)$  для деякої загальної щільності  $g$  з нульовим середнім значенням та параметром локалізації  $\mu_l$ . Припустимо також, що  $f_l$  та  $K$  є строго спадними функціями відстані від їх центрів симетрії. Тоді, для певного позитивного  $S$ , при  $n \rightarrow \infty$ , усереднена ймовірність помилкової класифікації класифікатора на основі ядерної оцінки щільності розподілу має тенденцію байесівського ризику за умови, що апіорні ймовірності рівні.

Дана теорема пояснює причину нелогічної поведінки  $\Xi(S)$ . Наступна теорема дає певне уявлення щодо поведінки класифікаторів на основі ядерної оцінки щільності розподілу для великих розмірів вибірок та великих смуг пропускання, коли щільності комплексів не обов'язково задовольняють умову симетричності.

**Теорема 2.** Припустимо, що  $f_l$  є функціями щільності, що задовольняють умову  $\int \|x\|^6 f_l(x) dx < \infty$  для всіх  $l = 1, 2, \dots, L$ , а ядро  $K$  є щільністю з режимом в 0 та обмежено третіми похідними. Тоді, якщо апіорні ймовірності є рівними при  $n, S \rightarrow \infty$ , усереднена ймовірність помилкової класифікації класифікатора на основі ядерної оцінки щільності розподілу має тенденцію лінійного класифікатора, що задається таким чином:

$$d_L(x) = \arg \min_l \left[ x' \nabla^2 K(0) E_{f_l}(X) - (1/2) E_{f_l} \{ X' \nabla^2 K(0) X \} \right].$$

Значимо, що коли ядро  $K$  є сферично-симетричною та строго спадною функцією норми її аргументу, тоді обмежений лінійний класифікатор, що отриманий в попередній теоремі для великої смуги пропускання є майже еквівалентним класифікатору, що класифікує спостереження  $x$  в клас  $l_0$ , що максимізує  $x' E_{f_l}(X) - (1/2) E_{f_l}(XX')$  (оскільки  $\nabla^2 K(0)$  є негативно визначеною функцією), або мінімізує  $E_{f_l}(\|x - X\|^2)$  для  $1 \leq l \leq L$ . Поведінка середньої ймовірності помилкової класифікації виявляється дещо іншою, коли апіорні ймовірності є різними для різних комплексів. Як приклад, розглядаються такі ж розподіли, що і на рис. 1-4 та рис. 5-8, однак встановлено апіорні ймовірності 0.6 та 0.4, відповідно для двох комплексів. Отримані результати узагальнені на рис. 9-10.

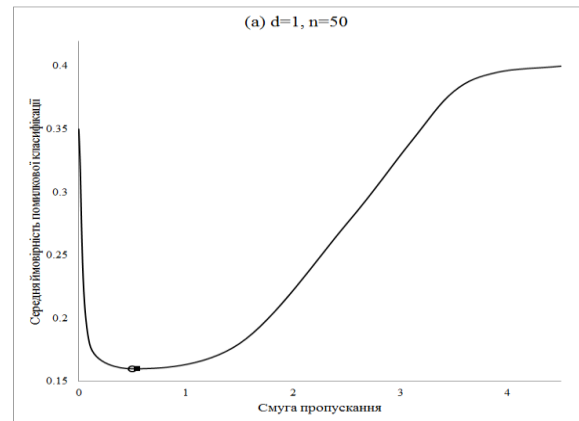


Рис. 9 Істинні  $\Xi$ -функції та оптимальні смуги пропускання (1 експеримент)

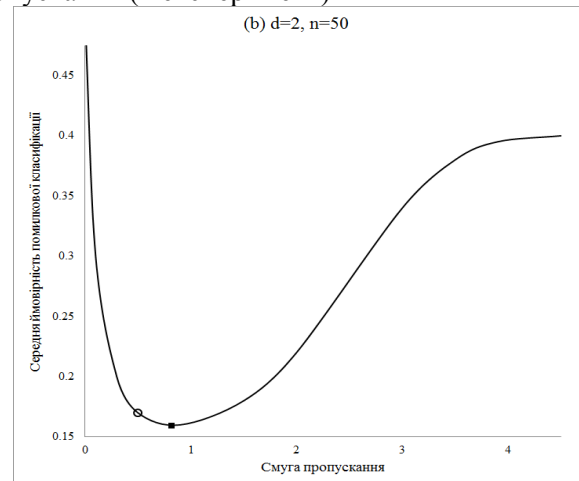


Рис. 10 Істинні  $\Xi$ -функції та оптимальні смуги пропускання (2 експеримент)

На попередньо представлених рисунках продемонстровано випадки нерівних апостеріорних ймовірностей.

У деяких випадках смуга пропускання, що мінімізує  $\Xi(S)$ , позначена '\*' та смуга пропускання, що мінімізує середньо-квадратичні інтегровані помилки для ядерних оцінок щільності розподілу, позначена 'o' виявляються різними. Однак важливим є те, що  $\Xi(S)$  тепер демонструє абсолютно іншу поведінку при зміні  $S$ . Після досягнення свого мінімального значення  $\Xi(S)$  значно збільшується, перш ніж стати гладкою. У даному випадку великі смуги пропускання не є належним вибором для задач класифікації.

### Висновки

Підводячи підсумки викладеного матеріалу зазначимо, що було досліджено деякі фундаментальні відмінності між оптимальною смугою пропускання, що мінімізує середньо-квадратичні інтегровані помилки ядерної оцінки щільності розподілу та оптимальною смугою

пропускання, що мінімізує середню ймовірність помилкової класифікації, коли ядерні оцінки щільності розподілу використовуються для дискримінантного аналізу. Один з основних результатів полягає в тому, що залежно від того, чи апріорні ймовірності для різних комплексів є рівними чи ні, усереднена ймовірність помилкової класифікації демонструє абсолютно різну поведінку для різних варіантів вибору параметра смуги пропускання. В той час як існує загальне переконання, що із зростанням об'єму вибірки слід поступово зменшувати смугу пропускання до нуля для досягнення більшої продуктивності ядерної оцінки щільності розподілу, було показано, що коли усі апріорні ймовірності є рівними, залежно від природи функцій щільності комплексів та ядра, великі смуги пропускання можуть призводити до майже оптимальних коефіцієнтів помилкової класифікації.

### Список використаних джерел

1. Scott D. Multivariate Density Estimation: Theory, Practice and Visualization / D. Scott. – Wiley, New York.
2. Wand M. Kernel Smoothing / M. Wand, M Jones. – London: Chapman and Hall, 1995.
3. Ripley B. Pattern Recognition and Neural Networks / B. Ripley. – Cambridge: Cambridge University Press, 1996.
4. Duda R. Pattern Classification / R. Duda, P. Hart, D. Stork. – New York: Wiley, 2000.
5. Lachenbruch P. Estimation of error rates in discriminant analysis / P. Lachenbruch, M. Mickey. – Technometrics. –1968. – 10. – P. 1-11.
6. Chaudhuri P. SiZer for exploration of structures in curves / P. Chaudhuri, J. Marron // Journal of the American Statistical Association. – 1999. – 94. – P. 807-823.
7. Chaudhuri P. Scale space view of curve estimation / P. Chaudhuri, J. Marron // The Annals of Statistics. – 2000. – 28. – P.408-428.

### References

1. SCOTT, D. (1992) *Multivariate Density Estimation: Theory, Practice and Visualization*. Wiley, New York.
2. WAND, M., JONES, M. (1995) *Kernel Smoothing*. London: Chapman and Hall.
3. RIPLEY, B. (1996) *Pattern Recognition and Neural Networks*. Cambridge: Cambridge University Press.
4. DUDA, R., HART, P. and STORK, D. (2000) *Pattern Classification*. New York: Wiley.
5. LACHENBRUCH, P. & MICKEY, M. (1968) Estimation of error rates in discriminant analysis. *Technometrics*. 10. p.1-11.
6. CHAUDHURI, P. & MARRON, J. (1999) SiZer for exploration of structures in curves. *Journal of the American Statistical Association*. 94. p.807-823.
7. CHAUDHURI, P. & MARRON, J. (2000) Scale space view of curve estimation. *The Annals of Statistics*. 28. p.408-428.

Надійшла до редколегії 29.09.14