

DOI: <https://doi.org/10.36910/6775-2524-0560-2021-42-24>

УДК 004.021

<sup>1</sup>Марценюк Василь Петрович, д.т.н., професор<https://orcid.org/0000-0001-5622-1038><sup>2</sup>Андрушак Ігор Євгенович, д.т.н., професор<https://orcid.org/0000-0002-8751-4420><sup>3</sup>Мілян Назар Васильович, аспірант<https://orcid.org/0000-0003-0825-1384><sup>1</sup>Університет Бельсько-Бяли, Польща<sup>2</sup>Луцький національний технічний університет<sup>3</sup>Тернопільський національний технічний університет імені Івана Пулюя

## МІНІМАКСНИЙ ПІДХІД В МАШИННОМУ НАВЧАННІ: СУЧАСНИЙ СТАН ТА ПЕРСПЕКТИВИ РОЗВИТКУ

Марценюк В.П., Андрушак І.С., Мілян Н.В. **Мінімаксний підхід в машинному навчанні: сучасний стан та перспективи розвитку.** В статті представлено значення машинного навчання у сучасному світі. Звернуто особливу увагу на використання алгоритмів машинного навчання в медицині, зокрема використання різноманітних моделей, починаючи від регресії, SVM, випадкових лісів для контрольованого навчання та PCA для неконтрольованого. Підкреслюються основні невизначеності та завдання машинного навчання, що виникають у основних медичних додатках (діагностика, лікування та профілактика). Математично описано проблеми машинного навчання в медичних дослідженнях. Оптимізація є важливою частиною машинного навчання. Основна увага приділена мінімаксному підходу у машинному навчанні. Розглянуто ряд мінімаксних підходів таких, як: Minimax Probability Machine (MPM), Generalized Hidden-Mapping Minimax Probability Machine (GHM-MPM), Minimum Error Minimax Probability Machine (MEMPM), парна мінімаксна ймовірність екстремального нахилу машини (TMPELM), машина подвійної мінімаксної ймовірності (TWMPM) та деякі інші.

**Ключові слова:** машинне навчання, оптимізація, мінімаксний підхід, Minimax Probability Machine.

Марценюк В.П., Андрушак І.С., Мілян Н.В. **Минимаксных подход в машинном обучении: современное состояние и перспективы развития.** В статье представлены значения машинного обучения в современном мире. Обращено особое внимание на использование алгоритмов машинного обучения в медицине, в частности использование различных моделей, начиная от регрессии, SVM, случайных лесов для контролируемого обучения и PCA для неконтролируемого. Подчеркиваются основные неопределенности и задачи машинного обучения, возникающие в основных медицинских приложениях (диагностика, лечение и профилактика). Математически описано проблемы машинного обучения в медицинских исследованиях. Оптимизация является важной частью машинного обучения. Основное внимание уделено подходу минимакс в машинном обучении. Рассмотрен ряд минимаксных подходов таких, как: Minimax Probability Machine (MPM), Generalized Hidden-Mapping Minimax Probability Machine (GHM-MPM), Minimum Error Minimax Probability Machine (MEMPM), парная минимаксная вероятность экстремального наклона машины (TMPELM), машина двойной минимаксной вероятности (TWMPM) и некоторые другие.

**Ключевые слова:** машинное обучение, оптимизация, минимаксных подход, Minimax Probability Machine.

Martsenyuk V.P., Andrushchak I.Ye., Milian N.V. **Features of information technology distribution of radio waves by frequency bands.** The article presents the importance of machine learning in the modern world. Particular attention is paid to the use of machine learning algorithms in medicine, in particular the use of various models, such as regression, SVM, random forests for controlled learning and PCA for uncontrolled. The main uncertainties and tasks of machine learning that arise in the main medical applications (diagnosis, treatment and prevention) are emphasized. The problems of machine learning in medical research are described mathematically. Optimization is an important part of machine learning. The main attention is paid to the minimax approach in machine learning. A number of minimax approaches are considered, such as: Minimax Probability Machine (MPM), Generalized Hidden-Mapping Minimax Probability Machine (GHM-MPM), Minimum Error Minimax Probability Machine (MEMPM), pair minimax probability of extreme tilt machine (TMPELM), double minima machine probabilities (TWMPM) and some others.

**Keywords:** machine learning, optimization, minimax approach, Minimax Probability Machine.

**Постановка проблеми.** Разом зі стрімким накопиченням інформації швидкими темпами розвиваються і технології аналізу даних. Якщо ще кілька років тому, наприклад, було можливо лише сегментувати клієнтів на групи зі схожими перевагами, то тепер можливо будувати моделі для кожного клієнта в режимі реального часу, аналізуючи, наприклад, його переміщення по мережі Інтернет для пошуку конкретного товару. Інтереси споживача можуть бути проаналізовані, і відповідно до побудованої моделі виведена відповідна реклама або конкретні пропозиції. Модель також може налаштовуватися і перебудовуватися в режимі реального часу.

Із зростанням обсягу даних відбувся швидкий розвиток машинного навчання (ML), який зробив багато теоретичних проривів та застосовується у різних галузях. Оптимізація є важливою частиною ML.

Суть більшості алгоритмів ML полягає у побудові моделі оптимізації та вивченні параметрів у цільовій функції з поданих даних. В епоху великих даних ефективність та дієвість алгоритмів чисельної оптимізації різко впливає на популяризацію та застосування моделей ML. З метою сприяння розвитку ML було висунуто низку ефективних методів оптимізації, які покращили продуктивність та ефективність методів ML.

**Аналіз останніх досліджень та публікацій.** ML успішно застосовується в медицині і може вирішити широкий спектр проблем. В даний час за алгоритмами ML аналізуються не тільки дані об'єктивного обстеження та спостережень пацієнта, але й результати аналізів та обстежень на медичному обладнанні, таких як ЕКГ, ЕЕГ, КТ, МРТ та ін. Використовуються різноманітні моделі, починаючи від регресії, SVM, випадкових лісів для контрольованого навчання та PCA для неконтрольованого [2,17,18]. Поглиблене навчання дозволяє розробити розумну систему моніторингу охорони здоров'я [4], дозволяє використовувати носимі датчики та дані соціальних мереж [1], підтримує обробку природних мов [3]. Застосування таких засобів підвищує ефективність лікаря, позбавляючи його від виконання ряду рутинних операцій. Однією із суттєвих проблем використання ML у медицині є підготовка правильних різноманітних медичних даних з різних джерел для навчальних алгоритмів, що враховують невизначеності.

Невизначеність залишається важливим фактором, що впливає на процес прийняття рішень у всіх галузях; це суттєво впливає на точність, чутливість та специфіку застосовуваних алгоритмів. Багато застосувань машинного навчання (ML) залежать від хорошої оцінки невизначеності [5]: прогнозування [6]; прийняття рішень [7]; навчання з обмежених, зашумлених та відсутніх даних [8]; вивчення складних персоніфікованих моделей [9]; стиснення даних [10]; автоматичне наукове моделювання [11] та проектування експерименту [12].

Зокрема, підкреслюються основні невизначеності та завдання машинного навчання, що виникають у основних медичних додатках (діагностика, лікування та профілактика). Моделі, що описують хвороби та патологічні процеси, слід враховувати, щоб справлятися з невизначеностями [13]. Ось чому доводиться аналізувати різні типи завдань машинного навчання в медичних дослідженнях (як з вчителем, так і без вчителя) також з точки зору впливу невизначеності.

Отже, більшість даних, які використовуються в ML для медичних досліджень, надходять з різних сенсорних пристроїв. На основі отриманих спостережень є прагнення зробити певні висновки або прийняти рішення, передбачити певні показники. Очевидно, що в одних і тих же спостереженнях потрібно приймати однакові рішення або передбачати однакові значення. З іншого боку, вони використовують різні типи вхідних даних з різних причин (наприклад, через зменшення розмірів даних). Крім того, для оцінки стану пацієнта можуть використовуватися різні діагностичні пристрої. Такі типи невизначеностей називаються алеаторними та пов'язані з невизначеностями даних.

Слід зазначити, що вибір моделі машинного навчання також впливає на прийняті рішення та прогнозовані медичні показники. Це пов'язано з поняттями чутливості та специфічності, так само, як і при медичній диференціальній діагностиці, порівнюючи певні методи медичного дослідження. Такий тип невизначеності, пов'язаний із моделлю, відомий як епістемічний. Водночас ці два види невизначеності не повинні впливати на прийняті рішення чи прогнозовані значення.

**Виклад основного матеріалу дослідження.** Математичний опис контрольованої проблеми машинного навчання в медичних дослідженнях було представлено в [17,18]. Тут формулюється це у випадку мінімаксу.

Математично проблеми машинного навчання в медичних дослідженнях базуються на наступних даних. Є набір  $D$ , що містить  $N$  кортежів

$$D = \{X_i \mid i = \overline{1, N}\} \quad (1)$$

З метою моделювання алеаторних невизначеностей розглянемо контрольовану ML щодо розподілу навчальних кортежів. У межах класу всіх підмножин  $D$  визначимо  $\Gamma$ , розподіли класів навчальних та тестових наборів

$$\Gamma = \left\{ (D_{train,j}, D_{test,j}) \subset D \times D \mid D_{train,j} \cap D_{test,j} = \emptyset, D_{train,j} \cup D_{test,j} = D, j = \overline{1, 2^N} \right\} \quad (2)$$

Де  $D_{train,j}$  та  $D_{test,j}$  – це всі можливі набори для навчання та тестування відповідно. На практиці стратегії передискретизації – це розподіл класів кортежів, які найкращим чином характеризують алеаторні невизначеності. Вводиться опис стратегії передискретизації  $\gamma \subset \Gamma$

$$\gamma = \left\{ (D_{train,k}, D_{test,k}) \subset D \times D \mid D_{train,j} \cap D_{test,j} = \emptyset, \cup_k D_{train,k} = \cup_k D_{test,k} = D \right\}, \quad (3)$$

Як приклади стратегій передискретизації  $\gamma$  можна розглянути  $cv3, cv5, cv10$ , що відповідають кратній перехресній валідації для різних  $k$ .

Кожен  $i$ -й кортеж  $X_i = (x_1^i, x_2^i, \dots, x_p^i, Y^i)^T$  складається з вхідних даних  $(x_1^i, x_2^i, \dots, x_p^i)^T$  (їх також називають атрибутами) та вихідних даних  $Y^i$ .

Нехай вектор рядка  $x_j = (x_j^1, x_j^2, \dots, x_j^N)$  представляє значення  $j$ -го атрибута всіх  $N$  кортежів. Вихідний атрибут  $Y = (Y^1, Y^2, \dots, Y^N)$  приймає всі вихідні дані. Атрибути  $x_1, \dots, x_p$  та  $Y$  (залежно від завдання класифікації або регресії) можуть приймати як числові, так і категоріальні значення.

**Об'єкт.** У найпростішому випадку проблемою контрольованого навчання є передбачити, використовуючи певний предиктор, значення вихідного атрибута  $Y^{N+1}$  на основі значень атрибутів  $x_1^{N+1}, \dots, x_p^{N+1}$ . Предиктор повинен максимізувати точність прогнозування вихідного атрибута, а саме ймовірність  $P\{Y^j \text{ corresponds to } x_1^{N+1}, \dots, x_p^{N+1}\}$  для довільного  $j \in \{1, \dots, N\}^1$ .

Далі, застосовуючи підхід  $\text{minimax}$ , вводиться  $h \in \Psi$ , розглянутий клас моделей ML  $h(X, \gamma)$ , який можна навчити та налаштувати на дані  $X \subset D$  та оцінити з урахуванням деяких стратегій передискретизації  $\gamma$ .

Порівнюючи різні моделі ML, метою машинного навчання є мінімізація очікуваних втрат. Але також потрібно взяти до уваги стратегії передискретизації, за якими також повинні оцінювати функцію втрат. Таке формулювання проблеми ML розглядає два види невизначеності. А саме, невизначеність у передискретизації є алеаторичною, оскільки вона пов'язана з даними. У той же час невизначеність при виборі моделей є епістемічною. Математично мінімальна задача ML описується як пошук моделі  $h$  завдяки

$$\arg \min_{h \in \Psi} \max_{\gamma \in \Gamma} E[L(Y, h(X, \gamma))] \quad (4)$$

Мінімаксий підхід (разом із максиміном та максимаксом) традиційно застосовується для проблем регресії [14,15]. У випадку ML один із узагальнених мінімаксий підходів відомий як  $\text{Minimax Probability Machine (MPM)}$  [24]. Можна стверджувати, що MPM є класичним результатом з вивчення надійності інтелектуальних моделей [24,25], який можна розглядати як типовий метод класифікації надійності навчання. Завданням оптимізації MPM є мінімізація верхньої межі ймовірності неправильної класифікації вивчення параметрів моделі. Верхню межу ймовірності неправильної класифікації можна використовувати як явний показник для оцінки надійності класифікаційних моделей. Версія MPM з параметричною редукацією була запропонована в [24,25] для нелінійних задач класифікації. Більше того, декілька вдосконалених алгоритмів MPM були представлені з різних точок зору [26, 27, 28,29]. У [27,28] вказувалося, що в деяких випадках слід розрізняти ймовірність неправильної класифікації двох класів, оскільки один клас може бути важливішим за інший. У [29] MPM було розширено для регресії. У роботі [26] MPM було введено для підготовки нечіткого класифікатора TSK до більш прозорої та зрозумілої моделі класифікації. На додаток до MPM, вивчення надійності інтелектуальних моделей розглядалося з інших точок зору. Наприклад, поняття “конфлікт” та “незнання” були введені для позначення надійності класифікаційних моделей у [30, 31].

Щоб зробити метод прийняття мінімального ймовірного методу доступним для навчання додаткових інтелектуальних моделей та реалізації вивчення надійності цих моделей, пропонується

© Марценюк В.П., Андрущак І.Є., Мілян Н.В.

узагальнена прихована мінімальна машина ймовірностей (Generalized Hidden-Mapping Minimax Probability Machine (GHM-MPM)). У роботі [20] верхня межа ймовірності неправильної класифікації MPM була використана як явний показник для характеристики надійності класифікаційної моделі.

Рисунок 1 використовується для ілюстрації функції прийняття рішення. Вхідний вектор  $x$  можна відобразити у прихований простір через  $\beta(\cdot)$  і підставити в MPM. Якщо вихідний результат у дорівнює  $-1$ , тоді  $x$  присвоюється негативному класу; в іншому випадку він віднесений до позитивного класу. Тут зовнішній  $\beta(\cdot)$  є запасною частиною, тоді як внутрішній MPM є фіксованим. Різниця лише у  $\beta(\cdot)$ , а обробка в MPM однакова.

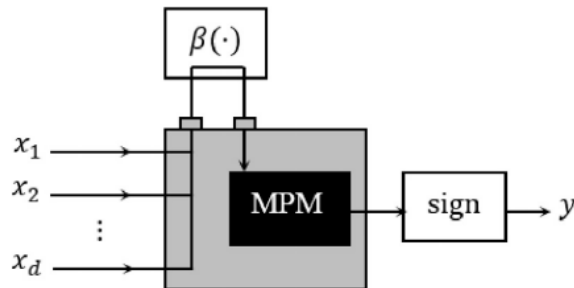


Рисунок 1. Функція прийняття рішення Generalized Hidden-Mapping Minimax Probability Machine (GHM-MPM)

Одним недоліком MPM є те, що він не включає термін регуляризації для побудови розділової гіперплощини. Регуляризація використовується для уникнення неправильно поставлених задач, а також для того, щоб добре підігнати навчальні зразки, одночасно знижуючи ризик перенавчання [32]. Наприклад, регуляризація Тихонова, або  $l_2$ -норма, широко використовується в навчальних машинах; стратегія, популяризована методом Support Vector Machine (SVM) [32]. Ця стратегія досі є одним з найпопулярніших методів класифікації та широко використовується в таких сферах, як комп'ютерний зір [33,34], медична діагностика [35, 36] та бізнес-аналітика [37, 38]. Сакета Натх (Saketha Nath) і Бхаттачарія (Bhattacharyya) [39] запропонували регульовану альтернативу MPM, в якій  $l_2$ -норма була мінімізована. Однак цей підхід використовує фіксовані значення для коефіцієнтів неправильної класифікації, а не оптимізує ці методи із завданням оптимізації.

Іншим недоліком MPM є те, що він передбачає унікальне відкликання класів для обох класів. Іншими словами, передбачається, що обидва класи однаково важливі. Машина мінімальних помилок мінімальної ймовірності (Minimum Error Minimax Probability Machine, MEMPM) [40] - це відповідне розширення MPM, яке мінімізує опуклу комбінацію коефіцієнтів неправильної класифікації, збалансовуючи два класи на основі їх попередніх ймовірностей.

Пропонується розглянути регуляризовані рецептури MPM та MEMPM, в яких  $l_2$  - норма везового вектора мінімізується разом із коефіцієнтами помилок. На відміну від моделі, запропонованої Saketha Nath та Bhattacharyya [39], ці показники є частиною проблеми оптимізації, що веде до нелінійних формулювань конусного програмування другого порядку (NSOCP).

Незважаючи на те, що MPM демонструє чудові результати узагальнення, він зазвичай вирішується вирішенням великої задачі програмування конусів другого порядку (SOCP). Коли кількість зразків збільшується, це збільшить складність його обчислення, що значно обмежує його придатність. Тому, спираючись на MPM та TELM, пропонується парна мінімаксна ймовірність екстремального нахилу машини (TMPELM) для розпізнавання шаблонів. Подібно до TELM, TMPELM також шукає дві непаралельні розділяючі гіперплощини в просторі об'єктів TELM. Для кожного гіперплану TMPELM намагається максимізувати ймовірність правильного класу класифікації зразків, тобто мінімізувати найгірший випадок (максимальну) ймовірність неправильної класифікації класу зразків, тоді як відстань до іншого класу якомога більша.

У роботі [23] [22] була запропонована машина подвійної мінімаксної ймовірності (TWMPM), яка поєднує в собі переваги як MPM, так і подвійної векторної машини підтримки. Ма та ін. вдосконалили цей підхід до MPM, який може безпосередньо оцінювати точність ймовірності, пов'язану з проблемами класифікації [19] та регресії [21], і таким чином робить модель ML більш прозорою.

У роботі [16] контрольована задача машинного навчання, що означає прогнозування  $Y$  за допомогою  $X$ , функції втрат  $L$  та набору розподілів ймовірностей  $\Gamma_{\text{on}}(X, Y)$ , була викладена як мінімакс-проблема щодо  $L$ , коли максимізація за всіма можливими розподілами  $G$  та мінімізація за правилами прийняття рішень  $\psi \in \Psi$ .

$$\arg \min_{\psi \in \Psi} \max_{P \in \Gamma} E[L(Y, \psi(X))] \quad (5)$$

де  $E[\bullet]$ - математичне сподівання.

Проблема (5) була вирішена шляхом введення узагальнення принципу максімументропії.

У роботі [41] розглянуто функціональний градієнтний підхід до імовірнісного мінімаксу активного навчання, де розглядається імовірнісний мінімакс активного навчання, зокрема нехай  $(x, y) \in \mathbf{R}^d \times \{-1, 1\}$  та  $Y, X$  - відповідні їм випадкові величини. Крім того, нехай  $p(Y = y | X = x)$  є умовною ймовірністю мітки  $y$  наведеного прикладу  $x$ . Простим класифікатором, заснованим на цій умовній ймовірності, є:

$$f(x) = \begin{cases} 1 & p(Y = 1 | X = x) \geq \frac{1}{2} \\ -1 & p(Y = 1 | X = x) < \frac{1}{2} \end{cases} \quad (6)$$

Прикладами з найбільшою невизначеністю є такі, що  $p(Y = 1 | X = x)$  найближче до  $\frac{1}{2}$ . Будь-які оцінки  $p(Y = 1 | X = x)$ , що дорівнюють  $\frac{1}{2}$ , свідчать про те, що мітка точки є випадковою. Враховуючи налаштування активного навчання, де маркованих даних мало і є багато немаркованих даних, неоптимально вибирати точки, оцінка яких  $p(Y = 1 | X = x)$  наближається до  $\frac{1}{2}$ . Бажано вибрати метод, заснований на обчисленні граничних значень для класифікації, а не просто поріг, що дорівнює  $\frac{1}{2}$  [42].

У прикладі  $x_i, i \in [1, n]$  найкращою міткою є  $y_i^b = \arg \max_{y_i} p(Y = y_i | X = x_i)$  відповідно до поточної оцінки умовної ймовірності; найгірша позначка –  $y_i^o = \arg \max_{y_i} p(Y = y_i | X = x_i)$  щодо поточної оцінки. Наприклад, коли умовні ймовірності цих двох міток далекі одна від одної, правильною міткою, безумовно, є  $\arg \max_{y_i} p(Y = y_i | X = x_i)$ . В іншому випадку правильна мітка не визначена. Одним із способів знайти такий приклад є пошук такого з максимальною ймовірністю гіршого позначення, тобто

$$q = \arg \max_q \min_{y_q} p(Y = y_q | X = x_q) \quad (7)$$

Однак ця модель не бере до уваги інші немарковані дані. Якщо припустити, що точність умовної оцінки ймовірності буде підвищена шляхом врахування немаркованих даних, інтуїтивно зрозумілою цільовою функцією ймовірнісного мінімаксного активного навчання (Probabilistic Minimax Active Learning, PMAL) є:

$$q = \arg \max_q \min_{y_q} \max_{y_{u-q}} p(y | x) \quad (8)$$

Де  $u$  - набір немаркованих даних,  $x = [x_1, x_2, \dots, x_n]$  та  $y = [y_1, y_2, \dots, y_n]$ .

Під час підготовки класифікатора Теварі та Бартлетт (Tewari and Bartlett) [43] доводять, що розрідженість не виникає, якщо умовні ймовірності можна точно оцінити. Для набору втрачених функцій вони забезпечують точний компроміс між можливістю оцінки умовної ймовірності та розрідженістю рішення. У статті [44] продемонстровано, що проблема класифікації визначає унікальну умовну ймовірність, яку опукла мінімізація ризику знаходить для неї. Крім того, дослідження [42] емпірично показує, що  $p[y=1|x]$  не може бути оцінено Logit-Boost та AdaBoost, оскільки ці два методи переважують навчання  $p[Y=1|X=x]$ , навіть незважаючи на те, що обидва ефективні як класифікатори. У [45] зазначено, що всі емпіричні алгоритми мінімізації ризиків, такі як SVM та логістична регресія, мінімізують функцію відстані оцінки умовної ймовірності даного прикладу міток до її справжнього невідомого  $p[Y=y|X=x]$ . Оскільки багато з цих алгоритмів, таких як SVM, є універсальними [46], PMAL також вважається універсальним, тобто він вибирає приклад, який мінімізує ризик класифікатора будь-якого даного розподілу даних.

AdaBoost був вперше представлений для підвищення точності набору простих правил великих вирівнювачів або слабких класифікаторів з метою отримання сильного класифікатора із задовільними властивостями узагальнення [47]. У багатьох проблемах навчання, порівняно легко знайти просте емпіричне правило з низькою точністю, дуже важко отримати класифікатор з високою точністю. Наприклад, якщо в документі є слово «футбол» як слабкий класифікатор, документ може стосуватися спорту. Грубо кажучи, в кожній ітерації AdaBoost знаходить слабкий класифікатор для неправильно класифікованих точок даних у навчальному наборі і додає цей слабкий класифікатор до зваженої суми раніше обраних слабких класифікаторів [47].

Цікаво, що AdaBoost можна трактувати по-різному. Кожне з цих уявлень про посилення забезпечує багате теоретичне розуміння навчальних проблем. Однією з них є перспектива функціонального градієнта Фрідмана [48], який пропонує більш загальне розуміння підсилення і може бути використаний для виведення його нових варіантів підсилення [49]. Дослідження Bühlmann та Yu [50] виявляє, що, хоча дисперсія або складність L2boosting для лінійних моделей зростає на кожній ітерації, величина збільшення дисперсії експоненційно зменшується у числах ітерацій. Це є причиною того, чому підсилувач стійкий до перенавчання. Хоча підсилення можна розглядати як підхід, заснований на секціях, такий як ансамбль або методи пакунків, він принципово відрізняється [51].

Мета авторів у роботі Mallapragada et al. [52] полягала у підвищенні точності будь-якого даного класифікатора, використовуючи немарковані дані. Для напівконтрольованої проблеми покращення вони пропонують алгоритм посилення, який називається SemiBoost [52]. SemiBoost використовує міру схожості між маркованими даними та немаркованими, щоб призначити псевдо мітку для будь-яких немаркованих даних. Присвоєння псевдоміток немаркованим даним було запропоновано в методах самонавчання, ASSEMBLE та Semisuperposed MarginBoost [52]. За таких підходів запас, як правило, покращується, але ніякої нової інформації класифікатор надавати не буде. Вибрані приклади для призначення псевдоміток насправді далекі від межі рішення. Перепідготовка з нещодавно присвоєними (псевдо) мітками підвищує рівень їхньої класифікації ще вище [52]. На відміну від них, автори у Mallapragada et al. [52] порівнюють приклади з маркованими та немаркованими даними та використовують подібність між прикладами, щоб призначити їм псевдо.

Цікавим підходом до маркованих даних є вибір немаркованих даних, у мітках яких класифікатор є найбільш впевненим у собі, і присвоєння цих міток, які називаються псевдо мітками, до цих вибраних немаркованих даних [52]. Всебічне дослідження методів, заснованих на цьому підході, їх властивості та експериментальні результати відображено у Triguero et al. [53]. У цій перспективі PMAL зображується як метод, який розглядає псевдо мітки для немаркованих даних і вибирає найгіршу мітку для вибраного прикладу за цільовою функцією.

**Висновки.** Із зростанням обсягу даних відбувся швидкий розвиток машинного навчання, який зробив багато теоретичних проривів та застосовується у різних галузях. Оптимізація є важливою частиною машинного навчання. Суть більшості алгоритмів машинного навчання полягає у побудові моделі оптимізації та вивченні параметрів у цільовій функції з поданих даних. В епоху великих даних ефективність та дієвість алгоритмів чисельної оптимізації різко впливає на популяризацію та застосування моделей машинного навчання. З метою сприяння розвитку машинного навчання було висунуто низку ефективних методів оптимізації, які покращили продуктивність та ефективність методів машинного навчання.

Основна увага приділена мінімаксовому підходу у машинному навчанні, сильні та слабкі сторони різних методів. Розглянуто ряд мінімаксових підходів таких, як: Minimax Probability Machine (MPM), Generalized Hidden-Mapping Minimax Probability Machine (GHM-MPM), Minimum Error Minimax Probability Machine (MEMPM), парна мінімаксна ймовірність екстремального нахилу машини (TMPELM), машина подвійної мінімаксової ймовірності (TWMPM) та деякі інші.

#### Список бібліографічних посилань

1. Farman Ali, Shaker El-Sappagh, S.M. Riazul Islam, Amjad Ali, Muhammad Attique, Muhammad Imran, and Kyung-Sup Kwak, "An intelligent healthcare monitoring framework using wearable sensors and social networking data" *Future Generation Computer Systems*, vol. 114, pp. 23-43, 2021, [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S0167739X1931605X>
2. A. Qayyum, J. Qadir, M. Bilal and A. Al-Fuqaha, "Secure and Robust Machine Learning for Healthcare: A Survey," in *IEEE Reviews in Biomedical Engineering*, doi: 10.1109/RBME.2020.3013489.
3. Spasic I, Nenadic G, "Clinical Text Data in Machine Learning: Systematic Review," *JMIR Med Inform*, vol.8(3):e17984, 2020, ;8(3):e17984. [Online]. Available: <https://medinform.jmir.org/2020/3/e17984>
4. Farman Ali, Shaker El-Sappagh, S.M. Riazul Islam, Daehan Kwak, Amjad Ali, Muhammad Imran, and Kyung-Sup Kwak, "A smart healthcare monitoring system for heart disease prediction based on ensemble deep learning and feature fusion," *Information Fusion*, vol. 63, pp. 208-222, 2020, [Online]. Available: <https://doi.org/10.1016/j.inffus.2020.06.008>
5. MI05\_uncertaintyindl.pdf, [https://cs.adelaide.edu.au/~javen/talk/ML05\\_Uncertainty\\_in\\_DL.pdf](https://cs.adelaide.edu.au/~javen/talk/ML05_Uncertainty_in_DL.pdf), (Accessed on 05/01/2020).
6. J. Noymanee and T. Theeramunkong, "Flood forecasting with machinelearning technique on hydrological modeling," *Procedia Computer Science*, vol. 156, pp. 377-386, 2019, [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S1877050919311330>.
7. P. Chemweno, L. Pintelon, P. N. Muchiri, and A. Horenbeek, "Risk assessment methodologies in maintenance decision making: A review of de-pendability modelling approaches," *Reliability Engineering System Safety*, vol. 173, pp. 64-77, 2018, [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S0951832016301934>.
8. X. Alameda-Pineda, T. M. Hospedales, E. Ricci, N. Sebe, and X. Wang, "Emerging topics in learning from noisy and missing data," in *Proceedings of the 2016 ACM on Multimedia Conference - MM '16*, ACM Press, 2016. [Online]. Available: <https://doi.org/10.1145/2964284.2986910>.
9. R. Pirracchio, M. J. Cohen, I. Malenica, J. Cohen, A. Chambaz, M. Cannesson, C. Lee, M. Resche-Rigon, and A. Hubbard, "Big data and targeted machine learning in action to assist medical decision in the icu," *Anaesthesia Critical Care Pain Medicine*, vol. 38, no. 4, pp. 377-384, 2019, [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S23525568183>
10. V. Cacciatore, A. Carullo, M. Parvis, and A. Vallan, "Uncertainty effects of data compression in measurement applications," in *2006 IEEE Instrumentation and Measurement Technology Conference Proceedings*, IEEE, Apr. 2006. [Online]. Available: <https://doi.org/10.1109/imtc.2006.328639>.
11. E. D. Mitra and W. S. Hlavacek, "Parameter estimation and uncertainty quantification for systems biology models," *Current Opinion in Systems Biology*, vol. 18, pp. 9-18, Dec. 2019. [Online]. Available: <https://doi.org/10.1016/j.coisb.2019.10.006>.
12. S.Wu and H. S. Mortveit, "A general framework for experimental design, uncertainty quantification and sensitivity analysis of computer simulation models," in *2015 Winter Simulation Conference (WSC)*, IEEE, Dec. 2015. [Online]. Available: <https://doi.org/10.1109/wsc.2015.7408240>.
13. A. G. Nakonechny and V. P. Marzeniuk, "Uncertainties in Medical Processes Control," in *Lecture Notes in Economics and Mathematical Systems*, vol. 581, pp. 185-192.
14. A. G. Nakonechnyi and A. B. Kachinskiy, "Minimax parameter estimators of a linear regression with multiplicative noises," *Journal of Automation and Information Sciences*, vol. 29, no. 2-3, pp. 98-104, 1997. [Online]. Available: <https://doi.org/10.1615/jautomatinfscien.v29.i2-3.130>.
15. J. Michálek and O. Nakonechny, "Minimax estimates of a linear parameter function in a regression model under restrictions on the parameters and variance-covariance matrix," *Journal of Mathematical Sciences*, vol. 102, no. 1, pp. 3790-3802, Oct. 2000. [Online]. Available: <https://doi.org/10.1007/bf02680236>.
16. F. Farnia and D. Tse, "A minimax approach to supervised learning," in *Proceedings of the 30th International Conference on Neural Information Processing Systems*, 2016, isbn: 9781510838819.
17. V. Martsenyuk, L. Babinets, Y. Dronyak, O. Paslay, O. Veselska, K. Warwas, I. Andrushchak, and A. Klos-Witkowska, "On development of machine learning models with aim of medical differential diagnostics of the comorbid states," in *2019 10th IEEE International Conference on Intelligent Data Acquisition and Advanced Computing Systems: Technology and Applications (IDAACS)*, IEEE, Sep. 2019. [Online]. Available: <https://doi.org/10.1109/idaacs.2019.8924345>.
18. V. Martsenyuk, V. Povoroznyuk, A. Semenets, and L. Martynyuk, "On an approach of the solution of machine learning problems integrated with data from the open-source system of electronic medical records: Application for fractures prediction," in *Artificial Intelligence and Soft Computing*, Springer International Publishing, 2019, pp. 228-239. [Online]. Available: [https://doi.org/10.1007/978-3-030-20915-5\\_21](https://doi.org/10.1007/978-3-030-20915-5_21).
19. Jun Ma, Liming Yang, Yakun Wen, and Qun Sun, "Twin minimax probability extreme learning machine for pattern recognition," *Knowledge-Based Systems*, vol. 187, Jan., pp. 104806, 2020. [Online]. Available: <https://doi.org/10.1016/j.knosys.2019.06.014>
20. Zhaohong Deng, Junyong Chen, Te Zhang, Longbing Cao, and Shitong Wang, "Generalized Hidden-Mapping Minimax Probability Machine for the training and reliability learning of several classical intelligent models," *Information Sciences*, vol. 436-437, pp. 302-319, 2018. [Online]. Available: <https://doi.org/10.1016/j.ins.2018.01.034>
21. Jun Ma and Jumei Shen, "A novel twin minimax probability machine for classification and regression," *Knowledge-Based Systems*, vol. 196, pp. 105703, 2020. [Online]. Available: <https://doi.org/10.1016/j.knosys.2020.105703>
22. Liming Yang, Yakun Wen, Min Zhang, and Xue Wang, "Twin minimax probability machine for pattern classification," *Neural Networks*, vol. 131, pp. 201-214, 2020. [Online]. Available: <https://doi.org/10.1016/j.neunet.2020.07.030>



23. Liming Yang, Boyan Yang, Shibo Jing, and Qun sun, "A minimax probability extreme machine framework and its application in pattern recognition," *Engineering Applications of Artificial Intelligence*, vol. 81, pp. 260-269, 2019. [Online]. Available: <https://doi.org/10.1016/j.engappai.2019.02.012>
24. G.R.G. Lanckriet, L.E. Ghaoui, C. Bhattacharyya, and M.I. Jordan, "A robust minimax approach to classification," *J. Mach. Learn. Res.*, vol. 3(3), pp. 555-582, 2003.
25. [G.R.G Lanckriet, L.E. Ghaoui, C. Bhattacharyya, et al., *Minimax probability machine*, in: 2001 Neural Information Processing Systems (NIPS), 2001, pp. 801-807.
26. Z. Deng, L. Cao, Y. Jiang, S. Wang, *Minimax probability tsk fuzzy system classifier: a more transparent and highly interpretable classification model*, *IEEE Trans. Fuzzy Syst.* 23 (4) (2015) 813-826.
27. K. Huang, H. Yang, I. King, et al., *The minimum error minimax probability machine*, *J. Mach. Learn. Res.* 5 (4) (2004) 1253-1286.
28. K. Huang, H. Yang, I. King, M.R. Lyu, *Imbalanced learning with a biased minimax probability machine*, *IEEE Trans. Syst. Man Cybern. Part B* 36 (4) (2006) 913-923.
29. T. Strohmann, G.Z. Grudic, *A formulation for minimax probability machine regression*, in: 2002 Neural Information Processing Systems (NIPS), 2002, pp. 769-776.
30. E. Lughofer, *Single-pass active learning with conflict and ignorance*, *Evolving Syst.* 3 (4) (2012) 251-271. [25] E. Lughofer, O. Buchtala, *Reliable all-pairs evolving fuzzy classifiers*, *IEEE Trans. Fuzzy Syst.* 21 (4) (2013) 625-641.
31. E. Lughofer, O. Buchtala, *Reliable all-pairs evolving fuzzy classifiers*, *IEEE Trans. Fuzzy Syst.* 21 (4) (2013) 625-641
32. V. Vapnik, *Statistical Learning Theory*, John Wiley and Sons, 1998.
33. W. Liu, L. Zhang, D. Tao, J. Cheng, *Support vector machine active learning by hessian regularization*, *J. Vis. Commun. Image Represent.* 49 (2017) 47-56.
34. D. Tao, L. Jin, W. Liu, X. Li, *Hessian regularized support vector machines for mobile image annotation on the cloud*, *IEEE Trans. Multimed.* 15 (4) (2013) 833-844.
35. B.D. Barkana, I. Saricicek, B. Yildirim, *Performance analysis of descriptive statistical features in retinal vessel segmentation via fuzzy logic, ann, svm, and classifier fusion*, *Knowl.-Based Syst.* 118 (2017) 165-176.
36. A. Wang, N. An, G. Chen, L. Liu, G. Alterovitz, *Subtype dependent biomarker identification and tumor classification from gene expression profiles*, *Knowl.-Based Syst.* 146 (2018) 104-117.
37. Z. -Y. Chen, Z. -P. Fan, *Distributed customer behavior prediction using multiplex data: A collaborative MK-SVM approach*, *Knowl.-Based Syst.* 35 (2012) 111-119.
38. S. Maldonado, C. Bravo, J. Pérez, J. López, *Integrated framework for profitbased feature selection and svm classification in credit scoring*, *Decis. Support Syst.* 104 (113-121) (2017).
39. J. Saketha Nath, C. Bhattacharyya, *Maximum margin classifiers with specified false positive and false negative error rates*, in: *Proceedings of the SIAM International Conference on Data Mining*, 2007.
40. K. Huang, H. Yang, I. King, M. Lyu, L. Chan, *The minimum error minimax probability machine*, in: *J. Mach. Learn. Res.*, *J. Mach. Learn. Res.* 5 (2004) 1253-1286.
41. Seyed Hossein Ghafarian, Hadi Sadoghi Yazdi, *Functional gradient approach to probabilistic minimax active learning*, in *Engineering Applications of Artificial Intelligence* 85 (2019) 21-32
42. Mease, D., Wyner, A., Buja, A., 2007. *Boosted classification trees and class probability/quantile estimation*. *J. Mach. Learn. Res.* 8, 409-439.
43. Tewari, A., Bartlett, P.L., 2007. *On the consistency of multiclass classification methods*. *J. Mach. Learn. Res.* 8, 1007-1025. [http://dx.doi.org/10.1007/11503415\\_10](http://dx.doi.org/10.1007/11503415_10).
44. Telgarsky, M., Dudík, M., Schapire, R., 2015. *Convex risk minimization and conditional probability estimation*. In: *JMLR: Workshop and Conference Proceedings*, vol. 40, pp. 1-54.
45. Zhang, T., 2004. *Statistical behavior and consistency of classification methods based on convex risk minimization*. *Ann. Statist.* 32 (1), 56-134. <http://dx.doi.org/10.1214/aos/1079120130>.
46. Steinwart, I., Scovel, C., 2007. *Fast rates for support vector machines using Gaussian kernels*. *Ann. Statist.* 35 (2), 575-607. <http://dx.doi.org/10.1214/009053606000001226>.
47. Schapire, R.E., Freund, Y., 2012. *Boosting: Foundations and Algorithms*. MIT Press, p. 544. <http://dx.doi.org/10.1017/CBO9781107415324.004>.
48. Friedman, J.H., 2001. *Greedy function approximation: A gradient boosting machine*. *Ann. Statist.* 29 (5), 1189-1232. <http://dx.doi.org/10.1214/aos/1013203451>.
49. Peter, B., Yu, B., Bühlmann, P., 2006. *Sparse boosting*. *J. Mach. Learn. Res.* 7, 1001-1024. <http://dx.doi.org/10.1109/ICASSP.2009.4959911>.
50. Bühlmann, P., Yu, B., 2003. *Boosting with the L2loss: Regression and classification*. *J. Amer. Statist. Assoc.* 98 (462), 324-339. <http://dx.doi.org/10.1198/016214503000125>.
51. Hastie, T., Tibshirani, R., Friedman, J., 2008. *The Elements of Statistical Learning Data*, second ed. Springer.
52. Mallapragada, P.K., Jin, R., Jain, A.K., Liu, Y., 2009. *SemiBoost: Boosting for semi-supervised learning*. *IEEE Trans. Pattern Anal. Mach. Intell.* 31 (11), 2000-2014.
53. Triguero, I., García, S., Herrera, F., 2015. *Self-labeled techniques for semi-supervised learning: Taxonomy, software and empirical study*. *Knowl. Inf. Syst.* 42 (2), 245-284. <http://dx.doi.org/10.1007/s10115-013-0706-y>.