МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ
КИЇВСЬКИЙ НАЦІОНАЛЬНИЙ УНІВЕРСИТЕТ ІМЕНІ ТАРАСА ШЕВЧЕНКА

О. С. Слабоспицький

ЗАДАЧІ КЛАСИФІКАЦІЇ

Навчальний посібник

Київ
Видавництво «Людмила»
2020

УДК 519.2(075.8)
С47

Рецензенти:
чл.-кор. НАН України, д-р техн. наук, проф. М. Ю. Кузнєцов
д-р техн. наук, проф. В. А. Заславський

Рекомендовано до друку
вченою радою факультету комп'ютерних наук та кібернетики
(протокол № 4 від 21 вересня 2020 року)

Слабоспицький О. С.

С47 Задачі класифікації : навч. посіб. / О. С. Слабоспицький. – К. : Видавництво «Людмила», 2020. – 43 с.

Розглянуто основні класи задач класифікації. Висвітлено базові методи дискримінантного та кластерного аналізу, які дозволяють розв'язувати ряд проблем із розпізнавання образів залежно від обсягу апріорної інформації про об'єкти. Досліджено випадки як відсутності навчальної вибірки так і її наявності.

Для студентів, які вивчають дисципліну "Аналіз даних" і споріднені з нею предмети.

УДК 519.2(075.8)

ВСТУП

Сучасні інформаційні технології дозволяють не тільки накопичувати різнотипні дані в неосяжних розмірах, але й отримувати оперативний доступ до них. Ефективне використання таких обсягів інформації важко уявити без наявності можливості її коректної обробки. Це дозволяють оперативно зробити можливості новітньої обчислювальної техніки з використанням широкого спектра сучасного програмного забезпечення з аналізу та обробки даних. Але головне не тільки провести обробку даних, але й на основі отриманих результатів обробки зробити кваліфіковані висновки.

Для цього буде доречним використання апарата сучасного аналізу даних. На озброєння можна взяти можливості попередньої обробки даних, кореляційного аналізу, дисперсійного аналізу, регресійного аналізу, коваріаційного аналізу, аналізу часових рядів і т. д. До цього слід додати, що все більшої актуальності набуває розв'язання проблем в області *розпізнавання образів* (*pattern recognition*). Тут стануть у нагоді методи як дискримінантного аналізу, так і кластерного аналізу. Саме вони дозволяють вважати об'єкти приналежними до того чи іншого класу об'єктів з близькими характеристиками. У роботі висвітлено розв'язання саме цих проблем залежно від обсягу апріорної інформації про об'єкти, які підлягають класифікації. Розглянуті випадки класифікації як із вчителем, так і без нього. Усе це може послужити фундаментом при розв'язанні задач у галузі штучного інтелекту, робототехніки, Data Science, Big Data (BD) analysis, Smart house, при створенні різноманітних безпілотних пристроїв тощо.

Автор щиро вдячний студентам і співробітникам факультету комп'ютерних наук та кібернетики Київського національного університету імені Тараса Шевченка, які сприяли покращенню цього посібника. Усі зауваження та побажання щодо посібника будуть із вдячністю сприйняті. Їх можна надіслати електронною поштою на адресу: *sl@univ.kiev.ua*.

1. ПРОБЛЕМА КЛАСИФІКАЦІЇ ОБ'ЄКТІВ

У цьому розділі розглянуто постановку задачі класифікації та її головні припущення. Описано основні класи цих задач залежно від обсягу апріорної інформації про об'єкти (процеси, явища, ситуації), які підлягають класифікації.

Такого типу проблеми виникають у багатьох галузях, а саме: у автомобілебудуванні, кораблебудуванні, літакобудуванні, виробництві обчислювальної техніки та мобільних телефонів, бізнесі, економіці, фінансах, медицині, біології, сільському господарстві, соціології тощо.

1.1. Постановка задачі класифікації та її основні припущення

Потреба розв'язання задач класифікації постійно виникає, коли необхідно обробити дані, які представляють об'єкти з деяких підмножин об'єктів, що мають спільні риси. Тому перш ніж виконувати обробку інформації, проводять розподіл об'єктів по цих підмножинах. Подальша обробка інформації після цього вже суттєво спрощується.

Наведемо деякі приклади задач класифікації.

Приклад 1. Перед батьками завжди постає задача навчити дитину розпізнавати літери абетки рукописні та друковані незалежно від почерку в рукописному тексті чи використаного шрифту в друкованому виданні.

Приклад 2. Рибалки повернулися з вдалої риболовлі. Їхній вилов необхідно відсортувати за видами риб. Наприклад: короп, лящ, плотва, щука, судак.

Приклад 3. Нехай необхідно з'ясувати, чи є певна особа тою за кого вона себе видає. Це можна зробити, розв'язуючи задачі *автентифікації особи за відбитками пальців рук* (*fingerprint authentication*), *відбитками долонь рук* (*palmprint authentication*), *обличчям* (*face authentication*), *райдужною оболонкою ока* (*iris authentication*), *голосом* (*voice authentication*), *рукописним почерком* (*handwriting authentication*) і т. ін. При потребі використовують декілька рівнів у цій процедурі: *двофакторну автентифікацію* (*two-factor authentication, 2FA*) або *багатофакторну автентифікацію* (*multi-factor authentication*). Усі ці підходи належать до *біометричної автентифікації* (*biometric authentication*).

Приклад 4. Сформулюємо тепер питання інакше. Необхідно з'ясувати, хто є ця особа? А це вже задача *ідентифікації особи* (*personal identification*), зазвичай *за біометричними характеристиками*, які вже згадувалися в попередньому прикладі. Наприклад, це може бути проблема ідентифікації водія-порушника на основі записів з камер відеофіксації дорожнього руху в місті. Або ідентифікації злочинця за відбитками пальців або долонь рук і т. ін.

Приклад 5. Припустимо, що по кожному з пацієнтів деякої медичної установи доступна інформація про результати ряду їхніх аналізів та обстежень. Необхідно на основі цих даних зробити висновок про наявність чи відсутність у них певної хвороби, що дозволить лікарю оперативно прийняти рішення про вибір найбільш доцільного курсу їхнього подальшого лікування.

Приклад 6. Розглянемо потік поштової кореспонденції, яка надходить на загальну поштову скриньку деякої установи. Необхідно організувати сортування поштових надходжень по відділах цієї установи.

Приклад 7. Нехай деякий агрегатор новин накопичує інформацію з багатьох джерел. Необхідно новини, які надійшли, оперативно відсортувати, наприклад, за такими розділами: політика, економіка, наука та технології, мистецтво, спорт, здоров'я, за кордоном, інше. Ще одна проблема пов'язана з вилученням з цього потоку інформації фейкових повідомлень.

Приклад 8. Злободенною є задача оперативного виявлення за доступною інформацією наближення якогось стихійного лиха. Наприклад: землетрусу, виверження вулкану, урагану, тайфуну, смерчу, цунамі тощо.

Приклад 9. Актуальною залишається задача розпізнавання небесних об'єктів, що наближаються до нашої планети. Це може бути астероїд, метеорит, комета або взагалі якийсь невпізнаний об'єкт.

1.2. Базові поняття та постановка задачі класифікації

Перейдемо до формальної постановки задачі класифікації. Нехай маємо деяку сукупність *об'єктів* (процесів, явищ, ситуацій) ω, які підлягають аналізу. За традицією ω ще називають *образом* (*pattern*). Позначимо через Ω множину всіх можливих таких об'єктів ω, тобто

$$\omega \in \Omega.$$

У подальшому Ω будемо називати *простором об'єктів* (*object space*), або *простором образів* (*pattern space*).

Нехай відомо, що простір усіх об'єктів Ω розбивається на m підмножин $\{\Omega_i\}_{i=1}^{m}$, які не перетинаються. Причому вважається, що кожна підмножина складається з об'єктів зі спільними рисами. Таким чином,

$$\Omega = \bigcup_{i=1}^{m} \Omega_i, \ \Omega_i \bigcap \Omega_j = \varnothing, \forall i \neq j.$$

Ці підмножини $\{\Omega_i\}_{i=1}^{m}$ будемо називати *класами*.

Нехай для кожного об'єкта ω відомий *вектор спостережень* (*observation vector*) $\vec{y} = \vec{y}(\omega)$ над його характерними властивостями. Множину всіх можливих таких векторів спостережень будемо називати *простором спостережень* (*observation space*) і позначати через Y , тобто

$$\vec{y} = \vec{y}(\omega) \in Y, \ \forall \omega \in \Omega .$$

Задача класифікації (*classification problem*) полягає в пошуку правила прийняття рішення, оптимального в деякому розумінні, про належність об'єкта ω до одного з класів $\{\Omega_i\}_{i=1}^{m}$ на базі вектора спостережень $\vec{y} = \vec{y}(\omega)$ та доступної апріорної інформації про об'єкт ω.

Якщо розмірність вектора спостережень $\vec{y}$ досить велика або він містить надлишкову інформацію, то має сенс провести пониження розмірності (редукцію розмірності) вектора $\vec{y}$, який використовується при класифікації. Тобто перейти від вектора спостережень $\vec{y}$ до *вектора ознак* (*feature vector*) $\vec{x} = \vec{x}(\omega)$ меншої розмірності, мінімізуючи втрати інформації про об'єкт ω, за допомогою деякого перетворення

$$R : Y \to X ,$$

де X – *простір ознак* (*feature space*), який складається з усіх можливих векторів ознак $\vec{x}$, тобто

$$\vec{x} = \vec{x}(\omega) \in X, \ \forall \omega \in \Omega .$$

У подальшому будемо вважати, що процедура пониження розмірності вектора спостережень $\vec{y}$ вже проведена, і тому при розв'язанні задача класифікації будемо використовувати замість вектора спостережень $\vec{y}$ вектор ознак $\vec{x}$.

А саму задачу класифікації будемо вирішувати шляхом розбиття простору ознак X на підмножини $\{X_i\}_{i=1}^{m}$, які не перетинаються. Причому X_i складається з векторів ознак, які відповідають об'єктам із класу $\Omega_i, i = \overline{1,m}$. Таким чином,

$$X = \bigcup_{i=1}^{m} X_i, \ X_i \bigcap X_j = \varnothing, \forall i \neq j.$$

Це розбиття простору ознак X зазвичай здійснюється за допомогою набору *дискримінантних функцій* $\{d_i(\vec{x})\}_{i=1}^{m}$, які будуються для цієї мети. Причому вони конструюються таким чином, щоб

$$X_j = \left\{ x : \; d_j\left(\vec{x}\right) > d_i\left(\vec{x}\right), \forall i \neq j \right\}, \;\; j = \overline{1,m}. \qquad (1.1)$$

Тоді алгоритм класифікації об'єкта ω (тобто віднесення об'єкта ω до певного класу) буде мати вигляд

$$\textit{якщо} \quad d_j\left(\vec{x}(\omega)\right) > d_i\left(\vec{x}(\omega)\right), \forall i \neq j, \quad \textit{то} \quad \omega \in \Omega_j.$$

Або, іншими словами, якщо взяти до уваги (1.1), останнє можна переписати так:

$$\textit{якщо} \quad \vec{x}(\omega) \in X_j, \quad \textit{то} \quad \omega \in \Omega_j.$$

Зауваження. У випадку двох класів достатньо мати у своєму розпорядженні тільки одну дискримінантну функцію, яка визначається таким чином:

$$d\left(\vec{x}\right) = d_1\left(\vec{x}\right) - d_2\left(\vec{x}\right). \qquad (1.2)$$

Тоді області $\left\{X_j\right\}_{j=1}^{2}$ будуть задаватися так:

$$X_1 = \left\{ x : \; d\left(\vec{x}\right) > 0 \right\},$$
$$X_2 = \left\{ x : \; d\left(\vec{x}\right) < 0 \right\}.$$

А безпосередньо алгоритм класифікації об'єкта ω буде мати вигляд

$$\left[\begin{array}{l} \textit{якщо}\; d\left(\vec{x}(\omega)\right) > 0, \quad \textit{то} \;\; \omega \in \Omega_1, \\ \textit{якщо}\; d\left(\vec{x}(\omega)\right) < 0, \quad \textit{то} \;\; \omega \in \Omega_2. \end{array}\right.$$

1.3. Основні типи задач класифікації

У подальшому будуть розглянуті деякі постановки задач класифікації залежно від обсягу апріорної інформації про вектор ознак $\vec{x} = \vec{x}(\omega) \in X$.

У найкращому випадку для вектора ознак $\vec{x}$ може бути доступна повна інформації про його розподіл.

Гірше, коли розподіл вектора ознак $\vec{x}$ заданий з точністю до деякого вектора невідомих параметрів. У такому випадку цю невизначеність зазвичай компенсують наявністю об'єктів з відомими класифікаціями, тобто вважають, що нам доступна *навчальна вибірка* (*training sample*). Тобто маємо задачу *класифікації з вчителем* (*supervised classification*). При розв'язуванні такого типу задач класифікації в нагоді стане *дискримінантний аналіз* (*discriminant analysis*).

У найгіршому випадку будь-яка інформація про розподіл вектора ознак $\vec{x}$ може бути взагалі відсутня, а апріорна інформація про об'єкти, які належить розподілити за класами, вичерпується тільки інформацією, що об'єкти в кожному з класів близькі (схожі) у деякому розумінні. Якщо припускається, що навчальна вибірка відсутня, тоді кажуть, що мають справу з задачею *класифікації без вчителя* (*unsupervised classification*). Цей спектр задач зазвичай розв'язують методами *кластерного аналізу* (*cluster analysis*).

2. ОСНОВИ ДИСКРИМІНАНТНОГО АНАЛІЗУ

Почнемо розгляд задач класифікації з випадку, коли для вектора ознак $\vec{x}$ доступна досить багата інформація, а саме, відома функція щільності для кожного класу $\Omega_i, i=\overline{1,m}$. Припустимо, що класи apriori визначені. Необхідно віднести кожен об'єкт ω до одного із цих класів.

2.1. Постановка задачі та основні припущення

Нехай маємо справу з об'єктами ω із m заданих класів $\{\Omega_i\}_{i=1}^{m}$, тобто

$$\Omega = \bigcup_{i=1}^{m} \Omega_i,\ \Omega_i \bigcap \Omega_j = \varnothing, \forall i \neq j.$$

Припустимо, що доступна така апріорна інформація:

I. для n-вимірного вектора ознак $\vec{x}=\vec{x}(\omega)\in X$ для кожного класу Ω_i відома функція щільності $p(\vec{x}/i), i=\overline{1,m}$; (2.1)

II. для кожного класу Ω_i відома апріорна ймовірність $p_i, i=\overline{1,m};\ \sum_{i=1}^{m} p_i = 1;$ (2.2)

III. відома невід'ємна *функція втрат* $C(j/i)$, яка дорівнює величині втрат (штрафу) при віднесенні об'єкта ω

до j-го класу, коли він насправді з i-го класу $\left(i, j = \overline{1,m}\right)$, причому $C\left(i / i\right) = 0, i = \overline{1,m}$. (2.3)

Необхідно побудувати правило прийняття рішення про віднесення об'єкта ω до одного з класів $\{\Omega_i\}_{i=1}^{m}$ на базі вектора ознак $\vec{x} = \vec{x}\left(\omega\right)$ так, щоб були найменшими середні втрати:

$$Q = \sum_{i=1}^{m} \left\{ p_i \left[\sum_{j=1}^{m} C\left(j / i\right) P\left(j / i\right) \right] \right\}, \tag{2.4}$$

де $P\left(j / i\right)$ – ймовірність віднесення об'єкта ω до j-го класу, коли він насправді з i-го класу $\left(i, j = \overline{1,m}\right)$.

Зауваження. Середні втрати при помилковій класифікації об'єкта з i-го класу визначаються виразом

$$\sum_{j=1}^{m} C\left(j / i\right) P\left(j / i\right).$$

Зрозуміло, що

$$\sum_{j=1}^{m} P\left(j / i\right) = 1,\ i = \overline{1,m}.$$

Причому необхідні ймовірності $P\left(j / i\right)$ можна підрахувати таким чином:

$$P\left(j / i\right) = \int\limits_{X_j} p\left(\vec{x} / i\right) d\vec{x}, \quad i, j = \overline{1,m}, \tag{2.5}$$

де множини X_j згідно з (1.1) мають вигляд

$$X_j = \left\{x:\ d_j\left(\vec{x}\right) > d_i\left(\vec{x}\right), \forall i \neq j\right\},\ j = \overline{1,m},$$

а $\left\{d_i\left(\vec{x}\right)\right\}_{i=1}^{m}$ – набір дискримінантних функцій, який побудований для цієї задачі.

Фактично треба буде знайти такий набір дискримінантних функцій $\left\{d_i\left(\vec{x}\right)\right\}_{i=1}^{m}$, який забезпечує розбиття простору ознак X на підмножини

$$\{X_i\}_{i=1}^{m} \quad \left(X = \bigcup_{i=1}^{m} X_i, \ X_i \bigcap X_j = \varnothing, \forall i \neq j \right),$$

які не перетинаються, причому так, щоб середні втрати Q були найменшими.

2.2. Побудова дискримінантних функцій

Переходимо до знаходження набору дискримінантних функцій $\{d_i(x)\}_{i=1}^{m}$ таких, що забезпечують потрібне розбиття простору ознак X:

$$X = \bigcup_{i=1}^{m} X_i, \ X_i \bigcap X_j = \varnothing, \forall i \neq j; \tag{2.6}$$

де $X_j = \{x : d_j(x) > d_i(x), \forall i \neq j\}, \ j = \overline{1,m}$.

Проаналізуємо критерій якості (2.4). Для цього перепишемо його в іншому вигляді, прийнявши до уваги (2.5):

$$Q = \sum_{i=1}^{m} \left\{ p_i \left[\sum_{j=1}^{m} C(j/i) P(j/i) \right] \right\} =$$

$$= \sum_{i=1}^{m} \left\{ p_i \left[\sum_{j=1}^{m} C(j/i) \int_{X_j} p(\vec{x}/i) d\vec{x} \right] \right\} =$$

$$= \sum_{i=1}^{m} \left\{ \sum_{j=1}^{m} \left[p_i C(j/i) \int_{X_j} p(\vec{x}/i) d\vec{x} \right] \right\} =$$

$$= \sum_{j=1}^{m} \left\{ \sum_{i=1}^{m} \left[p_i C(j/i) \int_{X_j} p(\vec{x}/i) d\vec{x} \right] \right\} =$$

$$=\sum_{j=1}^{m}\left\{\sum_{i=1}^{m}\int_{X_j}\left[p_i C\left(j/i\right)p\left(\vec{x}/i\right)\right]d\vec{x}\right\}=$$

$$=\sum_{j=1}^{m}\left\{\int_{X_j}\left(\sum_{i=1}^{m}\left[p_i C\left(j/i\right)p\left(\vec{x}/i\right)\right]\right)d\vec{x}\right\}=$$

$$=\sum_{j=1}^{m}\left\{\int_{X_j}g_j\left(\vec{x}\right)d\vec{x}\right\}, \qquad (2.7)$$

де в останньому перетворенні використано позначення

$$g_j\left(\vec{x}\right)=\sum_{i=1}^{m}\left[p_i C\left(j/i\right)p\left(\vec{x}/i\right)\right].$$

Припустимо, що для довільної підмножини індексів

$$1\le j_1<j_2<\ldots<j_q\le m$$

міра Лебега множини

$$\left\{\vec{x}:g_{j_1}\left(\vec{x}\right)=g_{j_2}\left(\vec{x}\right)=\ldots g_{j_q}\left(\vec{x}\right)\right\}$$

дорівнює нулеві.

Тоді враховуючи умову (2.6), неважко бачити, що вираз (2.7) досягає свого мінімуму, якщо підмножини $\left\{X_i\right\}_{i=1}^{m}$ вибрати у такому вигляді:

$$X_j=\left\{x:\ g_j\left(\vec{x}\right)<g_i\left(\vec{x}\right),\forall i\ne j\right\},\ j=\overline{1,m},$$

причому найменше значення функціонала буде дорівнювати

$$Q=\int_X\left[\min_{j=\overline{1,m}}g_j\left(\vec{x}\right)\right]d\vec{x}.$$

Це дозволяє взяти як потрібний набір дискримінантних функцій $\left\{d_i\left(\vec{x}\right)\right\}_{i=1}^{m}$ функції виду

$$d_j\left(\vec{x}\right)=-g_j\left(\vec{x}\right)=-\sum_{i=1}^{m}\left[p_i C\left(j/i\right)p\left(\vec{x}/i\right)\right],\ j=\overline{1,m}.$$

Висновок. Для задачі класифікації у випадку m класів $\left\{\Omega_i\right\}_{i=1}^{m}$ у припущеннях (2.1) – (2.3) шуканий набір дискримінантних функцій має вигляд

$$d_j\left(\vec{x}\right)=-\sum_{i=1}^{m}\left[p_i C\left(j/i\right)p\left(\vec{x}/i\right)\right],\; j=\overline{1,m}. \qquad (2.8)$$

2.3. Випадок використання простої функції втрат

Формула (2.8) задає вигляд дискримінантних функцій для задачі класифікації в припущеннях (2.1) – (2.3). Проаналізуємо розв'язання вищезгаданої задачі у випадку, коли як *приклад функції втрат* $C\left(j/i\right)$ вибрана така функція:

$$C\left(j/i\right)=c\left(1-\delta_{ji}\right), c>0, \qquad (2.9)$$

де δ_{ji} – символ Кронекера, тобто

$$\delta_{ji}=\begin{cases}1, & якщо\ j=i,\\ 0, & якщо\ j\neq i.\end{cases}$$

Почнемо з аналізу середніх втрат (2.4):

$$Q=\sum_{i=1}^{m}\left\{p_i\left[\sum_{j=1}^{m}C\left(j/i\right)P\left(j/i\right)\right]\right\}=$$

$$=\sum_{i=1}^{m}\left\{p_i\left[\sum_{j=1}^{m}c\left(1-\delta_{ji}\right)P\left(j/i\right)\right]\right\}=$$

$$=c\sum_{i=1}^{m}\left\{p_i\left[\sum_{\substack{j=1\\ j\neq i}}^{m}P\left(j/i\right)\right]\right\}=$$

$$= c\sum_{i=1}^{m}\left\{p_i\left[1 - P(i/i)\right]\right\} =$$

$$= c\left\{1 - \sum_{i=1}^{m} p_i P(i/i)\right\}.$$

Останній вираз приводить до висновку, що мінімізація функції середніх втрат (2.4) у випадку, коли функція втрат $C(j/i)$ має вигляд (2.9), звелася до мінімізації *ймовірності помилкової класифікації*

$$\sum_{i=1}^{m}\left\{p_i\left[\sum_{\substack{j=1\\ j\neq i}}^{m} P(j/i)\right]\right\} = 1 - \sum_{i=1}^{m} p_i P(i/i). \qquad (2.10)$$

У цьому разі замість дискримінантних функцій (2.8) можна використати набір дискримінантних функцій більш простого вигляду.

Для цього спочатку проаналізуємо набір дискримінантних функцій (2.8):

$$d_j(\vec{x}) = -\sum_{i=1}^{m}\left[p_i C(j/i)\, p(\vec{x}/i)\right] =$$

$$= -\sum_{i=1}^{m}\left[p_i c\left(1 - \delta_{ji}\right) p(\vec{x}/i)\right] =$$

$$= -c\sum_{\substack{i=1\\ i\neq j}}^{m}\left[p_i\, p(\vec{x}/i)\right] =$$

$$= -c\left[p(\vec{x}) - p_j\, p(\vec{x}/j)\right] =$$

$$= c\left[p_j\, p(\vec{x}/j) - p(\vec{x})\right],\ j = \overline{1,m}. \qquad (2.11)$$

Тут скористалися тим, що

$$p(\vec{x}) = \sum_{i=1}^{m}\left[p_i\, p(\vec{x}/i)\right],$$

де $p(\vec{x})$ – щільність вектора ознак.

Тоді у випадку, який розглядається, області X_j будуть задаватися таким чином:

$$X_j = \left\{ x:\ d_j\left(\vec{x}\right) > d_i\left(\vec{x}\right), \forall i \ne j \right\} =$$

$$= \left\{ x:\ c\left[p_j p\left(\vec{x} / j\right) - p\left(\vec{x}\right) \right] > c\left[p_i p\left(\vec{x} / i\right) - p\left(\vec{x}\right) \right], \forall i \ne j \right\} =$$

$$= \left\{ x:\ p_j p\left(\vec{x} / j\right) > p_i p\left(\vec{x} / i\right), \forall i \ne j \right\},\ j = \overline{1,m}. \qquad (2.12)$$

Представлення (2.12) для областей $\left\{X_i\right\}_{i=1}^{m}$ дає змогу стверджувати, що замість дискримінантних функцій виду (2.11) можна використовувати дискримінантні функції вигляду

$$d_j^{(1)}\left(\vec{x}\right) = p_j p\left(\vec{x} / j\right),\ j = \overline{1,m}, \qquad (2.13)$$

а області X_j будуть задаватися таким чином:

$$X_j = \left\{ x:\ d_j^{(1)}\left(\vec{x}\right) > d_i^{(1)}\left(\vec{x}\right), \forall i \ne j \right\},\ j = \overline{1,m}. \qquad (2.14)$$

Висновок 1. Для задачі класифікації у випадку m класів $\left\{\Omega_i\right\}_{i=1}^{m}$ згідно з критерієм мінімуму ймовірності помилкової класифікації (2.10) вигляд потрібного набору дискримінантних функцій $\left\{d_i^{(1)}\left(\vec{x}\right)\right\}_{i=1}^{m}$ спрощується і має представлення (2.13), а шукані області $\left\{X_i\right\}_{i=1}^{m}$ набувають вигляду (2.14).

Результат (2.12) дозволяє зробити також ще один корисний висновок.

Дійсно, згідно із формулою Байєса апостеріорна ймовірність $p_{j/\vec{x}}$ для j-го класу Ω_j, після отримання вектора ознак $\vec{x}$, задається таким виразом:

$$p_{j/\vec{x}} = \frac{p_j p\left(\vec{x} / j\right)}{p\left(\vec{x}\right)},\ j = \overline{1,m}.$$

Тоді з (2.12) отримуємо

$$X_j = \left\{ x:\ p_j p\left(\vec{x} / j\right) > p_i p\left(\vec{x} / i\right), \forall i \ne j \right\} =$$

$$= \left\{ x : \frac{p_j p(\vec{x}/j)}{p(\vec{x})} > \frac{p_i p(\vec{x}/i)}{p(\vec{x})}, \forall i \neq j \right\} =$$

$$= \left\{ x : p_{j/\vec{x}} > p_{i/\vec{x}}, \forall i \neq j \right\}.$$

У результаті можемо стверджувати таке.

Висновок 2. Задачу класифікації у випадку m класів $\{\Omega_i\}_{i=1}^m$ згідно з критерієм мінімуму ймовірності помилкової класифікації (2.10) можна розглядати як задачу класифікації для m класів згідно з критерієм *максимуму апостеріорної ймовірності*.

2.4. Випадок нормально розподілених векторів ознак

Додатково до припущень з попереднього розділу будемо вважати, що n-вимірні вектори ознак $\vec{x}$ є нормально розподіленими, а саме:

$$\vec{x} = \vec{x}(\omega) \sim \mathcal{N}(\vec{m}_i, V_i), V_i > 0, \ \textit{якщо} \ \omega \in \Omega_i, \ i = \overline{1,m},$$

де параметри $\{\vec{m}_i, V_i\}_{i=1}^m$ задані.

Тобто

$$p(\vec{x}/i) = \frac{1}{(2\pi)^{\frac{n}{2}} (\det(V_i))^{\frac{1}{2}}} e^{-\frac{1}{2}\|\vec{x}-\vec{m}_i\|^2_{V_i^{-1}}}, \ i = \overline{1,m},$$

де $\|\vec{x}\|_Q^2 = \vec{x}^T Q \vec{x}, Q > 0$.

Тоді згідно з (2.14) області X_j набувають вигляду

$$X_j = \left\{ x : d_j^{(1)}(\vec{x}) > d_i^{(1)}(\vec{x}), \forall i \neq j \right\} =$$

$$= \left\{ x : p_j p(\vec{x}/j) > p_i p(\vec{x}/i), \forall i \neq j \right\} =$$

$$=\left\{ x:\ \frac{p_j}{(2\pi)^{\frac{n}{2}}\left(\det\left(V_j\right)\right)^{\frac{1}{2}}} e^{-\frac{1}{2}\left\|\vec{x}-\vec{m}_j\right\|^2_{V_j^{-1}}} > \right.$$

$$\left. > \frac{p_i}{(2\pi)^{\frac{n}{2}}\left(\det\left(V_i\right)\right)^{\frac{1}{2}}} e^{-\frac{1}{2}\left\|\vec{x}-\vec{m}_i\right\|^2_{V_i^{-1}}}, \forall i \neq j \right\} =$$

$$=\left\{ x:\ \ln\left(p_j\right)-\frac{1}{2}\ln\left(\det\left(V_j\right)\right)-\frac{1}{2}\left\|\vec{x}-\vec{m}_j\right\|^2_{V_j^{-1}} > \right.$$

$$> \left\{ x: \ln\left(p_i\right)-\frac{1}{2}\ln\left(\det\left(V_i\right)\right)-\frac{1}{2}\left\|\vec{x}-\vec{m}_i\right\|^2_{V_i^{-1}}, \forall i \neq j \right\} =$$

$$=\left\{ x:\ 2\ln\left(p_j\right)-\ln\left(\det\left(V_j\right)\right)-\left\|\vec{x}-\vec{m}_j\right\|^2_{V_j^{-1}} > \right.$$

$$> \left\{ x: 2\ln\left(p_i\right)-\ln\left(\det\left(V_i\right)\right)-\left\|\vec{x}-\vec{m}_i\right\|^2_{V_i^{-1}}, \forall i \neq j \right\},\ \ j=\overline{1,m}.$$

Останнє дозволяє замість дискримінантних функцій (2.13) використовувати дискримінантні функції вигляду

$$d_j^{(2)}\left(\vec{x}\right)=-\left\|\vec{x}-\vec{m}_j\right\|^2_{V_j^{-1}}+2\ln\left(p_j\right)-\ln\left(\det\left(V_j\right)\right),\ j=\overline{1,m}\,. \quad (2.15)$$

Висновок 1. У випадку нормально розподілених векторів ознак $\vec{x}$ можна скористатися набором дискримінантних функцій (2.15), які вже є квадратичними функціями.

У свою чергу, області $\left\{X_j\right\}_{j=1}^{m}$ можна представити таким чином:

$$X_j=\left\{x:\ d_j^{(2)}\left(\vec{x}\right)>d_i^{(2)}\left(\vec{x}\right), \forall i\neq j\right\},\ \ j=\overline{1,m}\,.$$

Окремо звернемо увагу на випадок, коли коваріаційні матриці нормальних розподілів векторів ознак для всіх класів однакові, тобто

$$\vec{x}=\vec{x}\left(\omega\right)\sim\mathcal{N}\left(\vec{m}_i,V\right), V>0,\ \textit{якщо}\ \ \omega\in\Omega_i,\ i=\overline{1,m}\,.$$

Це дозволяє дещо спростити вигляд потрібних дискримінантних функцій. Дійсно,

$$X_j = \left\{ x:\ d_j^{(2)}(\vec{x}) > d_i^{(2)}(\vec{x}), \forall i \ne j \right\} =$$

$$= \left\{ x:\ 2\ln\left(p_j\right) - \ln\left(\det(V)\right) - \left\|\vec{x} - \vec{m}_j\right\|_{V^{-1}}^2 > \right.$$

$$\left. > 2\ln\left(p_i\right) - \ln\left(\det(V)\right) - \left\|\vec{x} - \vec{m}_i\right\|_{V^{-1}}^2, \forall i \ne j \right\} =$$

$$= \left\{ x:\ 2\ln\left(p_j\right) - \left\|\vec{x}\right\|_{V^{-1}}^2 + 2\vec{x}^T V^{-1}\vec{m}_j - \left\|\vec{m}_j\right\|_{V^{-1}}^2 > \right.$$

$$\left. > 2\ln\left(p_i\right) - \left\|\vec{x}\right\|_{V^{-1}}^2 + 2\vec{x}^T V^{-1}\vec{m}_i - \left\|\vec{m}_i\right\|_{V^{-1}}^2, \forall i \ne j \right\} =$$

$$= \left\{ x:\ \ln\left(p_j\right) + \vec{x}^T V^{-1}\vec{m}_j - \frac{1}{2}\left\|\vec{m}_j\right\|_{V^{-1}}^2 > \right.$$

$$\left. > \ln\left(p_i\right) + \vec{x}^T V^{-1}\vec{m}_i - \frac{1}{2}\left\|\vec{m}_i\right\|_{V^{-1}}^2, \forall i \ne j \right\},\ \ j = \overline{1,m}.$$

Таким чином, у цьому випадку замість дискримінантних функцій (2.15) можна використовувати дискримінантні функції спрощеного виду

$$d_j^{(3)}(\vec{x}) = \vec{x}^T V^{-1}\vec{m}_j + \ln\left(p_j\right) - \frac{1}{2}\left\|\vec{m}_j\right\|_{V^{-1}}^2,\ \ j = \overline{1,m}. \quad (2.16)$$

Висновок 2. Дискримінантні функції (2.16) у випадку, коли коваріаційні матриці вектора ознак для усіх класів однакові, вже є лінійними функціями від вектора ознак. Останнє суттєво спрощує їхню геометричну інтерпретацію.

А це, у свою чергу, дозволяє потрібні області $\left\{X_j\right\}_{j=1}^m$ представити в такому вигляді:

$$X_j = \left\{ x:\ d_j^{(3)}(\vec{x}) > d_i^{(3)}(\vec{x}), \forall i \ne j \right\},\ \ j = \overline{1,m}.$$

2.4.1. Випадок двох класів

Поглянемо, як можна спростити останні результати у випадку двох класів $\{\Omega_i\}_{i=1}^{2}$. Нехай припущення про нормальність вектора ознак $\vec{x}$ залишається в силі

$$\vec{x} = \vec{x}(\omega) \sim \mathcal{N}(\vec{m}_i, V_i), V_i > 0, \ \textit{якщо} \ \omega \in \Omega_i, i = \overline{1,2}, \qquad (2.17)$$

де параметри $\{\vec{m}_i, V_i\}_{i=1}^{2}$ вважаються відомими.

Для цього випадку, коли $m = 2$, дискримінантні функції мають вигляд (2.15), але можна обмежитися тільки однією дискримінантною функцією $d^{(2)}(\vec{x})$, яка згідно з (1.2) набуває вигляду

$$\begin{aligned} d^{(2)}(\vec{x}) &= d_1^{(2)}(\vec{x}) - d_2^{(2)}(\vec{x}) = \\ &= -\|\vec{x} - \vec{m}_1\|_{V_1^{-1}}^2 + 2\ln(p_1) - \ln(\det(V_1)) + \\ &\quad + \|\vec{x} - \vec{m}_2\|_{V_2^{-1}}^2 - 2\ln(p_2) + \ln(\det(V_2)) = \\ &= \|\vec{x} - \vec{m}_2\|_{V_2^{-1}}^2 - \|\vec{x} - \vec{m}_1\|_{V_1^{-1}}^2 + 2\ln\left(\frac{p_1}{p_2}\sqrt{\frac{\det(V_2)}{\det(V_1)}}\right). \end{aligned}$$

Тобто ця єдина дискримінантна функція $d^{(2)}(\vec{x})$ залишається квадратичною функцією від вектора ознак.

Тоді відповідний алгоритм класифікації об'єкта ω буде мати таке представлення:

$$\left[\begin{array}{l} \textit{якщо } d^{(2)}(\vec{x}(\omega)) > 0, \quad \textit{то } \ \omega \in \Omega_1, \\ \textit{якщо } d^{(2)}(\vec{x}(\omega)) < 0, \quad \textit{то } \ \omega \in \Omega_2. \end{array}\right.$$

А області $\{X_j\}_{j=1}^{2}$, у свою чергу, набувають вигляду

$$X_1 = \left\{x: \ d^{(2)}(\vec{x}) > 0\right\},$$

$$X_2 = \left\{x: \ d^{(2)}(\vec{x}) < 0\right\}.$$

Якщо додатково до припущення (2.17) вважати, що коваріаційні матриці вектора ознак для цих класів однакові, тобто

$$\vec{x} = \vec{x}(\omega) \sim \mathcal{N}(\vec{m}_i, V), V > 0, \textit{ якщо } \omega \in \Omega_i, i = \overline{1,2},$$

то відповідна єдина дискримінантна функція $d^{(3)}(\vec{x})$ набуде більш простого вигляду

$$\begin{aligned}
d^{(3)}(\vec{x}) &= d_1^{(3)}(\vec{x}) - d_2^{(3)}(\vec{x}) = \\
&= \vec{x}^T V^{-1} \vec{m}_1 + \ln(p_1) - \frac{1}{2}\|\vec{m}_1\|_{V^{-1}}^2 - \\
&\quad - \vec{x}^T V^{-1} \vec{m}_2 - \ln(p_2) + \frac{1}{2}\|\vec{m}_2\|_{V^{-1}}^2 = \\
&= \vec{x}^T V^{-1}(\vec{m}_1 - \vec{m}_2) + \ln\left(\frac{p_1}{p_2}\right) + \frac{1}{2}\left(\|\vec{m}_2\|_{V^{-1}}^2 - \|\vec{m}_1\|_{V^{-1}}^2\right) = \\
&= \vec{x}^T V^{-1}(\vec{m}_1 - \vec{m}_2) + \ln\left(\frac{p_1}{p_2}\right) + \frac{1}{2}(\vec{m}_2 - \vec{m}_1)^T V^{-1}(\vec{m}_2 + \vec{m}_1).
\end{aligned}$$

В останньому переході була використана тотожність, яка легко перевіряється:

$$\|\vec{x}_1\|_Q^2 - \|\vec{x}_2\|_Q^2 = (\vec{x}_1 - \vec{x}_2)^T Q(\vec{x}_1 + \vec{x}_2),$$

де $Q > 0$, $\vec{x}_1, \vec{x}_2$ – вектори відповідної розмірності.

А відповідний алгоритм класифікації об'єкта ω та вигляд областей $\{X_j\}_{j=1}^2$ записуються подібним чином, але вже через дискримінантну функцію $d^{(3)}(\vec{x})$.

2.4.2. Обчислення ймовірності помилкової класифікації у випадку двох класів

Якість розв'язання задачі класифікації можна оцінити за значенням імовірності помилкової класифікації.

Продемонструємо її обчислення у випадку, який був розглянутий останнім у попередньому розділі. Тобто у випадку двох класів $\left\{\Omega_i\right\}_{i=1}^{2}$, коли для вектора ознак $\vec{x}$ справедливо

$$\vec{x}=\vec{x}(\omega)\sim\mathcal{N}\left(\vec{m}_i,V\right),V>0,\ \textit{якщо}\ \ \omega\in\Omega_i,\, i=\overline{1,2}\,.$$

Імовірність помилкової класифікації у випадку двох класів згідно з (2.10) визначається таким чином:

$$p_1P(2/1)+p_2P(1/2).$$

Залишилося тільки обчислити ймовірності $P(2/1),P(1/2)$, які згідно з (2.5) задаються формулою

$$P(j/i)=\int\limits_{X_j}p(\vec{x}/i)\,d\vec{x},\quad i,j=\overline{1,2}\,,$$

де

$$p(\vec{x}/i)=\frac{1}{(2\pi)^{\frac{n}{2}}\left(\det(V)\right)^{\frac{1}{2}}}e^{-\frac{1}{2}\left\|\vec{x}-\vec{m}_i\right\|_{V^{-1}}^{2}},i=\overline{1,2}\,;$$

$$X_1=\left\{x:d^{(3)}(\vec{x})>0\right\},$$

$$X_2=\left\{x:d^{(3)}(\vec{x})<0\right\},$$

а $d^{(3)}(\vec{x})=\vec{x}^TV^{-1}\left(\vec{m}_1-\vec{m}_2\right)+\ln\left(\dfrac{p_1}{p_2}\right)+\dfrac{1}{2}\left(\vec{m}_2-\vec{m}_1\right)^TV^{-1}\left(\vec{m}_2+\vec{m}_1\right)$.

Згідно з останнім співвідношенням дискримінантна функція $d^{(3)}(\vec{x})$ є лінійним перетворенням нормально розподіленого вектора ознак $\vec{x}$, і тому теж буде мати нормальний розподіл (див. додаток), а саме:

$$d^{(3)}(\vec{x}) = d^{(3)}(\vec{x}(\omega)) \sim$$

$$\sim \mathcal{N}\left(\vec{m}_i^{\,T} V^{-1}(\vec{m}_1 - \vec{m}_2) + \ln\left(\frac{p_1}{p_2}\right) + \frac{1}{2}(\vec{m}_2 - \vec{m}_1)^T V^{-1}(\vec{m}_2 + \vec{m}_1), \right.$$

$$\left. (\vec{m}_1 - \vec{m}_2)^T V^{-1}(\vec{m}_1 - \vec{m}_2) \right), \ \textit{якщо} \ \omega \in \Omega_i, \, i = \overline{1,2}.$$

Або скорочено

$$d^{(3)}(\vec{x}(\omega)) \sim \mathcal{N}\left(\mu_i, \Delta^2\right), \ \textit{якщо} \ \omega \in \Omega_i, \, i = \overline{1,2},$$

де

$$\mu_i = \vec{m}_i^{\,T} V^{-1}(\vec{m}_1 - \vec{m}_2) + \ln\left(\frac{p_1}{p_2}\right) + \frac{1}{2}(\vec{m}_2 - \vec{m}_1)^T V^{-1}(\vec{m}_2 + \vec{m}_1), \, i = \overline{1,2},$$

$$\Delta^2 = \left\| \vec{m}_1 - \vec{m}_2 \right\|_{V^{-1}}^2 .$$

Означення. *Відстанню Махаланобіса* (*Mahalanobis distance*) між нормальними розподілами $\mathcal{N}(\vec{m}_1, V)$ та $\mathcal{N}(\vec{m}_2, V)$ зі спільною коваріаційною матрицею V $(V > 0)$ називається величина, яка визначається таким чином:

$$\Delta = \sqrt{\left\| \vec{m}_1 - \vec{m}_2 \right\|_{V^{-1}}^2} .$$

Тоді

$$\frac{d^{(3)}(\vec{x}(\omega)) - \mu_i}{\Delta} \sim \mathcal{N}(0,1), \ \text{якщо} \ \omega \in \Omega_i, \, i = \overline{1,2} .$$

Це дозволяє провести підрахунок потрібних ймовірностей.

Дійсно,

$$P(2/1) = \int_{X_2} p(\vec{x}/1) d\vec{x} = P\left(d^{(3)}(\vec{x}) < 0/1\right) =$$

$$= P\left(\frac{d^{(3)}(\vec{x}) - \mu_1}{\Delta} < \frac{-\mu_1}{\Delta} / 1 \right) = \Phi\left(\frac{-\mu_1}{\Delta}\right),$$

де

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

У свою чергу,

$$P(1/2)=\int_{X_1} p(\vec{x}/2)d\vec{x}=P\left(d^{(3)}(\vec{x})>0/2\right)=$$

$$=P\left(\frac{d^{(3)}(\vec{x})-\mu_2}{\Delta}>\frac{-\mu_2}{\Delta}\Big/2\right)=$$

$$=1-P\left(\frac{d^{(3)}(\vec{x})-\mu_2}{\Delta}<\frac{-\mu_2}{\Delta}\Big/2\right)=$$

$$=1-\Phi\left(\frac{-\mu_2}{\Delta}\right)=\Phi\left(\frac{\mu_2}{\Delta}\right),$$

де в останньому переході скористалися тотожністю

$$\Phi(z)+\Phi(-z)=1.$$

Таким чином, отримали, що *ймовірність помилкової класифікації у випадку двох класів* $\{\Omega_i\}_{i=1}^{2}$ має вигляд

$$p_1P(2/1)+p_2P(1/2)=p_1\Phi\left(\frac{-\mu_1}{\Delta}\right)+p_2\Phi\left(\frac{\mu_2}{\Delta}\right). \qquad (2.18)$$

Останній результат (2.18) можна записати й іншим чином, якщо взяти до уваги, що

$$\mu_i=\vec{m}_i^{\,T}V^{-1}(\vec{m}_1-\vec{m}_2)+\ln\left(\frac{p_1}{p_2}\right)+\frac{1}{2}(\vec{m}_2-\vec{m}_1)^T V^{-1}(\vec{m}_2+\vec{m}_1)=$$

$$=\frac{1}{2}(\vec{m}_2-\vec{m}_1)^T V^{-1}(-2\vec{m}_i)+\frac{1}{2}(\vec{m}_2-\vec{m}_1)^T V^{-1}(\vec{m}_2+\vec{m}_1)+\ln\left(\frac{p_1}{p_2}\right)=$$

$$=\frac{1}{2}(\vec{m}_2-\vec{m}_1)^T V^{-1}(\vec{m}_2+\vec{m}_1-2\vec{m}_i)+\ln\left(\frac{p_1}{p_2}\right)=$$

$$= \begin{bmatrix} \frac{1}{2}\Delta^2 + \ln\left(\frac{p_1}{p_2}\right), & \text{якщо } i = 1, \\ -\frac{1}{2}\Delta^2 + \ln\left(\frac{p_1}{p_2}\right), & \text{якщо } i = 2. \end{bmatrix}$$

Тоді *ймовірність помилкової класифікації у випадку двох класів* можна представити у вигляді

$$p_1 P(2/1) + p_2 P(1/2) =$$

$$= p_1 \Phi\left(-\frac{\Delta}{2} - \frac{1}{\Delta}\ln\left(\frac{p_1}{p_2}\right)\right) + p_2 \Phi\left(-\frac{\Delta}{2} + \frac{1}{\Delta}\ln\left(\frac{p_1}{p_2}\right)\right). \quad (2.19)$$

Висновок. Для підрахунку *ймовірності помилкової класифікації у випадку двох класів* $\{\Omega_i\}_{i=1}^2$ можна скористатися результатом (2.18) або (2.19).

2.5. Задача класифікації з вчителем

Розглянемо тепер постановку задачі класифікації об'єктів ω у випадку m класів

$$\{\Omega_i\}_{i=1}^m \quad \left(\Omega = \bigcup_{i=1}^m \Omega_i,\ \Omega_i \bigcap \Omega_j = \varnothing, \forall i \neq j\right),$$

але з ще меншим обсягом апріорної інформації.

А саме вважаємо, що:

I'. Для n-вимірного вектора ознак $\vec{x} = \vec{x}(\omega) \in X$ для кожного класу Ω_i функція щільності $p(\vec{x}, \alpha_i / i)$ відома з точністю до деякого вектора невідомих параметрів $\alpha_i, i = \overline{1,m}$.

III. Задана невід'ємна функція втрат $C\left(j/i\right)$ при віднесенні об'єкта ω до j-го класу, коли він насправді з i-го класу $\left(i,j=\overline{1,m}\right)$.

IV. Доступна навчальна вибірка S, тобто набір векторів ознак за їхніми відомими однозначними класифікаціями, а саме:

$$S=\bigcup_{i=1}^{m}S_i,\ \ S_i\bigcap S_j=\varnothing,\ \forall i\neq j\ ,$$

де

$$S_i=\left\{\vec{x}^{ij}=\vec{x}^{ij}\left(\omega\right):\omega\in\Omega_i,\ j=\overline{1,n_i}\right\},i=\overline{1,m},\quad N=\sum_{i=1}^{m}n_i\ .$$

Тут S_i – підвибірка навчальної вибірки S, яка складається з векторів ознак $\vec{x}^{ij}\left(\omega\right)$, що відповідають об'єктам ω з класу Ω_i, $i=\overline{1,m}$.

А апріорні ймовірності p_i для кожного класу Ω_i залишилися невідомими, $i=\overline{1,m};\ \sum_{i=1}^{m}p_i=1$.

Таким чином, наявна невизначеність у постановці задачі компенсована присутністю навчальної вибірки S. Тобто маємо задачу *класифікації з вчителем* (*supervised classification*).

Слід запропонувати таке правило прийняття рішення про віднесення об'єкта ω до одного з класів $\left\{\Omega_i\right\}_{i=1}^{m}$ на базі доступного вектора ознак $\vec{x}=\vec{x}\left(\omega\right)$, щоб середні втрати були найменшими.

Подібна задача класифікації, але при припущеннях (2.1) – (2.3) була розв'язана раніше. Скористаємося в подальшому цим результатом.

Для розв'язання поточної задачі пропонується використовувати *підстановочне розв'язувальне правило*. Суть його полягає в наступному.

Спочатку скористаємося навчальною вибіркою S для визначення:

- оцінки $\hat{p}_i = \dfrac{n_i}{N}$ для апріорної ймовірності p_i для кожного класу Ω_i, $i = \overline{1,m}$;
- оцінки $\hat{\alpha}_i$ методом максимальної правдоподібності для вектора невідомих параметрів α_i, $i = \overline{1,m}$.

У результаті середні втрати, які необхідно мінімізувати, набувають вигляду

$$\tilde{Q} = \sum_{i=1}^{m} \left\{ \hat{p}_i \left[\sum_{j=1}^{m} C(j/i) \tilde{P}(j/i) \right] \right\},$$

де

$$\tilde{P}(j/i) = \int_{X_j} p(\vec{x}, \hat{\alpha}_i / i) d\vec{x}, \quad i, j = \overline{1,m}.$$

Але ця задача була розглянута раніше і її розв'язок згідно з (2.8) має вигляд

$$\tilde{d}_j(\vec{x}) = -\sum_{i=1}^{m} \left[\hat{p}_i C(j/i) p(\vec{x}, \hat{\alpha}_i / i) \right], \; j = \overline{1,m}. \quad (2.20)$$

У підсумку для розв'язання задачі класифікації об'єктів ω із m класів $\{\Omega_i\}_{i=1}^{m}$ в умовах наявності припущень I', III, IV можна скористатися набором дискримінантних функцій (2.20).

Приклад. Продемонструємо використання підстановочного розв'язувального правила у випадку двох класів $\{\Omega_i\}_{i=1}^{2}$ з нормально розподіленими векторами ознак тобто, коли

$$\vec{x} = \vec{x}(\omega) \sim \mathcal{N}(\vec{m}_i, V_i), (V_i > 0), \; \textit{якщо} \; \omega \in \Omega_i, i = \overline{1,2}.$$

Як невідомі параметри у цьому випадку виступають пари $\{\vec{m}_i, V_i\}_{i=1}^{2}$.

Тоді, використовуючи навчальну вибірку S, маємо можливість визначити оцінки невідомих параметрів:

- $\hat{p}_i = \frac{n_i}{N}, i = \overline{1,2};$

- $\hat{\vec{m}}_i = \frac{1}{n_i}\sum_{j=1}^{n_i} \vec{x}^{ij}, \hat{V}_i = \frac{1}{n_i - 1}\sum_{j=1}^{n_i}\left(\vec{x}^{ij} - \hat{\vec{m}}_i\right)\left(\vec{x}^{ij} - \hat{\vec{m}}_i\right)^T, i = \overline{1,2}.$

Залишилось тільки підставити ці оцінки у представлення (2.20) для функцій $\tilde{d}_j\left(\vec{x}\right),\ j = \overline{1,m}$.

3. ЕЛЕМЕНТИ КЛАСТЕРНОГО АНАЛІЗУ

Познайомимося ще з одним колом задач розпізнавання, для яких доступний зовсім незначний обсяг апріорної інформації про об'єкти, які підлягають розподілу по підмножинах. А саме, відомо тільки те, що об'єкти в кожній такій підмножині близькі (схожі) між собою згідно з деякою мірою, а об'єкти, які не близькі (не схожі), належать до різних таких підмножин. Тобто фактично необхідно розбити множину всіх об'єктів на деякі підмножини (*кластери*), які складаються з близьких (схожих) між собою об'єктів. Розв'язанням проблем такого плану і займається *кластерний аналіз* (*cluster analysis*).

Отже, задача *кластерного аналізу* полягає в тому, що треба так розбити множину всіх об'єктів на такі *кластери* (підмножини), щоб до кожного кластера були віднесені тільки об'єкти близькі (схожі) між собою згідно з вибраною мірою близькості (схожості). Більш детальна інформація про такі міри близькості та схожості буде наведена далі.

Зауважимо, що в загальному випадку кількість кластерів може бути заздалегідь невідомою.

Не можна також забувати і про те, що належність об'єкта до певного кластера може залежати від вибору одиниць вимірювання компонент вектора ознак.

3.1. Вхідна інформація у задачі кластерного аналізу

Доступна інформація про об'єкти, які підлягають кластерізації, може задаватися різним чином. Наведемо деякі можливі варіанти її представлення.

1) Для кожного об'єкта ω_i відомий його n-вимірний вектор ознак $\vec{x}_i$, $i = \overline{1, N}$. Тобто фактично на вході маємо матрицю, яка складається з векторів ознак
$$\begin{pmatrix} \vec{x}_1 & \vec{x}_2 & \dots & \vec{x}_N \end{pmatrix}.$$
2) Для кожної пари об'єктів ω_i та ω_j з векторами ознак $\vec{x}_i$ та $\vec{x}_j$ відповідно, відома *відстань* (*distance*) d_{ij} між ними. Інакше кажучи, на вході отримуємо *матрицю відстаней* (*distance matrix*)
$$\begin{pmatrix} 0 & d_{12} & \dots & d_{1N} \\ d_{21} & 0 & \dots & d_{2N} \\ \vdots & \vdots & \ddots & \vdots \\ d_{N1} & d_{N2} & \dots & 0 \end{pmatrix},$$
де $d_{ij} = d\left(\vec{x}_i, \vec{x}_j\right)$ – відповідне значення *функції відстані* (*distance function*), яка є дійсною функцією, що задовольняє відомі аксіоми:
   1. $d\left(\vec{x}_i, \vec{x}_j\right) \geq 0, \forall \vec{x}_i, \vec{x}_j,$
   2. $d\left(\vec{x}_i, \vec{x}_j\right) = 0 \Leftrightarrow \vec{x}_i = \vec{x}_j,$
   3. $d\left(\vec{x}_i, \vec{x}_j\right) = d\left(\vec{x}_j, \vec{x}_i\right), \forall \vec{x}_i, \vec{x}_j,$
   4. $d\left(\vec{x}_i, \vec{x}_j\right) \leq d\left(\vec{x}_i, \vec{x}_k\right) + d\left(\vec{x}_k, \vec{x}_j\right), \forall \vec{x}_i, \vec{x}_k, \vec{x}_j.$

   Функцію відстані ще називають *метрикою* (*metric*).
3) Для кожної пари об'єктів ω_i та ω_j відоме значення функції схожості s_{ij} для них. Інакше кажучи, на вході маємо *матрицю схожості* (*similarity matrix*)
$$\begin{pmatrix} 1 & s_{12} & \dots & s_{1N} \\ s_{21} & 1 & \dots & s_{2N} \\ \vdots & \vdots & \ddots & \vdots \\ s_{N1} & s_{N2} & \dots & 1 \end{pmatrix},$$

де $s_{ij} = s\left(\vec{x}_i, \vec{x}_j\right)$ – відповідне значення *функції схожості* (*similarity function*), яка є дійсною функцією, що задовольняє аксіоми:

1. $0 \le s\left(\vec{x}_i, \vec{x}_j\right) \le 1, \quad \forall \vec{x}_i, \vec{x}_j$,
2. $s\left(\vec{x}_i, \vec{x}_j\right) = 1 \Leftrightarrow \vec{x}_i = \vec{x}_j$,
3. $s\left(\vec{x}_i, \vec{x}_j\right) = s\left(\vec{x}_j, \vec{x}_i\right), \forall \vec{x}_i, \vec{x}_j$.

Зауваження 1. Очевидно, що функції відстані та схожості взаємно пов'язані. Дійсно, якщо оперують функцією відстані d_{ij} , то вона дозволяє побудувати деяку функцію схожості s_{ij} , наприклад, таким чином:

$$s_{ij} = \frac{1}{1 + d_{ij}}, \quad \forall i, j.$$

Зауваження 2. На перший погляд здається, що можна користуватися і зворотним зв'язком

$$d_{ij} = \frac{1}{s_{ij}} - 1, \quad \forall i, j.$$

Тобто, володіючи деякою функцією схожості s_{ij} , запропонувати функцію відстані d_{ij} . Але з'ясувалося, що це в загальному випадку не так. А саме, не всі функції відстані, побудовані таким чином, задовольняють свою останню аксіому (нерівність трикутника), а тому не завжди можуть відігравати роль функції відстані.

Розглянемо деякі вектори ознак $\vec{x} = \left(x_1, x_2, \ \ldots, x_n\right)^T$ та $\vec{y} = \left(y_1, y_2, \ \ldots, y_n\right)^T$. Наведемо для них ряд прикладів функцій відстані (метрик).

1. *Манхеттенська відстань* (*Manhattan distance*)

$$d_1\left(\vec{x}, \vec{y}\right) = \sum_{i=1}^{n} \left| x_i - y_i \right| .$$

2. *Евклідова відстань* (*Euclidean distance*)

$$d_2\left(\vec{x},\vec{y}\right)=\sqrt{\sum_{i=1}^{n}\left(x_i-y_i\right)^2}\ .$$

3. *Відстань Чебишова* (*Chebyshev distance*)(рівномірна метрика)

$$d_\infty\left(\vec{x},\vec{y}\right)=\max_{i=\overline{1,n}}\left|x_i-y_i\right|.$$

4. *Відстань Мінковського* (*Minkowski distance*)

$$d_p\left(\vec{x},\vec{y}\right)=\left(\sum_{i=1}^{n}\left|x_i-y_i\right|^p\right)^{\frac{1}{p}},\ \ p\geq 1\,.$$

Зауважимо, що три попередніх приклади функцій відстані є частинними випадками функції відстані Мінковського $d_p\left(\vec{x},\vec{y}\right)$ при $p=1$, $p=2$ та $p=\infty$, відповідно.

5. *Відстань Махаланобіса* (*Mahalanobis distance*)

$$d_{V^{-1}}^{(M)}\left(\vec{x},\vec{y}\right)=\sqrt{\left\|\vec{x}-\vec{y}\right\|_{V^{-1}}^2}\ ,$$

де V – матриця відповідної розмірності, $V>0$.

3.2. Ієрархічний алгоритм

Одним із методів, який широко використовується на практиці при розв'язуванні задачі кластерного аналізу, є ієрархічний алгоритм. Саме він знайшов своє широке застосування з огляду на його прозорість та простоту реалізації.

Нехай треба провести процедуру кластерізації для об'єктів $\omega_i,\ i=\overline{1,N}$. Опишемо *ієрархічний алгоритм* (*hierarchical algorithm*) покроково.

1. Спочатку вважаємо, що кожен об'єкт ω_i являє собою окремий кластер Ω_i, $i = \overline{1, N}$. Тобто маємо $\Omega_i = \{\omega_i\}$, $i = \overline{1, N}$. А поточна кількість кластерів $\tilde{m}$ відповідно буде дорівнювати N.
2. Для довільної пари кластерів $\{\Omega_i, \Omega_j\}$ знаходимо міжкластерну відстань $D_{ij} = D(\Omega_i, \Omega_j), i < j$. (Приклади обчислення міжкластерної відстані будуть наведені в подальшому).
3. Знаходимо ту пару кластерів $\{\Omega_i, \Omega_j\}$ для якої міжкластерна відстань є найменшою й об'єднуємо ці кластери в один кластер.
4. Корегуємо поточну кількість кластерів $\tilde{m} := \tilde{m} - 1$.
5. Перевіряємо *умову зупинки ієрархічного алгоритму*. Якщо умова зупинки не виконується, то переходимо до кроку 2, інакше анулюємо останнє злиття, корегуємо значення $\tilde{m} := \tilde{m} + 1$ та зупиняємося. (Приклади умов зупинки ієрархічного алгоритму будуть наведені далі).

Зауваження 1. Якщо на третьому кроці алгоритму кількість пар кластерів $\{\Omega_i, \Omega_j\}$ із найменшою міжкластерною відстанню виявиться більшою, то об'єднувати можна і декілька таких пар кластерів з відповідною корекцією значення поточної кількості кластерів $\tilde{m}$.

Зауваження 2. Наглядну візуалізацію результату роботи ієрархічного алгоритму можна здійснити шляхом його графічного представлення у вигляді *дендрограми* (*dendrogram*).

Зауваження 3. Ієрархічні алгоритми поділяють на *агломеративні* (*agglomerative*) та *дивізимні* (*divisive*). Наведений вище ієрархічний алгоритм є агломеративним. А дивізимний ієрархічний алгоритм є не що інше як дзеркальне відображення агломеративного ієрархічного алгоритму. Тобто на його першому кроці вважають, що усі об'єкти ω_i входять в єдиний спільний кластер $\left(i = \overline{1, N}\right)$. А на наступних кроках цей кластер послідовно розщепляють на кластери з меншою кількістю об'єктів.

3.3. Приклади обчислення міжкластерної відстані

Залежно від методу, який використовується, конкретизується і формула для обчислення міжкластерної відстані. Наведемо деякі її приклади.

1. Метод найближчого сусіда

$$D_{ij} = D_{\min}\left(\Omega_i, \Omega_j\right) = \min_{\substack{\tilde{\omega}\in\Omega_i \\ \tilde{\tilde{\omega}}\in\Omega_j}} d\left(\vec{x}\left(\tilde{\omega}\right), \vec{x}\left(\tilde{\tilde{\omega}}\right)\right).$$

2. Метод найдальшого сусіда

$$D_{ij} = D_{\max}\left(\Omega_i, \Omega_j\right) = \max_{\substack{\tilde{\omega}\in\Omega_i \\ \tilde{\tilde{\omega}}\in\Omega_j}} d\left(\vec{x}\left(\tilde{\omega}\right), \vec{x}\left(\tilde{\tilde{\omega}}\right)\right).$$

3. Метод центроїдів

$$D_{ij} = D_c\left(\Omega_i, \Omega_j\right) = d\left(\bar{\vec{x}}_i, \bar{\vec{x}}_j\right),$$

де

$$\bar{\vec{x}}_k = \frac{1}{card\left(\Omega_k\right)} \sum_{\omega\in\Omega_k} \vec{x}\left(\omega\right), k = i, j.$$

4. Метод середнього зв'язку

$$D_{ij} = \bar{D}\left(\Omega_i, \Omega_j\right) = \frac{1}{card\left(\Omega_i\right) card\left(\Omega_j\right)} \sum_{\tilde{\omega}\in\Omega_i} \sum_{\tilde{\tilde{\omega}}\in\Omega_j} d\left(\vec{x}\left(\tilde{\omega}\right), \vec{x}\left(\tilde{\tilde{\omega}}\right)\right).$$

3.4. Варіанти умов зупинки ієрархічного алгоритму

В основі більшості умов зупинки ієрархічного алгоритму лежить міркування, що відстані між об'єктами всередині кластерів мають бути меншими, ніж між об'єктами з різних кластерів. Введемо деякі поняття.

Означення. *Середня внутрішньокластерна відстань кластера* Ω_i – це числова характеристика $\overline{d}_{in}(\Omega_i)$, яка дорівнює середній відстані між різними об'єктами з кластера Ω_i.

Означення. *Середня зовнішньокластерна відстань кластера* Ω_i – це числова характеристика $\overline{d}_{out}(\Omega_i)$, яка дорівнює середній відстані від об'єктів із кластера Ω_i до об'єктів з усіх інших кластерів.

Наведемо деякі приклади широко вживаних умов зупинки ієрархічного алгоритму.

1. Порушується умова, що для довільного кластера середня внутрішньокластерна відстань менша середньої зовнішньокластерної відстані, тобто умова зупинки має вигляд
$$\exists i \quad \overline{d}_{in}(\Omega_i) \geq \overline{d}_{out}(\Omega_i).$$
У цьому випадку кажуть, що мають справу з *кластерами типу згущення в середньому*.

2. Порушується умова, що для довільного кластера середня внутрішньокластерна відстань менша середньої зовнішньокластерної відстані більш ніж у k раз, тобто умова зупинки предстане в такому вигляді:
$$\exists i \quad \overline{d}_{in}(\Omega_i) \geq \frac{\overline{d}_{out}(\Omega_i)}{k}, \ (k > 1).$$
Вважають, що у цьому випадку стикаються із *сильними кластерами*.

3. Нехай apriori задана загальна кількість кластерів m_*. Тоді алгоритм зупиняється, якщо останнє злиття кластерів привело до загальної кількості кластерів, що менша ніж m_*. Тобто умова зупинки алгоритму набуває вигляду $\tilde{m} < m_*$.

ДОДАТОК

Лінійне перетворення нормально розподіленого вектора

Розглянемо деякий випадковий нормально розподілений вектор

$$\vec{\xi} \sim \mathcal{N}\left(\vec{m}, V\right), V > 0,\ \vec{\xi} \in \mathbb{R}^n,$$

тобто вважаємо, що його щільність має вигляд

$$p_{\vec{\xi}}\left(\vec{x}\right) = \frac{1}{\left(2\pi\right)^{\frac{n}{2}}\left(\det\left(V\right)\right)^{\frac{1}{2}}} e^{-\frac{1}{2}\left\|\vec{x}-\vec{m}\right\|_{V^{-1}}^{2}}.$$

Стосовно лінійних перетворень таких векторів справедливо таке твердження.

Т е о р е м а . Припустимо, що

- $\vec{\xi} \sim \mathcal{N}\left(\vec{m}, V\right), V > 0,\ \vec{\xi} \in \mathbb{R}^n,$
- $\vec{a} \in \mathbb{R}^m$, B – матриця розмірності $m \times n$,

тоді справедливо

$$\vec{a} + B\vec{\xi} \sim \mathcal{N}\left(\vec{a} + B\vec{m}, BVB^T\right).$$

Легко бачити, що остання теорема приводить до таких висновків.

Наслідок 1. Якщо $\vec{\xi} \sim \mathcal{N}\left(\vec{m}, V\right), V > 0,\ \vec{\xi} \in \mathbb{R}^n,$ то

$$V^{-\frac{1}{2}}\left(\vec{\xi} - \vec{m}\right) \sim \mathcal{N}\left(\theta_n, E_n\right)$$

де θ_n – нульовий вектор розмірності n, E_n – одинична матриця розмірності $n \times n$.

Наслідок 2. Нехай $\vec{\xi} \sim \mathcal{N}\left(\vec{m}, V\right), V > 0,\ \vec{\xi} \in \mathbb{R}^n,$ тоді

$$\left(\vec{\xi} - \vec{m}\right)^T V^{-1}\left(\vec{\xi} - \vec{m}\right) \sim \chi^2\left(n\right).$$

Наслідок 3. Якщо $\vec{\xi} \sim \mathcal{N}\left(\vec{m}, V\right), V > 0,\ \vec{\xi} \in \mathbb{R}^n,$ то його можна представити як лінійне перетворення стандартно нормально розподіленого вектора, а саме:

$$\vec{\xi} = \vec{m} + V^{\frac{1}{2}} \vec{\xi}_* ,$$

де $\vec{\xi}_* \sim \mathcal{N}\left(\theta_n, E_n\right)$.

СПИСОК ЛІТЕРАТУРИ

1. *Айвазян С. А.* Прикладная статистика и основы эконометрики / С. А. Айвазян, В. С. Мхитарян. – М. : ЮНИТИ, 1998.
2. *Андерсен Т.* Введение в многомерный статистический анализ / Т. Андерсен. – М. : Физматгиз, 1963.
3. *Афифи А.* Статистический анализ. Подход с использованием ЭВМ / А. Афифи, С. Эйзен. – М. : Мир, 1982.
4. *Бишоп К. М.* Распознавание образов и машинное обучение / К. М. Бишоп. – К. : Диалектика, 2020.
5. *Большев Л. Н.* Таблицы математической статистики / Л. Н. Большев, Н. В. Смирнов. – 3-е изд. – М. : Наука. Гл. редакция физ.-мат. лит-ры, 1983.
6. *Вапник В. Н.* Теория распознавания образов. Статистические проблемы обучения / В. Н. Вапник, А. Я. Червоненкис. – М. : Наука, Главная редакция физико-математической литературы, 1974.
7. *Гайдышев И.* Анализ и обработка данных: спец. справочник / И. Гайдышев. – СПб. : Питер, 2001.
8. *Дюран Б.* Кластерный анализ / Б. Дюран, П. Оделл. – М. : Статистика, 1977.
9. *Жамбю М.* Иерархический кластер-анализ и соответствия / М. Жамбю. – М. : Финансы и статистика, 1988.
10. Классификация и кластер / ред. Дж. Вэн Райзин. – М. : Мир, 1980.
11. *Лепский А. Е.* Математические методы распознавания образов : курс лекций / А. Е. Лепский, А. Г. Броневич. – Таганрог : Изд-во ТТИ ЮФУ, 2009.
12. *Мандель И. Д.* Кластерный анализ / И. Д. Мандель. – М. : Финансы и статистика, 1988.
13. *Патрик Э.* Основы теории распознавания образов / Э. Патрик. – М. : Сов. радио, 1980.
14. Прикладная статистика. Классификация и снижение размерности / С. А. Айвазян, В. М. Бухштабер, И. С. Енюков,

Л. Д. Мешалкин; под ред. С. А. Айвазяна. – М. : Финансы и статистика, 1989.
15. *Ту Дж.* Принципы распознавания образов / Дж. Ту, Р. Гонсалес. – М. : Мир, 1978.
16. Факторный, дискриминантный и кластерный анализ / Дж.-О. Ким, Ч. У. Мьюллер, У. Р. Клекка и др. ; под ред. И. С. Енюкова. – М. : Финансы и статистика, 1989.
17. Cluster Analysis / Brian S. Everitt, Sabine Landau, Morven Leese, Daniel Stahl. – 5th ed. – Chichester, United Kingdom : John Wiley & Sons, Ltd, 2011.
18. *Duda R.O.* Pattern Classification / R.O. Duda, P.E. Hart, D.G. Stork. – 2nd ed. – New York, NY : John Wiley & Sons, Inc., 2001.
19. *Fukunaga K.* Introduction to Statistical Pattern Recognition / Keinosuke Fukunaga. – 2nd ed. – San Diego, CA : Academic Press, Inc., 1990.
20. *Webb A. R.* Statistical Pattern Recognition / Andrew R. Webb, Keith D. Copsey. – 3rd ed. – Chichester, United Kingdom : John Wiley & Sons, Ltd, 2011.

ПРЕДМЕТНИЙ ПОКАЖЧИК

ПОКАЖЧИК ПОЗНАЧЕНЬ

ω – об'єкт (процес, явище, ситуація), що підлягає класифікації.

Ω – простір усіх об'єктів.

$\left\{\Omega_i\right\}_{i=1}^{m}$ – класи.

$\vec{y}=\vec{y}(\omega)$ – вектор спостережень.

Y – простір спостережень.

$\vec{x}=\vec{x}(\omega)$ – вектор ознак.

X – простір ознак.

$\left\{d_i(\vec{x})\right\}_{i=1}^{m}$ – набір дискримінантних функцій.

$\xi \sim \mathcal{N}\left(m,\sigma^2\right)$ – випадкова величина ξ нормально розподілена з параметрами m та σ^2.

$\vec{\xi} \sim \mathcal{N}(\vec{m},V)$ – випадковий вектор $\vec{\xi}$ нормально розподілений з параметрами $\vec{m}$ та V.

$\Phi(z)$ – функція розподілу стандартного нормального розподілу $\mathcal{N}(0,1)$.

$d_p(\vec{x},\vec{y})$ – відстань Мінковського між векторами $\vec{x}$ та $\vec{y}$, $p \geq 1$.

$d_{V^{-1}}^{(M)}(\vec{x},\vec{y})$ – відстань Махаланобіса між векторами $\vec{x}$ та $\vec{y}$, матриця $V>0$.

$card(A)$ – потужність множини A.

$\overline{d}_{in}(\Omega_k)$ – середня внутрішньокластерна відстань кластера Ω_k.

$\overline{d}_{out}(\Omega_k)$ – середня зовнішньокластерна відстань кластера Ω_k.

$\|\vec{x}\|_Q^2=\vec{x}^T Q\vec{x}, Q>0$.

ЗМІСТ

Навчальне видання

СЛАБОСПИЦЬКИЙ Олександр Сергійович

Задачі класифікації

Навчальний посібник

Друкується за авторською редакцією

Підписано до друку 27.10.2020. Формат 60х84/16.
Папір офсетний. Гарнітура Таймс. Друк офсетний.
Ум. друк. арк. 2,7. Наклад 100. Зам. № 421.

Надруковано в “Видавництво Людмила”.
Свідоцтво про внесення до Державного реєстру
суб’єктів видавничої справи серія ДК № 5303 від 02.03.2017.
“ Видавництво Людмила ”
03148, Київ, а/с 115.
Тел./факс: + 38 050 469 7485, 068 340 8332
E-mail: lesya3000@ukr.net