Issue 15, 2024

https://doi.org/10.23939/sisn2024.15.290

УДК 004.852

МАТЕМАТИЧНА МОДЕЛЬ ЛОГІСТИЧНОЇ РЕГРЕСІЇ ДЛЯ БІНАРНОЇ КЛАСИФІКАЦІЇ. Ч. 1. РЕГРЕСІЙНІ МОДЕЛІ УЗАГАЛЬНЕННЯ ДАНИХ

Петро Кравець[1], Володимир Пасічник[2], Микола Проданюк[3]

Національний університет "Львівська політехніка",
кафедра інформаційних систем та мереж, Львів, Україна,
[1] E-mail: Petro.O.Kravets@lpnu.ua, ORCID: 0000-0001-8569-423X;
[2] E-mail: Volodymyr.V.Pasichnyk@lpnu.ua, ORCID: 0000-0002-5231-6395;
[3] E-mail: Mykola.M.Prodaniuk@lpnu.ua, ORCID: 0000-0001-9544-3792

У цій статті виконано математичне обґрунтування логістичної регресії як ефективного і простого для реалізації методу машинного навчання.

Проведено огляд літературних джерел за напрямком статистичного опрацювання, аналізу та класифікації даних методом логістичної регресії, що підтвердило популярність застосування цього методу у різних предметних областях.

Виконано порівняння методу логістичної регресії з методами лінійної та пробіт-регресії щодо можливості прогнозування імовірностей подій. У цьому контексті відмічено недоліки лінійної регресії та переваги і спорідненість методів логіт та пробіт-регресії. Вказано, що можливість прогнозування імовірностей та бінарної класифікації методом логістичної регресії забезпечується використанням сигмоїдної функції з властивістю стискаючого перетворення аргумента з необмеженим числовим значенням в обмежене діапазоном від 0 до 1 дійсне значення функції. Описано виведення сигмоїдної функції двома різними способами: на основі моделі логарифма шансів подій та моделі логістичного зростання популяції.

На основі методу максимальної правдоподібності продемонстровано побудову логарифмічної функції втрат, використання якої дозволило перейти від багатоекстремальної задачі нелінійної регресії до задачі унімодальної оптимізації. Наведено методи регуляризації функції втрат для контролю складності та запобігання перенавчанню моделі логістичної регресії.

Ключові слова: математична модель, логістична регресія, бінарна класифікація, аналіз даних, машинне навчання, сигмоїдна функція, логарифм шансів, модель логістичного зростання, метод максимальної правдоподібності, логарифмічна функція втрат, методи регуляризації.

Вступ

Сучасні інформаційні системи опрацьовують надвеликі обсяги даних, для аналізу яких застосовують методи штучного інтелекту та машинного навчання. Тому, розроблення і застосування дієвих методів машинного навчання є актуальною науково-практичною задачею. Такі методи повинні бути достовірними, ефективними для опрацювання великих наборів даних, швидкими у роботі і простими у реалізації.

Машинне навчання (machine learning) полягає в узагальненні статистичної вибірки даних для наступної можливості правильно опрацьовувати, наприклад, класифікувати, нові дані, які раніше не аналізувалися обраним методом машинного навчання.

Існують такі основні групи методів машинного навчання:

- кероване навчання (supervised learning) або з учителем;
- некероване навчання (unsupervised learning) або без учителя;
- напівкероване навчання (semi-supervised learning);
- навчання з підкріпленням (reinforcement learning).

Методи керованого навчання отримують на вході розмічені тренувальні дані, які, крім характеристичних ознак (features) об'єкта, містять потрібну на виході відповідь.

Методи некерованого навчання отримують на вході тільки ознаки об'єкта і самостійно аналізують та виявляють залежності між вхідними нерозміченими даними.

У методах напівкерованого навчання не всі дані мають наперед встановлені тренувальні мітки. Також, у випадку слабокерованого навчання, ці мітки можуть бути неточними або зашумленими.

У навчанні з підкріпленням розглядається система, що складається із середовища та одного або декількох штучних автономних агентів, які своїми діями змінюють стани середовища у ході досягнення власної мети, а через зворотний зв'язок отримують за це певні заохочення або штрафи.

Процес керованого навчання здебільшого проходить швидше і дає кращі результати. До методів керованого навчання належать: k-найближчих сусідів, наївний байєсівський класифікатор, дерева рішень, опорних векторів, регресійні, штучні нейронні мережі та інші [1].

Методи керованого навчання здобули поширення для розв'язування багатьох практичних задач, серед яких важливе місце посідають задачі класифікації даних. Серед відомих методів класифікації даних важливу роль відіграє метод логістичної регресії. Логістична регресія – це статистичний регресійний метод керованого машинного навчання для бінарної класифікації даних з метою виявлення їх належності одному із двох класів, коли коли вхідна змінна задається набором ознак об'єкта, а вихідна змінна (результат аналізу) є двійковою величиною, тобто може набувати значення 0 або 1.

Метод логістичної регресії дає можливість розробляти швидкі та ефективні алгоритми машинного навчання, які можуть бути застосовані у багатьох галузях, наприклад, для виявлення емоційного забарвлення відгуків (позитивних або негативних) клієнтів у системах електронної комерції з метою покращення продажів товарів, виявлення переважаючої реакції користувачів соціальних мереж на соціально-політичні події (схвалення або осуд), у системах підтримки прийняття рішень (так – ні, дозволити – заборонити), фільтрування вхідної інформації (спам – не спам), діагностика захворювань на основі медичних даних (хворий – здоровий), прогнозування кредитного скорингу (надати кредит – не надавати кредит) та багатьох інших.

Популярність логістичної регресії пов'язана з такими її перевагами: простота реалізації та легка інтерпретованість вагових коефіцієнтів у вигляді величини впливу відповідних ознак на результат, ефективність роботи на великих вибірках даних з незначними обчислювальними затратами, добра робота із незбалансованими класами, можливість оцінювання імовірностей належності об'єкта до певного класу.

Серед недоліків логістичної регресії відзначають лінійну залежність між ознаками і логарифмом шансів подій та залежність результату від вхідної вибірки, а саме: негативний вплив корельованих між собою ознак на результати навчання моделі та відповідну інтерпретацію результатів, недостатню стійкість до аномальних відхилень значень ознак, необхідність попередньої обробки вхідних даних, такої як видалення рядків з відсутніми значеннями ознак та необхідність масштабування ознак.

Об'єктом дослідження у цій статті є процеси машинного навчання на основі методів регресійного аналізу даних. Предметом дослідження є математичні моделі машинного навчання, основані на методі логістичній регресії.

До нових результатів цієї статті належить проведена систематизація математичних засобів логістичної регресії у контексті машинного навчання та бінарної класифікації даних.

Практичне значення роботи полягає у тому, що її результати сприятимуть поглибленому розумінню проблем машинного навчання за рахунок впорядкованого, змістовного та деталізованого подання матеріалу.

Ця стаття складається з двох пов'язаних між собою частин. У першій частині розглянуто теоретичні аспекти логістичної регресії для бінарної класифікації даних. У другій частині описано схеми навчання, тестування та основні показники оцінювання якості моделей бінарної класифікації.

Аналіз останніх досліджень та публікацій

Логістична регресія є потужним інструментом аналізу даних та прийняття рішень, який застосовується у багатьох галузях для вирішення різноманітних завдань, що потребують моделювання імовірностей подій, аналізу, прогнозування або класифікації [1, 2]. Деякі із можливих застосувань логістичної регресії подані у таблиці.

Застосування логістичної регресії по галузях

Галузь діяльності	Основні застосування
Астрономія	Розпізнавання та класифікація зоряних об'єктів і систем на основі даних спостереження і вимірювання [3].
Банківська справа	Можливість надання кредитів клієнтам; прогнозування ризику неповернення кредитів; регулювання облікової чи депозитної ставок; прогнозування імовірності зростання або спадання курсу валют [4].
Бізнес	Прогнозування клієнтської відповіді на рекламу; визначення імовірності відтоку клієнтів; імовірнісне моделювання протікання бізнес-процесів [5].
Біологія та біотехнології	Аналіз впливу різних факторів на результати експериментів; прогнозування впливу природних та антропогенних факторів на розвиток популяцій живих організмів; класифікація організмів на основі їх характеристик; прогнозування ефективності лікування або розвитку хвороб; прогнозування результатів генної інженерії; ідентифікація генів, пов'язаних із певним захворюваннями або ознаками в генетичних дослідженнях; вплив інсектицидів та гербіцидів на біосферу [6].
Військова справа	Розпізнавання свій-чужий; розпізнавання та класифікація об'єктів на основі опрацювання зображень; аналіз стратегій, тактик та прийняття рішень; прогнозування подій; прогнозування ризиків на основі воєнно-стратегічних даних; аналіз інформаційної війни та прогнозування впливу різних інформаційних кампаній на суспільство та військові дії [7].
Геологія	Прогнозування ризику природних лих, наприклад, зсувів ґрунту чи повеней на основі різних географічних та кліматичних факторів [8].
Екологія та охорона довкілля	Прогнозування впливу абіотичних, біотичних та антропогенних факторів на природне середовище; прогнозування ризику забруднення довкілля; визначення оптимальних стратегій збереження природних ресурсів; вивчення впливу різних факторів на зміни клімату або для аналізу впливу екологічних ініціатив на підприємства [9].
Економіка та фінанси	Прогнозування макроекономічних показників; прогнозування зміни фінансових, промислових, споживчих та інших індексів; класифікація різних фінансових подій, зміни валютних курсів, цін на товари та послуги, рівень інфляції, зростання або спадання ціни акцій, екстремальні події на фондовому ринку, фінансові кризи; прогнозування ризиків у фінансовій сфері; передбачення імовірності дефолту позичальника на основі його кредитної історії та інших факторів; аналіз ефективності ринкової стратегії підприємства; аналіз ринкових тенденцій та прогнозування імовірності успіху нового продукту на ринку; прогнозування венчурного (пов'язаного з ризиками) бізнесу; вплив соціально-економічних факторів на рівень безробіття; ефективність програм соціального захисту [10].
Електронні сервіси та електронна комерція	Прогнозування покупок або використання послуг на основі попередніх та поведінкових даних користувачів; аналіз ефективності маркетингових кампаній та персоналізації пропозицій для клієнтів [11].

Прдовження табл.

Галузь діяльності	Основні застосування
Енергетика	Прогнозування виробництва та продажу електроенергії; аналіз енергоефективності та впливу різних факторів на споживання електроенергії [12].
Комп'ютерна лінгвістика	Опрацювання та класифікація текстової інформації; виявлення емоційного забарвлення тексту; розпізнавання інтонації у вимові; ідентифікація автора тексту на основі його стилістики [13].
Комп'ютерні та інформаційні технології	Передбачення відмови обладнання або програмного забезпечення; аналіз впливу різних факторів на рішення про закупівлю або використання конкретних технологій; розпізнавання спаму; усунення шумів на зображеннях, реконструкція відсутніх та пошкоджених відеоданих [14].
Кримінологія	Прогнозування імовірності вчинення злочину на основі різних соціальних, економічних та демографічних факторів; оцінка ефективності поліцейських стратегій та програм в протидії злочинності; прогнозування імовірності рецидиву злочину [15].
Маркетинг	Аналіз споживацької поведінки клієнтів; прогнозування попиту на товари та послуги; вивчення впливу різних маркетингових стратегій на імовірність покупки товару або послуги; визначення ефективності рекламних кампаній або акцій для залучення клієнтів [16].
Медицина і епідеміологія	Аналіз результатів клінічних досліджень; прогнозування ризику виникнення певних захворювань; виявлення факторів, що впливають на виникнення захворювань; діагностування захворювань, прогнозування результатів лікування [17].
Метеорологія	Прогнозування показників погоди на основі метеорологічних даних, таких як температура, вологість, тиск тощо; аналіз впливу різних факторів на погодні умови [18].
Освіта	Прогнозування успішності студентів на основі їхньої академічної діяльності та інших факторів; аналіз впливу різних програм та методів навчання на знання та вміння учнів; вивчення впливу соціально-економічних факторів на освітній процес [19].
Політологія та соціологія	Аналіз виборчих процесів, політичних орієнтацій та активностей, підтримки партій вплив соціальних чи економічних чинників на рішення щодо голосування на виборах; прогнозування результатів виборів; моделювання політичних ризиків; аналіз динаміки чисельності населення; вивчення впливу соціально-економічних факторів на демографічні процеси; та кандидатів [20].
Прийняття рішень та керування ризиками	Прогнозування імовірності позитивного результату від прийнятих рішень; прогнозування багатофакторних рішень; моделювання надзвичайних ситуацій; прогнозування катастроф; оцінка імовірності виникнення ризиків на основі різних факторів [21].
Сільське господарство	Прогнозування зміни врожаїв на основі кліматичних умов та агротехнічних параметрів; оцінки ризику захворювань рослин; аналіз ефективності різних методів обробітку землі [22].
Психологія	Аналіз впливу психологічних факторів на прийняття рішень або поведінку людей; аналіз впливу психологічних факторів на соціально-політичні процеси суспільства [23].
Транспорт та логістика	Прогнозування попиту на товари; прогнозування трафіку та шляхів доставки; аналізу впливу різних факторів на ланцюг постачання; оцінки ризику аварійності на дорогах [24].
Туризм	Прогнозування попиту на туристичні послуги на основі соціо-економічних та культурних факторів; аналіз впливу різних маркетингових стратегій на відвідуваність туристичних місць [25].

Найбільшу популярність метод логістичної регресії здобув у доказовій медицині, економетриці та соціальних науках. Наведений перелік галузей застосування логістичної регресії можна розширити. У таблиці подано посилання тільки на окремі приклади застосувань логістичної регресії у відповідних галузях. Детальнішу інформацію про наукові дослідження з використанням логістичної регресії можна знайти у наукових базах даних, таких як PubMed для медичних досліджень, Google Scholar або Web of Science для публікацій з різних наукових галузей.

Хоча основні принципи логістичної регресії були визначені ще в минулому столітті, вона продовжує активно вивчатися та застосовуватись в сучасному дослідницькому та практичному середовищі. Використання логістичної регресії зростало разом із поглибленням зацікавленості у статистичному аналізі даних та машинному навчанні.

У сучасному контексті логістична регресія широко використовується у багатьох галузях, таких як медицина, соціологія, економіка, маркетинг, біологія, комп'ютерна лінгвістика, політика, та інші. Вона є одним із найпопулярніших методів аналізу категоріальних (нечислових) даних та моделювання імовірностей. Логістична регресія є популярним методом у дослідженнях та практиці з таких причин:

- Ефективність і простота. Логістична регресія – це відносно простий, легко зрозумілий, але ефективний метод класифікації, який не потребує складних обчислень і зазвичай швидко навчається та реалізується.
- Гнучкість і розширюваність. Логістична регресія може бути легко адаптована до різноманітних завдань та досліджень з класифікації та аналізу даних, включаючи багатокласову класифікацію та регресійний аналіз. Вона може бути розширена для врахування багатьох факторів, включення взаємодії між змінними, врахування нелінійних взаємодій між змінними, інтервальних даних, кластеризованих даних тощо.
- Добра прогностична здатність. Логістична регресія зазвичай добре прогнозує імовірності подій у випадку правильної специфікації моделі та використання адекватних змінних.
- Інтерпретованість результатів. Результати логістичної регресії можуть бути легко інтерпретовані, оскільки вони виражені у вигляді імовірностей. Це дозволяє визначити вплив окремих змінних на результати та зробити висновки, які можуть бути використані в практичних діях.
- Ефективність та широке застосування на практиці. Логістична регресія часто виявляється дуже ефективною на практиці в багатьох галузях, таких як медицина, фінанси, маркетинг, біологія та інші, що робить цей метод корисним і універсальним інструментом аналізу даних. Вона може успішно вирішувати різні класифікаційні завдання та прогнозувати результати з високою точністю. Логістична регресія добре працює з реальними даними, навіть коли змінні не повністю задовольняють умови класичної лінійної регресії. Вона може ефективно моделювати імовірності неперервних подій з категоріальними або бінарними вихідними даними.
- Розвиток технологій та методів. Завдяки постійному розвитку комп'ютерних технологій і методів аналізу даних, логістична регресія продовжує залишатися актуальною. Вона може бути легко впроваджена в різні програмні середовища та бібліотеки, що дозволяє дослідникам та практикам використовувати її для розв’язування різноманітних завдань.
- Доступність програмних засобів. Існують різноманітні програмні пакети, такі як R, Python Scikit-learn, SAS, SPSS Statistics та інші, які надають зручні засоби для моделювання логістичної регресії, що робить цей метод доступним для широкого кола дослідників і практиків.

Отже, популярність логістичної регресії можна пояснити її ефективністю, простотою, гнучкістю та здатністю до інтерпретації результатів, які роблять її привабливим та популярним методом для вирішення класифікаційних завдань та аналізу даних серед дослідників і практиків.

Важливо продовжувати досліджувати цей метод і розвивати нові підходи для його застосування в різних галузях.

Додаткове дослідження та глибше пророблення методу логістичній регресії можливе у таких напрямах:

- Розвиток нових методів побудови та оцінки логістичних моделей, враховуючи нові методи статистичного аналізу, машинного навчання та складні взаємозв'язки між даними.

- Удосконалення методів навчання логістичної регресії з врахуванням нелінійних зв'язків між даними та часової залежності даних.
- Розробка методів логістичної регресії, які можуть ефективно працювати з великими за обсягом та різноформатними складними даними, такими як текст, звук, зображення, що надходять з давачів Інтернету речей, соціальних мереж тощо.
- Розробка методів аналізу неоднорідної вибірки даних, коли спостереження виходять за межі класичних умов логістичної регресії.
- Розробка та вдосконалення інструментів, пакетів, сервісів, методів і засобів автоматизації побудови моделей логістичної регресії для полегшення процесу аналізу даних та забезпечення його використання для широкого кола дослідників і практиків.
- Розширення областей застосування логістичної регресії, наприклад, для прийняття рішень в різних галузях, для прогнозування надзвичайних подій або ризиків, таких як фінансові кризи, природні катастрофи або епідемії.

Загалом, сучасні публікації з логістичної регресії роблять значний внесок в науку та практику, допомагаючи розвивати нові підходи до аналізу даних і прийняття рішень.

Однак, за винятком деяких монографічних досліджень, наприклад, [26 – 28], висвітлення математичних основ методу логістичної регресії у загальному масиві публікацій зі статистичного аналізу та машинного навчання є розрізненим, фрагментарним та проблемно-орієнтованим. Обмеження на обсяг статті у періодичних друкованих виданнях не дозволяє авторам цілісно висвітлити проблему класифікації методом логістичної регресії. Електронні публікації теж у своїй більшості фрагментарно описують проблему логістичної регресії. В існуючих публікаціях більша увага звернена на практичну реалізацію у конкретній предметній області, що дещо відволікає від загального розуміння проблеми.

У цій статті зроблена спроба подати матеріал у контексті машинного навчання систематизовано та послідовно, зробити його більш зрозумілим і хоча частково усунути наявні прогалини у відомих, особливо україномовних, публікаціях. Також методи регресійного аналізу та класифікації розглядаються у навчальних дисциплінах з машинного навчання та аналізу даних. Висвітлений у статті матеріал буде корисним для студентів, аспірантів та дослідників, які бажають глибше ознайомитися з регресійними методами прогнозування і класифікації. Можливо, що стаття доповнить існуючий масив публікацій з цієї проблеми, полегшить процес аналізу даних, зробить метод логістичної регресії ще доступнішим для широкого кола користувачів і популярнішим для практичних застосувань.

Мета роботи

Головною метою цієї роботи є математичне обґрунтування логістичної регресії як ефективного і простого для реалізації методу машинного навчання. Підпорядкованою метою є популяризація використання методу логістичної регресії у системах бінарної класифікації.

Для досягнення мети необхідно:

- Обґрунтувати застосування моделі логістичної регресії для бінарної класифікації ознак об’єктів.
- Провести математичну формалізацію логістичної регресії, описати її базові припущення та модельні параметри.
- Математично обґрунтувати вид логістичної функції на основі моделі шансів подій або логістичної моделі зростання та вивести логарифмічну функцію втрат на основі методу максимальної правдоподібності.

Регресійні моделі аналізу даних

Серед методів машинного навчання з учителем окремо виділимо групу методів регресійного аналізу. Термін “регресія” (спадання) вперше був введений Ф. Гальтоном (F. Galton) у 1877 р. для дослідження успадкування характеристик від одного покоління до іншого.

Регресійні моделі можна розглядати як математичні інструменти узагальнення даних, які дозволяють робити прогнози, розуміти залежності між різними змінними, виявляти закономірності та патерни в даних. Основна мета регресійного аналізу полягає в тому, щоб встановити зв'язок між залежними і незалежними змінними та використовувати цей зв'язок для прогнозування значень залежних змінних на основі нових значень незалежних змінних. Регресійні моделі дозволяють аналізувати та використовувати дані для розуміння та передбачення поведінки системи або явища.

Суть регресійного аналізу полягає в аналітичному або пошуковому визначенні коефіцієнтів ознак об’єктів так, щоб мінімізувати сумарну похибку відхилень модельної аналітичної функції від значень експериментальних даних по усій вхідній вибірці. Інакше, регресія визначає найкращу аналітичну апроксимацію вибірки даних, яка дозволяє прогнозувати значення вихідної змінної за новими значеннями вхідних даних.

Лінійна регресія

Лінійна регресія за методом найменших квадратів як засіб лінійної відповідності експериментальному набору точок була застосована А.-М. Лежандром (A.-M. Legendre) у 1805 р. і К. Ф. Гаусом (J. C. F. Gauss) у 1809 р. для передбачення руху планет. Пізніше у 1811 р. П.-С. Лаплас (P. S. Laplace) застосував метод найменших квадратів для розвитку теорії похибок оцінювання. Поширення цього методу на соціальні науки у 19 ст. виконав А. Кетле (L. A. J. Quetelet).

Цільова функція лінійної регресії визначається як мінімізація середньоквадратичної похибки між прогнозованим і актуальним значенням спостережуваного параметра:

$$F = \frac{1}{n}\sum_{i=1}^{n}\left(\hat{y}_i - y_i\right)^2 \to \min, \qquad (1)$$

Де $\hat{y}_i$ – прогнозоване (аналітичне) значення, лінійне відносно шуканих коефіцієнтів; y_i – актуальне (експериментальне) значення; n – довжина вибірки даних.

Загальна модель лінійної регресії має вигляд:

$$\hat{y} = w_0 + w_1x_1 + w_2x_2 + ... + w_mx_m = \sum_{i=0}^{m} w_ix_i = W^{\mathrm{T}}X \in R^1, \qquad (2)$$

де $W^{\mathrm{T}} = (w_0, w_1, w_2, ..., w_m)$ – ваги ознак об’єкта; $w_i \in R^1$; $X = (1, x_1, x_2, ..., x_m)^{\mathrm{T}}$ – числові ознаки (предиктори) об’єкта; $x_i \in R^1$, $i = 1..m$; m – кількість ознак об’єкта; T – символ транспонування.

Для нелінійної регресії функція $\hat{y}$ є нелінійною відносно ознак X або вагових коефіцієнтів W. Таку нелінійність можуть створювати самі значення ознак, наприклад, у вигляді їх різноманітних добутків, або існує нелінійна залежність функції регресії від вагових параметрів. Іноді можна перейти від нелінійної моделі до лінійної методом заміни змінних або відповідним перетворенням, наприклад, логарифмуванням. Якщо ж регресійну модель не можна звести до лінійної, то вона розв’язується наближеними ітеративними методами.

Узагальнена схема багатофакторної лінійної регресії зображена на рис. 1. У контексті машинного навчання лінійна регресія є методом навчання з учителем. Модель регресійного навчання виконує апроксимацію вхідних значень ознак (експериментальних даних) за допомогою лінійної аналітичної функції (2). Для цього розв’язується оптимізаційна задача знаходження коефіцієнтів W лінійної функції за методом найменших квадратів відхилення актуальних значень y_i залежної

змінної від обчислених за функцією цільових значень $\hat{y}_i$ по усій вибірці вхідних ознак $\{X_i\}_1^n$. Після тренування ця функція може використовуватися для прогнозування цільових значень для нових вхідних даних, які не були використані під час навчання. Подібний спосіб навчання стосується не лише лінійної регресії, але й інших регресійних методів, таких як логістична регресія та пробіт-регресія.

Мінімізація середньоквадратичної похибки (1) здійснюється належним визначенням ваг W ознак об'єктів. Для цього можна використати аналітичні або числові методи.

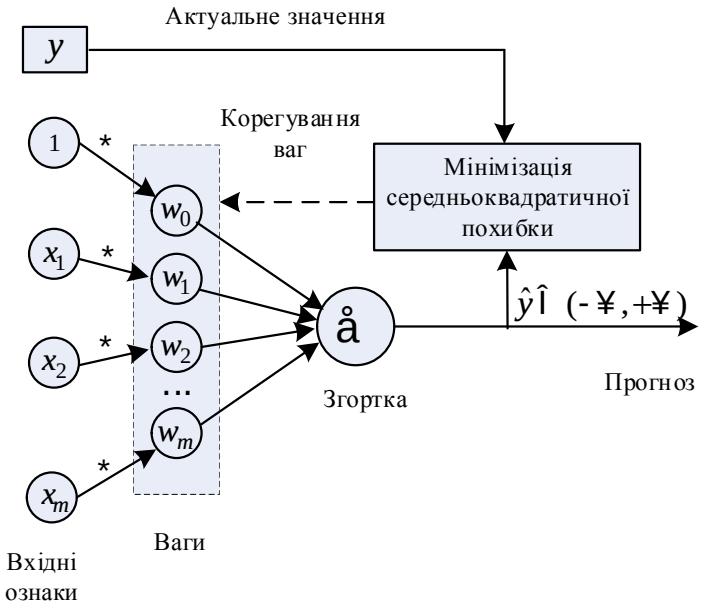


Рис. 1. Схема багатофакторної лінійної регресії

Для спрощення розглянемо однофакторну модель $\hat{y} = f(x)$:

$$\hat{y} = w_0 + w_1 x.$$

Тоді

$$F = \frac{1}{n}\sum_{i=1}^{n}\left(w_0 + w_1 x_i - y_i\right)^2 \to \min_W. \tag{3}$$

Задача мінімізації (3) полягає у знаходженні таких коефіцієнтів w_0, w_1, які забезпечують найменше середньоквадратичне відхилення $\hat{y}$ від y по усій експериментальній вибірці.

Виконаємо аналітичне визначення вагових коефіцієнтів w_0, w_1. У точці мінімуму виконуються умови: $\frac{\partial F}{\partial w_0} = 0$ та $\frac{\partial F}{\partial w_1} = 0$. За правилом обчислення похідної складеної функції отримаємо:

$$\frac{\partial F}{\partial w_0} = \frac{2}{n}\left(\sum_{i=1}^{n}\left(w_0 + w_1 x_i - y_i\right)\right) = 0,$$

$$\frac{\partial F}{\partial w_1} = \frac{2}{n}\left(\sum_{i=1}^{n}\left(w_0 + w_1 x_i - y_i\right)x_i\right) = 0.$$

Після нескладних перетворень матимемо:

$$\begin{cases} w_0 n + w_1 \sum_{i=1}^{n} x_i = \sum_{i=1}^{n} y_i, \\ w_0 \sum_{i=1}^{n} x_i + w_1 \sum_{i=1}^{n} x_i^2 = \sum_{i=1}^{n} x_i y_i. \end{cases} \quad (4)$$

Якщо детермінант $\Delta = n\sum_{i=1}^{n} x_i^2 - \left(\sum_{i=1}^{n} x_i\right)^2 \neq 0$, то система (4) має такі розв'язки:

$$w_0 = \frac{\Delta_0}{\Delta}, \; w_1 = \frac{\Delta_1}{\Delta},$$

де $\Delta_0 = \sum_{i=1}^{n} x_i^2 \sum_{i=1}^{n} y_i - \sum_{i=1}^{n} x_i \sum_{i=1}^{n} x_i y_i$, $\Delta_1 = n\sum_{i=1}^{n} x_i y_i - \sum_{i=1}^{n} x_i \sum_{i=1}^{n} y_i$.

Замість аналітичного визначення вагових коефіцієнтів характеристичних ознак об'єктів, особливо для багатопараметричної лінійної або нелінійної регресії, можна використати ітераційні методи оптимізації, які знаходять наближені розв'язки.

Методи регресії широко використовуються у біології, епідемології, економіці, фінансах, маркетингу, соціології, машинному навчанні для встановлення залежності між параметрами моделі, прогнозування значення та визначення тренду зміни цільової функції.

У лінійній регресії модельна функція $\hat{y} \in (-\infty, +\infty)$ може набувати довільні дійсні значення. У дихотомічних моделях $\hat{y} \in \{0,1\}$. В імовірнісних моделях $\hat{y} \in [0,1]$.

У задачах бінарної класифікації зручно оперувати імовірністю належності об'єкта до одного із двох класів. Порівняння такої імовірності з пороговим значенням дає можливість легко отримати прогнозований бінарний клас об'єкта.

Модель лінійної регресії не може гарантувати адекватні результати у термінах імовірностей подій, оскільки її прогнози можуть приймати значення від $-\infty$ до $+\infty$.

Для вирішення цієї проблеми необхідно зробити таке звужуюче перетворення:

$$p(E) = R(\hat{y}) = R(W^{\top} X) \in [0,1],$$

де $p(E)$ – імовірність події; R – деяка функція зі значеннями на відрізку [0, 1].

Якщо значення $\hat{y}$ лінійної регресії прямує до $-\infty$, то імовірність події $p(E)$ наближається до 0. Якщо ж прямує до $\hat{y}$ до $+\infty$, то імовірність події прямує до 1.

Для забезпечення таких властивостей у якості функції R найчастіше використовують:

- функцію нормального розподілу випадкової величини;
- функцію логістичного розподілу.

Відповідні моделі імовірності називають пробіт (probit) та логіт (logit) регресією. Вони є моделями бінарного вибору, що використовуються для прогнозування імовірностей виникнення деякої події. Для цього пробіт-регресія перетворює пряму лінійної регресії у криву нормального розподілу, а логістична регресія – в *S*-криву за допомогою сигмоїдної функції. Ці функції завжди приймають дійсні значення від 0 до 1.

Обидва методи, пробіт і логіт-регресія, широко використовуються для моделювання імовірностей бінарних подій. Вони є популярними і добре вивченими методами. Вибір між ними зазвичай залежить від конкретної ситуації та особливостей дослідження. Логіт-регресія часто використовується у статистичних моделях, тоді як пробіт-регресія може бути більш популярною в економетричних дослідженнях, особливо там, де базова теорія підтримує використання нормального розподілу.

Хоча обидва методи використовують лінійну комбінацію вхідних ознак і їх вагових коефіцієнтів, але для отримання імовірності використовуються різні функції: у логіт-регресії це

сигмоїдна функція, яка виводиться із логарифму шансів подій (відношення імовірностей подій), а у пробіт-регресії це нормальний кумулятивний (інтегральний) розподіл.

У логіт-регресії зміна вагових коефіцієнтів ознак об'єкта впливає на значення логарифму шансів подій, а у пробіт-регресії така зміна безпосередньо впливає на значення імовірності події.

По суті моделі пробіт та логіт-регресії є одношаровими штучними нейронними мережами, які відрізняються функціями активації.

Пробіт-регресія

Базова модель пробіту визначається законом Вебера-Фехнера (Weber-Fechner law) і датується 1860 роком. Дж. Гаддум (J. Gaddum) систематизував попередні роботи у 1933 р. Термін probit запропонував Ч. Бліс (C. Bliss) у 1934 р. Назва моделі probit походить від початку слова **prob**ability та закінчення слова un**it**.

Використовується в моделях бінарного або множинного вибору, для моделювання дефолту компаній, у банківській справі для оцінки імовірності неповернення кредиту, у токсикології та фармації для оцінки впливу дози чи концентрації речовини на біологічні об'єкти тощо.

Узагальнена схема пробіт-регресії зображена на рис. 2. Пробіт-регресія є методом навчанням з учителем.

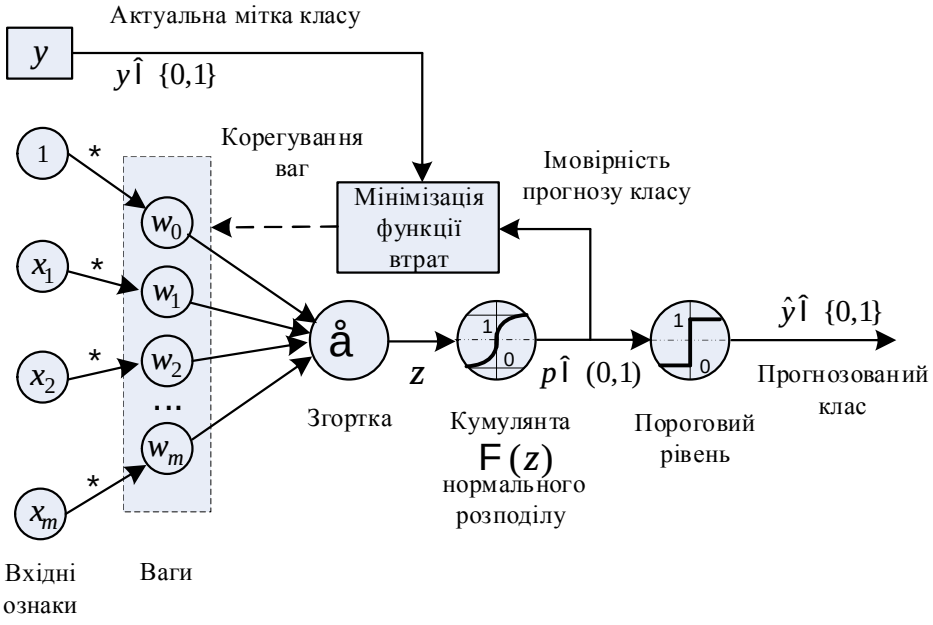


Рис. 2. Схема пробіт-регресії

Обґрунтування пробіт-регресії основано на такому.

Нехай вихід пробіт-регресії $\hat{y} \in \{0,1\}$ приймає тільки два значення: 0 та 1. На значення $\hat{y}$ впливає вектор-стовпець X ознак об'єкта: $X = (1, x_1, x_2, ..., x_m)^T$.

Припустимо, що існує випадкова величина $z = W^T X + u$, де $W^T = (w_0, w_1, w_2, ..., w_m)$ – вагові коефіцієнти; $W^T X$ – скалярний добуток двох векторів; $u : N(0,1)$ – нормально розподілена випадкова величина з нульовим математичним сподіванням та одиничною дисперсією. Тоді значення $\hat{y}$ можна розглядати як індикатор того, що прихована змінна z є додатною:

$$\hat{y} = \begin{cases} 1, & z > 0 \\ 0, & z \le 0 \end{cases}.$$

Нехай імовірність $P(\hat{y}=1|X)$ того, що $\hat{y}=1$ при спостереженні X визначається кумулятивною функцією стандартного нормального розподілу (законом розподілу Гауса):

$$\Phi(z=W^{\mathrm{T}}X)=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{z}e^{-\frac{t^2}{2}}dt.$$

Тоді $P(\hat{y}=1|X)=P(z>0)=P(W^{\mathrm{T}}X+u>0)=P(u>-W^{\mathrm{T}}X)$.

Завдяки симетрії нормального розподілу $P(u>-W^{\mathrm{T}}X)=P(u<W^{\mathrm{T}}X)$. З означення закону розподілу випадкової величини отримаємо $P(u<W^{\mathrm{T}}X)=\Phi(W^{\mathrm{T}}X)$ і як результат

$$P(\hat{y}=1|X)=\Phi(W^{\mathrm{T}}X). \qquad (5)$$

Застосування стандартного нормального розподілу не звужує загальності результату, оскільки інші значення математичного сподівання і середньоквадратичного відхилення можна скомпенсувати зміною параметрів моделі.

Функцію нормального розподілу Φ можна розглядати як активаційну функцію перетворення лінійної комбінації вхідних ознак $z=W^{\mathrm{T}}X$ в імовірність виникнення події.

Для моделювання імовірностей бінарних подій в статистичних моделях використовується пробіт-функція. Пробіт-функція визначається як обернена до функції $\Phi(z)$:

$$probit(p)=\Phi^{-1}(z)\in(-\infty,+\infty),$$

де $p\in(0,1)$ – імовірність; $z=W^{\mathrm{T}}X$ – згортка вхідних ознак X з вагами W.

Пробіт-функція дозволяє за значенням імовірності прогнозу p на виході моделі знайти значення $z=W^{\mathrm{T}}X$ для інтерпретації результату у термінах параметрів моделі. Це дозволяє оцінити важливість різних ознак X у прогнозуванні бінарної події шляхом визначення їх впливу на імовірність події p. Параметри моделі, що оцінюються за допомогою пробіт-функції, можуть бути інтерпретовані як фактори впливу вхідних ознак на імовірність виникнення події. Це дає змогу визначити, які вхідні змінні мають статистично значущий вплив на імовірність виникнення події та напрямок цього впливу.

У пробіт-регресії, вагові коефіцієнти можна інтерпретувати як величину зміни імовірності залежно від зміни незалежних ознак.

Застосуємо часткову похідну по x_i до імовірності (5):

$$\frac{\partial p}{\partial x_i}=\frac{\partial\Phi(z)}{\partial z}\times\frac{\partial z}{\partial x_i},$$

де $p=P(\hat{y}=1|X)$.

Враховуючи, що для стандартного нормального розподілу $\frac{\partial\Phi(z)}{\partial z}=\frac{1}{\sqrt{2\pi}}e^{-\frac{z^2}{2}}$ і $\frac{\partial z}{\partial x_i}=w_i$, отримаємо:

$$\frac{\partial p}{\partial x_i}=\frac{1}{\sqrt{2\pi}}e^{-\frac{z^2}{2}}\times w_i.$$

Отже зміна ознаки x_i за умови, що інші ознаки залишаються постійними, призводить до зміни імовірності події p на величину, пропорційну ваговому коефіцієнту w_i.

Для прикладу розглянемо пробіт-модель з однією незалежною змінною $z=w_0+w_1x_1$. Для спрощення приймемо $w_0=0$. Нехай $w_1=0.5$. Тоді

$$\frac{\partial p}{\partial x_1} = \frac{1}{\sqrt{2\pi}} e^{-\frac{(0.5x_1)^2}{2}} \times 0.5 \approx 0.199 e^{-\frac{(x_1)^2}{8}}.$$

Для прикладу задамо початкове значення ознаки $x_1 = 1$. Тоді одиничний приріст ознаки $\Delta x_1 = 1$ за умови, що інші ознаки приймають постійні значення, призведе до зміни імовірності на $\Delta p \approx 0.199 \times 0.882 \approx 0.176$.

Для додатних значень ваги $w_1 > 0$ одиничний приріст $\Delta x_1 = 1$ призведе до збільшення імовірності настання події: $\Delta p = p_2 - p_1 > 0$. Для від'ємних значень ваги $w_1 < 0$ одиничний приріст $\Delta x_1 = 1$ призведе до зменшення імовірності настання події: $\Delta p = p_2 - p_1 < 0$.

Залежність приросту Δp імовірності події від ваги w_1 є нелінійною, як це показано на рис. 3 для декількох значень x_1. Графіки функцій Δp проходять через точку (0, 0). Для кожного x_1 існує додатне значення $w_1^+ > 0$, для якого імовірність настання події зросте на максимальну величину Δp. Для значень $w_1 > w_1^+$ приріст імовірності Δp наближається до нуля зі сторони додатних значень Δp. Також існує симетричне від'ємне значення $w_1^- < 0$, для якого імовірність настання події зменшиться на максимальну величину $|\Delta p|$. Для значень $w_1 < w_1^-$ приріст імовірності також прямує до нуля, але зі сторони від'ємних значень Δp.

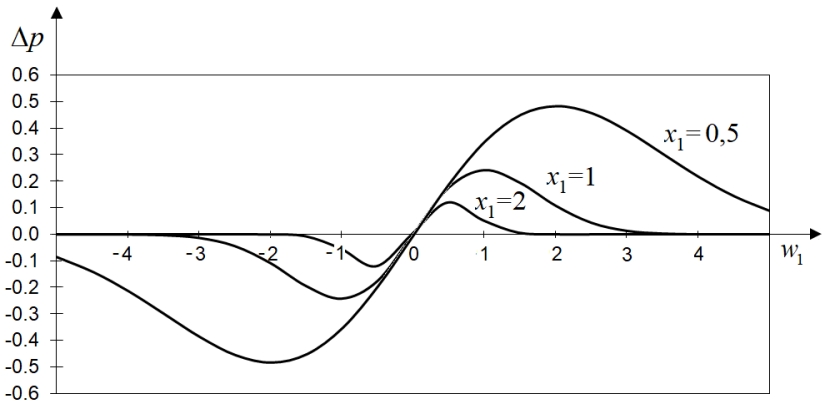


Рис. 3. Залежність приросту імовірності події від ваги ознаки для пробіт-регресії

Зростання абсолютного значення ознаки x_1 призводить до зменшення абсолютної величини приросту імовірності Δp. Така залежність може додатково вказувати на важливість нормалізації значень ознак.

Визначення оптимальних значень вагових коефіцієнтів моделі здійснюється у ході мінімізації логарифмічної функції втрат, яка виводиться з експерименту статистичного випробування Бернуллі.

Для окремого спостереження статистичного дихотомічного експерименту матимемо:

$$P(y = 1 | X) = \Phi(W^T X),$$

$$P(y = 0 | X) = 1 - \Phi(W^T X),$$

де y – актуальне значення мітки класу для набору ознак X.

Тоді імовірність одного спостереження можна обчислити за формулою Бернуллі:

$$P[Y = y | X] = \Phi\left(W^T X\right)^y \left(1 - \Phi\left(W^T X\right)\right)^{1-y}.$$

Нехай статистична вибірка складається з набору даних $\{X_i, y_i\}_{i=1}^n$. Для незалежних та однаково розподілених спостережень імовірність усієї вибірки даних дорівнює добутку імовірностей окремих спостережень:

$$L(W, X, Y) = \prod_{i=1}^{n} \sigma\left(W^{\top} X_i\right)^{y_i} \left(1 - \sigma\left(W^{\top} X_i\right)\right)^{1 - y_i},$$

де $y_i \in Y = \{0,1\}$, $W, X_i \in R^{m+1}$, $W^{\top} X_i$ – скалярний добуток двох векторів.

Тоді загальна логарифмічна функція правдоподібності дорівнює

$$\log L(W, X, Y) = \sum_{i=1}^{n} \left[y_i \times \log \sigma\left(W^{\top} X_i\right) + \left(1 - y_i\right) \times \log\left(1 - \sigma\left(W^{\top} X_i\right)\right) \right] \to \max_W. \qquad (6)$$

У машинному навчанні замість (6) прийнято мінімізувати функцію втрат. Враховуючи, що $\max\left(\log(z)\right) = \min\left(-\log(z)\right)$, після зміни знаку на протилежний отримаємо таку цільову функцію втрат:

$$F(W, X, Y) = -\log L(W, X, Y) = -\sum_{i=1}^{n} \left[y_i \times \log \sigma\left(W^{\top} X_i\right) + \left(1 - y_i\right) \times \log\left(1 - \sigma\left(W^{\top} X_i\right)\right) \right] \to \min_W.$$

Ця функція є гладкою та унімодальною, вона має один мінімум, тому стандандартні числові алгоритми її оптимізації швидко сходяться.

Логістична регресія

Логістична регресія (logistic regression) або логіт-регресія (logit regression) – це статистичний регресійний метод прогнозування залежної змінної з бінарними значеннями 0 або 1. У задачі бінарної класифікації ці значення позначають мітки класів: 0 – негативний, 1 – позитивний. Семантично ці мітки можна інтерпретувати як протилежні за значенням слова (антоніми), наприклад, “ні” – “так”, “проти” – “за”, або різні за значенням, наприклад, “обака” – “кіт”.

Логістична регресія – це метод навчання з учителем. Для формування прогнозованих (вихідних) міток класів використовується сигмоїдна функція обчислення імовірності належності об'єкта до певного класу з наступним порівнянням цієї імовірності із заданим порогом.

Підґрунтя методу логістичної регресії були закладені ще у 19 ст. в роботах К. Ф. Гаусса (J. C. F. Gauss) та П.-С. Лапласа (P. S. Laplace) у вигляді методу найменших квадратів.

Логістична модель (logit model) була вперше використана Е. Б. Вілсоном (E. B. Wilson) та Д. Вустер (J. Worcester) у 1943 р. як альтернатива пробіт-моделі в біоаналізі. Починаючи з 1944 р., значний розвиток логістичної моделі було зроблено у роботах Дж. Берксона (J. Berkson).

Назва logit утворена за аналогією probit і походить від терміну **log**istic un**it**, що вказує на використання функції logistic curve у цьому методі моделювання (Дж. Берксон, 1944). Функція логістичної кривої має вигляд S-подібної кривої, яка використовується для прогнозування імовірності виникнення події (наприклад, успіху чи невдачі) в залежності від значень вхідних змінних.

В узагальненому визначенні прогностична здатність логістичної регресії може задаватися більше ніж двома вихідними класами, наприклад, “проти” – “утримався” –“за”. Тоді така форма логістичної регресії називається поліноміальною (multinomial logistic regression або mlogit) [29]. Ця модель була незалежно представлена Коксом (D. R. Cox) у 1966 р. та Тейлом (H. Theil) у 1969 р. Поліноміальна модель підвищила популярність та поширила сферу застосування логіт-моделі.

Великий прорив у розвитку логістичної регресії стався в другій половині 20-го століття, коли були розроблені числові методи оцінювання параметрів моделі та відбувся стрімкий розвиток обчислювальної техніки, що дозволило швидко і точно обчислювати результати регресійного аналізу.

Схема логіт-регресії відповідає зображеній на рис. 2 схемі пробіт-регресії, тільки замість кумулятивної функції нормального розподілу використовується сигмоїдна функція активації, відома як логістична функція.

Визначальними факторами успішного застосування логістичної регресії у сучасних дослідженнях є легкість налаштування, висока швидкість навчання, надійність результатів класифікації, особливо, на лінійно роздільних даних, простота практичної реалізації. Логістична регресія не тільки прогнозує клас об'єкта, а ще й оцінює імовірність належності об'єкта до цього класу, що дозволяє визначити ризики роботи класифікатора.

Логістична регресія складається з таких кроків:

1. Підготовка даних: вхідні дані формуються у вигляді таблиці $(X_i, y_i)_{i=1}^n$, де кожен рядок (X_i, y_i) є результатом окремого спостереження; тут $X_i \in R^{m+1}$ є набором факторів (ознак об'єкта), а $y_i \in \{0,1\}$ – актуальна (дійсна) мітка класу цього об'єкта.

2. Навчання моделі: на вхід моделі подається вибірка навчальних даних $(X_i, y_i)_{i=1}^n$ для обчислення за допомогою сигмоїдної функції імовірності $p_i \in (0,1)$ прогнозованого класу для кожного елемента даних X_i з наступним використанням $(X_i, y_i, p_i)_{i=1}^n$ для мінімізації функції втрат за допомогою регульованих параметрів (ваг) $W \in R^{m+1}$ ознак об'єктів.

3. Оцінка моделі: визначаються метрики якості моделі на основі опрацювання вибірки тестових даних $(\hat{X}_i)_{i=1}^k$, які не були використані для навчання.

4. Прогнозування: після навчання та оцінювання якості прогнозу логістична модель використовується для визначення належності об'єктів до одного із бінарних класів $\hat{y}_i \in \{0,1\}$ порівнянням імовірності $p_i \in (0,1)$ прогнозу класу з пороговим значенням $t \in (0,1)$.

До базових понять логістичної регресії належать: логістична функція, логарифмічна модель шансів подій, логістична модель зростання, критерій максимальної правдоподібності, логарифмічна функція втрат, начання логістичної регресії, тестування логістичної регресії, матриця помилок прогнозування, міри якості логістичної регресії, порогова функція.

Логістична функція

Термін "логістична функція" введено бельгійським математиком П. Ф. Ферхюльстом (P. F. Verhulst) у серії робіт між 1838 та 1847 роками, який розробив її під керівництвом А. Кетле (A. Quetelet) як модель зростання населення у противагу експоненційній моделі.

У машинному навчанні логістична функція – це монотонно зростаюча функція з точкою перегину, яка визначає перехід області випуклості в увігнутість та двома областями насичення, коли дальша зміна аргумента (зменшення аргумента для випуклої області або збільшення для увігнутої області) не призводить до значної зміни значення функції.

Логістична функція або лог-функція, також відома як сигмоїдна функція, використовується у логістичній регресії для того, щоб відображати будь-яку дійсну числову величину $z \in R^1$ в діапазон між 0 і 1, а значення з цього діапазону можна було розглядати як імовірність події:

$$\sigma : z \to (0,1).$$

Аргумент сигмоїдної функції визначається як лінійна згортка

$$z = W^{\top} X = \sum_{i=0}^{m} w_i x_i \in R^1 \tag{7}$$

ваг $W^{\top} = (w_0, w_1, w_2, ..., w_m)$ та представлених у вигляді чисел $X = (1, x_1, x_2, ..., x_m)^{\top}$ ознак x_i об'єкта, де $w_i, x_i \in R^1$, $i = 0..m$; m – кількість ознак об'єкта.

Формула логістичної функції має вигляд:

$$\sigma(z) = p = \frac{1}{1 + e^{-z}} \in (0,1), \tag{8}$$

де $p = P(\hat{y} = 1) \in (0,1)$ – імовірність належності об'єкта до класу 1; $e \approx 2.71828$ – число Ейлера або експонента, основа натурального логарифма.

Значення w_0 позначає коефіцієнт зміщення (bias) зваженої суми числових ознак. У формулі лінійної згортки (7) додатково може використовуватись ще одне зміщення, яке не множиться на ваговий коефіцієнт. Зазвичай таке зміщення позначає стохастичну складову моделі.

Сигмоїдна функція має властивість симетрії:

$$1 - \sigma(z) = \sigma(-z).$$

Похідна сигмоїдної функції є додатною гладкою функцією, яка визначається через цю ж функцію:

$$\sigma'(z) = \sigma(z) \times (1 - \sigma(z)). \tag{9}$$

Щоб продемонструвати це, обчислимо

$$\frac{d\sigma(z)}{dz} = \frac{d}{dz}\left(\frac{1}{1+e^{-z}}\right).$$

Застосувавши правило похідної для частки, отримаємо:

$$\frac{d}{dz}\left(\frac{1}{1+e^{-z}}\right) = \frac{0 \times (1+e^{-z}) - 1 \times \frac{d}{dz}(1+e^{-z})}{(1+e^{-z})^2} = \frac{-1 \times \frac{d}{dz}(e^{-z})}{(1+e^{-z})^2}. \tag{10}$$

Ще раз застосуємо правило похідної для частки $e^{-z} = \frac{1}{e^z}$:

$$\frac{d}{dz}(e^{-z}) = \frac{0 \times e^z - 1 \times \frac{d}{dz}(e^z)}{(e^z)^2} = \frac{-e^z}{(e^z)^2}. \tag{11}$$

Підставивимо вираз (11) у вираз (10):

$$\frac{-1 \times \frac{d}{dz}(e^{-z})}{(1+e^{-z})^2} = \frac{e^z}{(e^z)^2(1+e^{-z})^2} = \frac{e^{-z}}{(1+e^{-z})^2}. \tag{12}$$

Додамо і віднімемо 1 у чисельнику виразу (12) . Після відповідного перетворення виразу остаточно отримаємо:

$$\frac{(1+e^{-z}) - 1}{(1+e^{-z}) \times (1+e^{-z})} = \frac{1}{1+e^{-z}}\left(1 - \frac{1}{1+e^{-z}}\right) = \sigma(z) \times (1 - \sigma(z)).$$

Графік сигмоїдної функції зображено на рис. 4, а графік її похідної – на рис. 5.

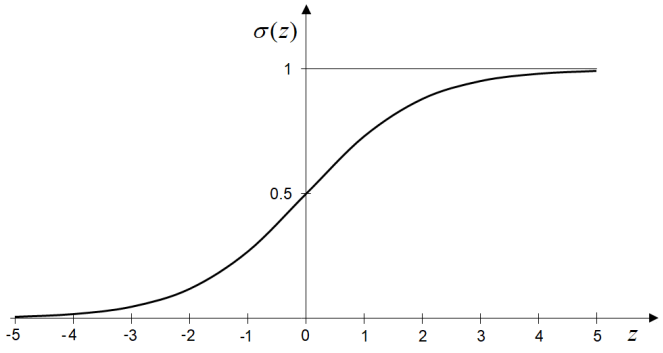


Рис. 4. Стандартна логістична функція

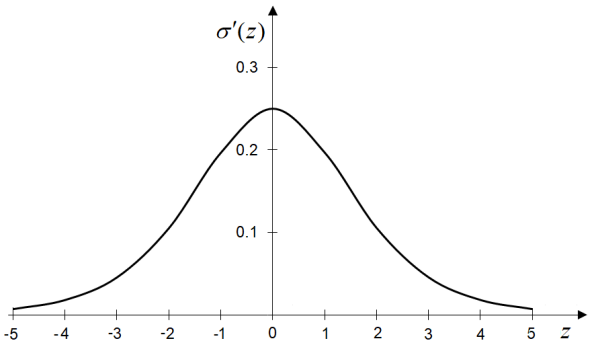


Рис. 5. Похідна логістичної функції

Нехай $z = w_0 + w_1 x_1$ і $w_0 = 0$. Тоді $\sigma(z) = \dfrac{1}{1+e^{-w_1 x_1}}$ і можна вивчати залежність сигмоїдної функції від параметрів w_1 та x_1. Сигмоїдна функція та її похідна мають оригінальний просторовий вигляд, як це видно на рис. 6 та рис. 7 відповідно. Значення функцій отримано для w_1 та x_1, які змінюються від -5 до 5 з кроком 0.5 . Поверхня сигмоїдної функції утворена S-кривими, а поверхня похідної сигмоїдної функції – дзвоноподібними кривими (рис. 5), накресленими на площинах, перпендикулярними до осі w_1, та площинах, перпендикулярними до осі x_1.

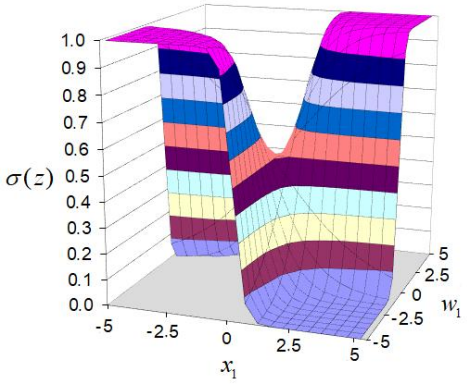


Рис. 6. Просторовий вигляд σ *-функції*

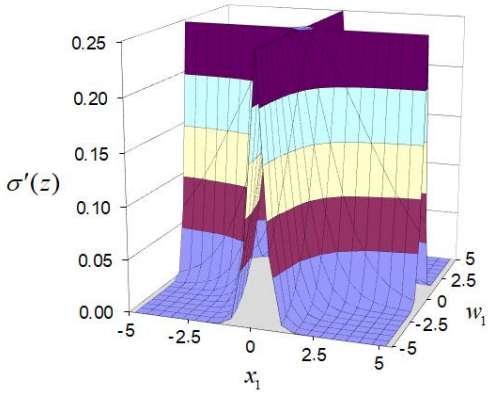


Рис. 7. Просторовий вигляд похідної σ *-функції*

Градієнт сигмоїдної функції $\sigma(z)$ за вагами w_j обчислюється за допомогою правила ланцюгового диференціювання:

$$\frac{\partial \sigma(z)}{\partial w_j} = \frac{\partial \sigma(z)}{\partial z} \times \frac{\partial z}{\partial w_j}, \quad j = 0..m. \tag{13}$$

Враховуючи (9), часткова похідна сигмоїдної функції дорівнює

$$\frac{\partial \sigma(z)}{\partial z} = \sigma(z)\left(1 - \sigma(z)\right). \tag{14}$$

Тепер обчислимо часткову похідну для $z = W^{\mathrm{T}} X$:

$$\frac{\partial z}{\partial w_j} = x_j. \tag{15}$$

Після підстановки значень часткових похідних (14) та (15) у вираз (13) отримаємо:

$$\frac{\partial \sigma(z)}{\partial w_j} = \sigma(z)\left(1 - \sigma(z)\right) x_j, \; j = 0..m.$$

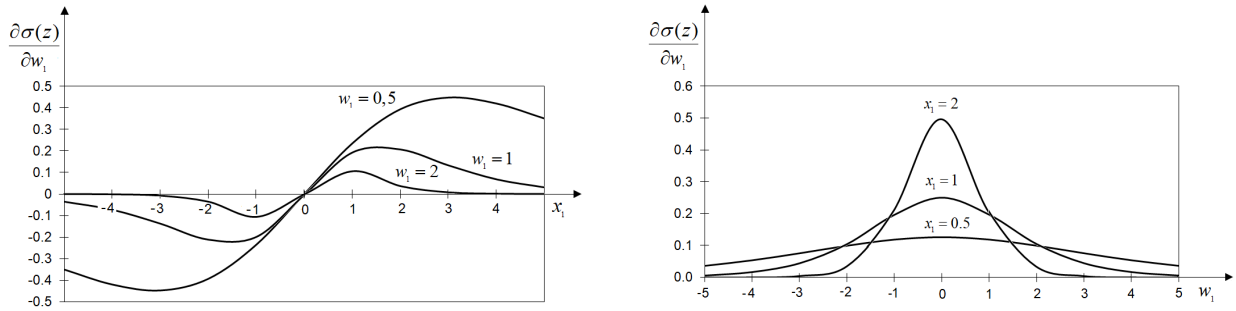
Для часткового випадку $z = w_0 + w_1 x_1$ і $w_0 = 0$ маємо:

$$\frac{\partial \sigma(z)}{\partial w_1} = \frac{1}{1+e^{-w_1 \times x_1}} \left(1 - \frac{1}{1+e^{-w_1 \times x_1}}\right) x_1 = \frac{e^{-w_1 \times x_1}}{\left(1+e^{-w_1 \times x_1}\right)^2} \times x_1. \tag{16}$$

Відповідні графіки залежності компоненти градієнта $\dfrac{\partial \sigma(z)}{\partial w_1}$ від параметрів x_1 та w_1 зображено на рис. 8 .

Оскільки похідна сигмоїдної функції завжди більша від 0 (рис. 5), то для від'ємних значень w_1 графіки також будуть мати вигляд, як на рис. 8а. Виходячи з (16), похідна домножу-

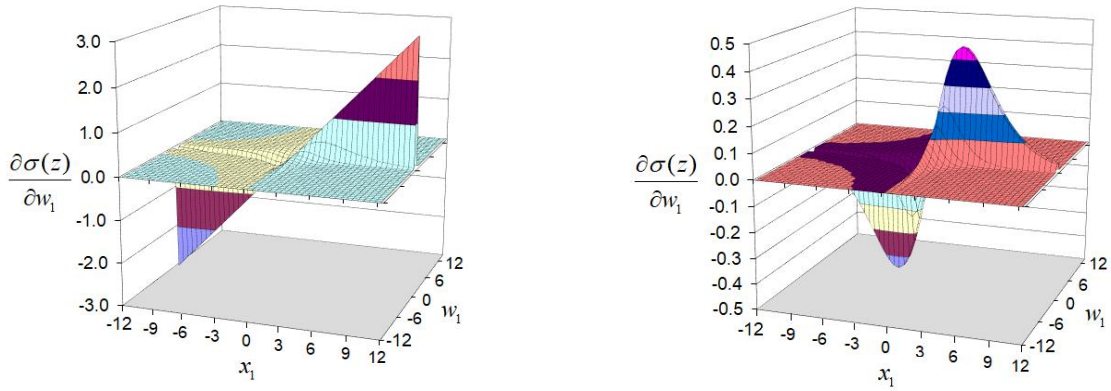
ється на значення x_1, тому для від'ємних значень x_1 графіки на рис. 8б будуть мати вигляд, симетричний відносно осі w_1.



а б) $x_1 \in \{0.5, 1, 2\}$

Рис. 8. Залежність компоненти градієнта логістичної функції від параметрів:
а – $w_1 \in \{0.5, 1, 2\}$ *б* – $x_1 \in \{0.5, 1, 2\}$

Повніша картина залежності складової $\frac{\partial s(z)}{\partial w_1}$ градієнта від параметрів подана на рис. 9 у вигляді поверхонь, які є апроксимацією комбінацій дискретних значень параметрів $x_1 \in [-12,12]$ та $w_1 \in [-12,12]$, що змінюються з кроком 0.5 .



а *б*

Рис. 9. Просторовий вигляд компоненти градієнта логістичної функції:
а – з урахуванням значення $w_1 = 0$ *; б – для значень* $w_1 \neq 0$

Загострений пік часткової похідної сигмоїдної функції на рис. 9а отримано для вагових коефіцієнтів $w_0 = w_1 = 0$, які дають лінійну залежність часткової похідної від параметра x_1 з коефіцієнтом 0.25 . Для нульових значень ваг ігноруються всі ознаки, за якими здійснюється класифікація об'єктів, тому такі ваги не є інформативними. Для порівняння, на рис. 9б зображено просторовий вигляд функції для значень $w_1 \neq 0$. Перерізи поверхні цієї функції площинами, перпендикулярними осі w_1 або x_1 відповідають графікам, зображеним на рис. 8.

Отже сигмоїдна функція має такі важливі властивості, які сприяють її використанню в оптимізаційних алгоритмах машинного навчання:

- легко обчислюється;

- забезпечує відображення діапазону $(-\infty,+\infty)$ у діапазон $(0,1)$;
- гладка, монотонно зростаюча нелінійна функція;
- у межах діапазону від -2 до 2 функція швидко зростає, мала зміна аргумента з цього діапазону призводить до суттєвого зростання функції; поза цим діапазоном зліва функція приймає близьке до нуля, а справа – близьке до одиниці значення; така властивість сигмоїдної функції робить її привабливою у системах класифікації для чіткого розділення даних на класи;
- має гладку похідну, яка запобігає різким стрибкам змінних на виході моделі.

Ці властивості дозволяють ефективно використовувати сигмоїдну функцію для прогнозування імовірностей у бінарній логістичній регресії.

Недоліком такої функції є проблема "зникнення" її градієнта для великих за модулем значень аргумента, коли градієнт приймає близькі до нуля значення. Результатом цього є сповільнення процесу навчання моделі логістичної регресії.

У випадку поліноміальної логістичної регресієї mlogit, коли є більше ніж два можливі класи для передбачення, для моделювання імовірностей належності до кожного з цих класів замість логістичної S-функції (8) використовують функцію softmax [29]:

$$P(\hat{y}=k)=\frac{e^{W_k^T X}}{\sum_{j=1}^{K} e^{W_j^T X}},$$

де $P(\hat{y}=k)$ – імовірність прогнозування класу $\hat{y}=k$; X – вектор ознак об'єкта; W_k – вектор вагових параметрів для класу k; K – загальна кількість класів.

Виведення формули сигмоїдної функції можна виконати на основі:

- моделі шансів подій;
- логістичної моделі зростання.

Модель шансів подій

Побудова сигмоїдної функції в контексті логістичної регресії може бути обґрунтована через інтерпретацію її результатів як імовірностей або шансів.

Логістична регресія часто використовується для розв'язання задачі бінарної класифікації, коли необхідно спрогнозувати імовірність того, що об'єкт належить до класу 1 (позитивний клас) або до класу 0 (негативний клас). Тобто, результат сигмоїдної функції можна інтерпретувати як імовірність відповідної події.

Нехай p – прогнозована імовірність належності об'єкта до класу 1. Тоді $(1-p)$ – імовірність протилежної події.

Щоб об'єкт можна було віднести до класу $y=1$, має виконуватись нерівність:

$$p(y=1|x) \geq p(y=0|x).$$

Тоді

$$\frac{p(y=1|x)}{p(y=0|x)}=\frac{p(y=1|x)}{1-p(y=1|x)} \geq 1,$$

де $p(y=0|x)=1-p(y=1|x) \geq 0$.

Відношення імовірності події до імовірності протилежної події

$$Odds(p)=\frac{p}{1-p}$$

називають шансом здійснення події.

Наприклад, якщо імовірність виникнення події дорівнює $p = 0.8$, то імовірність відносної події (або шанс здійснення події) буде $Odds(p) = \frac{0.8}{1 - 0.8} = 4$. Це означає, що шанс виникнення події в чотири рази більший, ніж її невиникнення.

Оскільки $p \in [0,1]$, то $Odds(p) \in [0,+\infty)$. Необхідно розширити цей діапазон до діапазону зміни ознак $x \in (-\infty,+\infty)$. Це можна зробити за допомогою функції логарифма шансів:

$$LogOdds(p) = logit(p) = \ln \frac{p}{1 - p}. \quad (17)$$

Значення функції $LogOdds(p) \in (-\infty,+\infty)$. Функція $LogOdds$ називається логарифмом шансів або логіт-функцією.

Отже, логарифм шансів вказує на логарифмічне співвідношення імовірності події та імовірності протилежної події, а його використання у логістичній регресії дозволяє моделі працювати з лінійними комбінаціями вхідних ознак, забезпечуючи при цьому результат у вигляді імовірностей у діапазоні від 0 до 1.

Логарифмічні коефіцієнти широко використовував у своїх роботах Ч. С. Пірс (C. S. Peirce) ще у кінці 19 ст. Загальновживаний термін Log Odds (логарифмічні шанси) ввів Г. А. Барнард (G. A. Barnard) у 1949 р.

У логістичній регресії припускається, що логарифм шансів $LogOdds$ є лінійною функцією від зваженої суми $z = W^{\top} X$ вхідних ознак:

$$\ln \frac{p}{1 - p} = z. \quad (18)$$

Це припущення використовується для спрощення математичної моделі логістичної регресії. При цьому залежність імовірності події p від згортки ознак z є нелінійною.

Застосувавши експоненту до обох частин рівняння (18), отримаємо:

$$e^{\ln\left(\frac{p}{1-p}\right)} = e^{z}.$$

Звідси $\frac{p}{1 - p} = e^{z}$. Розв'яжемо останнє рівняння відносно p:

$$p = \frac{e^{z}}{1 + e^{z}}.$$

Розділивши чисельник і знаменник на e^{z}, отримаємо логістичну функцію у вигляді (8). Отже, отримане значення p забезпечує лінійність логарифму шансів $LogOdds(p)$ як функції від зваженої суми $z = W^{\top} X$ вхідних ознак X.

Логістична функція (8) також називається сигмоподібною функцією або S-функцією. Зображений на рис. 4 графік сигмоїдної функції демонструє трансформацію прямої лінії $z = W^{\top} X$ в S-подібну криву.

Важливо відзначити, що сигмоїдна функція вносить нелінійність у модель, забезпечуючи можливість моделювання складних залежностей між вхідними ознаками X та імовірністю події. Однак сама вагова сума $z = W^{\top} X$ є лінійною функцією від зважених вхідних ознак. Такий підхід дозволяє легко використовувати методи оптимізації для оцінки ваг моделі. Вагові коефіцієнти W відшукуються так, щоб максимізувати імовірність правильної класифікації даних навчального набору.

Графіки залежності логарифма шансів (18) від параметрів p та z подані на рис. 10 та рис. 11 .

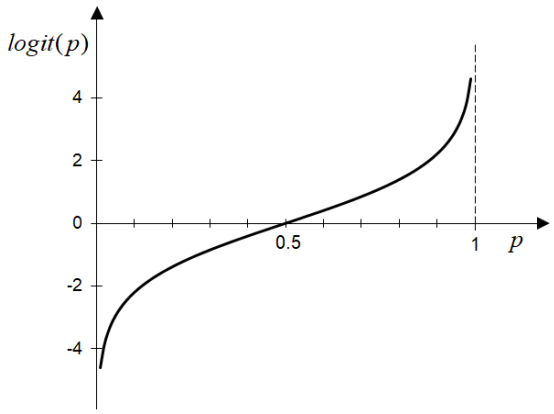


Рис. 10. Залежність логарифма шансів від p

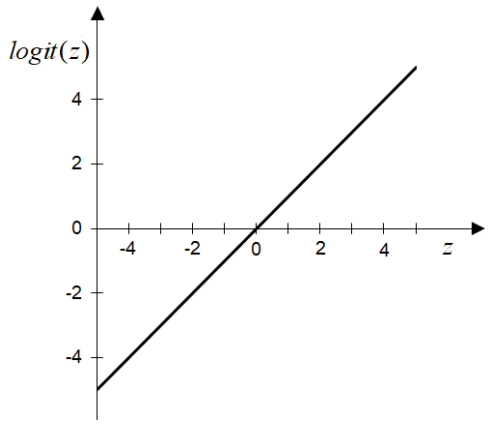
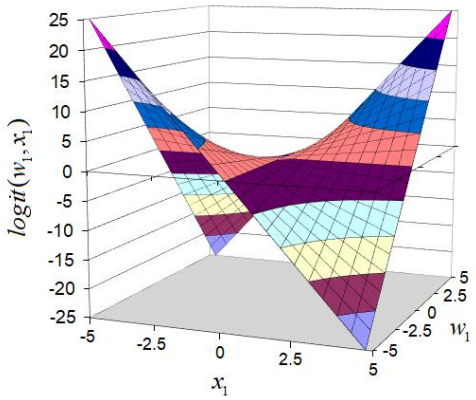


Рис. 11. Залежність логарифма шансів від z

Логарифм шансів як функція параметрів w_1 та x_1 має вигляд сідлової поверхні, твірною якої є пряма лінія $z = w_1 x_1$ (рис. 12). Функцію побудовано для значень w_1 та x_1 від -5 до 5 з кроком 0.5 .

Рис. 12. Просторовий вигляд logit*-функції*



Функція логарифма шансів є оберненою до стандартної логістичної (сигмоїдної) функції, тобто

$$logit(p) = \ln\frac{p}{1-p} = \sigma^{-1}(p) \in (-\infty, +\infty). \qquad (19)$$

З означення сигмоїдної функції маємо $p = \sigma(z) = \dfrac{1}{1+e^{-z}}$. Знайдемо обернену функцію $z = \sigma^{-1}(p)$. Після виокремлення експоненти та логарифмування виразу отримаємо:

$$z = -\ln\left(\frac{1-p}{p}\right) = \ln\left(\frac{p}{1-p}\right).$$

Підставивши знайдене значення z у лінійну залежність логарифма шансів від згортки z (18), отримаємо тотожність. Отже, справедливо (19), тобто логарифм шансів є оберненою сигмоїдною функцією, як це зображено на рис. 10.

Функція $logit(p)$ може бути використана для:

- обчислення значення сигмоїдної функції $p = \sigma(z) = \dfrac{1}{1+e^{-logit(p)}}$;

- перетворення імовірності прогнозу з виходу моделі у значення $z = \sigma^{-1}(p)$, яке може бути більш зрозумілим та інтерпретованим у термінах параметрів моделі;
- визначення, які змінні (ознаки) мають статистично значимий вплив на імовірність виникнення події та напрямок цього впливу.

Останнє вказує на те, що вивчати взаємозв'язки між змінними моделі та імовірностями можна безпосередньо за допомогою аналізу вагових коефіцієнтів.

Це випливає з наступного. У логістичній регресії модель логарифмічного шансу подій задається у вигляді:

$$\ln\left(\frac{p}{1-p}\right) = W^{\mathrm{T}} X \,. \tag{20}$$

Візьмемо часткову похідну обох частин рівняння (20) по x_i. Оскільки $\frac{\partial W^{\mathrm{T}} X}{\partial x_i} = w_i$, то

$$\frac{\partial \ln\left(\frac{p}{1-p}\right)}{\partial x_i} = w_i \,. \tag{21}$$

На основі (21) формально можна стверджувати, що коефіцієнти w_i перед незалежними змінними x_i показують, наскільки змінюється логарифм відношення шансів подій $\ln\left(\frac{p}{1-p}\right)$ в залежності від зміни незалежної змінної x_i.

Справедливість (21) також можна перевірити, використавши ланцюжок правил диференціювання:

$$\frac{\partial \ln\left(\frac{p}{1-p}\right)}{\partial x_i} = \frac{\partial \ln\left(\frac{p}{1-p}\right)}{\partial p} \times \frac{\partial p}{\partial z} \times \frac{\partial z}{\partial x_i} \,, \tag{22}$$

де $p = \sigma(z) = \frac{1}{1+e^{-z}}$ – сигмоїдна функція, $z = W^{\mathrm{T}} X$.

Знайдемо значення кожного співмножника:

1. $\frac{\partial \ln\left(\frac{p}{1-p}\right)}{\partial p} = \frac{1-p}{p} \times \frac{\partial}{\partial p}\left(\frac{p}{1-p}\right) = \frac{1-p}{p} \times \frac{1\times(1-p) - p(-1)}{(1-p)^2} = \frac{1}{p(1-p)}$.
2. $\frac{\partial p}{\partial z} = p(1-p)$.
3. $\frac{\partial z}{\partial x_i} = w_i$.

Підставивши значення співмножників у (22), отримаємо:

$$\frac{\partial \ln\left(\frac{p}{1-p}\right)}{\partial x_i} = \frac{1}{p(1-p)} \times p(1-p) w_i = w_i \,, \tag{23}$$

що співпадає з (21).

Запишемо (23) у вигляді різницевого рівняння:

$$\frac{\Delta \ln(Odds)}{\Delta x_i} = w_i, \tag{24}$$

де $Odds = \frac{p}{1-p}$ – шанс події.

Оскільки нам потрібно встановити загальну тенденцію впливу зміни ознаки x_i на логарифм шансів події, то таке спрощення допустиме з певною точністю, залежною від величини Δx_i.

Якщо у (24) задати $\Delta x_i = 1$, то $\Delta \ln(Odds) = w_i$, тобто одинична зміна ознаки x_i призведе до зміни логарифма шансу події на величину w_i.

Тепер визначимо вплив зміни ознаки x_i безпосередньо на зміну шансу події.

Нехай $\Delta \ln(Odds) = \ln(Odds_2(x_i + \Delta x_i)) - \ln(Odds_1(x_i))$. За правилом різниці логарифмів, отримаємо:

$$\ln\left(\frac{Odds_2}{Odds_1}\right) = w_i \Delta x_i. \tag{25}$$

Застосуємо експоненту з обох сторін рівняння (25):

$$K = \frac{Odds_2}{Odds_1} = e^{w_i \Delta x_i}.$$

Якщо $\Delta x_i = 1$, то за кожну одиницю зміни незалежної змінної x_i відношення шансів події змінюється в e^{w_i} разів.

Так, якщо ваговий коефіцієнт $w_i = 0.5$, то збільшення незалежної змінної x_i на 1 призведе збільшення шансу події в $e^{0.5} \approx 1.65$ разів.

Змінюючи значення ваги w_1, отримаємо інші значення коефіцієнта відношення шансів події для $\Delta x_i = 1$:

w_1	-3	-2	-1	0	1	2	3
K	0.05	0.14	0.37	1	2.72	7.39	20.09

Зростання ваги w_1 призводить до експоненціального зростання коефіцієнта відношення шансів K.

Логістична модель зростання

Для опису S-подібного зростання може бути використано декілька різних рівнянь, але найбільшу популярність отримало найпростіше з них – так зване логістичне.

Вперше запропоноване як модель зростання народонаселення в 1838 р. П. Ф. Ферхюльстом (P. F. Verhulst, 1838), воно було перевідкрито заново американськими дослідниками Р. Перлом і Л. Рідом (R. Pearl and L. J. Reed) в 1920 р., які згодом визнали пріоритет Ферхюльста.

Логістичну функцію можна визначити з розв'язку диференціального рівняння, яке характеризує логістичну модель зростання і має такий вигляд:

$$\frac{dp}{dt} = p(1-p), \tag{26}$$

де p – логістична функція, $\frac{dp}{dt}$ – її похідна за часом.

Логістичне рівняння зазвичай використовується для опису динаміки популяцій в біології, для прогнозування ринків в економіці, для машинного навчання та в інших галузях.

Нехай зростання популяції обмежено ресурсами та іншими факторами. У цьому випадку змінна p відповідає за зростання популяції. У контексті логістичної регресії значення p визначає імовірність такого зростання. Більшому p відповідає більший внесок у зростання популяції. Співмножник $1-p$ відповідає за обмеження росту популяції. Наближення p до 1 вказує на те, що популяція практично зайняла всі доступні ресурси, і тоді $1-p$ буде малим, обмежуючи зростання.

Загалом, добуток $p(1-p)$ відображає взаємодію між відносним рівнем зростанням популяції p та обмеженням її росту $1-p$. Якщо p близьке до 0, то зростання популяції буде прискореним. Однак, якщо p наближається до 1, то $1-p$ наближається до 0, що обмежує зростання популяції.

Логістична модель росту дозволяє враховувати асимптотичне насичення, коли популяція перестає зростати через обмеження ресурсів. Так, при $p=0$ або $p=1$ швидкість зміни $\frac{dp}{dt}$ буде нульовою, що вказує на асимптотичного насичення популяції.

Розв'язком рівняння (26) є логістична функція, що характеризує обмежене зростання або насичення популяції. У детальнішому формулюванні модель зростання має додаткові параметри, такі як початкова кількість популяції і розмір популяції, при якому наступає насичення.

Узагальнюючи, логістична модель росту надає математичний підхід до моделювання обмеженого зростання популяцій, враховуючи ефекти обмеження ресурсів на рівень зростання.

Розв'яжемо диференціальне рівняння (26) методом відокремлення змінних. Для цього проінтегруємо обидві частини рівняння:

$$\int \frac{1}{p(1-p)}dp = \int dt. \tag{27}$$

Враховуючи, що $\frac{1}{p(1-p)} = \frac{1}{p} + \frac{1}{1-p} = \frac{1}{p} - \frac{-1}{1-p}$, після інтегрування лівої частини рівняння (27) отримаємо:

$$\int \left(\frac{1}{p} - \frac{-1}{1-p}\right)dp = \ln|p| - \ln|1-p| + C_1.$$

Інтегрування правої частини (27) дає:

$$\int dt = t + C_2.$$

Після роздільного інтегрування та використання властивості різниці логарифмів спростимо вираз до такого:

$$\ln\left|\frac{p}{1-p}\right| = t + C, \tag{28}$$

де $C = C_2 - C_1$ – константа інтегрування.

Враховуючи, що $p \in (0,1)$, замість (28) запишемо:

$$\ln\frac{p}{1-p} = t + C.$$

Після цього візьмемо експоненту з обох боків отриманого рівняння:

$$\frac{p}{1-p} = e^{t+C}.$$

Розв'язавши останнє рівняння відносно p, отримаємо логістичну функцію:

$$p = \frac{C\phi^t}{1 + C\phi^t},$$

де $C\phi = e^C$.

Якщо задати $C\phi = 1$, що має місце при $C = 0$, то після ділення чисельника і знаменника на e^t остаточна формула логістичної функції матиме вигляд:

$$p(t) = \frac{1}{1 + e^{-t}}.$$

Покладаючи тут $t = z$, отримаємо логістичну функцію у вигляді (8).

Функція втрат логістичної регресії

Логістична регресія – це різновид контрольованої класифікації, у якій відома актуальна (правильна) мітка класу $y_i \in \{0,1\}$ для кожного спостереження X_i. Використовуючи сигмоїдну функцію, логістична модель виробляє імовірність $p_i = \sigma\left(W^{\mathrm{T}} X_i\right)$ прогнозу мітки класу $\hat{y}_i$. Необхідно визначити такі значення параметрів W моделі, які сформують імовірність p_i таку, що прогнозований клас $\hat{y}_i$ буде імовірнісно ближчим до справжнього y_i для кожного навчального спостереження.

Для цього необхідно визначити показник того, наскільки близький прогноз p_i до справжньої мітки класу y_i.

Метою логістичної регресії є мінімізація середньоквадратичної похибки, яка визначається різницею між прогнозованим і актуальним значенням параметра:

$$F = \frac{1}{n}\sum_{i=1}^{n}\left(\sigma\left(W^{\mathrm{T}} X_i\right) - y_i\right)^2 \to \min_W, \tag{29}$$

$y_i \in \{0,1\}$ – актуальне значення мітки класу.

Застосування такої середньоквадратичної похибки (29) дасть неопуклий графік з декількома локальними мінімумами. Це може значно утруднити пошук глобального мінімуму і призвести до помилки класифікації об'єктів за їх параметрами.

Для вирішення цієї проблеми використаємо іншу функцію втрат для логістичної регресії, яка називається логарифм втрат, і яка виводиться з методу оцінки максимальної правдоподібності.

Метод максимальної правдоподібності (maximum likelihood method) – це метод обчислення невідомих параметрів статистичної моделі шляхом максимізації функції правдоподібності, для яких отримані за домогою моделі дані є найбільш імовірними. Для логістичної регресії такими параметрами є ваги W ознак X вибірки об'єктів. Застосування критерію максимальної правдоподібності дозволяє отримати ефективні оцінки параметрів моделі, які мають добрі статистичні властивості. Цей метод є основоположним у статистиці та дослідженні параметричних моделей, включаючи логістичну регресію.

Витоки використання методу максимальної правдоподібності прослідковуються у роботах К. Ф. Гауса (J. C. F. Gauss), П.-С. Лапласа (P. S. Laplace) та інших. Значний внесок в аналіз та популяризацію цього методу зробив Р. Фішер (R. Fisher) на початку XX ст.

Критерій максимальної правдоподібності для оцінки параметрів логістичної регресії розробив Д. Кокс (D. Cox) у 1958 р., що стало важливим кроком у розвитку цього методу аналізу даних та відкрило шлях для застосування логістичної регресії у різних галузях.

Розвинули та удосконалили метод максимальної правдоподібності для оцінки параметрів моделі логістичної регресії Дж. Нельдер (J. A. Nelder) та Р. Уеддерберн (R. W. M. Wedderburn) у

1972 році. Одним із їх важливих досягнень була розробка методу псевдомаксимальної правдоподібності (Pseudo Maximum Likelihood Estimation, PMLE) для оцінки параметрів логістичних моделей. Цей метод є ефективним для оцінювання параметрів у логістичних моделях зі складною структурою даних, таких як кластеризація, кореляція чи інші випадкові залежності.

Виведення функції правдоподібності для логіт-регресії здійснюється на основі припущення про розподіл випадкової бінарної змінної (0 або 1) при відомих значеннях вхідних даних і параметрів моделі.

Для незалежних експериментів бінарної класифікації кожен вихідний клас можна інтерпретувати як значення випадкової змінної Бернуллі $y \in \{0,1\}$. Випадковий експеримент, результати якого мають тільки два значення, що отримуються з імовірностями p та $q = 1 - p$ відповідно, називається випробуванням Бернуллі. Якщо на вхід класифікатора було подано деяке значення $X = x$, то імовірність того, що на виході отримаємо значення $Y = y \in \{0,1\}$ визначається розподілом Бернуллі:

$$P[Y = y \mid X = x] = p^y (1 - p)^{1-y}.$$

Оскільки $p = \sigma\left(W^\top X\right)$, де $\sigma\left(W^\top X\right)$ – сигмоїдна функція, то

$$P[Y = y \mid X = x] = \sigma\left(W^\top X\right)^y \left(1 - \sigma\left(W^\top X\right)\right)^{1-y},$$

де $W, X \in R^{m+1}$, $y \in \{0,1\}$.

Імовірність спільного настання незалежних подій в експерименті бінарної класифікації визначається добутком імовірностей настання кожної окремої події. З урахуванням цього для n незалежних випробувань матимемо таке значення функції імовірності:

$$L(W) = \prod_{i=1}^{n} \sigma\left(W^\top X_i\right)^{y_i} \left(1 - \sigma\left(W^\top X_i\right)\right)^{1-y_i}, \tag{30}$$

де $y_i \in \{0,1\}$, $W, X_i \in R^{m+1}$, $W^\top X_i$ – скалярний добуток двох векторів, $\sigma\left(W^\top X_i\right) = p_i$.

Потрібно знайти таке значення ваг W, яке максимізує цю функцію імовірності. Для полегшення розрахунків застосуємо логарифм з обох сторін виразу (30), враховуючи, що логарифм добутку дорівнює сумі логарифмів співмножників:

$$\log L(W) = \sum_{i=1}^{n} \left[y_i \times \log \sigma\left(W^\top X_i\right) + (1 - y_i) \times \log\left(1 - \sigma\left(W^\top X_i\right)\right) \right] \to \max_W, \tag{31}$$

де log – загальний логарифм з додатною дійсною основою (крім 1), у логістичній регресії використовується натуральний логарифм.

Суму (31) можна усереднити для порівняльності результатів класифікації на різних за обсягом вибірках даних. Така нормалізація не змінить значення оптимальних ваг. З урахуванням цього та після зміни знаку функції (31) на протилежний отримаємо:

$$F(W) = -\log L(W) = -\frac{1}{n} \sum_{i=1}^{n} \left[y_i \times \log \sigma\left(W^\top X_i\right) + (1 - y_i) \times \log\left(1 - \sigma\left(W^\top X_i\right)\right) \right] \to \min_W, \tag{32}$$

де $y_i \in \{0,1\}$ позначає фактичну мітку класу для ознак X_i; $\sigma\left(W^\top X_i\right) = P[y_i = 1 \mid X = x_i] = p_i$ – це імовірність настання події $y_i = 1$; $\left(1 - \sigma\left(W^\top X_i\right)\right) = P[y_i = 0 \mid X = x_i] = (1 - p_i)$ – імовірність настання події $y_i = 0$.

Отриману функцію називають функцією логарифмічних втрат *LogLoss*, похибкою, логарифмічною функцією імовірності або сумою логарифмічної умовної імовірності.

Позначимо $-L_i(W) = -\left[y_i \times \log s\left(W^T X_i\right) + (1-y_i) \times \log\left(1-s\left(W^T X_i\right)\right)\right]$.

Нехай $y_i = 0$, тоді $-L_i(W;0) = -\log\left(1-s\left(W^T X_i\right)\right)$. Якщо $s\left(W^T X_i\right) = 0$, то $-L_i(W;0) = 0$. Якщо ж $s\left(W^T X_i\right) \to 1$, то $-L_i(W;0) \to \infty$.

Тепер нехай $y_i = 1$, тоді $-L_i(W;1) = -\log s\left(W^T X_i\right)$. Якщо $s\left(W^T X_i\right) \to 0$, то $-L_i(W;0) \to \infty$. Якщо ж $s\left(W^T X_i\right) = 1$, то $-L_i(W;0) = 0$.

Графіки складових $-L_i(W;0)$ та $-L_i(W;1)$ функції втрат $-L_i(W)$, отриманих відповідно при $y_i = 0$ та $y_i = 1$, зображено на рис. 13. Як видно на рис. 13, функції $-L_i(W;0)$ та $-L_i(W;1)$ є опуклими та гладкими.

У загальному випадку вхідна вибірка містить об'єкти з обох класів. Тоді, враховуючи, що $y_i \in \{0,1\}$, функцію втрат (32) запишемо у вигляді:

$$F(W) = -\log L(W) = -\left(\frac{1}{n_1}\sum_{y_i=0} L_i(W;0) + \frac{1}{n_2}\sum_{y_i=1} L_i(W;1)\right) = -\left(L(W;0) + L(W;1)\right) \to \min_W, \qquad (33)$$

де $n = n_1 + n_2$ – загальний обсяг вхідної вибірки; n_1 – кількість об'єктів вхідної вибірки, для яких $y_i = 0$; n_2 – кількість об'єктів вхідної вибірки, для яких $y_i = 1$.

Функції $L(W;0)$ та $L(W;1)$ в (33), як суми гладких опуклих функцій, будуть мати такі ж властивості. Крім того, функція логарифмічних втрат $F(W)$, як сума монотонно спадної опуклої функції $L(W;1)$ та монотонно зростаючої опуклої функції $L(W;0)$, буде унімодальною, вона матиме один мінімум. Ілюстративний вигляд функції середніх втрат (33) для добре збалансованої вибірки зображено на рис. 14.

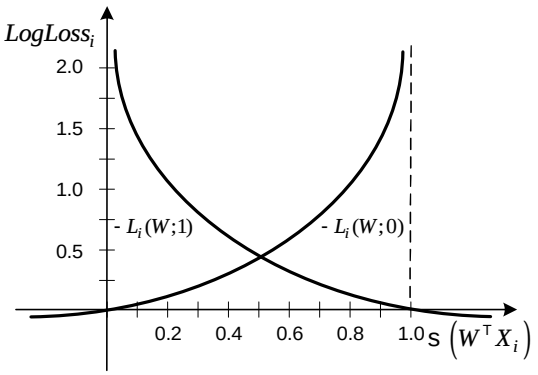


Рис. 13. Складові функції втрат

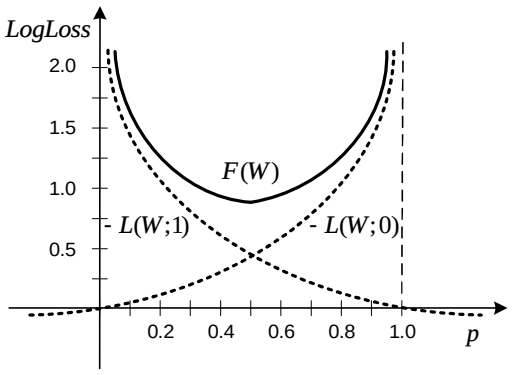


Рис. 14. Функція втрат для збалансованої вибірки

Для пошуку мінімуму унімодальної функції зручно використати метод градієнтного спуску. Перед застосуванням цього методу необхідно виконати належну підготовку вхідних даних.

Регуляризація функції втрат

Регуляризація – це накладання штрафу на модель за її надмірну складність. Основними показниками складності моделі логістичної регресії є кількість характеристичних ознак об'єктів, які використовуються для прогнозування класу, наявність кореляції між вхідними даними, розмір навчального набору даних, вигляд функції втрат.

Показником складності моделі можна вважати величини ваг ознак об'єктів, сформовані під час навчання моделі. Зазвичай складні регресійні моделі характеризуються великими значеннями таких ваг. Великі значення ваг можуть свідчити про те, що модель враховує багато деталей та взаємозв'язків у вхідних даних, що може призвести до перенавчання. Малі значення ваг можуть вказувати на недонавчання моделі, коли вона недостатньо враховує важливі закономірності в даних.

Оптимальна модель має баланс ваг, які адекватно відображають важливі закономірності в даних без надмірного перенавчання чи недонавчання. Регуляризація функції втрат може контролювати складність моделі, обмежуючи аномальні відхилення ваг та тримаючи їх у певному діапазоні значень.

Для контролю складності моделі та запобігання перенавчанню використовуються такі основні методи регуляризації: L1 (Lasso Regression), L2 (Ridge Regression) та Elastic Net.

У методі регуляризації L1 до функції втрат додається штрафний член, що дорівнює сумі абсолютних значень ваг параметрів моделі:

$$F_{lasso}(W) = -\log L(W) + \lambda \sum_{j=1}^{m} |w_j| \to \min_W .$$

Параметр $\lambda \in (0,\infty)$ контролює величину штрафу. У ході мінімізації функції втрат малі значення λ можуть призвести до перенавчання, коли вагові коефіцієнти значно зростають. Великі значення λ призводять до недонавчання, коли ваги зменшуються аж до нуля. Тим самим можна позбавитися від непотрібних для моделі ознак об'єкта. Вибір оптимального значення λ можна виконати, аналізуючи метрики якості навчання моделі.

Штраф L1 вперше застосували Ф. Сантоза (F. Santosa) та В. Саймс (W. Symes) у 1986 році у роботі з геофізичних досліджень.

У методі регуляризації L2 до функції втрат додається штрафний член, що дорівнює квадрату суми значень ваг параметрів моделі:

$$F_{ridge}(W) = -\log L(W) + \lambda \sum_{j=1}^{m} w_j^2 \to \min_W .$$

Параметр λ у моделі L2 працює аналогічно як у моделі L1, але вагові коефіцієнти ніколи не приймуть нульових значень, хоча можуть бути досить малими.

Регуляризація L2 була винайдена незалежно у багатьох різних контекстах. Вона стала широко відома із робіт Д. Л. Філіпса (D. L. Phillips, 1962) та А. Тихонова (A. Tikhonov, 1963) завдяки застосуванню до інтегральних рівнянь.

Обидва методи регуляризації L1 та L2 допомагають контролювати складність моделі та покращують її узагальнюючу здатність на нових даних, зменшуючи ризик перенавчання. Вибір між L1 та L2 регуляризацією залежить від специфіки задачі та вимог до моделі.

Два види регуляризації L1 та L2 одночасно використовуються в регресійній моделі логістичної еластичної сітки (Logistic Elastic Net, LEN) з такою функцією втрат:

$$F_{len}(W) = -\log L(W) + \lambda \sum_{j=1}^{m} \left(\alpha w_j^2 + (1-\alpha)|w_j|\right) \to \min_W ,$$

де $\alpha \in (0,1)$.

Регуляризація методом еластичної сітки запропонована у роботі Х. Зоу (H. Zou) та Т. Хасті (T. Hastie) у 2005 р.

Логістична еластична сітка поєднує у собі кращі сторони обох методів, вона є надійнішою за кожен з них окремо, особливо при роботі з великими вибірками даних, може опрацьовувати корельовані змінні та змінні з різними масштабами. Подібно до методу L1 вона може зводити вагові коефіцієнти нерелевантних ознак об'єктів до нуля, що призводить до моделі з меншою кількістю змінних, які простіше інтерпретувати, та з меншою схильністю до перенавчання. Ця регресія добре працює із зашумленими, неповними або пошкодженими даними.

Висновки

У статті виконано систематизацію та математичне обґрунтування методу логістичної регресії для бінарної класифікації даних, що полегшує розуміння засадничих основ машинного навчання.

Проведений аналіз літературних джерел показав популярність методу логістичної регресії для прогнозування імовірностей подій та бінарної класифікації у різних галузях діяльності завдяки надійності результатів, ефективності та простоті практичної реалізації.

Показано відмінність логістичної регресії від методу лінійної регресії, яка полягає у здатності логіт-регресії звужувати область значення залежної змінної до діапазону від 0 до 1, елементи якого можна інтерпретувати у термінах імовірностей. Відмічено, що таку звужуючу властивість має сигмоїдна функція логістичної регресії, яка перетворює числові або закодовані числами категоріальні ознаки об'єкта, які можуть приймати значення на усій осі дійсних чисел, у значення з діапазону від 0 до 1. На основі цієї властивості показана спорідненість логіт-регресії з методом пробіт-регресії.

Продемонстровано виведення сигмоїдної функції логістичної регресії двома способами: на основі моделі логарифма шансів подій та моделі логістичного зростання популяції.

Описано спосіб побудови функції втрат логістичної регресії методом максимальної правдоподібності, що дозволяє перейти від багатоекстремальної до унімодальної оптимізаційної задачі.

Застосування цих фундаментальних функцій логістичної регресії для навчання і тестування даних розглядається у другій частині цієї роботи.

Подана у статті інформація зможе покращити розуміння логістичної регресії як методу машинного навчання в освітньому процесі, наукових дослідженнях та у практичній побудові простих та ефективних систем аналізу і класифікації даних у багатьох галузях діяльності людини, від медицини до фінансів.

Список літератури

1. Басюк, Т. М., Литвин, В. В., Захарія, Л. М., & Кунанець, Н. Е. (2019). *Машинне навчання: навч. посіб.* Львів: Видавництво "Новий Світ – 2000".

2. Kumar, P. P., Vairachilai, S., Sirisha, P., & Mohanty, S. N. (2021). *Recommender Systems: Algorithms and Applications*. Boca Raton, London, New York: CRC Press. DOI: https://doi.org/10.1201/9780367631888.

3. Haghighi, M. H. Z. (2023). Analyzing astronomical data with machine learning techniques. *Astronomical & Astrophysical Transactions, 33*(3), 323–336. DOI: https://doi.org/10.48550/arXiv. 2302.11573.

4. Матвійчук, А., & Артюх, О. (2022) Оцінювання кредитних ризиків малих і середніх підприємств методами інтелектуального аналізу даних. *Наукові записки Національного університету Острозька академія, Серія "Економіка": науковий журнал, 26*(54), 114–120. DOI: 10.25264/2311-5149-2022-26(54)-114-120.

5. Головач, К. С., Оліфір, І. А., & Головач О. П. (2022). Розпізнавання кризових явищ та методика їх виявлення. *Бізнес-навігатор: наук.-виробнич. журнал, 1*(68), 155–159. DOI: https://doi.org/10.32847/business-navigator.68–24.

6. Wang, Z., Sun, X., Wang, B., Shi, S., & Chen, X. (2023). Lasso-Logistic regression model for the identification of serum biomarkers of neurotoxicity induced by strychnos alkaloids. *Toxicology Mechanisms and Methods, 33*(1), 65–72. DOI: https://doi.org/10.1080/15376516.2022.2086088.

7. Nottingham, Q. J., Birch, J. B., & Bodt, B. A. (2000). Local logisitic regression an application to army penetration data. *Journal of Statistical Computation and Simulation, 66*(1), 35–50, DOI: https://doi.org/10.1080/00949650008812010.

8. Madani, N., Maleki, M., & Soltani-Mohammadi, S. (2022). Geostatistical modeling of heterogeneous geoclusters in a copper deposit integrated with multinomial logistic regression: An exercise on resource estimation. *Ore Geology Reviews, 150*, 105132, 1–22. DOI: https://doi.org/10.1016/ j.oregeorev.2022.

9. Yaseliani, M., & Khedmati, M. (2023). Prediction of Heart Diseases Using Logistic Regression and Likelihood Ratios. *International Journal of Industrial Engineering & Production Research, 34*(1), 1–15. DOI: https://doi.org/10.22068/ijiepr.34.1.5.

10. Hu, X., Luo, H., Guo, M., & Wang, J. (2022). Ecological technology evaluation model and its application based on Logistic Regression. *Ecological Indicators, 136* (108641), 1–11. DOI: https://doi.org/10.1016/j.ecolind.2022.108641.

11. Зомчак, Л. М., & Старчевська, І. М. (2022). Моделювання економічного зростання України за допомогою логістичної регресії. *Науковий вісник Полтавського університету економіки і торгівлі. Серія "Економічні науки", 2*(106), 78–83. DOI: https://doi.org/10.37734/2409-6873-2022-2-11.

12. Ahn, Y. H., Park, K. R., Kim, D. H., & Cho, H. J. (2021). Logistic Regression Algorithm-Based Product Recommendation System Model. *Journal of Computational and Theoretical Nanoscience, 18*(5), 1429–1435. DOI: https://doi.org/10.1166/jctn.2021.9619.

13. Hernández, J., Etemadi, A., Roberts-Baca, S., & Muthyapu, V. K. (2021, April). Developing a logistic regression method for valuation of grid-level energy storage systems. *In 2021 IEEE Conference on Technologies for Sustainability (SusTech)*, 1–8. DOI: https://doi.org/10.1109/SusTech51236.2021. 9467419.

14. Tan, K. L., Lee, C. P., & Lim, K. M. (2023). A survey of sentiment analysis: Approaches, datasets, and future research. *Applied Sciences*, *13*(7), 4550. DOI: https://doi.org/10.3390/app13074550.

15. Indu, R., & Dimri, S. C. (2023). Detecting Spam E-mails with Content and Weight-Based Binomial Logistic Model. *Journal of Web Engineering, 22*(7), 939–959. DOI: https://doi.org/10.13052/jwe1540-9589.2271.

16. Berezka, K. M., Kovalchuk, O. Ya., Banakh, S. V., Zlyvko, S. V., & Hrechaniuk, R. (2022). A Binary Logistic Regression Model for Support Decision Making in Criminal Justice. *Folia Oeconomica Stetinensia, 22*(1), 1–17. DOI: https://doi.org/10.2478/foli-2022-0001.

17. Zhang, L. (2022). Smart Marketing Data Collection and Analysis based on Logistic Regression Algorithm. *3rd International Conference on Smart Electronics and Communication (ICOSEC), Trichy, India*, 1611–1614. DOI: https://doi.org/10.1109/ICOSEC54921.2022.9951974.

18. Fayaz, S. A., Zaman, M., & Butt, M. A. (2021). An application of logistic model tree (LMT) algorithm to ameliorate Prediction accuracy of meteorological data. *International Journal of Advanced Technology and Engineering Exploration, 8*(84), 1424–1440. DOI: https://doi.org/10.19101/IJATEE. 2021.874586.

19. Niu, L. (2020). A review of the application of logistic regression in educational research: common issues, implications, and suggestions. *Educational Review, 72*(1), 41–67. DOI: https://doi.org/10.1080/00131911.2018.1483892.

20. Rivera, P. P., & Garashchuk, A. (2023). Strategic partner election: proposal for a Binary Logistic Model for the European Union. *Humanities and Social Sciences Communications, 10*(1), 1–13. DOI: https://doi.org/10.1057/s41599-023-02121-y.

21. Velu, A. (2021). Application of logistic regression models in risk management. *International Journal of Innovations in Engineering Research and Technology*, *8*(04), 251–260. Retrieved from https://repo.ijiert.org/index.php/ijiert/article/view/2594.

22. Gai, R., & Zhang, H. (2023). Prediction model of agricultural water quality based on optimized logistic regression algorithm. *EURASIP Journal on Advances in Signal Processing, 21*, 1–14, DOI: https://doi.org/10.1186/s13634-023-00973-9.

23. Chen, Q. (2022). Research on identifying psychological health problems of college students by logistic regression model based on data mining. *Applied Mathematics and Nonlinear Sciences, 8*(1), 2253–2262. DOI: https://doi.org/10.2478/amns.2021.2.00195.

24. Borucka, A. (2020). Logistic regression in modeling and assessment of transport services. *Open Engineering, 10*, 26–34. DOI: https://doi.org/10.1515/eng-2020-0029.

25. Kang, R. (2020). Using logistic regression for persona segmentation in tourism: A case study. Social Behavior and Personality: an international *journal, 48*(4), 1–16. DOI: https://doi.org/ 10.2224/sbp.8793.

26. Christensen, R. (1997). *Log-Linear Models and Logistic Regression*. Springer. ISBN 10: 0387982477 / ISBN 13: 9780387982472.

27. Hosmer, D. W., & Lemeshow, S. (2000*). Applied Logistic Regression*. John Wiley & Sons, Inc. DOI: https://doi.org/10.1002/0471722146.

28. Hilbe, J. M. (2009). *Logistic Regression Models (1st ed.)*. Chapman and Hall/CRC. DOI: https://doi.org/10.1201/9781420075779.

29. Cramer, J. S. (2003). The standard multinomial logit model. *In Logit Models from Economics and Other Fields, Chapter 7*. Cambridge: Cambridge University Press, 104–125. DOI: https://doi.org/10.1017/CBO9780511615412.008.

References

1. Basyuk, T. M., Lytvyn, V. V., Zakharia, L. M., & Kunanets, N. E. (2019). *Machine learning: a study guide (in Ukrainian)*. Lviv: "Novyy Svit – 2000" Publishing House.

2. Kumar, P. P., Vairachilai, S., Sirisha, P., & Mohanty, S. N. (2021). *Recommender Systems: Algorithms and Applications*. Boca Raton, London, New York: CRC Press. DOI: https://doi.org/10.1201/9780367631888.

3. Haghighi, M. H. Z. (2023). Analyzing astronomical data with machine learning techniques. *Astronomical & Astrophysical Transactions, 33*(3), 323–336. DOI: https://doi.org/10.48550/arXiv. 2302.11573.

4. Matviychuk, A., & Artyukh, O. (2022) Assessment of credit risks of small and medium-sized enterprises by methods of intellectual data analysis (in Ukrainian). *Scientific Notes of the National University of Ostroh Academy, "Economics" Series: scientific journal, 26*(54), 114–120. DOI: 10.25264/2311-5149-2022-26(54)-114-120.

5. Golovach, K. S., Olifir, I. A., & Golovach, O. P. (2022). Recognition of crisis phenomena and methods of their detection (in Ukrainian). *Business navigator: science and production. magazine*, 1(68), 155–159. DOI: https://doi.org/10.32847/business-navigator.68-24.

6. Wang, Z., Sun, X., Wang, B., Shi, S., & Chen, X. (2023). Lasso-Logistic regression model for the identification of serum biomarkers of neurotoxicity induced by strychnos alkaloids. *Toxicology Mechanisms and Methods, 33*(1), 65–72. DOI: https://doi.org/10.1080/15376516.2022.2086088.

7. Nottingham, Q. J., Birch, J. B., & Bodt, B. A. (2000). Local logisitic regression an application to army penetration data. *Journal of Statistical Computation and Simulation, 66*(1), 35–50, DOI: https://doi.org/10.1080/00949650008812010.

8. Madani, N., Maleki, M., & Soltani-Mohammadi, S. (2022). Geostatistical modeling of heterogeneous geo-clusters in a copper deposit integrated with multinomial logistic regression: An exercise on resource estimation. *Ore Geology Reviews, 150*, 105132, 1–22. DOI: https://doi.org/10.1016/ j.oregeorev.2022.

9. Yaseliani, M., & Khedmati, M. (2023). Prediction of Heart Diseases Using Logistic Regression and Likelihood Ratios. *International Journal of Industrial Engineering & Production Research, 34*(1), 1–15. DOI: https://doi.org/10.22068/ijiepr.34.1.5.

10. Hu, X., Luo, H., Guo, M., & Wang, J. (2022). Ecological technology evaluation model and its application based on Logistic Regression. *Ecological Indicators, 136* (108641), 1–11. DOI: https://doi.org/10.1016/j.ecolind.2022.108641.

11. Zomchak, L. M., & Starchevska, I. M. (2022). Modeling the economic growth of Ukraine using logistic regression (in Ukrainian). *Scientific Bulletin of the Poltava University of Economics and Trade. Series "Economic Sciences", 2*(106), 78–83. DOI: https://doi.org/10.37734/2409-6873-2022-2-11.

12. Ahn, Y. H., Park, K. R., Kim, D. H., & Cho, H. J. (2021). Logistic Regression Algorithm-Based Product Recommendation System Model. *Journal of Computational and Theoretical Nanoscience, 18*(5), 1429–1435. DOI: https://doi.org/10.1166/jctn.2021.9619.

13. Hernández, J., Etemadi, A., Roberts-Baca, S., & Muthyapu, V. K. (2021, April). Developing a logistic regression method for valuation of grid-level energy storage systems. *In 2021 IEEE Conference on Technologies for Sustainability (SusTech)*, 1–8. DOI: https://doi.org/10.1109/SusTech51236.2021. 9467419.

14. Tan, K. L., Lee, C. P., & Lim, K. M. (2023). A survey of sentiment analysis: Approaches, datasets, and future research. *Applied Sciences*, *13*(7), 4550. DOI: https://doi.org/10.3390/app13074550.

15. Indu, R., & Dimri, S. C. (2023). Detecting Spam E-mails with Content and Weight-Based Binomial Logistic Model. *Journal of Web Engineering, 22*(7), 939–959. DOI: https://doi.org/10.13052/jwe1540-9589.2271.

16. Berezka, K. M., Kovalchuk, O. Ya., Banakh, S. V., Zlyvko, S. V., & Hrechaniuk, R. (2022). A Binary Logistic Regression Model for Support Decision Making in Criminal Justice. *Folia Oeconomica Stetinensia, 22*(1), 1–17. DOI: https://doi.org/10.2478/foli-2022-0001.

17. Zhang, L. (2022). Smart Marketing Data Collection and Analysis based on Logistic Regression Algorithm. *3rd International Conference on Smart Electronics and Communication (ICOSEC), Trichy, India*, 1611-1614. DOI: https://doi.org/10.1109/ICOSEC54921.2022.9951974.

18. Fayaz, S. A., Zaman, M., & Butt, M. A. (2021). An application of logistic model tree (LMT) algorithm to ameliorate Prediction accuracy of meteorological data. *International Journal of Advanced Technology and Engineering Exploration, 8*(84), 1424–1440. DOI: https://doi.org/10.19101/IJATEE. 2021.874586.

19. Niu, L. (2020). A review of the application of logistic regression in educational research: common issues, implications, and suggestions. *Educational Review, 72*(1), 41–67. DOI: https://doi.org/10.1080/ 00131911.2018.1483892.

20. Rivera, P. P., & Garashchuk, A. (2023). Strategic partner election: proposal for a Binary Logistic Model for the European Union. *Humanities and Social Sciences Communications, 10*(1), 1–13. DOI: https://doi.org/10.1057/s41599-023-02121-y.

21. Velu, A. (2021). Application of logistic regression models in risk management. *International Journal of Innovations in Engineering Research and Technology*, *8*(04), 251–260. Retrieved from https://repo.ijiert.org/index.php/ijiert/article/view/2594.

22. Gai, R., & Zhang, H. (2023). Prediction model of agricultural water quality based on optimized logistic regression algorithm. *EURASIP Journal on Advances in Signal Processing, 21*, 1–14, DOI: https://doi.org/10.1186/s13634-023-00973-9.

23. Chen, Q. (2022). Research on identifying psychological health problems of college students by logistic regression model based on data mining. *Applied Mathematics and Nonlinear Sciences, 8*(1), 2253–2262. DOI: https://doi.org/10.2478/amns.2021.2.00195.

24. Borucka, A. (2020). Logistic regression in modeling and assessment of transport services. *Open Engineering, 10*, 26–34. DOI: https://doi.org/10.1515/eng-2020-0029.

25. Kang, R. (2020). Using logistic regression for persona segmentation in tourism: A case study. *Social Behavior and Personality: an international journal, 48*(4), 1–16. DOI: https://doi.org/10.2224/sbp.8793.

26. Christensen, R. (1997). *Log-Linear Models and Logistic Regression*. Springer. ISBN 10: 0387982477 / ISBN 13: 9780387982472.

27. Hosmer, D. W., & Lemeshow, S. (2000*). Applied Logistic Regression*. John Wiley & Sons, Inc. DOI: https://doi.org/10.1002/0471722146.

28. Hilbe, J. M. (2009). *Logistic Regression Models (1st ed.)*. Chapman and Hall/CRC. DOI: https://doi.org/10.1201/9781420075779.

29. Cramer, J. S. (2003). The standard multinomial logit model. *In Logit Models from Economics and Other Fields, Chapter 7*. Cambridge: Cambridge University Press, 104–125. DOI:https://doi.org/10.1017/CBO9780511615412.008.

MATHEMATICAL MODEL OF LOGISTIC REGRESSION FOR BINARY CLASSIFICATION. PART 1. REGRESSION MODELS OF DATA GENERALIZATION

Petro Kravets[1], Volodymyr Pasichnyk[2], Mykola Prodaniuk[3]

Lviv Polytechnic National University,
Information Systems and Networks Department, Ukraine, Lviv,
[1] E-mail: Petro.O.Kravets@lpnu.ua, ORCID: 0000-0001-8569-423X;
[2] E-mail: Volodymyr.V.Pasichnyk@lpnu.ua, ORCID: 0000-0002-5231-6395;
[3] E-mail: Mykola.M.Prodaniuk@lpnu.ua, ORCID: 0000-0001-9544-3792

In this article, the mathematical justification of logistic regression as an effective and simple to implement method of machine learning is performed.

A review of literary sources was conducted in the direction of statistical processing, analysis and classification of data using the logistic regression method, which confirmed the popularity of this method in various subject areas.

The logistic regression method was compared with the linear and probit regression methods regarding the possibility of predicting the probabilities of events. In this context, the disadvantages of linear regression and the advantages and affinity of logit and probit regression methods are noted. It is indicated that the possibility of forecasting probabilities and binary classification by the method of logistic regression is provided by the use of a sigmoid function with the property of compressive transformation of an argument with an unlimited numerical value into a limited range from 0 to 1 real value of the function. The derivation of the sigmoid function in two different ways is described: based on the model of the logarithm of the odds of events and the model of logistic population growth.

Based on the method of maximum likelihood, the construction of a logarithmic loss function was demonstrated, the use of which made it possible to move from a multi-extremal nonlinear regression problem to a unimodal optimization problem. Methods of regularization of the loss function are presented to control the complexity and prevent retraining of the logistic regression model.

Key words: mathematical model, logistic regression, binary classification, data analysis, machine learning, sigmoid function, log odds, logistic growth model, maximum likelihood method, log loss function, regularization methods.