УДК 004.93

М. І. КРИВОШЕЯ

ДОСЛІДЖЕННЯ ФЕНОМЕНУ ПОДВІЙНОГО СПУСКУ ТА ПОРІВНЯННЯ МІНІМАКСНОЇ АПРОКСИМАЦІЇ З L2-РЕГУЛЯРИЗАЦІЄЮ

Вінницький національний технічний університет,
21021, вул. Хмельницьке шосе, 95, м. Вінниця, Україна

Анотація. У цій роботі досліджено феномен подвійного спуску та запропоновано використання мінімаксної апроксимації (L∞-норма) як альтернативу L2-регуляризації для покращення якості апроксимації моделей. Подвійний спуск описує залежність похибки від складності моделі: похибка спершу зменшується, потім зростає через перенавчання, а далі знову знижується. Проте в експериментах із моделлю без регуляризації було виявлено переважно зростаючу тенденцію похибки із короткими періодами спаду, що свідчить про неповний прояв феномену. Це, ймовірно, пов'язано з аномальними точками в даних, які спричинили експоненційне зростання похибки на високих ступенях. Було розглянуто три підходи: класичну модель без регуляризації, модель із L2-регуляризацією та мінімаксну апроксимацію. L2-регуляризація додала штраф за велику норму коефіцієнтів, що дозволило стабілізувати похибку та запобігти перенавчанню, особливо на високих ступенях полінома (200+). Мінімаксна апроксимація мінімізувала максимальну похибку, завдяки чому забезпечувала кращу стійкість до аномалій і перевершувала L2-регуляризацію на низьких ступенях (до 50). Результати підтвердили, що мінімаксна апроксимація є більш ефективною для задач із аномаліями, тоді як L2-регуляризація краще працює на складних моделях із високими ступенями полінома. Отримані висновки сприяють розширенню розуміння феномену подвійного спуску й показують практичну користь різних підходів у залежності від особливостей даних і вимог до моделі.

Ключові слова: подвійнний спуск, L2-регуляризація, мінімаксна апроксимація, поліноміальні моделі, машинне навчання, аномалії.

Abstract. This paper investigates the phenomenon of double descent and proposes the use of minimax approximation (L∞-norm) as an alternative to L2-regularization to improve the quality of model approximation. Double descent describes the dependence of the error on the complexity of the model: the error first decreases, then increases due to overfitting, and then decreases again. In contrast, in experiments with a model without regularization, a predominantly increasing trend of the error with short periods of decline was found, which is observed for an incomplete manifestation of the phenomenon. This is probably due to anomalous points in the data that caused an exponential increase in the error at high powers. Three approaches were noted: a classical model without regularization, a model with L2-regularization, and minimal approximation. L2 regularization added a penalty for large coefficient norms, which stabilized the error and prevented overfitting, especially at high polynomial degrees (200+). Minimax approximation minimized the error, thereby providing better maximum anomaly robustness and outperforming L2 regularization at low degrees (up to 50). The results confirmed that minimax approximation is more effective for problems with anomalies, while L2 regularization performs better on complex models with high polynomial degrees. The findings contribute to the understanding of the double descent phenomenon and show the practicality of applying different approaches due to data features and model requirements.

Keywords: double descent, L2 regularization, minimax approximation, polynomial models, machine learning, anomalies.

DOI: 10.31649/1681-7893-2025-49-1-36-43

ВСТУП

У сучасному машинному навчанні прогнозування складних залежностей у даних є важливим завданням, яке дозволяє розширити можливості моделей у різних галузях, включаючи економіку, медицину та інженерію. Одним із викликів є подолання проблем, пов'язаних із перенавчанням, шумом у даних та аномаліями. Одним із феноменів у цьому контексті є феномен подвійного спуску. Цей ефект демонструє два мінімуми похибки, між якими спостерігається різке зростання через перенавчання, що викликане надмірною складністю моделі.[4] У цій залежності спостерігається зменшення похибки на початкових етапах, її різке зростання через перенавчання, а також подальше зниження при збільшенні складності моделі. Феномен подвійного спуску має практичне значення у багатьох сферах.

У фінансовому секторі складні моделі прогнозування ризиків використовуються для виявлення фінансових шахрайств, але перенавчання може призводити до неправильної класифікації транзакцій.

У медицині при аналізі медичних зображень (наприклад, МРТ чи КТ) використання складних моделей глибокого навчання іноді викликає перенавчання, що знижує точність діагностики. Робота з шумовими даними та аномаліями є критичною для підвищення ефективності систем підтримки прийняття рішень.

У сфері самокерованих автомобілів складні алгоритми обробки даних сенсорів (камер, лідарів) стикаються з викликами шуму та аномальних даних, що може впливати на здатність правильно реагувати на нестандартні ситуації.

У реальних умовах дана теоретична поведінка часто змінюється через наявність шуму та аномалій у вибірках, що призводить до нестабільності моделей і зростання похибок. Класичні методи, такі як L2-регуляризація (Ridge), допомагають зменшити перенавчання, але виявляються менш ефективними для задач, де дані містять аномальні точки. Натомість мінімаксна апроксимація, яка базується на мінімізації максимальної похибки (L∞-норма), пропонує ефективніший підхід до роботи з даними, забрудненими аномаліями.[2]

Актуальність дослідження полягає у порівнянні двох підходів: L2-регуляризації та мінімаксної апроксимації, у контексті задач регресії, що містять аномальні точки і шум. Проведені експерименти дозволяють оцінити ефективність кожного методу на різних рівнях аномалій (10%, 30%, 50%, 70%, 90%) та шуму ($\sigma = 0.1, 0.3, 0.5, 0.7, 0.9$).

Метою роботи є дослідження феномену подвійного спуску в реальних умовах із аномаліями та шумом, а також розробка рекомендацій щодо вибору оптимального підходу для регуляризації. Результати дослідження спрямовані на підвищення стійкості та точності моделей, що є критично важливим у задачах машинного навчання та інженерії.

РЕЗУЛЬТАТИ ДОСЛІДЖЕННЯ

Для проведення експерименту було створено навчальну вибірку з 70 точок, цільова залежність моделювала синусоїду з додаванням випадкового шуму. Щоб змоделювати реальні умови, у вибірку додали аномальні точки, які становили 40% від загальної кількості. Значення аномалій створювалися шляхом додавання випадкових зсувів із діапазону $[-15,-5]\cup[5,15]$ до координат y, що дозволило зімітувати ситуації, в яких дані містять помітні викиди. Такий підхід дав змогу оцінити стійкість різних методів регресії до впливу аномальних точок. [3]

Дослідження виконано на трьох моделях:

Модель без регуляризації. Розрахунок коефіцієнтів здійснювався шляхом мінімізації середньоквадратичної похибки (MSE):

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2, \tag{1}$$

де y_i – реальне значення, $\hat{y}_i$ – прогноз моделі

Алгоритм виконання для моделі без регуляризації:

1. Згенерувати навчальну вибірку з цільовою залежністю, що моделює синусоїду з шумом.
2. Використати метод псевдообернення матриці для обчислення вагових коефіцієнтів:

$$w = (X^T X)^{-1} X^T y$$

де X – матриця ознак.

3. Застосувати отримані коефіцієнти w до тестових даних для прогнозування $\hat{y}_i$
4. Обчислити середньоквадратичну похибку між прогнозами та реальними значеннями тестової вибірки.

Модель із L2-регуляризацією (Ridge). Ця модель додає штраф за велику норму коефіцієнтів:

$$\min_w \|Xw - y\|_2^2 + \alpha \|w\|_2^2, \tag{2}$$

де X - матриця ознак, w - вектор ваг моделі, y - вектор цільових значень, $\|Xw - y\|_2^2$ – середньоквадратична похибка (MSE), $\|w\|_2^2$ - штраф за норму ваг моделі, α - коефіцієнт регуляризації. [5], [9]

Алгоритм виконання для L2-регуляризації:

1. Застосувати поліноміальні ознаки до навчальних і тестових даних для створення матриць X_{train} *і* X_{test}
2. Виконати стандартизацію даних для покращення стійкості моделі до масштабів.
3. Навчити модель Ridge-регресії, використовуючи регуляризацію з параметром α, що контролює величину штрафу.
4. Отримати прогнози на тестовій вибірці та обчислити середньоквадратичну похибку.

Мінімаксна апроксимація (L∞-норма). Ця модель мінімізує максимальну похибку:

$$\min_w \max_i |(Xw - y)_i|, \tag{3}$$

де $\max_i |(Xw - y)_i|$ – максимальна похибка серед усіх точок. [6]

Алгоритм виконання для мінімаксної апроксимації:

1. Використати бібліотеку CVXPY для формулювання задачі опуклої оптимізації:
   - Визначити змінну w як вектор ваг.
   - Задати функцію цілі: мінімізувати максимальну абсолютну похибку.
2. Розв’язати задачу оптимізації за допомогою ітеративного методу, обираючи точність та кількість ітерацій.
3. Застосувати знайдені коефіцієнти w для прогнозування значень тестової вибірки.
4. Обчислити середньоквадратичну похибку для оцінки якості моделі.

Для порівняння даних моделей було здійснено тестування. Для тестування було розроблено та імплементовано код за допомогою мови програмування Python [7] із використанням нижче зазначених бібліотек:

- NumPy для роботи з масивами та математичними операціями.
- Matplotlib для побудови графіків і візуалізації результатів.
- scikit-learn для реалізації моделей поліноміальної регресії та L2-регуляризації (Ridge).
- CVXPY для розв'язання оптимізаційної задачі мінімаксної апроксимації.

АЛГОРИТМ ПРОГРАМИ

1. Генерація даних

Алгоритм створює навчальну та тестову вибірки для моделювання залежності подвійного спуску.

1) Використовуються випадкові значення $x \in [0, 1]$, для яких цільова залежність задається як $y = \sin(2\pi x) + \varepsilon$, де ε — нормальний шум із нульовим середнім і дисперсією σ^2.
2) У навчальну вибірку додаються аномалії, які становлять певний відсоток від загальної кількості даних. Аномальні точки формуються шляхом додавання випадкових зсувів із діапазону $[-15, -5] \cup [5, 15]$ до значень y.
3) Тестова вибірка створюється як рівномірний розподіл точок у діапазоні $x \in [0, 1]$ без додавання шуму.

2. Навчання моделей

Кожна модель тренується на поліноміальних ознаках різного ступеня (від 1 до заданого максимуму).

1) Поліноміальні ознаки: Для кожного ступеня полінома генерується матриця ознак X_{poly} за допомогою PolynomialFeatures.
2) Масштабування: Дані стандартизуються для забезпечення стабільності моделі під час навчання.
3) Методи регресії:
   - Без регуляризації: Використовується псевдообернення матриці для знаходження ваг w.
   - L2-регуляризація: Використовується Ridge, який додає штраф за велику норму ваг.
   - Мінімаксна апроксимація: Розв’язується оптимізаційна задача з мінімізацією максимальної абсолютної похибки, формулюється у вигляді опуклого завдання через cvxpy.

3. Оцінка моделей

1) Для кожного ступеня полінома розраховується середньоквадратична похибка (MSE) на тестовій вибірці.
2) Результати зберігаються для подальшого аналізу.

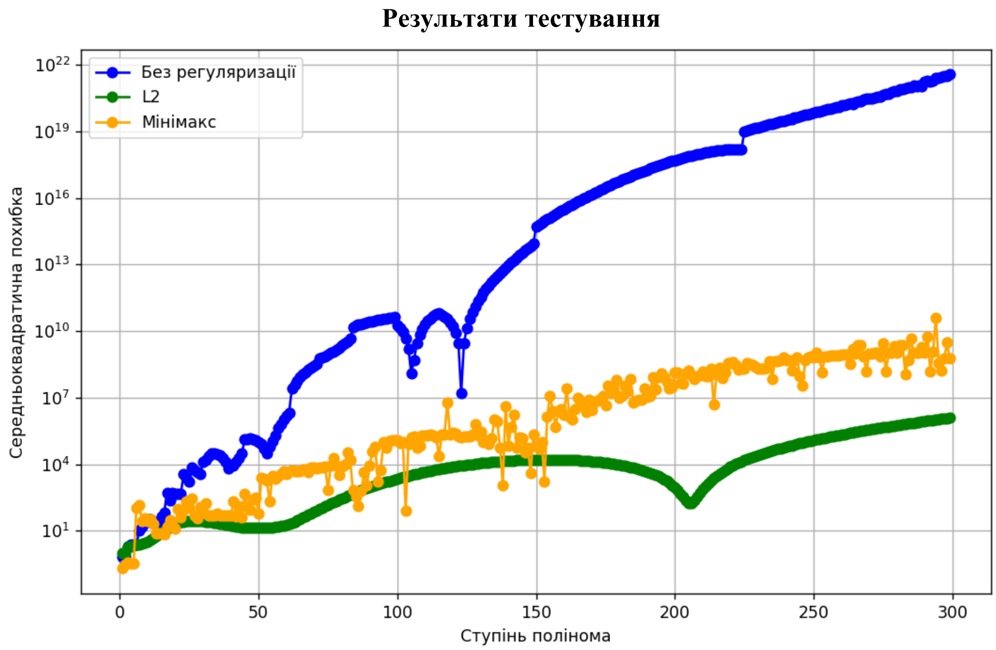

Рисунок 1 – Порівняння без регуляризації, L2 і мінімаксної апроксимації

Для моделі без регуляризації (синя лінія на графіку рис.1) спостерігається переважно зростаюча похибка із короткими локальними спадами на певних діапазонах ступенів. Це свідчить про те, що модель не досягає стадії, де похибка суттєво знижується на високих ступенях, як це характерно для подвійного спуску. Отже феномен подвійного спуску не проявляється повною мірою, як це описано в теоретичній моделі. Замість цього відбувається домінантне експоненційне зростання похибки. Така поведінка може бути пов'язана з тим, що модель без регуляризації чутливо реагує на аномалії та шум у даних, що посилюється зі збільшенням складності.

L2-регуляризація (зелена лінія на рис.1) демонструє стабільну похибку, яка поступово збільшується з ростом складності моделі. Вона не має експоненційного зростання, як у моделі без регуляризації, і показує найкращі результати на високих ступенях (200+), що відповідає її здатності справлятися з перенавчанням.

Мінімаксна апроксимація (помаранчева лінія на рис. 1) має схожу поведінку із L2-регуляризацією, але демонструє дещо гірші результати на високих ступенях. Проте, на низьких ступенях (до 50) вона є кращою за L2-регуляризацію завдяки своїй стійкості до аномальних точок.

Дані щодо зменшення середньоквадратичної похибки для різних підходів (L2-регуляризація та мінімаксна апроксимація) були проаналізовані в таблиці (Додаток 1).

З огляду на тестування, результати якого відображені на рис. 1, можна дійти висновків, що на низьких ступенях полінома (1–10) мінімаксна апроксимація демонструє перевагу, забезпечуючи суттєве зниження середньоквадратичної похибки. На середніх ступенях полінома (10–50) L2-регуляризація починає працювати стабільніше та ефективніше. Обидва методи регуляризації (L2 та мінімакс) забезпечують майже однакове зменшення похибки, проте L2-регуляризація демонструє трохи вищу стабільність на дуже високих ступенях.

Для подальшого дослідження ефективності методів регуляризації було проведено додаткові тестування із різними рівнями аномалій та шуму.

1. *Порівняльний аналіз для різних рівнів аномалій*

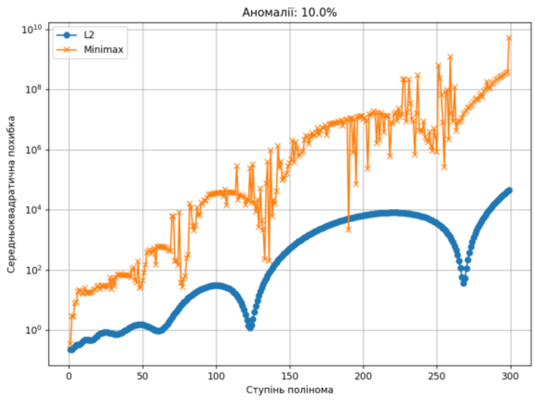

Рисунок 2

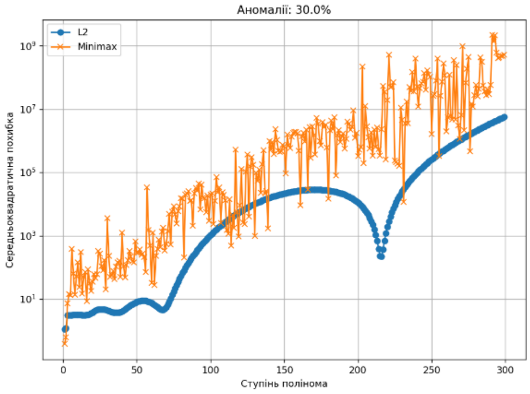

Рисунок 3

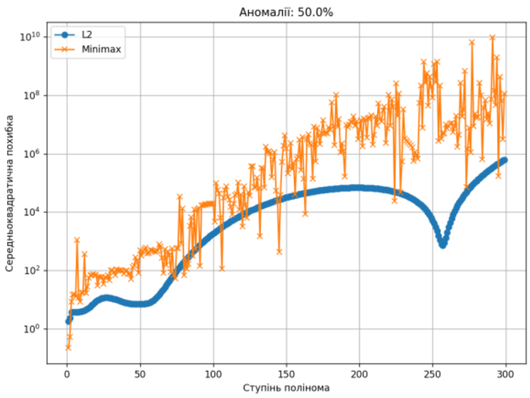

Рисунок 4

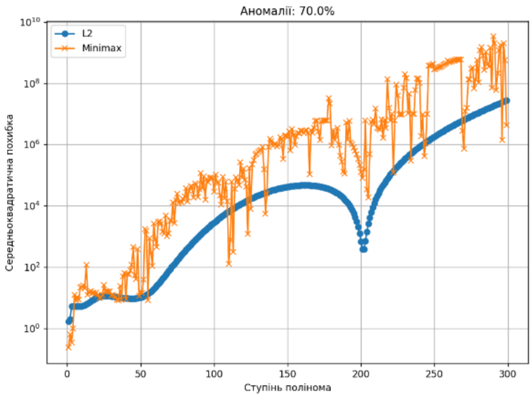

Рисунок 5

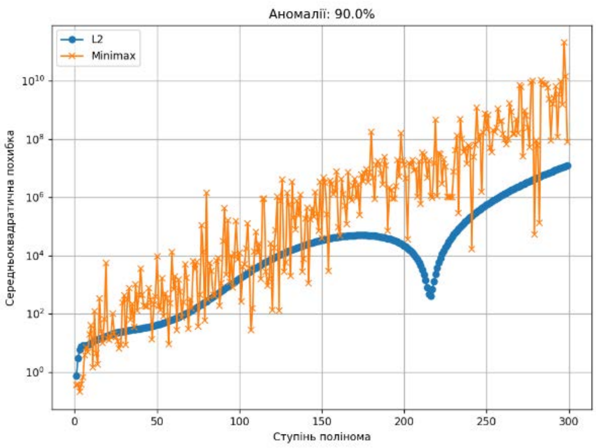

Рисунок 6 - Результати для частки аномалій. Порівняльний аналіз для різних рівнів аномалій

Було побудовано п’ять графіків (рис. 2-6), які демонструють результати для частки аномалій 10%, 30%, 50%, 70% та 90% відповідно. Це дало змогу проаналізувати, як збільшення кількості аномальних точок впливає на точність методів L2-регуляризації та мінімаксної апроксимації.

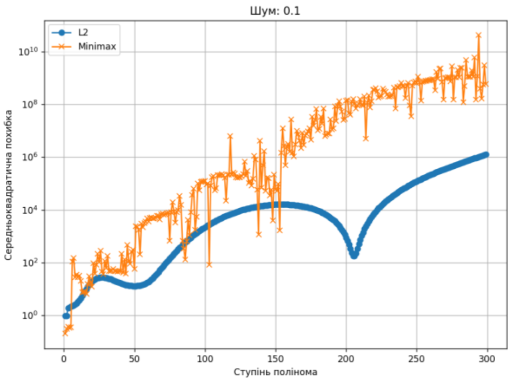

Рисунок 7

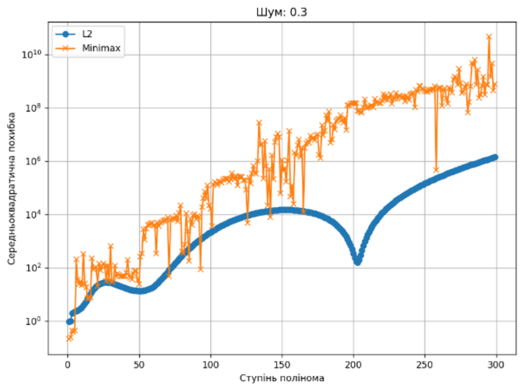

Рисунок 8

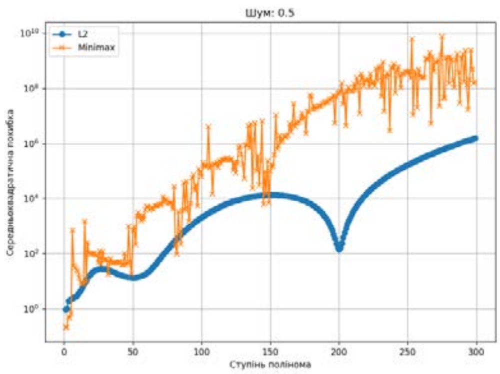

Рисунок 9

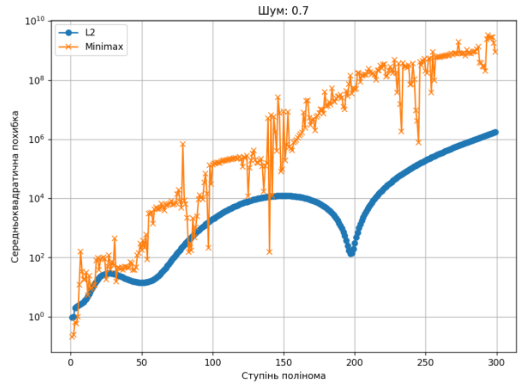

Рисунок 10

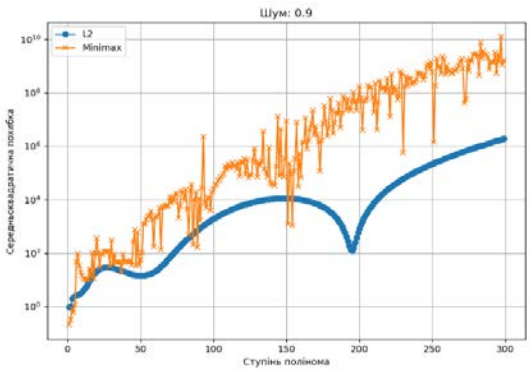

Рисунок 11- Результати впливу рівня шуму

На графіках видно, що мінімаксна апроксимація (помаранчева лінія) втрачає свою ефективність зі збільшенням рівня аномалій. При 10% аномалій вона демонструє низькі похибки на малих ступенях полінома, значно перевершуючи L2-регуляризацію. Зі збільшенням частки аномалій (30–50%) мінімаксна апроксимація поступово втрачає свою ефективність. Похибка на високих ступенях починає зростати, а результати стають менш стабільними. L2-регуляризація, натомість, демонструє стабільну похибку навіть за наявності 50% аномальних точок. На високому рівні аномалій (70–90%) мінімаксна апроксимація повністю втрачає свою ефективність на високих ступенях. Похибка росте експоненційно і стає нестабільною. L2-регуляризація зберігає стійкість до аномалій, демонструючи плавне зростання похибки зі збільшенням складності моделі. L2-регуляризація (синя лінія), навпаки, демонструє більш стійкі результати при будь-якому рівні аномалій. Її похибка повільно зростає зі збільшенням складності моделі, але вона залишається стабільною і не проявляє різких коливань, як у випадку мінімаксної апроксимації.

1. *Порівняльний аналіз для різних рівнів зашумленості*

Для подальшого аналізу було зафіксовано частку аномалій на рівні 40%, де мінімаксна апроксимація показала помірну ефективність на попередніх етапах дослідження. Щоб дослідити вплив рівня шуму на точність обох методів, було побудовано п’ять графіків (Рис. 7–11) для значень шуму 0.1, 0.3, 0.5, 0.7 та 0.9. Це дозволило оцінити стійкість методів до додаткової варіативності даних.

Проведений аналіз показав, що L2-регуляризація є значно стійкішою до підвищеного рівня шуму у даних. Її похибка зростає поступово зі збільшенням шуму, але залишається стабільною навіть на високих ступенях полінома. Мінімаксна апроксимація, навпаки, демонструє високу чутливість до шуму: похибка швидко зростає, особливо на середніх і високих ступенях, і стає нестабільною з підвищенням варіативності даних. Таким чином, для задач із високим рівнем шуму L2-регуляризація є кращим вибором завдяки своїй здатності стабілізувати модель і контролювати перенавчання. Мінімаксна апроксимація може бути ефективною лише при низькому рівні шуму та на малих ступенях полінома.

ВИСНОВОК

Проведене мною дослідження продемонструвало вплив аномалій і шуму на феномен подвійного спуску та порівняло ефективність L2-регуляризації й мінімаксної апроксимації. Мінімаксна апроксимація показала високу ефективність на низьких та середніх ступенях полінома, особливо при невеликій кількості аномалій у даних. Її стійкість до аномальних точок дозволяє суттєво знижувати похибку на цих етапах, проте зі зростанням рівня аномалій її стабільність знижується, що призводить до різких коливань і експоненційного зростання похибки на високих ступенях.

Натомість L2-регуляризація продемонструвала стійку поведінку за різних умов, зокрема при високій кількості аномалій та шуму. Вона ефективно контролює перенавчання та забезпечує стабільне зростання похибки без різких коливань, що робить її кращим вибором для моделей високої складності.

Збільшення рівня шуму у даних показало, що обидва методи втрачають точність, але L2-регуляризація залишається стійкішою, у той час як мінімаксна апроксимація демонструє значні коливання.

Отже, отримані результати підтверджують, що ефективність методів залежить від характеру даних. Мінімаксна апроксимація є доцільною для задач із низьким рівнем аномалій та шуму, тоді як L2-регуляризація є більш ефективною для задач зі складними моделями, високим рівнем аномалій і шуму. Це підкреслює важливість адаптивного вибору методів регуляризації залежно від специфіки даних і завдання, що є ключовим для побудови стійких і точних моделей машинного навчання.

СПИСОК ВИКОРИСТАНОЇ ЛІТЕРАТУРИ

1. Головач, А. Б. Математичні основи моделювання складних систем. Київ: Наукова думка, 2018. – с. 286
2. Н.В. Бурєннікова, О.В. Зелінська, І.М. Ушкаленко, Ю.Ю. Бурєнніков "Оптимізаційні методи та моделі" – Вінниця: ВНТУ, 2019. – 121 с.
3. Огірко О. І., Галайко Н. В. "Теорія ймовірностей та математична статистика: навчальний посібник". – Львів: Львівський державний університет внутрішніх справ, 2017. – 292 с.
4. Belkin, M., Hsu, D., Ma, S., & Mandal, S. Reconciling Modern Machine Learning and the Bias-Variance Trade-off. PNAS, 2019. – с. 23
5. Ng, A. Y. Feature Selection, L1 vs. L2 Regularization, and Rotational Invariance. Proceedings of the 21st International Conference on Machine Learning, 2004. – с. 8
6. Boyd, S., & Vandenberghe, L. Convex Optimization. Cambridge University Press, 2004. – с. 732
7. Chen, T., & Guestrin, C. XGBoost: A Scalable Tree Boosting System. KDD, 2016. – с. 785–790
8. Bishop, C. M. Pattern Recognition and Machine Learning. Springer, 2006. – с. 738

REFERENCES

1. Golovach, A. B. Mathematical foundations of modeling complex systems. Kyiv: Naukova Dumka, 2018. – p. 286
2. N.V. Burennikova, O.V. Zelinska, I.M. Ushkalenko, Y.Yu. Burennikov "Optimization methods and models" – Vinnytsia: VNTU, 2019. – 121 p.

3. Ogirko O. I., Galayko N. V. "Probability theory and mathematical statistics: a textbook". – Lviv: Lviv State University of Internal Affairs, 2017. – 292 p.
4. Belkin, M., Hsu, D., Ma, S., & Mandal, S. Reconciling Modern Machine Learning and the Bias-Variance Trade-off. PNAS, 2019. – p. 23
5. Ng, A. Y. Feature Selection, L1 vs. L2 Regularization, and Rotational Invariance. Proceedings of the 21st International Conference on Machine Learning, 2004. - p. 8
6. Boyd, S., & Vandenberghe, L. Convex Optimization. Cambridge University Press, 2004. - p. 732
7. Chen, T., & Guestrin, C. XGBoost: A Scalable Tree Boosting System. KDD, 2016. - p. 785–790
8. Bishop, C. M. Pattern Recognition and Machine Learning. Springer, 2006. - p. 738

Надійшла до редакції 17.05.25 р.

КРИВОШЕЯ МИХАЙЛО ІГОРОВИЧ — аспірант кафедри автоматизації та інтелектуальних інформаційних технологій, факультет інтелектуальних інформаційних технологій та автоматизації, Вінницький національний технічний університет, Вінниця. ***e-mail:*** ***mishakryvoshea@gmail.com***

MYKHAILO KRYVOSHEIA

STUDY OF THE DOUBLE TRIGGER PHENOMENON AND COMPARISON OF MINIMAX APPROXIMATION WITH L2-REGULARIZATION

Vinnytsia National Technical University, Vinnytsia