N.A.Ignatev

Intellektual tizimlar bo‘yicha masalalar to‘plami

Uslubiy qo‘llanma

Сборник заданий по интеллектуальным системам

Методическое пособие

Тошкент 2023

Mundarija (Оглавнение)

Kirish (Введение)

Intellektual tizimlar bo‘yicha masalalar to‘plami kompyuterli ko‘rish, nutqni anglash, ko‘p o‘lchamli berilganlarni vizuallashtirish va obrazlarni anglash masalalarini echishda nazariy bilimlarni mustahkamlash hamda alomatlar fazosini shakllantirish jarayonini avtomatlashtirishda amaliy ko‘nikmalarni egallashga qaratilgan.

Masalalarni amalga oshirish, tenzorlar ustida amallar, alomatlar qiymatlarini o‘zaro kesishmaydigan intervallarga bo‘lish, to‘plamga tegishlilik funksiyasi qiymatini hisoblash va klassifikatorlar ansamblini qurish usullaridan foydalanishga asoslangan.

Alomatlar fazosini shakllantirish muammosi intellekual tizimlarning BigData, berilganlarni intellektual tahlili va sun’iy neyron to‘rlarni chuqur o‘rgatish sohalarida ayniqsa dolzarbdir.

Сборник заданий по интеллектуальным системам (ИС) предназначен для закрепления теоретических знаний и приобретения практических навыков по автоматизatsii процесса формирования признакового пространства в задачах компьютерного зрения, распознавания речи, визуализации многомерных данных, распознавания образов.

Практическая реализация заданий основываются на использовании операtsiй над тензорами, разбиении значений признаков на непересекающиеся интервалы, вычислении значений функции принадлежности к множествам, построении ансамблей классификаторов.

Особую актуальность решение проблемы формирования признакового пространства имеет в таких областях ИС как BigData, интеллектуальный анализ данных (Data minning), глубокое обучение искусственных нейронных сетей.

1.Berilganlarni razvedka tahlili (Разведочный анализ данных)

Berilganlarni razvedka tahlili (ingl. Exploratory data analysis, EDA) – bu berilganlarning asosiy xossalarini tahlili bo'lib, berilganlardagi umumiy qonuniyatlarni, taqsimotlar va anomaliyalar aniqlash, ayrim hollarda vizuallashtirish vositalaridan foydalangan holda.

Ushbu tushuncha matematik Djonom Tyuki tomonidan kiritilgan. U bunday tahlilning maqsadini quyidagicha shakllantirib berdi:

- maksimal ravishda berilganlarga «kirib borish»;
- asosiy tuzilmalarni yuzaga chiqarish;
- eng muhim o'zgaruvchilarni tanlab olish;
- cheklanishlar va anomaliyalarni aniqlash;
- asosiy gipotezalarni tekshirish;
- boshlang'ich modellarni ishlab chiqish.

Razvedka tahlilining asosiy vositalari – bu o'zgaruvchilarni taqsimlanish ehtimolligini o'rganish, korrelyatsiya matritsalarini qurish va tahlil qilish, faktor tahlil, ko'p o'lchamli shkalalash.

Разведочный анализ данных (англ. *Exploratory data analysis, EDA*) — анализ основных свойств данных, нахождение в них общих закономерностей, распределений и аномалий, построение начальных моделей, зачастую с использованием инструментов визуализatsiи.

Понятие введено математиком Джоном Тьюки, который сформулировал цели такого анализа следующим образом:

- максимальное «проникновение» в данные,
- выявление основных структур,
- выбор наиболее важных переменных,
- обнаружение отклонений и аномалий,
- проверка основных гипотез,
- разработка начальных моделей.

Основные средства разведочного анализа–изучение вероятностных распределений переменных, построение и анализ корреляционных матриц, факторный анализ, дискриминантный анализ, многомерное шкалирование.

1.1 Berilgan n ta test savolga m ta javob beruvchilarning javoblari $X = \{x_{ij}\}_{m\times n}$ matritsa ko'rinishida saqlanadi, agar $x_{ij} = 1$ bo'lsa, i-chi savolga j-chi javob beruvchi to'g'ri javob berganini anglatadi va $x_{ij} = 0$

bo‘lsa noto‘g‘ri javob bo‘ladi. Har bir k- savolning W_k vaznini va p-javob beruvchining R_p nisbiy bahosini hisoblansin:

$$W_k = \frac{q_k}{m_k}\sum_{i=1}^{m} x_{ik}\exp\left(-\left(1-\frac{n_i}{n}\right)\right),$$

$$R_p = \exp\left(-\left(1-\frac{n_p}{n}\right)\right)\sum_{j=1}^{n} x_{pj}\, q_j.$$

Bu erda $q_i = 1-\frac{m_i}{m+1}$, $m_i = \sum_{j=1}^{m} x_{ji}$, $n_i = \sum_{j=1}^{n} x_{ij}$.

1.1 Ответы m тестируемых на n тестовых задач (вопросов) хранятся в виде матрицы $X = \{x_{ij}\}_{m\times n}$, где $x_{ij} = 1$ означает, что - ый тестируемый правильно ответил на j-ый вопрос, $x_{ij} = 0$ соответственно неправильно. Вычислить вес каждой -ой задачи как $W_k = \frac{q_k}{m_k}\sum_{i=1}^{m} x_{ik}\exp\left(-\left(1-\frac{n_i}{n}\right)\right)$ и относительную оценку p –го тестируемого как

$$R_p = \exp\left(-\left(1-\frac{n_p}{n}\right)\right)\sum_{j=1}^{n} x_{pj}\, q_j,$$

где $q_i = 1-\frac{m_i}{m+1}$, $m_i = \sum_{j=1}^{m} x_{ji}$, $n_i = \sum_{j=1}^{n} x_{ij}$.

1.2. Tekislikda berilgan $S_1, ..., S_n, n > 2$, nuqtalarni 2 ta kesishmaydigan sinfga har xil ajratishlar sonini hisoblovchi algoritm amalga oshirilsin (dixotomiya masalasi).

1.2. Реализовать алгоритм вычисления числа различных разбиений точек $S_1, ..., S_n, n > 2$, на плоскости на 2 непересекающихся класса (задача дихотомии).

1.3. Genom molekulalari tavsiflash uchun to'rtta ‘A’, ‘B’, ‘C’, ‘D’ harflarning turli kombinatsiyalari ko'rinishidagi belgilar zanjiridan (ketma-ketligidan) foydalaniladi. Berilgan m ketma-ketlikning har biri $n_i, i = 1, \dots, m$ belgilardan iborat bo'lib, ulardan harflarning ma'lum bir ko'rinishlari – ‘A’, ‘B’, ‘C’, ‘D’, ‘AB’, ‘AC’, ‘AD’, ‘BA’, ‘BC’, ‘BD’, ‘CA’, ‘CB’, ‘CD’, ‘DA’, ‘DB’, ‘DC’ bo'yicha alomatlari aniqlangan T “*ob'ekt-xossa*” jadvali shakllantirilgan. Alomatlar qiymatlari belgilar va ularning birlashmalarining ketma-ketlikka kirishlari miqdori bo’ladi. Berilgan T matritsasi bo’yicha $R = \{r_{ij}\}$ korrelyatsiya matritsasi hisoblansin va maksimal qiymatli alomatlar juftligi ajratib ko’rsatilsin.

1.3. Для описания молекулы генома используются цепочки символов (последовательности) в виде различных комбинатsий из четырёх букв ‘A’, ‘B’, ‘C’, ‘D’. По заданным m последовательностям, каждая из которых состоит из $n_i, i = 1, \dots, m$ символов, формируется таблица “*объект-свойство*” T по определяемому перечню признаков: ‘A’, ‘B’, ‘C’, ‘D’, ‘AB’, ‘AC’, ‘AD’, ‘BA’, ‘BC’, ‘BD’, ‘CA’, ‘CB’, ‘CD’, ‘DA’, ‘DB’, ‘DC’. Значениями признаков являются количество их (символов или их сочетаний) вхождений в состав каждой последовательности. Требуется по таблице T вычислить значения матрицы корреляций $R = \{r_{ij}\}$ и выделить пару признаков с максимальным значением.

1.4. Berilgan $A = \{S_1, \ldots, S_m\}$ matritsada har bir $S_i \in A$ obyekt n ta nominal alomatlar bilan tavsiflanadi $S_i = (a_{i1}, \ldots, a_{in})$. Har bir (x, y) nominal alomatlar juftligi uchun bog‘liqlik $N(x, y)$ matritsasini quring va

$$\sum_{i=1}^{l_x}\sum_{j=1}^{l_y}\frac{\left(n_{ij}-\frac{n_{i.}n_{.j}}{n}\right)^2}{\frac{n_{i.}n_{.j}}{n}}$$

formula bo‘yicha hisoblanadigan kvadratik bog‘liqlik koeffitsiyenti X^2 qiymatini hisoblansin. Bu erda $n_{.j} = \sum_{i=1}^{l_x} n_{ij}$; $n_{i.} = \sum_{j=1}^{l_y} n_{ij}$, l_x, l_y – mos ravishda x va y alomatlarning gradatsiyalar soni.

1.4. В заданной матрице $A = \{S_1, \ldots, S_m\}$ каждый объект $S_i \in A$ описывается набором из n номинальных признаков $S_i = (a_{i1}, \ldots, a_{in})$. Для каждой пары номинальных признаков (x, y) требуется построить матрицу сопряжённости $N(x, y)$ и вычислить значения коэффициента квадратичной сопряжённости X^2 по формуле

$$\sum_{i=1}^{l_x}\sum_{j=1}^{l_y}\frac{\left(n_{ij}-\frac{n_{i.}n_{.j}}{n}\right)^2}{\frac{n_{i.}n_{.j}}{n}},$$

где $n_{.j} = \sum_{i=1}^{l_x} n_{ij}$; $n_{i.} = \sum_{j=1}^{l_y} n_{ij}$, l_x, l_y – количество градаций соответственно признаков x и y.

1.5. m obyektdan tashkil topgan va n miqdoriy alomatlar tavsiflangan ob'ektlar to'plami $A = \{S_1, \ldots, S_m\}$ berilgan. K yaqin qo'shni usuli bilan berilgan $S_i \in A$ obyekt uchun taqsimot zichligi baholansin. Evklid va Chebishev metrikalari bo'yicha hisoblangan taqsimot zichliglari solishtirilsin.

1.5. Множество $A = \{S_1, \ldots, S_m\}$ из m объектов описывается набором из n количественных признаков. Оценить плотность распределения для заданного объекта $S_i \in A$ методом k ближайших соседей. Сравнить значения плотности по метрике Евклида и Чебышева.

1.6. m obyektdan tashkil topgan n o'lchovli alomatlar fazosida $A = \{S_1, \ldots, S_m\}$ tanlanma berilgan. Berilgan h kenglik bo'yicha parzen darchasi usuli bilan berilgan $S_i \in A$ obyekt uchun taqsimot zichligini baholansin.

1.6. Множество $A = \{S_1, \ldots, S_m\}$ из m объектов описывается набором из n количественных признаков. Оценить плотность распределения для заданного объекта $S_i \in A$ методом парзеновского окна по заданной ширине h.

1.7. S_i va S_j obyektlar o'rtasidagi $S_i\, R\, S_j$, $i, j \in \{1, \ldots, m\}$ munosabat $A = \{a_{ij}\}_{m*m}$ kvadratik matritsa ko'rinishida beriladi, agar $S_i\, R\, S_j$ rost bo'lsa $a_{ij} = 1$, aks holda $a_{ij} = 0$. A matritsa uchun R munosabat bo'yicha tranzitivlik xossasi bajariladimi? Agar bajarilsa, tranzitivlik xossasi rost bo'ladigan obyektlar uchligi ko'rsatilsin.

1.7. Отношение $S_i\, R\, S_j$, $i, j \in \{1, \ldots, m\}$ между объектами S_i и S_j задаётся в виде квадратной матрицы $A = \{a_{ij}\}_{m\times m}$. $a_{ij} = 1$ если $S_i\, R\, S_j$ истинно и $a_{ij} = 0$ если $S_i\, R\, S_j$ ложно. По матрице A определить выполняется ли свойство транзитивности по отношению R. Если да, то указать все тройки объектов для которых отношение транзитивности истинно.

1.8. S_i va S_j obyektlar o'rtasidagi $S_i\, R\, S_j$, $i, j \in \{1, \ldots, m\}$ munosabat $A = \{a_{ij}\}_{m*m}$ kvadratik matritsa ko'rinishida beriladi, agar $S_i\, R\, S_j$ rost bo'lsa $a_{ij} = 1$, aks holda $a_{ij} = 0$. A matritsa uchun R munosabat xossasi ekvivalentlik sinfini hosil qiladimi?

1.8. Отношение $S_i\ R\ S_j$, $i,j \in \{1,\dots,m\}$ между объектами S_i и S_j задаётся в виде квадратной матрицы $A = \{a_{ij}\}_{m\times m}$. $a_{ij} = 1$ если $S_i\ R\ S_j$ истинно и $a_{ij} = 0$ если $S_i\ R\ S_j$ ложно. Определить образуют ли свойства отношения R класс эквивалентности.

1.9. S_i va S_j obyektlar o‘rtasidagi $S_i\ R_1\ S_j$, $i,j \in \{1,\dots,m\}$ munosabat $A = \{a_{ij}\}_{m*m}$ kvadratik matritsa ko‘rinishida, $S_i\ R_2\ S_j$, $i,j \in \{1,\dots,m\}$ munosabat esa $B = \{b_{ij}\}_{m*m}$ ko‘rinishda berilgan, agar $S_i\ R_1\ S_j$ rost bo‘lsa $a_{ij} = 1$, aks holda $a_{ij} = 0$. Xuddi shunday B matritsa elementlari R_2 munosabat bo‘yicha aniqlanadi. Aniqlansin, munosabatlarning $R_1 \cup R_2$ o‘rami qat’iy tartiblangan sinf hosil qiladimi?

1.9. Отношение $S_i\ R_1\ S_j$, $i,j \in \{1,\dots,m\}$ между объектами S_i и S_j задаётся в виде квадратной матрицы $A = \{a_{ij}\}_{m\times m}$, отношение $S_i\ R_2\ S_j$, $i,j \in \{1,\dots,m\}$ в виде матрицы $B = \{b_{ij}\}_{m*m}$. $a_{ij} = 1$ если $S_i\ R_1\ S_j$ истинно и $a_{ij} = 0$ если $S_i\ R_1\ S_j$ ложно. Аналогично определяются элементы матрицы B по отношению R_2. Определить, образует ли свёртка отношений $R_1 \cup R_2$ класс строгий порядок.

1.10. Uchta ekspertlarning S_i va S_j obektlar o’rtasidagi R munosabat bo’yicha $S_i\ R\ S_j$, $i,j \in \{1,\dots,m\}$ xulosalari $A_1 = \{a^1_{ij}\}_{m\times m}$, $A_2 = \{a^2_{ij}\}_{m\times m}$ и $A_3 = \{a^3_{ij}\}_{m\times m}$ko’rinishidagi matritsalar bilan berilgan. Xulosalari bir-biriga eng yaqin bo’lgan ikkita ekspert aniqlansin.

1.10. Заключения трёх экспертов по отношению R, $S_i\ R\ S_j$, $i,j \in \{1,\dots,m\}$ между объектами S_i и S_j задаётся в виде квадратных матриц. Считается, что $a_{ij} = 1$ если $S_i\ R\ S_j$ истинно и $a_{ij} = 0$ если $S_i\ R\ S_j$ ложно. Определить двух экспертов, мнения которых наиболее близки друг к другу.

1.11. S_i va S_j obyektlar o‘rtasidagi $S_i\ R_1\ S_j$, $i,j \in \{1,\dots,m\}$ munosabat $A = \{a_{ij}\}_{m*m}$ kvadratik matritsa ko‘rinishida, $S_i\ R_2\ S_j$, $i,j \in \{1,\dots,m\}$ munosabat esa $B = \{b_{ij}\}_{m*m}$ ko‘rinishda berilgan, agar $S_i\ R_1\ S_j$ rost bo‘lsa aij=1, aks holda $a_{ij} = 0$. Xuddi shunday B matritsa elementlari R_2 munosabat bo‘yicha aniqlanadi. Aniqlansin, $R_1 \cap R_2$ munosabatlar o‘rami antirefleksivlik xossasiga egami?

1.11. Отношение $S_i\ R_1\ S_j$, $i,j \in \{1,\dots,m\}$ между объектами S_i и S_j задаётся в виде квадратной матрицы $A = \{a_{ij}\}_{m*m}$, отношение

$S_i\, R_2\, S_j$, $i, j \in \{1, \dots, m\}$ в виде матрицы $B = \{b_{ij}\}_{m*m}$. $a_{ij} = 1$ если $S_i\, R_1\, S_j$ истинно и $a_{ij} = 0$ если $S_i\, R_1\, S_j$ ложно. Аналогично определяются элементы матрицы B по отношению R_2. Определить, обладает ли свёртка отношений $R_1 \cap R_2$ свойством антирефлексивности.

1.12. Berilgan $A = \{S_1, \dots, S_m\}$ tanlanma bo‘yicha $S \in A$, $S = (x_1, \dots, x_n)$ obyektlarni R^n fazodan sonlar o‘qiga

$$P(S) = w_1 x_1 + \cdots + w_n x_n.$$

Formula bilan akslantirilsin. Koeffitsiyentlar sifatida

$$w_i = 1 - \frac{\sum_{u=1}^{m} \sum_{v=1}^{m} \rho^i(S_u, S_v)}{\sum_{u=1}^{m} \sum_{v=1}^{m} \rho(S_u, S_v)}$$

qiymatlaridan foydalanilsin. Bu erda $\rho(S_u, S_v)$, $\rho^i(S_u, S_v)$ – mos ravishda S_u va S_v obyektlar orasidagi evklid metrikasi bo‘yicha R^n va R^{n-1} fazodagi masofa (i- alomat o’chirilgan jadval bo’yicha).

1.12. По заданной выборке $A = \{S_1, \dots, S_m\}$ произвести отображение описаний объектов $S \in A$, $S = (x_1, \dots, x_n)$ из R^n на числовую прямую по формуле

$$P(S) = w_1 x_1 + \cdots + w_n x_n.$$

В качестве коэффициентов использовать значения

$$w_i = 1 - \frac{\sum_{u=1}^{m} \sum_{v=1}^{m} \rho^i(S_u, S_v)}{\sum_{u=1}^{m} \sum_{v=1}^{m} \rho(S_u, S_v)},$$

где $\rho(S_u, S_v)$, $\rho^i(S_u, S_v)$ расстояние по евклидовой метрике между объектами S_u и S_v соответственно в R^n и R^{n-1} (по таблице из которой удалён -й признак).

1.13. n ta miqdoriy alomat bilan tavsiflanadigan har bir $S_i \in E_0$ obyekt uchun $E_0 = \{S_1, \dots, S_m\}$ tanlanma bo‘yicha tasodifiy μ va $\beta_1, \dots, \beta_t$ kattaliklar guruhlar o‘rtasidagi to‘plamli korrelatsiya koeffitsiyentini hisoblash dasturi tuzilsin va u

$$R = \sqrt{1 - \frac{D}{D_\beta}}$$

formuladan aniqlanadi, bu erda D – $\mu, \beta_1, \dots, \beta_t$ kattaliklarning korrelatsiya matritsa aniqlovchisi, D_β ham xuddi shunday, faqat u $\beta_1, \dots, \beta_t$ qiymatlar uchun.

1.13. По заданной выборке $E_0 = \{S_1, \dots, S_m\}$ в которой каждый объект $S_i \in E_0$ описывается n количественными признаками из набора $H(n) = (x_1, \dots, x_n)$ реализовать вычисление значения коэффициента множественной корреляции между случайной величиной μ и группой случайных величин $\beta_1, \dots, \beta_t$ как

$$R = \sqrt{1 - \frac{D}{D_\beta}},$$

где D – определитель матрицы корреляций величин $\mu, \beta_1, \dots, \beta_t$, D_β то же самое, но только для величин $\beta_1, \dots, \beta_t$.

1.14. Satrlari bo‘yicha n miqdoriy alomatlar bilan tavsiflangan m obyektlar joylashgan $A = \{a_{ij}\}_{m\times n}$ jadval berilgan bo‘lib, alomatlardan biri (tanlash orqali) maqsad, qolganlari bog‘liq alomatlar hisoblanadi. $Y = (y_1, \dots, y_m)$ – maqsad va $A_1 = (a_{11}, \dots, a_{1,n-1}), \dots, A_m = (a_{m1}, \dots, a_{m,n-1})$ bog‘liq vektorlar qiymatlari $y_1 = F(a_{i1}, \dots, a_{i,n-1})$ funksional bilan bog‘langan bo‘lsin.

Bog‘liq ko‘rsatkichlar vektori $B = (b_1, \dots, b_{n-1})$ bo‘yicha prognoz qilinuvchi y qiymati $y = \frac{1}{k}(y_{i_1} + \dots + y_{i_k})$ ko‘rinishida hisoblanadi. Bu erda $y_{i_1}, \dots, y_{i_k}$ – $A_{i_1}, \dots, A_{i_k}$ bog‘liq alomatlar bo‘yicha berilgan $\rho(x, y)$ metrika bilan hisoblangan k yaqinlarning maqsad ko‘rsatkich qiymatlari.

Quyidagi mezon bo‘yicha sirpanuvchi usul bilan k parametrining optimal qiymati tanlanadi:

$$F(k) = \sum_{i=1}^{m} (y_i - \hat{y}_i)^2 \to min$$

Bu erda $\hat{y}_i$ – prognoz qilinuvchi qiymat.

1.14. В заданной таблице $A = \{a_{ij}\}_{m\times n}$ по строкам которой содержатся m объектов, описываемых n количественными признаками, один из признаков (по выбору) считается целевым, остальные зависимыми. Считается, что значения целевого вектора $Y = (y_1, \dots, y_m)$ и зависимых $A_1 = (a_{11}, \dots, a_{1,n-1}), \dots, A_m = (a_{m1}, \dots, a_{m,n-1})$ связаны функциональной зависимостью $y_1 = F(a_{i1}, \dots, a_{i,n-1})$.

Прогнозируемое значение y по вектору зависимых показателей $B = (b_1, \dots, b_{n-1})$ вычисляются как

$$y = \frac{1}{k}(y_{i_1} + \dots + y_{i_k})$$

где $y_{i_1}, \dots, y_{i_k}$ – значения целевых показателей у k ближайших по заданной метрике $\rho(x, y)$ векторов зависимых признаков $A_{i_1}, \dots, A_{i_k}$.

Оптимальное значение k выбирается методом скользящего экзамена по критерию

$$F(k) = \sum_{i=1}^{m} (y_i - \widehat{y_i})^2 \to min$$

где $\widehat{y_i}$ – прогнозируемое значение.

1.15. Berilgan $A=\{a_{uv}\}_{m*n}$ jadvalida m ta $S_1,\ldots,S_m$ ob'ektlar to'plamini $X(n)=(x_1,\ldots,x_n)$ miqdoriy alomatlar bilan tavsifi berilgan. Maqsad alomat sifatida $x_i \in X(n)$ alomat tanlangan bo'lib, $X(n\text{-}1)= X(n)\backslash\{x_i\}$ alomatlar bog'liq hisoblanadi. $S_1,\ldots,S_m$ ob'ektlar $y_1,\ldots,y_m$ sifatida identifikatsiya qilinuvchi maqsad alomatlari uchun $y_i=F(X(p))$ ko'rinishidagi funksional bog'liqlik aniqlangan, bu erda $X(p)\subset X(n\text{-}1)$, $2\leq p\leq n\text{-}1$. Prognoz qilinuvchi y alomatning $X(p)$ bog'liq alomatlar bo'yicha qiymati

$$y(X(p)) = \frac{1}{k}\left(y_{i_1} + \ldots \quad + y_{i_k}\right)$$

ko'rinishida hisoblandi, bu erda $y_{i_1},\ldots,y_{i_k}$ – $X(p)$ alomatlar bo'yicha evklid metrikasi bo'yicha eng yaqin k ta $S_{i_1},\ldots,S_{i_k}$ ob'ektlar maqsad alomatlarining qiymatlari.

Alomatlarning $X(p)$ to'plam ostisi bo'yicha k ning optimal qiymati sirpanuvchi imtihon usuli bilan

$$F(k) = \sum_{i=1}^{m}\left(y_i - y_i(X(p))\right)^2 \to \min,$$

mezon bo'yicha tanlanadi, bu erda $y_i(X(p))$ – prognoz qilinayotgan qiymat.

O'zaro kesishmaydigan $X(p_1)$, $X(p_2)$, $X(p_3)$ to'plam ostilari bo'yicha k ning optimal qiymatini topish talab qilinadi va aniqlik quyidagicha hisoblansin:

$$R(X(p_1), X(p_2), X(p_3)) = \sum_{i=1}^{m}\left(y_i - \frac{y_i(X(p_1)) + y_i(X(p_2)) + y_i(X(p_3))}{3}\right)^2.$$

1.15. Таблица $A=\{a_{uv}\}_{m*n}$ содержит описания множества из m объектов $S_1,\ldots,S_m$ по набору количественных признаков $X(n)=(x_1,\ldots,x_n)$. Признак $x_i \in X(n)$ выбран в качестве целевого, $X(n\text{-}1)= X(n)\backslash\{x_i\}$ являются зависимыми. Для значений целевого признака объектов $S_1,\ldots,S_m$, идентифицируемых как $y_1,\ldots,y_m$, определена функциональная зависимость $y_i=F(X(p))$, где $X(p)\subset X(n\text{-}1)$, $2\leq p\leq n\text{-}1$.

Прогнозируемое значение y по значениям набора зависимых признаков $X(p)$ вычисляются как

$$y(X(p)) = \frac{1}{k}\left(y_{i_1} + \quad + y_{i_k}\right)$$

где $y_{i_1},\ldots,y_{i_k}$ – значения целевых показателей у k ближайших по евклидовой метрике объектов $S_{i_1},\ldots,S_{i_k}$ на наборе $X(p)$.

Оптимальное значение k на наборе $X(p)$ выбирается методом скользящего экзамена по критерию

$$F(k) = \sum_{i=1}^{m}\left(y_i - y_i(X(p))\right)^2 \to \min,$$

где $y_i(X(p))$ – прогнозируемое значение.

Требуется получить оптимальные значения k по трём несовпадающим наборам $X(p_1)$, $X(p_2)$, $X(p_3)$ и оценить точность как

$$R(X(p_1), X(p_2), X(p_3)) = \sum_{i=1}^{m}\left(y_i - \frac{y_i(X(p_1)) + y_i(X(p_2)) + y_i(X(p_3))}{3}\right)^2.$$

1.16. Berilgan $A=\{a_{uv}\}_{m*n}$ jadvalida m ta $S_1,\ldots,S_m$ ob’ektlar to‘plamini $X(n)=(x_1,\ldots,x_n)$ miqdoriy alomatlar bilan tavsifi berilgan. Har bir $(x_i, x_j)\in X(n)$ alomatlar juftligi bo‘yicha aniqlash talab etiladi:

– mumkin bo‘lgan intervallar $[z_1,z_2]$ chegarasini, bu erda, $z_1 = \min\limits_{E_0} R(i, j)$, $z_2 = \max\limits_{E_0} R(i, j)$, $R(i,j)=\{a_{ki}/P_i - a_{kj}/P_j\}_{k\in\{1,\ldots,m\}}$, $S_k=(a_{k1},\ldots,a_{kn})$, P_i,P_j – (x_i,x_j) alomatlarning E_0 to‘plamdagi qiymatlarining matematik kutilmasi;

– E_0 to‘plamdagi qiymatlari z_1 va z_2 intervallar chegaralari bo‘lgan ob’ektlar nomerlari.

1.16. В виде таблицы $A=\{a_{uv}\}_{m*n}$ задана выборка из множества объектов $E_0=\{S_1,\ldots,S_m\}$, описываемых набором количественных признаков $X(n)=(x_1,\ldots,x_n)$. По каждой паре признаков $(x_i, x_j)\in X(n)$ требуется определить:

– границы допустимых интервалов $[z_1,z_2]$, по

$$R(i,j)=\{a_{ki}/P_i - a_{kj}/P_j\}_{k\in\{1,\ldots,m\}},$$

где $S_k= (a_{k1},\ldots,a_{kn})$, P_i,P_j – математическое ожидание значений признаков (x_i, x_j) на множестве E_0, $z_1 = \min\limits_{E_0} R(i, j)$, $z_2 = \max\limits_{E_0} R(i, j)$;

– номера объектов из E_0, со значениями границ интервалов z_1 и z_2.

1.17. Berilgan $A=\{a_{uv}\}_{m*n}$ jadvalida m ta $S_1,\ldots,S_m$ ob'ektlar to'plamini $X(n)=(x_1,\ldots,x_n)$ miqdoriy alomatlar bilan tavsifi berilgan. Har bir $(x_i, x_j)\in X(n)$ alomatlar juftligi bo'yicha aniqlash talab etiladi:

– mumkin bo'lgan intervallar $[z_1,z_2]$ chegarasini, bu erda, $z_1 = \min_{E_0} R(i, j)$, $z_2 = \max_{E_0} R(i, j)$, $R(i,j)=\{a_{ki}/P_i - a_{kj}/P_j\}_{k\in\{1,\ldots,m\}}$, $S_k=(a_{k1},\ldots,a_{kn})$, P_i,P_j – (x_i,x_j) alomatlarning E_0 to'plamdagi qiymatlarining medinanasi;

– E_0 to'plamdagi qiymatlari z_1 va z_2 intervallar chegaralari bo'lgan ob'ektlar nomerlari.

1.17. В виде таблицы $A=\{a_{uv}\}_{m*n}$ задана выборка из множества объектов $E_0=\{S_1,\ldots,S_m\}$, описываемых набором количественных признаков $X(n)=(x_1,\ldots,x_n)$. По каждой паре признаков $(x_i, x_j)\in X(n)$ требуется определить:

– границы допустимых интервалов $[z_1,z_2]$, по

$$R(i,j)=\{a_{ki}/P_i - a_{kj}/P_j\}_{k\in\{1,\ldots,m\}},$$

где $S_k= (a_{k1},\ldots,a_{kn})$, P_i,P_j – значения медиан признаков (x_i, x_j) на множестве E_0, $z_1 = \min_{E_0} R(i, j)$, $z_2 = \max_{E_0} R(i, j)$;

– номера объектов из E_0, со значениями границ интервалов z_1 и z_2.

1.18. Berilgan $A=\{a_{ij}\}_{mn}$ jadvalda $E_0 = (S_1,\ldots,S_m)$ ob'ektlarning n miqdoriy alomatlar bo'yicha tavsifi keltirilgan. Ikkita $S_u,S_v\in E_0$ ob'ektlar o'rtasidagi masofa $\rho(x,y)$ evklid metrikasi bilan hisoblanadi. Tanlanmaning S ob'ektining k ta eng yaqin qo'shnilarini $N_k(S)$ bilan, k–$dis(S)$ orqali S ob'ektdan k – yaqin qo'shnigacha bo'lgan masofani belgilaylik.

S_v ob'ektdan S_u ob'ektgacha erishiluvchi masofa $RD_k(S_u,S_v)= \max\{k\text{–}dis(S_v), \rho(S_u,S_v)\}$ sifatida aniqlanadi (1.a-rasmga qarang).

S_u ob'ektga erishuvchanlikning lokal zichligi

$$lrd(S_u)=1/\left(\frac{\sum_{S_v\in N_k(S_u)} RD_k(S_u,S_v)}{|N_k(S_u)|}\right)$$

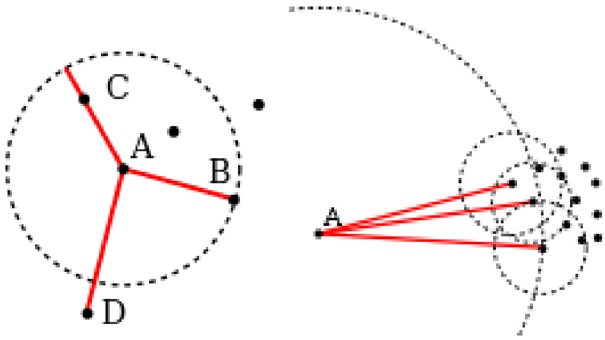
formula bilan aniqlanadi va S_u ob'ektga uning qo'shnilaridan erishish masofalarining o'rta arifmetikiga nisbatan teskari hisoblanadi. Erishishning lokal zichligi qo'shnilarning erishish lokal zichligi bilan

$$LOF_k(S_u)=\frac{\sum_{S_v\in N_k(S_u)} lrd(S_v)}{|N_k(S_u)|}/lrd(S_u)$$

ko‘rinishda taqqoslanadi, ya’ni qo‘shnilarga erishishning o‘rtacha lokal zichligini ob’ektning o‘zining erishish lokal zichligiga bo‘lish orqali.

Agar $LOF_k(S_u)$, 1 yaqin yaqin qiymat teng bo‘lsa, S_u ob’ektni qo‘shnilar bilan qiyoslash mumkin bo‘ladi (u holda bu ob’ekt anomal emas (sachratqi)). Birdan kichik qiymatlar zich sohani anglatadi (u ichki soha bo‘lishi mumkin), birdan etarlicha katta bo‘lgan qiymatlar ob’ektning anomalligidan guvoh beradi.

Berilgan E_0 tanlanmaning k=3,5,7 holatlar uchun barcha anomal ob’ektlari aniqlansin.



a) b)

1-rasm. Anomal ob’ektlarni aniqlash

(1-рис. Обнаружение аномальных объектов)

1.a – rasmda V va S ob’ektlari bir xil erishish masofasiga ega (k=3) bo‘lgan holda D ob’ekt k- yaqin qo‘shni bo‘lmaydi. 1.b-rasmda A nuqta boshqa qo‘shnilariga nisbatan kam zichlikka ega.

1.18. В заданной таблице $A=\{a_{ij}\}_{m*n}$ содержатся описание объектов $E_0=(S_1,\ldots,S_m)$ по n количественным признакам. Расстояние между объектами $S_u,S_v \in E_0$ вычисляется по метрике Евклида $\rho(x,y)$. Обозначим через $N_k(S)$ – множество из k ближайших соседей к объекту S и $k\text{-}dis(S)$ – расстояние от S до k-го ближайшего соседа. Достижимое расстояние объекта S_u из S_v определяется как $RD_k(S_u,S_v)= \max\{k\text{-}dis(S_v), \rho(S_u,S_v)\}$ (см. рис.1.а).

Локальная плотность достижимости объекта S_u определяется как

$$lrd(S_u)=1/\left(\frac{\sum\limits_{S_v\in N_k(S_u)} RD_k(S_u,S_v)}{|N_k(S_u)|}\right)$$

и является обратным значением среднему расстоянию достижимости объекта S_u из его соседей. Локальная плотность достижимости сравниваются с локальными плотностями достижимости соседей

$$LOF_k\left(S_u\right)=\frac{\sum\limits_{S_v\in N_k(S_u)} lrd\left(S_v\right)}{\left|N_k\left(S_u\right)\right|}/\,lrd\left(S_u\right)$$
,

которая есть средняя локальная плотность достижимости соседей, делённая на локальную плотность достижимости самого объекта.

Значение $LOF_k(S_u)$, примерно равное 1, означает, что объект S_u сравним с его соседями (а тогда он не является выбросом). Значение меньше 1 означает плотную область (которая может быть внутренностью), а значения, существенно большие 1, свидетельствуют о выбросах.

В рис. 1.a объекты B и C имеют одно и то же расстояние достижимости (k=3), в то время как D не является k-ближайшим соседом, а в рис.1.b точка A имеет меньшую плотность по сравнению с соседями.

Определить все объекты E_0, являющиеся выбросами при k=3,5,7.

1.19. Ixtiyoriy, mumkin bo'lgan obektlar tavsifidagi $X(n)$, $n>2$ alomatlar to'plami uchun expert yo'li bilan informativlik bo'yicha tartiblangan

$$\omega_1(t_1),\ldots,\omega_g(t_g),\ldots,\omega_n(t_n),\ t_i\in\{1,\ldots,n\}$$

ketma-ketligi aniqlangan. Bu erda $\omega_g(t_g)$ элементning g indeksi alomat rangi sifatida qaraladi.

K ta ($k<n$) o'lchangan alomat bilan tavsiflangan S obyekt haqidagi ma'lumot $\alpha(S)=(\alpha_1,\ldots,\ \alpha_i,\ldots,\ \alpha_n)$ xarakteristika vektori ko'rinishida berilgan bo'lib, agar α_i =0 – alomat o'lchanmaganligini, α_i =1 – o'lchanganligini anglatadi.

O'lchangan k alomatlar rangini $D(S,k)$ bilan belgilaylik. Yig'indining qiymatlar diapazoni $k(1+k)/2 \le D(S,k) \le k(2n-k+1)/2$ oraliqda yotadi.

Talab qilinadi:

– S obektning informativlik o'lchami

$$\Omega(S,k)=1-\left(\frac{D(S,k)-\alpha}{\beta-\alpha}\right)\left(1-\frac{k}{n}\right),$$

formula bilan hisoblansin/ Bu erda $\alpha= k(1+k)/2$, $\beta= k(2n-k+1)/2$;

– o'lchamga лингвистик izoh berilsin: (0;0.5] – "*yomon*"; (0.5;0.8] – "*qoniqarli*"; (0.8;1] – "*yaxshi*".

1.19. Для набора признаков $X(n)$, $n>2$ в описании произвольных допустимых объектов экспертным путём определена упорядоченная по информативности последовательности признаков

$$\omega_1(t_1),\ldots,\omega_g(t_g),\ldots,\omega_n(t_n),\ t_i \in\{1,\ldots,n\}$$

Индекс g элемента $\omega_g(t_g)$ интерпретируется как ранг признака.

Информация об объекте S с k измеренными значениями признаков ($k<n$) задаётся в виде характеристического вектора $\alpha(S)=(\alpha_1,\ldots, \alpha_i,\ldots, \alpha_n)$, в котором $\alpha_i=0$ – признак неизмерен, $\alpha_i=1$ – измерен.

Пусть $D(S,k)$ – сумма рангов k измеренных значений признаков объекта S. Диапазон значений суммы находится между $k(1+k)/2 \le D(S,k) \le k(2n-k+1)/2$.

Требуется:

– вычислить меру информативности объекта S как

$$\Omega(S,k)=1-\left(\frac{D(S,k)-\alpha}{\beta-\alpha}\right)\left(1-\frac{k}{n}\right),$$

где $\alpha= k(1+k)/2$, $\beta= k(2n-k+1)/2$;

– дать лингвистическую интерпретацию меры: (0;0.5] – "*плохой*"; (0.5;0.8] – "*удовлетворительный*"; (0.8;1] – "*хороший*".

1.20. ("*Sachratqilar*"). Berilgan $E_0=(S_1,\ldots,S_m)$ tanlanma ob'ektlari $X(n)=(x_1,\ldots,x_n)$, $n>2$ – miqdoriy alomatlar bilan tavsiflangan. Tanlanmaning S ob'ekti asosiy deyiladi, agar hech bo'lmaganda k ta ob'ekt undan ε masofada joylashgan bo'lsa (ε – S ob'ektga qo'shnichilik radiusining maksimali), o'zini ham hisobga olgan holda. Ular S ob'ektdan erishish mumkin ob'ektlar hisoblanadi. Asosiylardan erishish mumkin bo'lmagan barcha ob'ektlar sachratqi ob'ektlar hisoblanadi.

Talab qilinadi:

1. Berilgan k va ε bo'yicha sachratqi bo'lgan barcha ob'ektlar topilsin;
2. Evklid va Chebishev metrikalari bo'yicha aniqlangan sachratqilar to'plami o'zaro taqqoslansin.

1.20. ("*Выбросы*"). Дано описание множества объектов $E_0=(S_1,\ldots,S_m)$ с помощью набора количественных признаков $X(n)=(x_1,\ldots,x_n)$. Объект S является основным, если по крайней мере k точек находятся на расстоянии ε (ε является максимальным радиусом соседства от S) включая сам объект. Считается, что эти объекты достижимы из S. Все объекты не достижимые из основных точек считаются выбросами.

Требуется:

1. По заданным k и ε определить все объекты, являющиеся выбросами;

Сравнить множества выбросов по метрикам Евклида и Чебышева.

2. O'qituvchili klassifikatsiya (Классификация с учителем)

2.1. Berilgan $E_0 = \{x_1, \dots, x_m\}$ tanlanma obyektlari 2 ta kesishmaydigan K_1 va K_2 sinf vakillaridan iborat. Har bir $x_i \in E_0$ obyekt n miqdoriy alomatlar bilan tavsiflanadi. O‘rgatishning “*rag‘batlantirish – jazolash*” tamoyilidan foydalangan holda iterativ algoritm (perseptron algoritmi) orqali 2 ta sinf orasidagi chegarani aniqlang. Vaznlar vektori $w(1)$ va sozlovchi orttirma $c\ (0 < c < 1)$ ning boshlang‘ich qiymatlari berish talab qilinadi. Agar p- qadamda $x(p) \in K_1 \cap E_0$ va $w(p)x(p) \leq 0$ bo‘lsa, u holda $w(p+1) = w(p) + cx(p)$. Agar $x(p) \in K_2$ va $w(p)x(p) \geq 0$ bo‘lsa, u holda $w(p+1) = w(p) - cx(p)$.

2.1. Выборка объектов $E_0 = \{x_1, \dots, x_m\}$, содержит представителей двух непересекающихся классов K_1 и K_2. Каждый объект $x_i \in E_0$ описывается n количественными признаками. Используя принцип “*подкрепления – наказания*” определить границу между двумя классами с помощью итеративного алгоритма (алгоритма перцептрона). Требуется задать начальное значение вектора весов $w(1)$ и $c\ (0 < c < 1)$ – корректирующее приращение. Если на -ом шаге итерatsiи $x(p) \in K_1 \cap E_0$ и $w(p)x(p) \leq 0$, то $w(p+1) = w(p) + cx(p)$. Если $x(p) \in K_2$ и $w(p)x(p) \geq 0$, то $w(p+1) = w(p) - cx(p)$.

2.2. O‘zaro kesishmaydigan l ta $K_1, \dots, K_l$ sinflarga bo‘lingan $E_0 = \{S_1, \dots, S_m\}$ o‘rgatuvchi tanlanma bo‘yicha eng yaqin qo‘shni algoritmi uchun optimal k ning qiymatini hisoblash algoritmini amalga oshiruvchi dastur tuzilsin. Har bir $S \in E_0$ obyekt n ta miqdoriy alomat bilan tavsiflanadi. Obyektlar orasida yaqinlik o‘lchovi sifatida evklid metrikasidan foydalanilsin.

2.2. Реализовать алгоритм вычисления оптимального значения k ближайших соседей по обучающей выборке $E_0 = \{S_1, \dots, S_m\}$, разделённой на l непересекающихся классов $K_1, \dots, K_l$. Считается, что каждый объект $S \in E_0$ описывается n количественными признаками. В качестве меры близости между объектами использовать евклидову метрику.

2.3. O‘zaro kesishmaydigan l ta $K_1,\dots,K_l$ sinflarga bo‘lingan $E_0=\{S_1,\dots,S_m\}$ o‘rgatuvchi tanlanma bo‘yicha obyektlar turg‘unligi qiymatlarini hisoblovchi dastur tuzilsin. Har bir $S\in E_0$ – obyekt n ta miqdoriy alomat bilan tavsiflanadi va masofani topish uchun Chebishev metrikasidan foydalanilsin. K_j sinfdagi $S_i\in K_j, j=\overline{1,l}, i=\overline{1,m}$ obyektning λ_i^j turg‘unligi:

$$\lambda_i^j=\frac{d_i^j}{2\min\limits_{1\le j\le l}|K_j|-3}$$

bo‘yicha topiladi, bu erda, d_i^j – E_0 tanlanmadagi S_i obyektning k ta $\left(k=1,\dots,2\min\limits_{1<j<l}|K_i|-3\right)$ eng yaqin qo‘shnilarining ko’pchiligi $K_j\cap E_0$ sinfdan bo‘lgandagi holatlar soni.

2.3. Реализовать вычисления значений устойчивости объектов обучающей выборке $E_0=\{S_1,\dots,S_m\}$, разделённой на l непересекающихся классов $K_1,\dots,K_l$. Считается, что каждый объект $S\in E_0$ описывается n количественными признаками и для определения расстояния используется метрика Чебышева. Устойчивость λ_i^j объекта $S_i\in K_j, j=\overline{1,l}, i=\overline{1,m}$ в классе K_j определяется как

$$\lambda_i^j=\frac{d_i^j}{2\min\limits_{1\le j\le l}|K_j|-3},$$

где d_i^j - число событий в E_0, когда среди k ближайших к S_i объектов, $k=1,\dots,2\min\limits_{1\le j\le l}|K_j|-3$ большинство составляют объекты из $K_j\cap E_0$.

2.4. O‘zaro kesishmaydigan l ta $K_1,\dots,K_l$ sinflarga bo‘lingan $E_0=\{S_1,\dots,S_m\}$ o‘rgatuvchi tanlanma bo‘yicha Evklid metrikasidan foydalangan holda chiziqli qobiqni hisoblash algoritmi dasturi tuzilsin. Har bir $S\in E_0$ – obyekt n ta miqdoriy alomat bilan tavsiflanadi. Chiziqli qobiq obyektlari $\rho(S_{i_p},S_{i_r})=\min\limits_{S_{i_t}\in O(S_i)}\rho(S_{i_p},S_{i_t})$ formuladan aniqlanadi, bu erda $O(S_i)$ – S_i obyektning atrofi, ya’ni obyektning qarama-qarshi sinfning eng yaqin obyektigacha bo‘lgan, o‘z sinf obyektlaridan tashkil topgan atrofi.

2.4. Реализовать алгоритм вычисление линейных оболочек непересекающихся классов объектов $K_1,\dots,K_l$ по метрике Евклида. Считается, что представители l классов заданы в выборке $E_0=$

$\{S_1, \dots, S_m\}$ и каждый объект $S \in E_0$ описывается с помощью n количественных признаков. Объекты линейной оболочки определяются с помощью формулы $\rho(S_{i_p}, S_{i_r}) = \min_{S_{i_t} \in O(S_i)} \rho(S_{i_p}, S_{i_t})$, где $O(S_i)$ – окрестность объекта S_i, содержащая все ближайшие объекты одного с ним класса до первого объекта из противоположного класса.

2.5. O‘zaro kesishmaydigan l ta $K_1, \dots, K_l$ sinflarga bo‘lingan $E_0 = \{S_1, \dots, S_m\}$ o‘rgatuvchi tanlanma bo‘yicha Chebishev metrikasidan foydalangan holda chiziqli qobiqni hisoblash algoritmi dasturi tuzilsin. Har bir $S \in E_0$ – obyekt n ta miqdoriy alomat bilan tavsiflanadi. Chiziqli qobiq obyektlari $\rho(S_{i_p}, S_{i_r}) = \min_{S_{i_t} \in O(S_i)} \rho(S_{i_p}, S_{i_t})$ formuladan aniqlanadi, bu erda $O(S_i)$ – S_i obyektning atrofi, ya’ni obyektning qarama-qarshi sinfning eng yaqin obyektigacha bo‘lgan, o‘z sinf obyektlaridan tashkil topgan atrofi.

2.5. Реализовать алгоритм вычисление оболочки непересекающихся классов объектов $K_1, \dots, K_l$ по метрике Чебышева. Считается, что представители l классов заданы в выборке $E_0 = \{S_1, \dots, S_m\}$ и каждый объект $S \in E_0$ описывается с помощью n количественных признаков. Объекты оболочки определяются с помощью формулы $\rho(S_{i_p}, S_{i_r}) = \min_{S_{i_t} \in O(S_i)} \rho(S_{i_p}, S_{i_t})$, где $O(S_i)$ – окрестность объекта S_i, содержащая все ближайшие объекты одного с ним класса по метрике Чебышева до первого объекта из противоположного класса.

2.6. O‘zaro kesishmaydigan l ta $K_1, \dots, K_l$ sinflarga bo‘lingan $E_0 = \{S_1, \dots, S_m\}$ o‘rgatuvchi tanlanma bo‘yicha Juravlev metrikasidan foydalangan holda chiziqli qobiqni hisoblash algoritmi dasturi tuzilsin.

Har bir $S \in E_0$ obyekt n ta turli toyifadagi alomatlar bilan tavsivlangan. Juravlyov metrikasi quyidagi formuladan hisoblanadi:

$$\rho(x,y) = \sum_{j \in I} |x_j - y_j| + \sum_{j \in J} \begin{cases} 1, & x_j \neq y_j, \\ 0, & x_j = y_j. \end{cases}$$

Bu erda I va J – mos ravishda miqdoriy va nominal alomatlar indekslar to‘plami. Bu erda o‘lchov masshtablarini unifikatsiyalash maqsadida miqdoriy alomatlar qiymatlari kasr-chiziqli almashtirish orqali [0,1] oraliqqa akslantiriladi.

Chiziqli qobiq obyektlari $\rho(S_{i_p}, S_{i_r}) = \min_{S_{i_t} \in O(S_i)} \rho(S_{i_p}, S_{i_t})$ formuladan aniqlanadi, bu erda $O(S_i)$ – S_i obyektning atrofi, ya'ni obyektning qarama-qarshi sinfning eng yaqin obyektigacha bo'lgan, o'z sinf obyektlaridan tashkil topgan atrofi.

2.6. Реализовать алгоритм вычисление оболочки непересекающихся классов объектов $K_1, \dots, K_l$ по метрике Журавлёва. Считается, что представители l классов заданы в выборке $E_0 = \{S_1, \dots, S_m\}$ и каждый объект $S \in E_0$ описывается с помощью n разнотипных признаков. Метрика Журавлёва вычисляется по формуле

$$\rho(x,y) = \sum_{j \in I} \left| x_j - y_j \right| + \sum_{j \in J} \begin{cases} 1, & x_j \neq y_j, \\ 0, & x_j = y_j, \end{cases}$$

где I и J множество индексов соответственно количественных и номинальных признаков. Объекты оболочки определяются с помощью формулы $\rho(S_{i_p}, S_{i_r}) = \min_{S_{i_t} \in O(S_i)} \rho(S_{i_p}, S_{i_t})$, где $O(S_i)$ – окрестность объекта S_i, содержащая все ближайшие объекты одного с ним класса по метрике Журавлёва до первого объекта из противоположного класса.

2.7. O'zaro kesishmaydigan l ta $K_1, \dots, K_l$ sinflarga bo'lingan $E_0 = \{S_1, \dots, S_m\}$ o'rgatuvchi tanlanmada quyidagi mezon bo'yicha miqdoriy alomatlar qiymatlarini kesishmaydigan intervallarga optimal bo'luvchi algoritmni amalga oshirilsin:

$$\left(\frac{\sum_{p=1}^{l} \sum_{i=1}^{l} (u_i^p - 1) u_i^p}{\sum_{i=1}^{l} |K_i| (|K_i| - 1)} \right) \left(\frac{\sum_{p=1}^{l} \sum_{i=1}^{l} u_i^p \left(m - |K_i| - \sum_{j=1}^{l} u_j^p + u_i^p \right)}{\sum_{i=1}^{l} |K_i| (m - |K_i|)} \right) \to \max_{\{A\}}$$

bu erda $A = (a_0, \dots, a_l)$ – butun sonlardan tashkil topgan vektor va uning elementlari quyidagi shartlarni qanoatlantiradi: $a_0 = 0, a_l = m$, $a_r < a_{r+1}, r = \overline{1, l-1}$ va obyektlar nomeri miqdoriy alomatlar qiymatlarini o'sish tartibida beriladi, u_i^p – alomatning p – intervaldagi K_i sinfdagi qiymatlari miqdori.

2.7. На заданной выборке $E_0 = \{S_1, \dots, S_m\}$, содержащей представителей l непересекающихся классов объектов $K_1, \dots, K_l$ реализовать алгоритм оптимального разбиения значений количественных признаков на непересекающиеся интервалы по критерию

$$\left(\frac{\sum_{p=1}^{l}\sum_{i=1}^{l}\left(u_i^p-1\right)u_i^p}{\sum_{i=1}^{l}|K_i|\left(|K_i|-1\right)}\right)\left(\frac{\sum_{p=1}^{l}\sum_{i=1}^{l}u_i^p\left(m-|K_i|-\sum_{j=1}^{l}u_j^p+u_i^p\right)}{\sum_{i=1}^{l}|K_i|\left(m-|K_i|\right)}\right)\to\max_{\{A\}},$$

где $A=(a_0,\dots,a_l)$ - целочисленный вектор, элементы которого удовлетворяют условиям: $a_0=0, a_l=m,\ \ a_r<a_{r+1}, r=\overline{1,l-1}$ и задают номера объектов упорядоченных по возрастанию значений количественного признака, u_i^p – количество значений признака из класса K_i в p -ом интервале.

2.8. O‘zaro kesishmaydigan l ta $K_1,\dots,K_l$ sinflarga bo‘lingan $E_0=\{S_1,\dots,S_m\}$ o‘rgatuvchi tanlanma bo‘yicha Evklid metrikasidan foydalanib “*eng yaqin qo‘shni*” anglash algoritmi amalga oshirilsin. Har bir $S\in E_0$ – obyekt n ta miqdoriy alomat bilan tavsiflanadi.

2.8. На заданной выборке $E_0=\{S_1,\dots,S_m\}$, содержащей представителей l непересекающихся классов объектов $K_1,\dots,K_l$, реализовать алгоритм распознавания по правилу “*ближайший сосед*”. Считается, что каждый объект $S\in E_0$ описывается с помощью n количественных признаков и для вычисления расстояния используется евклидова метрика.

2.9. O‘zaro kesishmaydigan l ta $K_1,\dots,K_l$ sinflarga bo‘lingan $E_0=\{S_1,\dots,S_m\}$ o‘rgatuvchi tanlanma berilgan. Har bir $S\in E_0$ – obyekt n ta nominal alomat bilan tavsiflanadi. 2 ta alomat orqali $K_i, i=\overline{1,l}$ sinfda mavjud va boshqalarda bo‘lmagan barcha mumkin bo‘lgan qiymatlar kombinatsiyalari topilsin (mantiqiy qonuniyat).

2.9. Выборка объектов $E_0=\{S_1,\dots,S_m\}$, содержит представителей l классов $K_1,\dots,K_l$. Каждый объект $S\in E_0$ описывается n номинальными признаками. Из всевозможных комбинatsiй по 2 признака требуется определить те, которые есть у одного класса $K_i, i=\overline{1,l}$ и нет у представителей других классов (логическая закономерность).

2.10. $E_0=\{S_1,\dots,S_m\}$ tanlanma obyektlari 2 ta K_1,K_1 kesishmaydigan sinf vakillaridan iborat.

$$\left|\frac{d_1(u,v)}{|K_1|}-\frac{d_2(u,v)}{|K_2|}\right|\to max$$

mezon bo‘yicha miqdoriy alomatlar ustunlik intervalini hisoblaydigan algoritmni amalga oshirilsin. Bu erda $d_1(u,v)$, $d_2(u,v)$, $u \le v$ – mos ravishda miqdoriy alomatlarning $r_1, r_2, \dots, r_m$ o‘sish tartibidagi u va v – o‘rinlar oralig‘idagi K_1, K_2 sinflar vakillarining miqdori.

2.10. Выборка объектов $E_0 = \{S_1, \dots, S_m\}$ содержит представителей двух непересекающихся классов K_1, K_2. Каждый объект описывается с помощью количественных признаков. Реализовать алгоритм вычисления интервалов доминирования значений количественных признаков по критерию

$$\left|\frac{d_1(u,v)}{|K_1|} - \frac{d_2(u,v)}{|K_2|}\right| \to max$$

где $d_1(u,v)$, $d_2(u,v)$, $u \le v$ число представителей классов K_1, K_2 в упорядоченной по возрастанию последовательности значений $r_1, r_2, \dots, r_m$ количественного признака с u-й по -ю позицию.

2.11. Kesishmaydigan 2 ta K_1 va K_2 sinfga bo‘lingan $E_0 = \{S_1, \dots, S_m\}$ o‘rgatuvchi tanlanma berilgan. Har bir $S \in E_0$ – obyekt n ta miqdoriy alomat bilan tavsiflanadi. Obyektlarni ajratish uchun

$$R(S) = \sum_{i=1}^{n} t_i w_i x_i$$

diskriminant funksiyadan foydalaniladi, bu erda w_i - (0,1] intervalda berilgan vazn, $T = (t_1, \dots, t_n)$, $t_i \in [-1,1]$ vektori

$$\min_{S_i \in K_1} R(S_i) - \max_{S_i \in K_2} R(S_i) \to max$$

bo‘yicha aniqlanadi.

2.11. Выборка объектов $E_0 = \{S_1, \dots, S_m\}$, содержит представителей двух непересекающихся классов K_1 и K_2. Каждый объект $S_i \in E_0$ описывается n количественными признаками. Для разделения объектов используется дискриминантная функция

$$R(S) = \sum_{i=1}^{n} t_i w_i x_i,$$

где $S = (x_1, \dots, x_n)$, w_i – заданные значения весов в интервале (0,1]. Определить значения элементов вектора $T = (t_1, \dots, t_n)$, $t_i \in [-1,1]$ при которых $\min_{S_i \in K_1} R(S_i) - \max_{S_i \in K_2} R(S_i) \to max$. Сравнить значения результатов при выборе разных начальных приближений T.

2.12. $E_0 = \{x_1, \dots, x_m\}$ tanlanma obyektlari 2 ta kesishmaydigan K_1

va K_2 sinf vakillaridan iborat. Har bir $x_i \in E_0$ obyekt n har xil toifali alomatlar bilan tavsiflanadi. $\omega E_0 - E_0$ tanlanmaning ω qismi hisoblanadi, $\omega = \{\omega_1, \dots, \omega_n\}$, agar i-chi alomat klassifikatsiyada qatnashsa $\omega_i = 1$, aks holda $\omega_i = 0$. Berilgan ω vektor bo‘yicha ωE_0 ilojsiz obyektlar (K_1 va K_2 sinfda ajralmaydigan obyektlar) aniqlansin.

2.12. Выборка объектов $E_0 = \{S_1, \dots, S_m\}$ содержит представителей двух непересекающихся классов K_1 и K_2. Каждый объект $S_i \in E_0$ описывается n разнотипными признаками $X(n) = (x_1, \dots, x_n)$. ωE_0 является ω частью выборки E_0, определяемой характеристическим вектором $\omega = \{\omega_1, \dots, \omega_n\}$, $\omega_i = 1$ если признак x_i используется для классификации и $\omega_i = 0$ если не используется. По заданным значениям элементов вектора ω определить не образуются с их помощью на ωE_0 тупики – неразличимые объектов из классов K_1 и K_2.

2.13. Mumkin bo‘lgan obyektlarning ikkita o‘zaro kesishmaydigan K_1, K_2 sinflarga bo‘lingan $E_0 = \{S_1, \dots, S_m\}$ tanlanma berilgan. Tanlanma obyektlari n ta $X(n) = (x_1, \dots, x_n)$ miqdoriy alomatlar bilan tavsiflangan. E_0 tanlanmada $\rho(x, y)$– Evklid metrikasi berilgan bo‘lsin.

Tanlanmadagi $S_d \in K_p, p = 1{,}2$ obyektning turg‘unligi $X(k) = (x_1, \dots, x_k)$, $k \le n$ bo‘yicha hisoblash uchun foydalaniladi:

– o‘zaro munosabati $\rho(S_{d_i}, S_d) \le \rho(S_{d_{i+1}}, S_d)$ tengsizlik bilan aniqlangan E_0 tanlanma obyektlarining $S_{d_0}, \dots, S_{d_{m-1}}, S_{d_0} = S_d$ tartiblangan ketma-ketligi;

– $F(S_d, X(k)) = \max\limits_{0 \le i \le m-1} \left(\dfrac{z_p(i)}{|K_p|} - \dfrac{z_{3-p}(i)}{|K_{3-p}|} \right)$ formula bilan aniqlanuvchi funksional qiymati. Bu erda $z_p(i), z_{3-p}(i)$ – mos ravishda $P(S_d, X(k)) = \{S_{d_0}, \dots, S_{d_i}\}$ to‘plamdagi K_p, K_{3-p} sinflarga tegishli obyektlar soni.

Berilgan $X(k) = (x_1, \dots, x_k)$ alomatlar to‘plami bo‘yicha $S_d \in E_0$ obyekt turg‘unligi hisoblansin.

2.13. Рассматривается множество (выборка) $E_0 = \{S_1, \dots, S_m\}$ допустимых объектов, разбитая на 2 непересекающихся подмножества (класса) K_1, K_2. Объекты выборки описываются с помощью n количественных признаков $X(n)=(x_1,...,x_n)$. На множестве объектов E_0 определена Евклидова метрика $\rho(x,y)$.

Для вычисления устойчивости объекта $S_d \in K_p, p=1,2$ по набору признаков $X(k)=(x_1,...,x_k)$, $k \le n$ используется:

– упорядоченная последовательность объектов E_0, $S_{d_0},..., S_{d_{m-1}}, S_{d_0}=S_d$, отношения между которыми определяются неравенствами вида $\rho(S_{d_i},S_d) \le \rho(S_{d_{i+1}},S_d)$;

– значение функционала $F(S_d,X(k))=\max\limits_{0\le i\le m-1}\left(\frac{z_p(i)}{|K_p|}-\frac{z_{3-p}(i)}{|K_{3-p}|}\right)$, где $z_p(i), z_{3-p}(i)$ – число объектов в $P(S_d,X(k))=\{S_{d_0},...,S_{d_i}\}$ соответственно из класса K_p, K_{3-p}.

По заданному набору признаков $X(k)=(x_1,...,x_k)$ определить значение устойчивости объекта $S_d \in E_0$.

2.14. O‘zaro kesishmaydigan l ta $K_1,...,K_l$ sinflar vakillarini o‘z ichiga olgan $E_0=\{S_1,...,S_m\}$ o‘rgatuvchi tanlanma berilgan. Har bir $S\in E_0$ obyekt n ta $X(n)=(x_1,...x_n)$ miqdoriy alomatlar bilan tavsiflangan. Berilgan $\rho(x,y)$ metrika bo’yicha chegaraviy obyektlar to’plam ostisi sifatida qaraluvchi shovqin obyektlar to’plami $B\subset E_0$ quydagicha aniqlanadi

$$B=\left\{S\in E_0 \,\middle|\, \rho(S_i,S)=\min_{S_i\in K_j, S_d\in CK_j}\rho(S_i,S_d)\right\}$$

Tanlanmaning $S\in B\cap K_j$, $j=1,\ldots,l$ obyekti K_j sinfning D_j shovqin obyektlar to’plamiga tegishli deyiladi, agar

$$\left|\left\{S_i\in E_0 \,\middle|\, \rho(S_i,S)=\min_{S_i\in CK_j, S_d\in K_j}\rho(S_i,S_d)\right\}\right| > \left|\left\{S_i\in K_j \,\middle|\, \rho(S_i,S)<\min_{S_i\in K_j, S_d\in CK_j}\rho(S_i,S_d)\right\}\right|$$

shart o’rinli bo’lsa.

Har bir K_i sinf uchun Evklid, Hemming va Chebishev metrikalari bo‘yicha shovqin obyektlar to‘plami D_i aniqlansin.

2.14. Обучающая выборка $E_0=\{S_1,...,S_m\}$ содержит представителей l непересекающихся классов $K_1,...,K_l$. Каждый объект $S\in E_0$ описывается набором из n количественных признаков $X(n)=(x_1,...x_n)$. Множество шумовых объектов рассматривается как подмножество граничных объектов классов $B \subset E_0$ по заданной метрике $\rho(x, y)$, которое определяется как

$$B=\left\{S\in E_0 \middle| \rho(S_i,S)=\min_{S_i\in K_j, S_d\in CK_j}\rho(S_i,S_d)\right\}.$$

Объект $S \in B \cap K_j$, $j=1,\ldots,l$ принадлежит множеству шумовых объектов D_j класса K_j, если

$$\left|\left\{S_i\in E_0 \middle| \rho(S_i,S)=\min_{S_i\in CK_j, S_d\in K_j}\rho(S_i,S_d)\right\}\right| > \left|\left\{S_i\in K_j \middle| \rho(S_i,S)<\min_{S_i\in K_j, S_d\in CK_j}\rho(S_i,S_d)\right\}\right|.$$

Требуется определить все множества шумовых объектов D_i для каждого класса K_i по метрикам Евклида, Хэмминга, Чебышева.

2.15. O‘zaro kesishmaydigan l ta $K_1, \ldots, K_l$ sinflar vakillarini o‘z ichiga olgan $E_0 = \{S_1, \ldots, S_m\}$ o‘rgatuvchi tanlanma berilgan. Har bir $S \in E_0$ obyekt n ta $X(n) = (x_1, \ldots x_n)$ miqdoriy alomatlar bilan tavsiflangan. Obyektlar o‘rtasidagi masofa $\rho(x, y)$ metrikasi bilan o‘lchanadi. S obyekt $\rho(x, y)$ metrika bo‘yicha chegaraviy obyektlar to‘plami L tegishli bo‘ladi, agar $\rho(S_i, S) = \min\limits_{S \in CK_d} \rho(S_i, S)$ bilan aniqlanuvchi shunday $S_i \in K_d$ mavjud bo‘lsa. Ikkita $S_i, S_j \in K_d$ obyektlar R munosabat bog‘langan, ya’ni $S_i\, R\, S_j$ rost bo‘ladi, agar $S \in L \cap K_d$mavjud bo‘lsaki, uning uchun $\rho(S_i, S) < r_i$ va $\rho(S_j, S) < r_j$ o‘rinli bo‘lsa. Bu erda $r_i = \min\limits_{S_\mu \in CK_d} \rho(S_i, S_\mu)$ va $r_j = \min\limits_{S_\mu \in CK_d} \rho(S_j, S_\mu)$.

Har bir K_d sinf uchun R munosabat bo‘yicha Evklid, Xemming va Chebishev metrikalaridan foydalangan holda o‘zaro kesishmaydigan G_{d1}, $\ldots, G_{dp}$, $1 \le p < |K_d|$, guruhlarga ajratilsin. Har bir $S_i, S_j \in G_{dt}$ juftlik uchun $S_i\, R S_\mu R\, \ldots\, S_\tau R\, S_j$ rost bo‘lgan yo‘l (G_{dt} dan obyektlar zanjiri) mavjud deb hisoblanadi.

2.15. Выборка объектов $E_0 = \{S_1, \ldots, S_m\}$ содержит представителей l непересекающихся классов $K_1, \ldots, K_l$. Каждый объект $S \in E_0$ описывается набором из n количественных признаков $X(n) = (x_1, \ldots x_n)$. Расстояние между объектами определяется по метрике $\rho(x, y)$. Объект $S \in E_0$ принадлежит к множеству граничных объектов L по метрике $\rho(x, y)$, $S \in L$ если существует такой $S_i \in K_d$, что $\rho(S_i, S) = \min\limits_{S \in CK_d} \rho(S_i, S)$. Объекты $S_i, S_j \in K_d$ связаны отношением R, т.е. $S_i\, R\, S_j$ истинно, если существует $S \in L \cap$

K_d, $\rho(S_i, S) < r_i$ и $\rho(S_j, S) < r_j$ где $r_i = \min_{S_\mu \in CK_d} \rho(S_i, S_\mu)$ и $r_j = \min_{S_\mu \in CK_d} \rho(S_j, S_\mu)$. Требуется по отношению R объекты каждого класса K_d разбить на непересекающиеся группы $G_{d1},\dots, G_{dp}$, $1 \le p < |K_d|$, используя в качестве меры близости метрики Эвклида, Хэмминга, Чебышева. Считается, что для каждой пары $S_i, S_j \in G_{dt}$ существует путь (цепочка объектов из G_{dt}), для которого $S_i\, RS_\mu R \,...\, S_\tau R\, S_j$ истинно.

2.16. Yechimlar daraxtini qurish uchun ikkita o‘zaro kesishmaydigan K_1, K_2 sinflarga ajratilgan obyektlar $E_0 = \{S_1, \dots, S_m\}$ tanlanmasidan foydalaniladi. Har bir $S \in E_0$ obyekt n ta nominal alomatlar $X(n) = (x_1, \dots x_n)$ bilan tavsiflangan bo‘lib, p gradatsiyaga ega x_i alomat $\{1, \dots, p\}$ to‘plamidan qiymat qabul qiladi. Boshlang‘ich E_0 tanlanmaning $T_1, T_2, \dots, T_c$ to‘plam ostilariga bo‘linishi aniqlagan bo‘lsin. $K_d, d = 1{,}2$ sinfga tegishli obyektning $T \subset E_0$ to‘plam ostida paydo bo‘lish ehtimolligi $P_d = \frac{\mu(d,T)}{|T|}$ formulasi bilan aniqlanadi. Bu erda $\mu(d, T) -$ T to‘plam ostisidagi K_d sinf obyektlari soni. Sinfdagi obyektni aniqlash uchun zarur ma’lumotlar miqdori quyidagicha hisoblanadi.

$$Info(T) = -\frac{\mu(1,T)}{|T|}\log_2\left(\frac{\mu(1,T)}{|T|}\right) - \frac{\mu(2,T)}{|T|}\log_2\left(\frac{\mu(2,T)}{|T|}\right).$$

T to‘plam ostisi bo‘yicha $p(i)$ gradatsiyali x_i alomatning bahosi

$$Info(T, x_i) = \sum_{r=1}^{p(i)} T_r/|T|\ Info(T_r)$$

orqali hisoblanadi. Bu erda T_r $- r \in \{1, \dots, p(i)\}$ gradatsiyali T kiruvchi obyektlar to‘plami. $T = E_0$ shartida daraxt ildiziga joylashtirish uchun

$$Gain(x_i, T) = \max_{x_t \in X(n)} \big(Info(T) - Info(T, x_t)\big).$$

Bo‘yicha $x_i \in X(n)$ alomat topilsin.

2.16. Для построения дерева решений используется выборка объектов $E_0 = \{S_1, \dots, S_m\}$, разделённая на 2 непересекающихся класса K_1, K_2. Каждый объект $S \in E_0$ описывается набором из n номинальных признаков $X(n) = (x_1, \dots x_n)$. Признак x_i с p градациями принимает значения из множества $\{1, \dots, p\}$. На E_0 определяется разбиение на подмножества $T_1, T_2, \dots, T_c$. Вероятность появления объекта класса $K_d, d = 1{,}2$ в подмножестве $T \subset E_0$

определяется как $P_d = \frac{\mu(d,T)}{|T|}$, где $\mu(d,T) -$ количество объектов класса K_d в T. Среднее количество информatsiи, необходимое для определения объекта в классе

$$Info(T) = -\frac{\mu(1,T)}{|T|}\log_2\left(\frac{\mu(1,T)}{|T|}\right) - \frac{\mu(2,T)}{|T|}\log_2\left(\frac{\mu(2,T)}{|T|}\right).$$

Оценка признака x_i с $p(i)$ градациями по T вычисляется как

$$Info(T, x_i) = \sum_{r=1}^{p(i)} T_r/|T| \; Info(T_r),$$

где T_r – множество объектов из T с градацией $r \in \{1, \dots, p(i)\}$. Определить признак $x_i \in X(n)$ для размещения в корень дерева при $T = E_0$, для которого

$$Gain(x_i, T) = \max_{x_t \in X(n)} \big(Info(T) - Info(T, x_t)\big).$$

2.17. Yechimlar daraxtini qurish uchun ikkita o‘zaro kesishmaydigan K_1, K_2 sinflarga ajratilgan obyektlar $E_0 = \{S_1, \dots, S_m\}$ tanlanmasidan foydalaniladi. Har bir $S \in E_0$ obyekt n ta miqdoriy alomatlar $X(n) = (x_1, \dots x_n)$ bilan tavsiflangan. Har bir $x_i \in X(n)$ alomat uchun aniqlanuvchi $T \subset E_0$ to‘plam bo‘yicha $(-\infty, a_i], (a_i, +\infty)$ intervallarga bo‘lishdagi a_i chegara

$$Gini(i, T) = 1 - (P_1^2 + P_2^2),$$

mezoni bo‘yicha hisoblanadi. Bu erda $P_d = \frac{\mu(d,T)}{\varphi(a_i,T)}$, $\mu(d,T)$ – $(-\infty, a_i]$ intervaldagi K_d sinf vakillarining soni, $\max_d \mu(d,T) \geq 3$, $\varphi(a_i,T)$–x_i alomatining qiymatlari $(-\infty, a_i]$ intervalda bo‘lgan T to‘plamdagi obyektlar soni.

$T = E_0$ shartida daraxt ildiziga joylashtirish uchun

$$Gini(i, T) = \min_{1 \leq r \leq n} Gini(r, T).$$

Bo‘yicha $x_i \in X(n)$ alomat topilsin.

2.17. Для построения дерева решений используется выборка объектов $E_0 = \{S_1, \dots, S_m\}$, разделённая на 2 непересекающихся класса K_1, K_2. Объект $S \in E_0$ описывается набором из n количественных признаков $X(n) = (x_1, \dots x_n)$. Для каждого признака $x_i \in X(n)$ граница a_i разбиения на интервалы $(-\infty, a_i], (a_i, +\infty)$ по определяемому множеству $T \subset E_0$ вычисляется по критерию

$$Gini(i, T) = 1 - (P_1^2 + P_2^2),$$

где $P_d = \frac{\mu(d,T)}{\varphi(a_i,T)}$, $\mu(d,T)$ – число представителей класса K_d в $(-\infty, a_i]$, $\max_d \mu(d,T) \geq 3$. $\varphi(a_i,T)$ − число объектов из T со значениями признака x_i в $(-\infty, a_i]$. Требуется определить признак $x_i \in X(n)$ для размещения его в корень дерева при $T = E_0$, для которого

$$Gini(i,T) = \min_{1\leq r\leq n} Gini(r,T).$$

2.18. Berilgan $E_0 = \{S_1,...,S_m\}$ ob'ektlar to'plami ikkita o'zaro kesishmaydigan K_1 va K_2 sinflarga bo'lingan. To'plamning $S_u \in E_0$ ob'ekti $X(n)=(x_1,\ldots,x_n)$ nominal alomatlar to'plami bilan tavsiflanadi. Har bir $x_i \in X(n)$ nominal alomatning mumkin bo'lgan qiymatlari (gradasiyalari) $1,2,\ldots,p_i$, $2\leq p_i \leq m/2$ sonlari bilan beriladi. Nominal x_i alomatning $\mu \in \{1,2,\ldots,p_i\}$ gradasiyalari bo'yicha K_1 sinfga tegishlilik funksiyasi qiymati quyidagi ko'rinishda hisoblanadi

$$f_i(\mu) = \frac{d_{1i}(\mu)/|K_1|}{d_{1i}(\mu)/|K_1| + d_{2i}(\mu)/|K_2|}.$$

Bu erda $d_{1\mu}$ ($d_{2\mu}$) – x_i alomatining qiymati μ bo'lgan $K_1(K_2)$ sinf ob'ektlari soni (to'plami).

Yuqorida keltirilgan formula $A = \{a_{uv}\}_{m\times n}$ matritsadagi E_0 ob'ektlari tavsifidagi nominal alomatlar qiymatlarini o'girishda foydalaniladi. Har bir $S_u = (x_{u1},...,x_{un})(S_u \in E_0)$ ob'ekt jadvalda $(a_{u1},...,a_{un})$ ko'rinishdagi satr bilan yoziladi, bu erda $a_{uv} = f_v(x_{uv})$.

Sirpanuvchi nazorat usuli bilan E_0 tanlanmada yaqin qo'shni qoidasi bo'yicha anglash aniqligi hisoblansin. Usulni amalga oshirishda ketma-ket ravishda har bir $S_j \in E_0$ ob'ekt nazorat, $E_0 \backslash \{S_j\}$ – o'rgatuvchi sifatida taqdim etiladi. A jadvalidagi ob'ektlari tavsifi uchun masofa o'lchovi sifatida Juravlev metrikasidan foydalanilsin.

2.18. Множество объектов $E_0 = \{S_1,...,S_m\}$ разделено на два непересекающихся класса K_1 и K_2. Для описания объектов используется набор номинальных признаков $X(n)=(x_1,\ldots,x_n)$. Множеством допустимых значений (градаций) каждого признака $x_i \in X(n)$ является $1,2,\ldots,p_i$, $2\leq p_i \leq m/2$. Значение функции принадлежности $f_i(\mu)$ к классу K_1 по градации $\mu \in \{1,2,\ldots,p_i\}$ признака x_i вычисляется как

$$f_i(\mu)=\frac{d_{1i}(\mu)/|K_1|}{d_{1i}(\mu)/|K_1|+d_{2i}(\mu)/|K_2|},$$

где $d_{1\mu}$ ($d_{2\mu}$) – множество объектов класса K_1 (K_2), значением признака x_i у которых является μ. Выше изложенная формула используется для преобразования исходных описаний объектов E_0 к таблице $A=\{a_{uv}\}_{m\times n}$. Объект $S_u=(x_{u1},...,x_{un})(S_u\in E_0)$ запишется в виде строки таблицы как $(a_{u1},...,a_{un})$, в которой $a_{uv}=f_v(x_{uv})$. Требуется методом скользящего экзамена определить точность распознавания на E_0 по правилу ближайший сосед. При реализatsiи метода каждый объект $S_j\in E_0$ последовательно представляет контроль, $E_0\backslash\{S_j\}$ – обучение. В качестве меры расстояния между описаниями объектов в A используется метрика Журавлёва.

2.19. Berilgan $E_0=\{S_1,\ldots,S_m\}$ ob'ektlar to'plami ikkita o'zaro kesishmaydigan K_1 va K_2 sinflarga bo'lingan. To'plamning $S_u\in E_0$ ob'ekti $X(n)=(x_1,\ldots,x_n)$ turli toifadagi alomatlar to'plami bilan tavsiflanadi. Alomatlar qiymatlari $A=\{a_{uv}\}_{m*n}$ matritsa satrini hosil qiladi. $X(n)$ to'plamda β ta alomatlar nominal, $n-\beta$ tasi miqdoriy shkalalarda o'lchanadi. Har bir $x_i\in X(n)$ nominal alomatning mumkin bo'lgan qiymatlari (gradasiyalari) $1,2,\ldots,p_i$, $2\le p_i\le m/2$ sonlari bilan beriladi. Nominal x_i alomatning $\mu\in\{1,2,\ldots,p_i\}$ gradasiyalari bo'yicha K_1 sinfga tegishlilik funksiyasi qiymati quyidagi ko'rinishda hisoblanadi

$$f_i(\mu)=\frac{d_{1i}(\mu)/|K_1|}{d_{1i}(\mu)/|K_1|+d_{2i}(\mu)/|K_2|}.$$

Bu erda $d_{1\mu}$ ($d_{2\mu}$) – x_i alomatining qiymati μ bo'lgan $K_1(K_2)$ sinf ob'ektlari soni (to'plami).

Yuqorida keltirilgan formula $A=\{a_{uv}\}_{m*n}$ matritsadagi E_0 ob'ektlari tavsifidagi nominal alomatlar qiymatlarini o'girishda foydalaniladi. Agar $S_u=(a_{u1},\ldots,a_{un})$ ($S_u\in E_0$) ob'ektda a_{uv} qiymat $x_v\in X(n)$ alomat gradasiyasi bo'lsa, u holda $a_{uv}=f_v(a_{uv})$. Har bir K_i, i=1,2 sinflar uchun A matritsa bo'yicha $M_i=(b_{i1},\ldots,b_{in})$ matematik kutilma vektori hisoblansin. Mos ravishda M_1 va M_2 vektorlarni K_1 va K_2 sinflar etalonlari deb hisoblagan holda yaqin qo'shni qoidasi bo'yicha E_0 ob'ektlarini anglash aniqligi hisoblansin. M_i, i=1,2 etalonlar va A tavsiflangan $S_u\in E_0$ ob'ekt o'rtasidagi masofa o‘lchovi sifatida $\rho\left(S_u, M_i\right) = \sum_{k=1}^{n} \frac{\left|a_{uk} - b_{ik}\right|}{\left|a_{uk} + b_{ik}\right|}$ – Kanberra metrikasidan foydalanilsin.

2.19. Множество объектов $E_0=\{S_1,\dots,S_m\}$ разделено на два непересекающихся класса K_1 и K_2. Объект $S_u \in E_0$ описываются набором разнотипных признаков $X(n)=(x_1,\dots,x_n)$, значения которых представлены в виде строки таблицы $A=\{a_{uv}\}_{\text{m*n}}$. В наборе $X(n)$ β признаков измеряются в номинальной, n – β в количественной шкале измерений. Множеством допустимых значений (градаций) каждого номинального признака $x_i \in X(n)$ является $1,2,\dots,p_i$, $2 \le p_i \le m/2$. Значение функции принадлежности $f_i(\mu)$ к классу K_1 по градации $\mu \in \{1,2,\dots,p_i\}$ признака x_i вычисляется как

$$f_i(\mu) = \frac{d_{1i}(\mu)/|K_1|}{d_{1i}(\mu)/|K_1| + d_{2i}(\mu)/|K_2|},$$

где $d_{1\mu}$ ($d_{2\mu}$) – множество объектов класса K_1 (K_2), значением признака x_i у которых является μ. Приведенная формула используется для преобразования значений номинальных признаков объектов E_0 в таблице $A=\{a_{uv}\}_{\text{m*n}}$. Если в объекте $S_u=(a_{u1},\dots,a_{un})$ ($S_u \in E_0$), a_{uv} представляет градацию номинального признака $x_v \in X(n)$, то $a_{uv}=f_v(a_{uv})$.

Требуется для каждого класса K_i, $i=1,2$ по таблице A вычислить вектор математического ожидания $M_i=(b_{i1},\dots,b_{in})$. Считая M_1 и M_2 эталонами классов K_1 и K_2 определить точность распознавания объектов E_0 по правилу ближайший сосед. В качестве меры расстояния между описанием объекта $S_u \in E_0$ в A и эталоном M_i, $i=1,2$ использовать метрику Канберра

$$\rho(S_u, M_i) = \sum_{k=1}^{n} \frac{\left|a_{uk} - b_{ik}\right|}{\left|a_{uk} + b_{ik}\right|}.$$

2.20. Berilgan $E_0=\{S_1,\dots,S_m\}$ ob’ektlar to‘plami ikkita o‘zaro kesishmaydigan K_1 va K_2 sinflarga bo‘lingan. To‘plamning $S_u \in E_0$ ob’ekti $X(n)=(x_1,\dots,x_n)$ turli toifadagi alomatlar to‘plami bilan tavsiflanadi. Alomatlar qiymatlari $A=\{a_{uv}\}_{m*n}$ matritsa satrini hosil qiladi. $X(n)$ to‘plamda β ta alomatlar nominal, n – β tasi miqdoriy shkalalarda o‘lchanadi. Har bir $x_i \in X(n)$ nominal alomatning mumkin bo‘lgan qiymatlari (gradasiyalari) $1,2,\dots,p_i$, $2 \le p_i \le m/2$ sonlari bilan beriladi. Nominal x_i alomatning $\mu \in \{1,2,\dots,p_i\}$ gradasiyalari bo‘yicha K_1 sinfga tegishlilik funksiyasi qiymati quyidagi ko‘rinishda hisoblanadi

$$f_i(\mu)=\frac{d_{1i}(\mu)/|K_1|}{d_{1i}(\mu)/|K_1|+d_{2i}(\mu)/|K_2|}.$$

Bu erda $d_{1\mu}$ ($d_{2\mu}$) – x_i alomatining qiymati μ bo‘lgan K_1(K_2) sinf ob’ektlari soni (to‘plami).

Yuqorida keltirilgan formula $A=\{a_{uv}\}_{m*n}$ matritsadagi E_0 ob’ektlari tavsifidagi nominal alomatlar qiymatlarini o‘girishda foydalaniladi. Agar $S_u=(a_{u1},\ldots,a_{un})$ ($S_u\in E_0$) ob’ektda a_{uv} qiymat $x_v\in X(n)$ alomat gradasiyasi bo‘lsa, u holda $a_{uv}=f_v(a_{uv})$. Har bir K_i, i=1,2 sinflar uchun A matritsa bo‘yicha $M_i=(b_{i1},\ldots,b_{in})$ matematik kutilma vektori hisoblansin. Mos ravishda M_1 va M_2 vektorlarni K_1 va K_2 sinflar etalonlari deb hisoblagan holda ixtiyoriy mumkin bo‘lgan $S=(a_1,\ldots,a_n)$ ob’ekt uchun

$$F(S)=w_1a_1+w_2a_2+ + w_n a_n + w_0,$$

ko‘rinishidagi chiziqli qaror qilish funksiyasining analitik ko‘rinishi topilsin. Bu erda $w_1=b_{11}$, ,$w_n=b_{1n}$, $w_0=(-)\frac{1}{2}\sum_{i=1}^{n}w_i^2$.

2.20. Множество объектов $E_0=\{S_1,\ldots,S_m\}$ разделено на два непересекающихся класса K_1 и K_2. Объект $S_u\in E_0$ описываются набором разнотипных признаков $X(n)=(x_1,\ldots,x_n)$, значения которых представлены в виде строки таблицы $A=\{a_{uv}\}_{m*n}$. В наборе $X(n)$ β признаков измеряются в номинальной, n – β в количественной шкале измерений. Множеством допустимых значений (градаций) каждого номинального признака $x_i\in X(n)$ является $1,2,\ldots,p_i$, $2\le p_i\le m/2$. Значение функции принадлежности $f_i(\mu)$ к классу K_1 по градации $\mu\in\{1,2,\ldots,p_i\}$ признака x_i вычисляется как

$$f_i(\mu)=\frac{d_{1i}(\mu)/|K_1|}{d_{1i}(\mu)/|K_1|+d_{2i}(\mu)/|K_2|},$$

где $d_{1\mu}$ ($d_{2\mu}$) – множество объектов класса K_1 (K_2), значением признака x_i у которых является μ.

Приведенная формула используется для преобразования значений номинальных признаков объектов E_0 в таблице $A=\{a_{uv}\}_{m*n}$. Если в объекте $S_u=(a_{u1},\ldots,a_{un})$ ($S_u\in E_0$), a_{uv} представляет градацию номинального признака $x_v\in X(n)$, то $a_{uv}=f_v(a_{uv})$. Требуется для каждого класса K_i, i=1,2 по таблице A вычислить вектор математического ожидания $M_{1i}=(b_{i1},\ldots,b_{in})$. Считая M_1 и M_2 эталонами классов K_1 и K_2 получить аналитический вид линейной решающей функции для произвольного допустимого объекта $S=(a_1,\ldots,a_n)$

$$F(S)=w_1a_1+ w_2a_2+ + w_n\, a_n +w_0,$$

где $w_1=b_{11}$, $,w_n=b_{1n}$, $w_0=(-)\frac{1}{2}\sum_{i=1}^{n} w_i^2$.

2.21. Berilgan $E_0=\{S_1,\dots,S_m\}$ ob'ektlar to'plami ikkita o'zaro kesishmaydigan K_1 va K_2 sinflarga bo'lingan. To'plamning $S_u\in E_0$ ob'ekti $X(n)=(x_1,\dots,x_n)$ nominal alomatlar to'plami bilan tavsiflanadi. Alomatlar qiymatlari $A=\{a_{uv}\}_{m*n}$ matritsa satrini hosil qiladi. Har bir $x_i\in X(n)$ nominal alomatning mumkin bo'lgan qiymatlari (gradasiyalari) $1,2,\dots,p_i$, $2\le p_i\le m/2$ sonlari bilan beriladi. Nominal x_i alomatning $\mu\in\{1,2,\dots,p_i\}$ gradasiyalari bo'yicha K_1 sinfga tegishlilik funksiyasi qiymati quyidagi ko'rinishda hisoblanadi

$$f_i(\mu)=\frac{d_{1i}(\mu)/|K_1|}{d_{1i}(\mu)/|K_1|+d_{2i}(\mu)/|K_2|}.$$

Bu erda $d_{1\mu}$ ($d_{2\mu}$) – x_i alomatining qiymati μ bo'lgan K_1 (K_2) sinf ob'ektlari soni (to'plami).

Nominal $x_i\in X(n)$ alomat turg'unligi quyidagi formula bilan hisoblanadi:

$$U(i)=\frac{1}{m}\sum_{u=1}^{m}\begin{cases} f_i(a_{ui}), & f_i(a_{ui})>0.5,\\ 1-f_i(a_{ui}), & f_i(a_{ui})<0.5.\end{cases}$$

Berilgan E_0 to'plamining ob'ektlarini tavsiflovchi nominal alomat gradasiyalarini $a_{uv}=f_v(a_{uv})$ asoslangan holda $A=\{a_{uv}\}_{m*n}$ jadvaldagi qiymatlarga o'girishni amalga oshiring. Turg'unlik qiymatlarining kamayishi bo'yicha tartiblangan alomatlar ro'yxati chop etilsin.

2.21. Множество объектов $E_0=\{S_1,\dots,S_m\}$ разделено на два непересекающихся класса K_1 и K_2. Объект $S_u\in E_0$ описываются набором номинальных признаков $X(n)=(x_1,\dots,x_n)$, значения которых представлены в виде строки таблицы $A=\{a_{uv}\}_{m*n}$. Множеством допустимых значений (градаций) каждого номинального признака $x_i\in X(n)$ является $1,2,\dots,p_i$, $2\le p_i\le m/2$. Значение функции принадлежности $f_i(\mu)$ к классу K_1 по градации $\mu\in\{1,2,\dots,p_i\}$ признака x_i вычисляется как

$$f_i(\mu)=\frac{d_{1i}(\mu)/|K_1|}{d_{1i}(\mu)/|K_1|+d_{2i}(\mu)/|K_2|},$$

где $d_{1\mu}$ ($d_{2\mu}$) – множество объектов класса K_1 (K_2), значением признака x_i у которых является μ. Устойчивость признака $x_i\in X(n)$ вычисляется как

$$U(i)=\frac{1}{m}\sum_{u=1}^{m}\begin{cases} f_i(a_{ui}),\ f_i(a_{ui})>0.5,\\ 1-f_i(a_{ui}),\ f_i(a_{ui})<0.5.\end{cases}$$

Требуется произвести преобразование значений градаций признаков объектов E_0 в таблице $A=\{a_{uv}\}_{m*n}$ как $a_{uv}=f_v(a_{uv})$. Вывести список признаков, упорядоченных в порядке убывания значений устойчивости.

2.22. Berilgan $E_0=\{S_1,\ldots,S_m\}$ ob'ektlar to'plami ikkita o'zaro kesishmaydigan K_1 va K_2 sinflarga bo'lingan. To'plamning $S_u\in E_0$ ob'ekti $X(n)=(x_1,\ldots,x_n)$ turli toifadagi alomatlar to'plami bilan tavsiflanadi. Alomatlar qiymatlari $A=\{a_{uv}\}_{m*n}$ matritsa satrini hosil qiladi. $X(n)$ to'plamda β ta alomatlar nominal, n – β tasi miqdoriy shkalalarda o'lchanadi. Har bir $x_i\in X(n)$ nominal alomatning mumkin bo'lgan qiymatlari (gradasiyalari) $1,2,\ldots,p_i$, $2\le p_i\le m/2$ sonlari bilan beriladi. Nominal x_i alomatning $\mu\in\{1,2,\ldots,p_i\}$ gradasiyalari bo'yicha K_1 sinfga tegishlilik funksiyasi qiymati quyidagi ko'rinishda hisoblanadi

$$f_i(\mu)=\frac{d_{1i}(\mu)/|K_1|}{d_{1i}(\mu)/|K_1|+d_{2i}(\mu)/|K_2|}.$$

Bu erda $d_{1\mu}$ ($d_{2\mu}$) – x_i alomatining qiymati μ bo'lgan K_1 (K_2) sinf ob'ektlari soni (to'plami).

Yuqorida keltirilgan formula $A=\{a_{uv}\}_{m*n}$ matritsadagi E_0 ob'ektlari tavsifidagi nominal alomatlar qiymatlarini o'girishda orqali $X(n)$ to'plamidan yangi $H(n)=(y_1,\ldots,y_n)$ to'plamini tanlash uchun foydalaniladi. Agar $S_u=(a_{u1},\ldots,a_{un})$ ($S_u\in E_0$) ob'ektda a_{uv} qiymat $x_v\in X(n)$ alomat gradasiyasi bo'lsa, u holda $a_{uv}=f_v(a_{uv})$.

Har bir K_i, i=1,2 sinflar uchun A matritsa bo'yicha $M_i=(b_{i1},\ldots,b_{in})$ matematik kutilma vektorini hisoblash talab qilinadi. Har bir $y_j\in H(n)$ alomat bo'yicha fazoni qisqartirish uchun sinf ichidagi yaqinlik

$$\theta_j=\frac{1}{m}\sum_{i=1}^{2}\sum_{S_u\in K_i}\left|a_{ui}-b_{ij}\right|$$

va sinflar o'rtasidagi farqlanish

$$\gamma_j=\frac{1}{m}\sum_{i=1}^{2}\sum_{S_u\in K_i}\left|a_{ui}-b_{3-i,j}\right|$$

formula bilan hisoblanadi.

O‘chirishga nomzod sifatida $H(r)$, r=2,…,n to‘plamidan $\frac{\theta_j}{\gamma_j}=\max\limits_{y_i\in H(r)}$ shartni qanoatlantiruvchi $y_j \in H(r)$ tanlanadi. Alomatlar fazosini n dan $n-p$, $p>0$ gacha qisqartirish, o‘chirilgan alomatlar tartibini ko‘rsatgan holda amalga oshirilsin.

2.22. Множество объектов $E_0=\{S_1,\ldots,S_m\}$ разделено на два непересекающихся класса K_1 и K_2. Объект $S_u \in E_0$ описываются набором разнотипных признаков $X(n)=(x_1,\ldots,x_n)$, значения которых представлены в виде строки таблицы $A=\{a_{uv}\}_{m*n}$. В наборе $X(n)$ β признаков измеряются в номинальной, $n-\beta$ в количественной шкале измерений. Множеством допустимых значений (градаций) каждого номинального признака $x_i \in X(n)$ является 1,2,…,p_i, $2 \le p_i \le m/2$. Значение функции принадлежности $f_i(\mu)$ к классу K_1 по градации $\mu \in \{1,2,\ldots,p_i\}$ признака x_i вычисляется как

$$f_i(\mu)=\frac{d_{1i}(\mu)/|K_1|}{d_{1i}(\mu)/|K_1|+d_{2i}(\mu)/|K_2|}, \qquad (2.1)$$

где $d_{1\mu}$ ($d_{2\mu}$) – множество объектов класса K_1(K_2), значением признака x_i у которых является μ. Формула (2.1) используется для выбора нового набора $H(n)=(y_1,\ldots,y_n)$ из $X(n)$ путём преобразования значений номинальных признаков объектов E_0 в таблице $A=\{a_{uv}\}_{m*n}$. Если в объекте $S_u=(a_{u1},\ldots,a_{un})$ ($S_u \in E_0$), a_{uv} представляет градацию номинального признака $x_v \in X(n)$, то $a_{uv}=f_v(a_{uv})$.

Требуется для каждого класса K_i, i=1,2 по таблице A вычислить вектор математического ожидания $M_i=(b_{i1},\ldots,b_{in})$. Для сокращения размерности пространства по каждому признаку $y_j \in H(n)$ вычисляется внутриклассовая близость

$$\theta_j=\frac{1}{m}\sum_{i=1}^{2}\sum_{S_u\in K_i}\left|a_{ui}-b_{ij}\right|$$

и межклассовое различие

$$\gamma_j=\frac{1}{m}\sum_{i=1}^{2}\sum_{S_u\in K_i}\left|a_{ui}-b_{3-i,j}\right|.$$

В качестве кандидата на удаление из набора H®, r=2,…,n выбирается $y_j \in H$® с $\frac{\theta_j}{\gamma_j}=\max\limits_{y_i\in H(r)}$. Требуется сократить размерность признакового пространства с n до $n-p$, $p>0$ с указанием порядка удаляемых признаков.

2.23. Berilgan $E_0=\{S_1,\ldots,S_m\}$ ob’ektlar to‘plami ikkita o‘zaro kesishmaydigan K_1 va K_2 sinflarga bo‘lingan. To‘plamning $S_u \in E_0$ ob’ekti $X(n)=(x_1,\ldots,x_n)$ miqdoriy alomatlar to‘plami bilan tavsiflangan bo‘lib, ularning qiymatlari $A=\{a_{uv}\}_{m*n}$ matritsa satrini aniqlaydi. Matematik kutilma $M=(b_1,\ldots,b_n)$ vektori elementlarining qiymatlari A matritsa ustunlari bo‘yicha o‘rta arifmetik sifatida hisoblandi. M vektor yordamida $X(n)$ to‘plamida tavsiflangan ob’ektlar kengaytirilgan $Y(\mu)$, $\mu=n(n+1)/2$ to‘plamdagi tavsiflanishga o‘tkaziladi. Ixtiyoriy $S_u \in E_0$ ob’ekt $D=\{d_{uk}\}_{m*\mu}$ matritsa satri ko‘rinishida yoziladi, unda $k=1,\ldots,n$ alomatlar qiymatlari 2 ikkita gradasiyaga

$$d_{uk}=\begin{cases}1, & a_{ui}<b_i,\\ 2, & a_{ui}>b_i,\end{cases}$$

$k=n+1,\ldots, n(n+1)/2-4$ uchun

$$d_{uk}=\begin{cases}1, & a_{ui}<b_i \quad \text{va} \quad a_{uj}<b_j,\\ 2, & a_{ui}>b_i \quad \text{va} \quad a_{uj}>b_j,\\ 3, & a_{ui}>b_i \quad \text{va} \quad a_{uj}<b_j,\\ 4, & a_{ui}<b_i \quad \text{va} \quad a_{uj}>b_j.\end{cases}$$

Ixtiyoriy $y_i \in Y(\mu)$ alomatning $\delta \in \{1,\ldots,4\}$ gradasiyasi bo‘yicha $f_i(\delta)$ – K_1 sinfga tegishlilik funksiyasining qiymati

$$f_i(\delta)=\frac{z_{1i}(\delta)/|K_1|}{z_{1i}(\delta)/|K_1|+z_{2i}(\delta)/|K_2|}$$

formulasi bilan hisoblanadi, bu erda $z_{1\mu}$ ($z_{2\mu}$) – y_i alomatining qiymati δ bo‘lgan K_1 (K_2) sinf ob’ektlari soni (to‘plami). Yuqorida keltirilgan formula $D=\{d_{uk}\}_{m*\mu}$ matritsaga $d_{uk}=f_k(d_{uk})$ orqali E_0 ob’ektlari tavsifidagi nominal alomatlar qiymatlarini o‘girishda foydalaniladi. Har bir K_i, $i=1,2$ sinf uchun matematik kutilma $M_i=(m_{i1},\ldots,m_{i\mu})$ vektorini hisoblash talab qilinadi. Qaysi $y_j \in Y(\mu)$ alomat bo‘yicha $|m_{1j}-m_{2j}|=max$ ekanligi aniqlansin.

2.23. Множество объектов $E_0=\{S_1,\ldots,S_m\}$ разделено на два непересекающихся класса K_1 и K_2. Объект $S_u \in E_0$ описываются набором количественных признаков $X(n)=(x_1,\ldots,x_n)$, значения которых представлены в виде строки таблицы $A=\{a_{uv}\}_{m*n}$. Определены значения элементов вектора математического ожидания $M=(b_1,\ldots,b_n)$ как среднего арифметического по столбцам матрицы A. С помощью вектора M производится переход от представления объектов в $X(n)$ к представлению по расширенному набору $Y(\mu)$, $\mu=n(n+1)/2$. Объект $S_u \in E_0$ записывается в виде строки матрицы $D=\{d_{uk}\}_{m*\mu}$ в которой значения признаков с $k=1,\ldots,n$ имеют 2 градации

$$d_{uk}=\begin{cases}1, & a_{ui}<b_i,\\ 2, & a_{ui}>b_i,\end{cases}$$

с $k=n+1,\ldots, n(n+1)/2-4$

$$d_{uk}=\begin{cases}1, & a_{ui}<b_i \quad \text{и} \quad a_{uj}<b_j,\\ 2, & a_{ui}>b_i \quad \text{и} \quad a_{uj}>b_j,\\ 3, & a_{ui}>b_i \quad \text{и} \quad a_{uj}<b_j,\\ 4, & a_{ui}<b_i \quad \text{и} \quad a_{uj}>b_j.\end{cases}$$

Значение функции принадлежности $f_i(\delta)$ к классу K_1 по градации $\delta\in\{1,\ldots,4\}$ признака $y_i\in Y(\mu)$ вычисляется как

$$f_i(\delta)=\frac{z_{1i}(\delta)/|K_1|}{z_{1i}(\delta)/|K_1|+z_{2i}(\delta)/|K_2|},$$

где $z_{1\mu}$ ($z_{2\mu}$) – множество объектов класса K_1 (K_2), значением признака y_i у которых является δ. Приведенная формула используется для преобразования значений номинальных признаков объектов E_0 по таблице $D=\{d_{uk}\}_{m*\mu}$ как $d_{uk}=f_k(d_{uk})$. Требуется для каждого класса K_i, i=1,2 по таблице D вычислить вектор математического ожидания $M_i=(m_{i1},\ldots,m_{i\mu})$. Определить по какому признаку $y_j\in Y(\mu)$ справедливо $|m_{1j}-m_{2j}|=max$.

2.24. Ob’ektlar to‘plami $E_0=\{S_1,\ldots,S_m\}$ ikkita o‘zaro kesishmaydigan K_1 va K_2 sinflarga ajratilgan. Ob’ektlarni tavsiflash uchun nominal alomatlarning $X(n)=(x_1,\ldots,x_n)$ majmuasidan foydalanilgan. Har bir $x_i\in X(n)$ nominal alomatning mumkin bo‘lgan qiymatlar (gradasiyalar) to‘plami $1,2,\ldots,p_i$, $2\le p_i\le m/2$ qiymatlardan iborat. Nominal x_i alomatning $\mu\in\{1,2,\ldots,p_i\}$ gradasiyasi bo‘yicha K_1 sinfga tegishlilik funksiyasi

$$f_i(\mu)=\frac{d_{1i}(\mu)/|K_1|}{d_{1i}(\mu)/|K_1|+d_{2i}(\mu)/|K_2|}$$

ko‘rinishida hisoblanadi. Bu erda $d_{1\mu}$ ($d_{2\mu}$) – x_i alomat qiymati μ bo‘lgan K_1 (K_2) sinf ob’ektlari to‘plami.

Har bir $x_i \in X(n)$ alomat bo‘yicha sinflar o‘rtasidagi chegara $G_i=(s_1+s_2)/2$ orqali hisoblansin. Bu erda

$$s_2=\min_{\{0.5-f_i(\mu)>0\}}\left(0.5-f_i\left(\mu\right)\right) \text{ va } s_1=\max_{\{1-f_i(\mu)<0.5\}}\left(1-f_i\left(\mu\right)\right).$$

Berilgan E_0 tanlanmaning boshlang‘ich ob’ektlarini $A=\{a_{uv}\}_{m*n}$ jadval ko‘rinishida, uning kengaymasini $B=\{b_{rc}\}_{\mathrm{m*p}}$, $p=n(n-1)/2$ taqdim etilsin. Ixtiyoriy $S_u=(x_{u1},\dots,x_{un})$ ($S_u \in E_0$) ob’ekt A jadvalida $(a_{u1},\dots,a_{un})$ satr sifatida yoziladi, bu erda

$$a_{uv}=\begin{cases}1, & f_v\left(x_{uv}\right)<G_v,\\ 2, & f_v\left(x_{uv}\right)>G_v.\end{cases}$$

O‘z navbatida V jadvalning har elementi A jadvalining ikkita satridan $b_{uc}=2a_{ui}+a_{uj}-2$, $i \neq j$, $c=2(i-1)+j$ formulalar orqali shakllantiriladi.

2.24. Множество объектов $E_0=\{S_1,\dots,S_m\}$ разделено на два непересекающихся класса K_1 и K_2. Для описания объектов используется набор номинальных признаков $X(n)=(x_1,\dots,x_n)$. Множеством допустимых значений (градаций) каждого признака $x_i \in X(n)$ является $1,2,\dots,p_i$, $2 \le p_i \le m/2$. Значение функции принадлежности $f_i(\mu)$ к классу K_1 по градации $\mu \in \{1,2,\dots,p_i\}$ признака x_i вычисляется как

$$f_i(\mu)=\frac{d_{1i}(\mu)/|K_1|}{d_{1i}(\mu)/|K_1|+d_{2i}(\mu)/|K_2|},$$

где $d_{1\mu}$ ($d_{2\mu}$) – множество объектов класса K_1 (K_2), значением признака x_i у которых является μ. По каждому признаку $x_i \in X(n)$ определить значение границы между классами как $G_i=(s_1+s_2)/2$, где

$$s_1=\max_{\{1-f_i(\mu)<0.5\}}\left(1-f_i\left(\mu\right)\right) \text{ и } s_2=\min_{\{0.5-f_i(\mu)>0\}}\left(0.5-f_i\left(\mu\right)\right).$$

Требуется представить исходное описание объектов E_0 в таблице $A=\{a_{uv}\}_{\mathrm{m*n}}$ и её расширении $B=\{b_{rc}\}_{\mathrm{m*p}}$, $p=n(n-1)/2$. Объект $S_u=(x_{u1},\dots,x_{un})$ ($S_u \in E_0$) в A запишется как строка таблицы $(a_{u1},\dots,a_{un})$, в которой

$$a_{uv}=\begin{cases}1, & f_v\left(x_{uv}\right)<G_v,\\ 2, & f_v\left(x_{uv}\right)>G_v.\end{cases}.$$

Каждый элемент таблицы B для объекта S_u формируется из двух строк A как $b_{uc}=2a_{ui}+a_{uj}-2$, $i \neq j$, $c=2(i-1)+j$.

2.25. Ob’ektlar to‘plami $E_0=\{S_1,\ldots,S_m\}$ ikkita o‘zaro kesishmaydigan K_1 va K_2 sinflarga ajratilgan. Ob’ektlarni tavsiflash uchun nominal alomatlarning $X(n)=(x_1,\ldots,x_n)$ majmuasidan foydalanilgan. Har bir $x_i\in X(n)$ nominal alomatning mumkin bo‘lgan qiymatlar (gradasiyalar) to‘plami $1,2,\ldots,p_i$, $2\le p_i\le m/2$ qiymatlardan iborat. Nominal x_i alomatning $\mu\in\{1,2,\ldots,p_i\}$ gradasiyasi bo‘yicha K_1 sinfga tegishlilik funksiyasi

$$f_i(\mu)=\frac{d_{1i}(\mu)/|K_1|}{d_{1i}(\mu)/|K_1|+d_{2i}(\mu)/|K_2|}$$

ko‘rinishida hisoblanadi. Bu erda $d_{1\mu}$ ($d_{2\mu}$) – x_i alomat qiymati μ bo‘lgan K_1 (K_2) sinf ob’ektlari to‘plami.

Har bir $x_i\in X(n)$ alomatning gradasiyasi bo‘yicha $r_i=(r_{i1},\ldots,r_{ip(i)})$, $r_{i\mu} = f_i(\mu)$ vektor hosil qilinsin va sinflar o‘rtasidagi G_i=(s1 + s2)/2 ko‘rinishidagi chegaraviy qiymat hisoblansin. Bu erda s2=max$\{f_c(\mu)|\ 0.5 - f_c(\mu)>0\}$ va s1=min$\{f_c(\mu)|\ 1 - f_c(\mu)<0.5\}$.

Mumkin bo‘lgan $S=(b_1,\ldots,b_n)$ ob’ekt tavsifida ikki bosqichli o‘girish amalga oshirilsin. Bu erda $c=b_i$, $b_i^*=r_{ic}$, $b_i^{**}=\begin{cases}1, & b_i^*<G_i,\\ 2, & b_i^*>G_i.\end{cases}$

2.25. Множество объектов $E_0=\{S_1,\ldots,S_m\}$ разделено на два непересекающихся класса K_1 и K_2. Для описания объектов используется набор номинальных признаков $X(n)=(x_1,\ldots,x_n)$. Множеством допустимых значений (градаций) каждого признака $x_i\in X(n)$ является $1,2,\ldots,p(i)$, $2\le p(i)\le)m/2$. Значение функции принадлежности $f_i(\mu)$ к классу K_1 по градации $\mu\in\{1,2,\ldots, p(i)\}$ признака x_i вычисляется как

$$f_i(\mu)=\frac{d_{1i}(\mu)/|K_1|}{d_{1i}(\mu)/|K_1|+d_{2i}(\mu)/|K_2|},$$

где $d_{1\mu}$ ($d_{2\mu}$) – множество объектов класса K_1 (K_2), значением признака x_i у которых является μ.

По градациям каждого признака $x_i\in X(n)$ необходимо сформировать вектор $r_i=(r_{i1},\ldots,r_{ip(i)})$, $r_{i\mu} = f_i(\mu)$ и вычислить значение границы между классами как G_i=(s1 + s2)/2, где s2=max$\{f_c(\mu)|\ 0.5 - f_c(\mu)>0\}$ и s1=min$\{f_c(\mu)|\ 1 - f_c(\mu)<0.5\}$.

Требуется произвести двухэтапное преобразование описаний произвольного допустимого объекта $S=(b_1,\ldots,b_n)$, $c=b_i$, $b_i^*=r_{ic}$,

$$b_i^{**}=\begin{cases}1, & b_i^*<G_i,\\ 2, & b_i^*>G_i.\end{cases}$$

2.26. Ob'ektlar to'plami $E_0=\{S_1,\ldots,S_m\}$ ikkita o'zaro kesishmaydigan K_1 va K_2 sinflarga ajratilgan. Ob'ektlarni tavsiflash uchun miqdoriy alomatlarning $X(n)=(x_1,\ldots,x_n)$ majmuasidan foydalanilgan bo'lib, ularning qiymatlari $A=\{a_{uv}\}_{m*n}$ matritsaning satri ko'rinishida taqdim etilgan. Matematik kutilma – $M=(b_1,\ldots,b_n)$ vektorining elementlari A matritsa ustunlari bo'yicha o'rta arifmetik sifatida aniqlangan bo'lsin. M vektor yordamida ob'ektlarning tavsifini $X(n)$ alomatlar majmuasidan kengaytirilgan $Y(\mu)$, $\mu=n(n+1)/2$ majmuasiga o'tkazish amalga oshiriladi.

Ixtiyoriy $S_u\in E_0$ ob'ekt $D=\{d_{uk}\}_{m*\mu}$ matritsa satri

$$d_{uk}=\begin{cases}1, & a_{ui}<b_i,\\ 2, & a_{ui}>b_i,\end{cases}$$

ko'rinishida yozilib, unda $k=1,\ldots,n$ alomatlar qiymatlari ko'rinishida hisoblanadigan 2 gradasiyaga, tartib nomerlari $k=n+1,\ldots,\ n(n+1)/2$ bo'lganlar

$$d_{uk}=\begin{cases}1, & a_{ui}<b_i \quad \text{va} \quad a_{uj}<b_j,\\ 2, & a_{ui}>b_i \quad \text{va} \quad a_{uj}>b_j,\\ 3, & a_{ui}>b_i \quad \text{va} \quad a_{uj}<b_j,\\ 4, & a_{ui}<b_i \quad \text{va} \quad a_{uj}>b_j.\end{cases}$$

formula orqali hisoblanuvchi 4 gradasiyaga ega bo'ladi.

Ixtiyoriy $y_i\in Y(\mu)$ alomatning $\delta\in\{1,\ldots,4\}$ gradasiyalari bo'yicha K_1 sinfiga $f_i(\delta)$ tegishlilik funksiyasining qiymati quyidagicha hisoblanadi:

$$f_i(\delta)=\frac{z_{1i}(\delta)/|K_1|}{z_{1i}(\delta)/|K_1|+z_{2i}(\delta)/|K_2|}.$$

Bu erda $z_{1i}(\delta)(z_{2i}(\delta))$ – y_i alomatning qiymati δ bo'lgan K_1 (K_2) sinf ob'ektlari to'plami. Tegishlilik funksiyasini qiymatini hisoblash formulasi E_0 tanlanmadagi ob'ektlarning nominal alomatlar qiymatlarini $D=\{d_{uk}\}_{m*\mu}$ jadvali bo'yicha $d_{uk}=f_k(d_{uk})$ sifatida o'girish uchun foydalaniladi. D jadval bo'yicha

$$U(i)=\frac{1}{m}\sum_{u=1}^{m}\begin{cases}f_i(d_{ui}), & f_i(d_{ui})>0.5,\\ 1-f_i(d_{ui}), & f_i(d_{ui})<0.5.\end{cases}$$

orqali har bir $y_i \in Y(\mu)$ alomat turg'unligini hisoblash talab qilinadi. $Y(\mu)$ majmuasidagi alomatlaridan turg'unligi maksimal qiymatiga ega bo'lgan 5 ta alomat nomerlari ro'yxati chop etilsin.

2.26. Множество объектов $E_0=\{S_1,\ldots,S_m\}$ разделено на два непересекающихся класса K_1 и K_2. Объект $S_u \in E_0$ описываются набором количественных признаков $X(n)=(x_1,\ldots,x_n)$, значения которых представлены в виде строки таблицы $A=\{a_{uv}\}_{m*n}$. Определены значения элементов вектора математического ожидания $M=(b_1,\ldots,b_n)$ как среднего арифметического по столбцам матрицы A. С помощью вектора M производится переход от представления объектов в $X(n)$ к представлению по расширенному набору $Y(\mu)$, $\mu=n(n+1)/2$. Объект $S_u \in E_0$ записывается в виде строки матрицы $D=\{d_{uk}\}_{m*\mu}$ в которой значения признаков с $k=1,\ldots,n$ имеют 2 градации

$$d_{uk}=\begin{cases}1, & a_{ui}<b_i,\\ 2, & a_{ui}>b_i,\end{cases}$$

с $k=n+1,\ldots,\ n(n+1)/2$ – 4 градации

$$d_{uk}=\begin{cases}1, & a_{ui}<b_i \quad \text{и} \quad a_{uj}<b_j,\\ 2, & a_{ui}>b_i \quad \text{и} \quad a_{uj}>b_j,\\ 3, & a_{ui}>b_i \quad \text{и} \quad a_{uj}<b_j,\\ 4, & a_{ui}<b_i \quad \text{и} \quad a_{uj}>b_j.\end{cases}$$

Значение функции принадлежности $f_i(\delta)$ к классу K_1 по градации $\delta \in \{1,\ldots,4\}$ признака $y_i \in Y(\mu)$ вычисляется как

$$f_i(\delta)=\frac{z_{1i}(\delta)/|K_1|}{z_{1i}(\delta)/|K_1|+z_{2i}(\delta)/|K_2|},$$

где $z_{1i}(\delta)(z_{2i}(\delta))$ – множество объектов класса K_1 (K_2), значением признака y_i у которых является δ. Приведенная формула используется для преобразования значений номинальных признаков объектов E_0 по таблице $D=\{d_{uk}\}_{m*\mu}$ как $d_{uk}=f_k(d_{uk})$. Требуется по таблице D вычислить устойчивость каждого признака $y_i \in Y(\mu)$ как

$$U(i)=\frac{1}{m}\sum_{u=1}^{m}\begin{cases}f_i(d_{ui}), & f_i(d_{ui})>0.5,\\ 1-f_i(d_{ui}), & f_i(d_{ui})<0.5.\end{cases}$$

Вывести список 5 номеров признаков из $Y(\mu)$, имеющих максимальные значения устойчивости.

2.27. Ob’ektlar tanlanmasi $E_0=\{S_1,\ldots,S_m\}$ ikkita o‘zaro kesishmaydigan K_1 va K_2 sinflarga ajratilgan. Har bir $S\in E_0$ ob’ekt, mumkin bo‘lgan qiymatlari to‘plami {1,2} gradasiyalari iblorat $X(n)=(x_1,\ldots,x_n)$ nominal alomatlar majmuasi bilan tavsiflangan.

Mos ravishda K_1 va K_2 sinflar ob’ektlari tavsifidagi $x_c\in X(n)$ alomatning gradasiyalar qiymatlarining miqdorlarini g_{1c}^j, g_{2c}^j orqali belgilaylik ($j\in\{1,2\}$).

Nominal x_c alomat bo‘yicha sinflararo farqlanish

$$\lambda_c = 1 - \frac{\sum_{j=1}^{2} g_{1c}^j g_{2c}^j}{|K_1||K_2|}. \tag{1}$$

kattalik ko‘rinishida aniqlanadi.

Quyidagi formula bilan K_1,K_2 sinflar bo‘yicha alomat gradasiyalarining qiymatlari asosida bir jinslilik darajasi (sinf ichidagi o‘xshashlik o‘lchovi) β_c hisoblanadi:

$$\beta_c = \frac{\sum_{j=1}^{2} g_{1c}^j (g_{1c}^j - 1) + g_{2c}^j (g_{2c}^j - 1)}{|K_1|(|K_1|-1) + |K_2|(|K_2|-1)}. \tag{2}$$

Yuqorida keltirilgan (1) va (2) formulalar yordamida $x_c\in X(n)$ alomatning nominal shkaladagi vazni sinf ichidagi o‘xshashlik va sinflar o‘rtasidagi farqlanishning ko‘paytmasi ko‘rinishida aniqlangan:

$$v_s=\beta_s\lambda_s. \tag{3}$$

Nominal $x_d\in X(n)$ alomatning $j\in\{1,2\}$ gradasiyalar va (3) vazn bo‘yicha hissasi quyidagi formula bilan aniqlangan:

$$\mu_d(j) = v_d \left(\frac{\alpha_{dj}^1}{|K_1|} - \frac{\alpha_{dj}^2}{|K_2|} \right), \tag{4}$$

bu erda $\alpha_{dj}^1, \alpha_{dj}^2$ – mos ravishda K_1 i K_2 sinflardagi x_d alomatning j gradasiyasining qiymatlari miqdori. Har bir $S_r\in E_0$ ob’ektning nominal shkaladagi o‘lchami $S_r=(x_{r1},\ldots,x_{rn})$ va (4) hissa bo‘yicha umumlashgan bahosi

$$Z(S_r) = \sum_{i=1}^{n} \mu_i(a_{ri}).$$

ko‘rinishida hisoblanadi.

Aniqlash talab qilinadi:

$$p_1=|\{S_i\in K_1 \mid Z(S_i)<0\}| \text{ va } p_2=|\{S_i\in K_2 \mid Z(S_i)>0\}|.$$

2.27. Выборка объектов $E_0=\{S_1,\ldots,S_m\}$ разделено на два непересекающихся класса K_1 и K_2. Каждый объект $S\in E_0$ описываются набором номинальных признаков $X(n)=(x_1,\ldots,x_n)$, множеством допустимых значений которых являются градации $\{1,2\}$.

Обозначим через g^j_{1c}, g^j_{2c} – количество значений градации $j\in\{1,2\}$ признака $x_c\in X(n)$ в описании объектов соответственно класса K_1 и K_2. Межклассовое различие по признаку x_c определяется как величина

$$\lambda_c = 1 - \frac{\sum_{j=1}^{2} g^j_{1c} g^j_{2c}}{|K_1||K_2|}. \qquad (2.1)$$

Степень однородности (мера внутриклассового сходства) β_c значений градаций признака по классам K_1, K_2 вычисляется по формуле:

$$\beta_c = \frac{\sum_{j=1}^{2} g^j_{1c}(g^j_{1c}-1) + g^j_{2c}(g^j_{2c}-1)}{|K_1|(|K_1|-1)+|K_2|(|K_2|-1)}. \qquad (2.2)$$

С помощью (2.1),(2.2) вес признака $x_c\in X(n)$ в номинальной шкале определяется как произведение внутриклассового сходства и межклассового различия

$$v_c=\beta_c\lambda_c. \qquad (2.3)$$

Вклад признака $x_d\in X(n)$ по градации $j\in\{1,2\}$ и весу (2.3) определяется как

$$\mu_d(j) = v_d\left(\frac{\alpha^1_{dj}}{|K_1|} - \frac{\alpha^2_{dj}}{|K_2|}\right), \qquad (2.4)$$

где $\alpha^1_{dj}, \alpha^2_{dj}$ – количество значений градации j признака x_d соответственно в классах K_1 и K_2. Обобщённая оценка объекта $S_r\in E_0$ по его описанию в номинальной шкале измерений $S_r=(x_{r1},\ldots,x_{rn})$ и вкладам (2.4) вычисляется как $Z(S_r)=\sum_{i=1}^{n}\mu_i(a_{ri})$.

Требуется определить $p_1=|\{S_i\in K_1 \ |Z(S_i)<0\}|$ и $p_2=|\{S_i\in K_2 \ |Z(S_i)>0\}|$.

2.28. (*Konkurient o'xshashlikni hisoblash*). Berilgan $E_0=(S_1,\ldots,S_m)$ tanlanma ob'ektlari $X(n)=(x_1,\ldots,x_n)$, $n>2$ – miqdoriy alomatlar bilan tavsiflangan. Tanlanma ob’ektlari ikkita o‘zaro kesishmaydigan K_1 va K_2 sinflarga bo‘lingan. Ob’ektlar o‘rtasidagi yaqinlik $\rho(x,y)$ metrika bilan aniqlangan bo‘lsin.

Berilgan $S \in K_t$, t=1,2 ob’ekt uchun talab qilinadi:

- barcha $S_j \in K_{3-t}$ ob’ektlar bo‘yicha $S_i \in K_t$ ob’ektga nisbatan konkurient o‘xshashlik o‘lchami

$$F(S,S_i|S_j) = (\rho(S,S_j) - \rho(S,S_i))/(\rho(S,S_j) + \rho(S,S_i))$$

formula orqali aniqlansin;

- $F(S,S_i|S_j)$ qiymati maksimal va minimal bo‘lgan ob’ektlar ajratib ko‘rsatilsin;

- olingan natijalar Evklid va Chebishev metrikalari bo‘yicha taqqoslansin.

2.28. (*Вычисление конкурентного сходства*). Дано описание множества объектов $E_0=(S_1,\ldots,S_m)$ с помощью набора количественных признаков $X(n)=(x_1,\ldots,x_n)$. Объекты E_0 разделены на два непересекающихся класса K_1 и K_2. Считается, что близость между объектами определяется по метрике $\rho(x,y)$. Требуется для заданного объекта $S \in K_t$, t=1,2:

– определить меру конкурентного сходства относительно объекта $S_i \in K_t$ по всем объектам $S_j \in K_{3-t}$, используя формулу

$F(S,S_i|S_j) = (\rho(S,S_j) - \rho(S,S_i))/(\rho(S,S_j) + \rho(S,S_i))$;

– выделить объекты с максимальным и минимальным значениями $F(S,S_i|S_j)$;

– сравнить результаты по метрике Евклида и Чебышева.

2.29. (*Taqsimot zichligi*). Turli toifadagi alomatlar $X(n)=(x_1,\ldots,x_n)$ bilan tavsiflangan ob’ektlar to‘plami $E_0=(S_1,\ldots,S_m)$ berilgan. E_0 ob’ektlari l ta o‘zaro kesishmaydigan $K_1,\ldots,K_l$ sinflarga bo‘lingan. Juravlev metrikasi $\rho(x,y)$ bo‘yicha chegaraviy ob’ektlar to‘plami

$$B \subset E_0,\quad B=\left\{S_i \,\middle|\, S_i \in K_t, \rho(S_i,S_j)=\min_{S_r \in CK_t} \rho(S_i,S_r), t=1,2\right\}$$

ajratib olingan bo‘lsin.

Har bir $S \in B \cap K_t$ ob’ekt uchun aniqlansin:

– $r_s = \arg\min_{S_j \in CK_t} \rho(S,S_j)$ qiymatlari;

– taqsimot zichligi $P(S)=\sum_{\rho(S_i,S)<r_s}\left(1-\frac{\rho(S_i,S)}{r_s}\right)/k$, bu erda $k=|\{S_i|\rho(S,S_i<r_s)\}|$.

2.29. (*Плотность распределения*). Дано описание множества объектов $E_0=(S_1,\ldots,S_m)$ с помощью набора разнотипных признаков $X(n)=(x_1,\ldots,x_n)$. Объекты E_0 разделены на l непересекающихся класса $K_1,\ldots,K_l$. Считается, что по метрике Журавлёва $\rho(x,y)$ выделено множество граничных объектов $B\subset E_0$, $B=\left\{S_i \middle| S_i\in K_t, \rho(S_i,S_j)=\min\limits_{S_r\in CK_t}\rho(S_i,S_r), t=1,2\right\}$. Требуется для каждого $S\in B\cap K_t$ определить:

– значение $r_s=\arg\min\limits_{S_j\in CK_t}\rho(S,S_j)$;

– плотность распределения $P(S)=\sum\limits_{\rho(S_i,S)<r_s}\left(1-\frac{\rho(S_i,S)}{r_s}\right)/k$, где $k=|\{S_i|\rho(S,S_i<r_s)\}|$.

3. Берилганларнинг кластер таҳлили (Кластерный анализ данных)

3.1. Berilgan $A=\{S_1,\ldots,S_m\}$ tanlanmada berilgan $k\ (2\le k<m)$ uchun “*sinf ichidagi o‘rtacha*” algoritmi orqali klasterlash amalga oshirilsin. Obyektlar o‘rtasidagi masofani hisoblash uchun Evklid va Chebishev metrikasidan foydalanilsin.

3.1. На заданной выборке $A=\{S_1,\ldots,S_m\}$ произвести группировку объектов на $k\ (2\le k<m)$ кластеры с помощью алгоритма “*внутригрупповых средних*”. Для вычисления расстояния между объектами, описываемых n, $n\ge 2$ количественными признаками, использовать метрику Евклида и Чебышева.

3.2. Berilgan R bo‘sag‘a qiymat va $\rho(x,y)$ metrika bo‘yicha $A=\{S_1,\ldots,S_m\}$ tanlanmani kesishmaydigan guruhlarga ajratishlar sonini hisoblansin. Masofani hisoblash uchun Hemming va Chebishev metrikasidan foydalanilsin.

3.2. По заданному пороговому значению R и метрике $\rho(x,y)$ вычислить число разбиений выборки $A=\{S_1,\ldots,S_m\}$ в R^n, $n\ge 2$ на непересекающиеся группы. Для вычисления расстояния использовать метрики Хемминга и Чебышева.

3.3. Obektlari R^n, $n\ge 2$ miqdoriy alomatlar fazosida tavsiflangan $A=\{S_1,\ldots,S_m\}$ tanlanma uchun Evklid metrikasidan foydalangan holda *maksmin* masofa algoritmi amalga oshirilsin. To‘xtash mezoni uchun quyidagi shartni qabul qiling: guruh markazlarigacha bo‘lgan minimal masofalardan maksimali berilgan d qiymatidan kichik.

3.3. Реализовать алгоритм максиминного расстояния по евклидовой метрике для выборки $A = \{S_1, \dots, S_m\}$ в R^n, $n \geq 2$. За критерий останова принять условие: максимальное из минимальных расстояний до центров групп меньше заданного значения d.

3.4. Obektlari R^n, $n \geq 2$ turli toifadagi (miqdoriy va nominal) alomatlar fazosida tavsiflangan $A = \{S_1, \dots, S_m\}$ tanlanma uchun Juravlev metrikasidan foydalangan holda *maksmin* masofa algoritmi amalga oshirilsin. To‘xtash mezoni uchun quyidagi shartni qabul qiling: guruh markazlarigacha bo‘lgan minimal masofalardan maksimali berilgan d qiymatidan kichik.

3.4. Реализовать алгоритм максиминного расстояния по метрике Журавлёва для выборки $A = \{S_1, \dots, S_m\}$ с описанием объектов набором разнотипных (количественных и номинальных) признаков $X(n)=(x_1,\dots,x_n)$. За критерий останова принять условие: максимальное из минимальных расстояний до центров групп меньше заданного значения R.

3.5. $E_0 = \{S_1, \dots, S_m\}$ o‘rgatuvchi tanlanma berilgan. Har bir $S \in E_0$ obyekt n ta miqdoriy alomat bilan tavsiflanadi. Berilgan $\rho(x, y)$ metrika bo‘yicha $S_1, \dots, S_m$ obyektlar orasidagi eng qisqa yopiqmas yo‘l qurilsin. $\rho(x, y)$ metrika sifatida Evklid, Hemming, Chebishev metrikalari bo‘lgan variantlar ko‘rib chiqilsin.

3.5. По заданной выборке $A=\{S_1, \dots, S_m\}$, каждый объект которой описывается с помощью n количественных признаков, и заданной метрике $\rho(x, y)$ построить кратчайший незамкнутый путь (КНП) между объектами $S_1, \dots, S_m$. В качестве $\rho(x, y)$ использовать метрики Евклида, Хемминга, Чебышева.

3.6. O‘zaro kesishmaydigan l ta $K_1, \dots, K_l$ sinflarga bo‘lingan $E_0 = \{S_1, \dots, S_m\}$ o‘rgatuvchi tanlanma berilgan. Har bir $S \in E_0$ – obyekt n ta miqdoriy alomatlar bilan tavsiflangan. “*Yaqin qo’shni*” algoritmi uchun “*ketma-ket o‘chirish*” usulini qo’llash orqali tanlanmani lokal-optimal qoplovchisi obyekt-etalonlarini izlashni amalga oshirilsin. Yaqinlik o’lchovi sifatida Evklid metrikasi ishlatilsin.

3.6. Реализовать поиск объектов-эталонов локально-оптимального покрытия обучающей выборки $E_0 = \{S_1, \dots, S_m\}$, разделённой на l непересекающихся классов $K_1, \dots, K_l$ методом “последовательного исключения” для алгоритма “ближайший сосед”. Считается, что каждый объект $S \in E_0$ описывается с помощью n количественных признаков и в качестве меры близости между объектами используется метрика Евклида.

3.7. Berilgan $E_0=\{S_1,\ldots,S_m\}$ o‘rgatuvchi ob’ektlar to‘plami ikkita o‘zaro kesishmaydigan K_1 va K_2 sinflarga bo‘lingan. To‘plamning $S_u \in E_0$ ob’ekti $X(n)=(x_1,\ldots,x_n)$ turli toifadagi n ta alomatlar bilan tavsiflangan bo‘lib, ularning p tasi miqdoriy, $n-p$ tasi nominal shkalalarda o‘lchanadi. Berilgan $l, l \geq 2$ son bo‘yicha x_j, $x_j \in X(n)$ alomat qiymatlarini l ta o‘zaro kesishmaydigan $[a_0, a_1], (a_1, a_2], \ldots, (a_{l-1}, a_l]$ intervallarga bo‘linsin. Bu erda $a_0 = \min_{S_i \in E_0} x_{ij}$, $a_l = \max_{S_i \in E_0} x_{ij}$, $S_i = (x_{i1}, \ldots, x_{in})$, $h=(a_0+a_l)/l$ qadam bilan. $(a_{d-1}, a_d]$, $1 \leq d \leq l$ interval bo‘yicha tanlanma ob’ektlarining K_1 sinfga tegishlilik funksiyasi $n_{1d}+n_{2d}>0$ sharti bilan

$$f_d = \frac{n_{1d}/|K_1|}{n_{1d}/|K_1| + n_{2d}/|K_2|},$$

ko‘rinishda hisoblanadi. Bu erda $n_{1d}(n_{2d})$- $(a_{d-1}, a_d]$ intervaldagi $K_1(K_2)$ sinf vakillari soni. Quyidagi formula bilan $x_j \in X(n)$ alomat qiymatlarini l ta intervallarga bo‘lish turg‘unligi hisoblansin:

$$\beta_j = \frac{1}{\sigma} \sum_{|\{(a_{d-1}, a_d]\}|>0} \begin{cases} f_d, f_d > 0.5, \\ 1 - f_d, f_d < 0.5. \end{cases}$$

Bu erda $\sigma - n_{1d} + n_{2d} > 0$ shartni qanoatlantiruvchi intervallar soni.

3.7. В обучающей выборке $E_0 = \{S_1, \ldots, S_m\}$ содержаться представители двух непересекающихся классов K_1, K_2. Каждый объект выборки описывается набором $X(n)=(x_1,\ldots,x_n)$ из n разнотипных признаков, p из которых измеряются в количественной, $n-p$ номинальной шкалах измерений. По заданному числу $l, l \geq 2$ произвести разбиения значений количественного признака x_j $x_j \in X(n)$ на l непересекающихся интервалов $[a_0, a_1], (a_1, a_2], \ldots, (a_{l-1}, a_l]$,…где $a_0 = \min_{S_i \in E_0} x_{ij}$, $a_l = \max_{S_i \in E_0} x_{ij}$, $S_i = (x_{i1}, \ldots, x_{in})$ с шагом $h=(a_0+a_l)/l$. Функция принадлежности объектов к классу K_1 по интервалу $(a_{d-1}, a_d]$ при условии $n_{1d}+n_{2d}>0$ вычисляется как

$$f_d = \frac{\frac{n_{1d}}{|K_1|}}{n_{1d}/|K_1| + n_{2d}/|K_2|},$$

где $n_{1d}(n_{2d})$ – количество объектов класса $K_1(K_2)$ в интервале $(a_{d-1}, a_d]$. Определить значение устойчивости разбиения на l интервалов по признаку $x_j \in X(n)$

$$\beta_j = \frac{1}{\sigma} \sum_{|\{(a_{d-1}, a_d]\}|>0} \begin{cases} f_d, f_d > 0.5, \\ 1 - f_d, f_d < 0.5, \end{cases}$$

где σ – количество интервалов с $n_{1d} + n_{2d} > 0$.

3.8. m obyektdan tashkil topgan n o‘lchovli alomatlar fazosida $A = \{S_1, \dots, S_m\}$ ko‘rinishda to‘plam berilgan. Berilgan k parametr bo‘yicha $R_1, \dots, R_m$ qiymatlarini hisoblansin. R_i ning qiymati markazi S_i da bo‘lgan va A to‘plamdan S_i ga $\rho(x, y)$ metrika bo‘yicha eng yaqin k ta qo‘shnini o‘z ichiga oladigan gipersharning radiusiga teng. O‘sish tartibidagi $R_{i_1}, R_{i_2} \dots, R_{i_m}$ ketma-ketlikdan S_{i_1} obyektni va unga eng yaqin k ta qo‘shnisi o‘chirilsin. O‘chirilgan obyektlar G_1 ga qo‘shiladi.

Keyingi guruhning xos vakili sifatida tartiblangan tanlanmadagi qolgan obyektlarining birinchisi e’lon qilinadi, unga o’xshash obyektlar G_2 guruhga kiritiladi, ular ichida oldingi qadamda o’chirilmaganlari tanlanmadan o’chiriladi. Jarayon tanlanmaning bir qismi yoki to’laligicha klassifikatsiya qilinmaguncha davom etadi.

3.8. Множество из m объектов в n-мерном признаковом пространстве представлено в виде $A = \{S_1, \dots, S_m\}$. С помощью заданного параметра k произвести вычисление значений $R_1, \dots, R_m$. Значение R_i определяет радиус гипершара по метрике $\rho(x, y)$ с центром в S_i в окрестности которого содержится k ближайших к S_i объектов множества A. Из упорядоченной по возрастанию последовательности $R_{i_1}, R_{i_2} \dots, R_{i_m}$ удалить объект S_{i_1} и k ближайших к нему объектов. Удаленные объекты включаются в группу G_1 .

Типичным представителем следующей группы объявляется первый из оставшихся объектов упорядоченной выборки, а кортеж похожих на него объектов включается в группу G_2 , те же из них, которые не были удалены на предыдущем шаге, удаляются из выборки. Процесс продолжается до тех пор, пока часть или вся выборка не будут расклассифицированы.

3.9. Satrlari bo‘yicha n miqdoriy alomatlar bilan tavsiflangan m obyektlar joylashgan $A = \{a_{ij}\}_{m \times n}$ jadval berilgan. Iyerarxik guruhlash algoritmi quyidagi sxema bo‘yicha amalga oshirilsin:

1-qadam. Uchburchak jadval shaklida $d(i, j)$ masofalar hisoblansin;

2-qadam. $d(p,q) = \inf(d(i,j)) \forall i,j$ sharti asosida:

– jadvaldan barcha q kattaliklar o‘chirilsin;

– p belgilashlar r bilan almashtirilsin;

– jadvalning qolgan qismida $d(i,r) \forall i$ hisoblansin;

3-qadam. Agar qisqartirilgan jadvaldagi satr va ustunlar sonlari $k, 2 \le k < m$ bo‘lsa, 2-qadamga o‘tilsin, aks holda tamom.

Bu erda $d(i,r) = a_r d(i,p) + a_q\, d(i,q) + bd(p,q) + c|d(i,p) - d(i,q)|^p$ umumiy formula bo‘yicha hisoblanadi.

Iyerarxiyanig quyidagi variantlarini ko‘rib chiqaylik:

1-variant: $a_p = a_q = \frac{1}{2}; b = 0; c = -\frac{1}{2},\ d(i,r) = inf[d(i,p), d(i,q)]$.

2-variant: $a_p = a_q = \frac{1}{2}; b = 0; c = \frac{1}{2},\ d(i,r) = sup[d(i,p), d(i,q)]$.

3-variant: $a_p = \frac{K_p}{K_p + K_q}\ a_q = \frac{K_q}{K_p + K_q}; b = c = 0,$ bu erda K_p va $K_q - p$ va q guruhdagi obyektlar soni.

3.9. В заданной таблице $A = \{a_{ij}\}_{m \times n}$ по строкам содержатся m объектов, описываемых n количественными признаками.

Реализовать алгоритм иерархической группировки по следующей схеме:

1-шаг: вычислить расстояния $d(i,j)$ в форме треугольной таблицы;

2-шаг: пусть $d(p,q) = inf(d(i,j))\ \forall\ i,j$;

удалить из таблицы все величины q;

заменить p на r;

вычислить $d(i,r)\ \forall\ i,$ оставщихся в таблице.

3-шаг: Если число столбцов и строк в получившейся сокращённой таблице больше $k, 2 \le k < m$ то идти к 2, иначе конец.

Расстояния $d(i,r)$ вычисляются по общей формуле

$d(i,r) = a_r d(i,p) + a_q\, d(i,q) + bd(p,q) + c|d(i,p) - d(i,q)|^p$.

Рассмотреть следующие варианты иерархий.

1-вариант: $a_p = a_q = \frac{1}{2}; b = 0; c = -\frac{1}{2}, d(i,r) = inf[d(i,p), d(i,q)]$.

2-вариант: $a_p = a_q = \frac{1}{2}; b = 0; c = \frac{1}{2}, d(i,r) = sup[d(i,p), d(i,q)]$.

3-вариант: $a_p = \frac{K_p}{K_p+K_q}$ $a_q = \frac{K_q}{K_p+K_q}$; $b = c = 0$, где K_p и K_q число объектов в p-ой и -ой группе.

3.10. Har bir obyekti n ta miqdoriy alomat bilan tavsiflangan A={ S_1,...,S_m}, tanlanma berilgan. Berilgan $\rho(x,y)$ metrika bilan S_1,...,S_m. obyektlar o‘rtasidagi eng qisqa yopiqmas yo‘l qurilsin. Berilgan $K, 1 \leq K < m$ bo‘yicha yo‘lning K ta eng uzun qirralarini o‘chirish orqali tanlanma $K+1$ guruhlarga bo‘linsin. Metrika sifatida Evklid metrikasi ishlatilsin.

3.10 По заданной выборке A={S_1,...,S_m}, каждый объект которой описывается с помощью n количественных признаков, и заданной метрике $\rho(x,y)$ построить кратчайший незамкнутый путь (КНП) между объектами S_1,...,S_m. С помощью заданного $K, 1 \leq K < m$ получить разбиение на $K+1$ группу путём удаления K самых длинных рёбер. В качестве $\rho(x,y)$ использовать метрику Евклида.

3.11. Har bir obyekti n ta miqdoriy alomat bilan tavsiflangan A={ S_1,...,S_m}, tanlanma berilgan bo’lib, u ikkita o’zaro kesishmaydigan K_1, K_2 sinflarga bo’lingan. Berilgan $\rho(x,y)$ metrika bilan S_1,...,S_m. obyektlar o‘rtasidagi eng qisqa yopiqmas yo‘l qurilsin. Quyidagi masala echilsin: agar turli sinf obyektlarini tutashtiruvchi yo’l (eng qisqa yopiqmas yo‘l qirralari) mavjud bo’lsa, undagi eng uzun yoy uchirilsin. Metrika sifatida Hemming metrikasi ishlatilsin.

3.11. По заданной выборке A={S_1,...,S_m}, каждый объект которой описывается с помощью n количественных признаков, и заданной метрике $\rho(x,y)$ построить кратчайший незамкнутый путь (КНП) между объектами S_1,...,S_m.

На множестве $\{S_1,...,S_l\} \in A$, $l < m$ определено разбиение на два класса K_1, K_2. Если существует путь, соединяющий объекты (вершины КНП) из разных классов, то удалить самое длинное ребро в нём. В качестве $\rho(x,y)$ использовать метрику Хемминга.

3.12. Berilgan $E_0 = \{S_1,\ldots,S_m\}$ tanlanma obyektlarini approksimatsiya qilish uchun k kodli W_i vektorlar bilan vektorli kvantlash masalasi echilsin. Kodli vektorlar uchun qandaydir, tugunlarning yaqinlik o‘lchamining simmetrik matritsasi berilgan bo‘lsin: har bir $(i,j), i,j = 1,\cdots,k$ juftlik uchun $\eta_{ij}(0 \leq \eta_{ij} \leq 1)$ soni aniqlangan hamda jadvalning barcha diagonal elementlari birga teng $\left(\eta_{ij} = 1\right)$. Har bir $S \in E_0$ obyekt n ta miqdoriy alomatlar bilan tavsiflangan.

Kiruvchi s signallar vektorlari alohida ishlanadi, har biri uchun eng yaqin kod vektori $W_{j(S)}$ topiladi ("*g'olib barchasiga ega*" tamoyili). Shundan keyin $\eta_{ij} \neq 0$ barcha W_i kod vektorlari

$$W_i^{new} = W_i^{old}\left(1 - \eta_{j(s)}\theta\right) + s\eta_{j(s)}\theta$$

formulasi bilan qayta hisoblanadi. Bu erda $\theta \in (0..1)$ – o'rganish qadami. G'olib kod vektorining qo'shnilari (aprior ravishda yaqinlik jadvali orqali berilgan) o'lchov birligiga proporsional ravishda ushbu vektor siljigan tomonga suriladi. Masofa funksiyasi sifatida Evklid metrikasi ishlatilsin.

3.12. Решить задачу векторного квантования для аппроксиматsiи объектов выборки данных $E_0 = \{S_1, \dots, S_m\}$ k кодовыми векторами W_i. Для кодовых векторов задана некоторая симметричная таблица мер близости узлов: для каждой пары $(i,j), i,j = 1,\cdots,k$ определяется число $\eta_{ij}(0 \leq \eta_{ij} \leq 1)$ при этом диагональные элементы таблицы равны единице $\left(\eta_{ij} = 1\right)$. Каждый объект $S \in E_0$ описывается n количественными признаками.

Векторы входных сигналов s обрабатываются по одному, для каждого из них находится ближайший кодовый вектор («победитель», который «забирает всё») $W_{j(s)}$. После этого все кодовые векторы W_i, для которых $\eta_{ij} \neq 0$, пересчитываются по формуле

$$W_i^{new} = W_i^{old}\left(1 - \eta_{j(s)}\theta\right) + s\eta_{j(s)}\theta,$$

где $\theta \in (0,1)$ – шаг обучения. Соседи кодового вектора – победителя (по априорно заданной таблице близости) сдвигаются в ту же сторону, что и этот вектор, пропорционально мере близости. В качестве функции расстояния использовать евклидову метрику.

3.13. O'zaro kesishmaydigan l ta $K_1, \dots, K_l$ sinflar vakillarini o'z ichiga olgan $E_0 = \{S_1, \dots, S_m\}$ o'rgatuvchi tanlanma berilgan. Har bir $S \in E_0$ obyekt n ta $X(n) = (x_1, \dots x_n)$ miqdoriy alomatlar bilan tavsiflangan. Berilgan $\rho(x,y)$ metrika bo'yicha tanlanmaning shovqin obyektlari sinflarning chegaraviy obyektlari quyidagicha aniqlanuvchi $B \subset E_0$ to'plamning to'plam ostisi sifatida qaraladi:

$$B = \left\{ S \in E_0 \middle| \rho(S_i, S) = \min_{S_i \in K_j, S_d \in CK_j} \rho(S_i, S_d) \right\}$$

Tanlanmaning $S \in B \cap K_j$, $j=1,\dots,l$ obyekti K_j sinfining D_j shovqin obyektlari to'plamiga tegishli hisoblanadi, agar

$$\left|\left\{S_i \in E_0 \middle| \rho(S_i, S) = \min_{S_i \in CK_j, S_d \in K_j} \rho(S_i, S_d)\right\}\right| > \left|\left\{S_i \in K_j \middle| \rho(S_i, S) < \min_{S_i \in K_j, S_d \in CK_j} \rho(S_i, S_d)\right\}\right|.$$

sharti o'rinli bo'lsa.

Har bir K_i sinf uchun Evklid, Xemming va Chebishev metrikalari bo'yicha shovqin obyektlar to'plami D_i aniqlansin.

3.13. Обучающая выборка $E_0 = \{S_1, \dots, S_m\}$ содержит представителей l непересекающихся классов $K_1, \dots, K_l$. Каждый объект $S \in E_0$ описывается набором из n количественных признаков $X(n) = (x_1, \dots x_n)$. Множество шумовых объектов рассматривается как подмножество граничных объектов классов $B \subset E_0$ по заданной метрике $\rho(x, y)$, которое определяется как

$$B = \left\{S \in E_0 \middle| \rho(S_i, S) = \min_{S_i \in K_j, S_d \in CK_j} \rho(S_i, S_d)\right\}.$$

Объект $S \in B \cap K_j$, $j = 1, \dots, l$ принадлежит множеству шумовых объектов D_j класса K_j, если

$$\left|\left\{S_i \in E_0 \middle| \rho(S_i, S) = \min_{S_i \in CK_j, S_d \in K_j} \rho(S_i, S_d)\right\}\right| > \left|\left\{S_i \in K_j \middle| \rho(S_i, S) < \min_{S_i \in K_j, S_d \in CK_j} \rho(S_i, S_d)\right\}\right|.$$

Требуется определить все множества шумовых объектов D_i для каждого класса K_i по метрикам Евклида, Хэмминга, Чебышева.

3.14. O'zaro kesishmaydigan l ta $K_1, \dots, K_l$ sinflar vakillarini o'z ichiga olgan $E_0 = \{S_1, \dots, S_m\}$ o'rgatuvchi tanlanma berilgan. Har bir $S \in E_0$ obyekt n ta $X(n) = (x_1, \dots x_n)$ turli toyifadagi (miqdoriy va nominal) alomatlar bilan tavsiflangan. Berilgan $\rho(x, y)$ metrika bo'yicha tanlanmaning shovqin obyektlari sinflarning chegaraviy obyektlari quyidagicha aniqlanuvchi $B \subset E_0$ to'plamning to'plam ostisi sifatida qaraladi:

$$B = \left\{S \in E_0 \middle| \rho(S_i, S) = \min_{S_i \in K_j, S_d \in CK_j} \rho(S_i, S_d)\right\}$$

Tanlanmaning $S \in B \cap K_j$, $j = 1, \dots, l$ obyekti K_j sinfining D_j shovqin obyektlari to'plamiga tegishli hisoblanadi, agar

$$\left|\left\{S_i \in E_0 \middle| \rho(S_i, S) = \min_{S_i \in CK_j, S_d \in K_j} \rho(S_i, S_d)\right\}\right| > \left|\left\{S_i \in K_j \middle| \rho(S_i, S) < \min_{S_i \in K_j, S_d \in CK_j} \rho(S_i, S_d)\right\}\right|.$$

sharti o'rinli bo'lsa.

Har bir K_i sinf uchun Juravlev metrikalari bo'yicha shovqin obyektlar to'plami D_i aniqlansin.

3.14. Обучающая выборка $E_0 = \{S_1, \dots, S_m\}$ содержит представителей l непересекающихся классов $K_1, \dots, K_l$. Каждый объект $S \in E_0$ описывается набором из n разнотипных (количественных и номинальных) признаков $X(n) = (x_1, \dots x_n)$. Множество шумовых объектов рассматривается как подмножество граничных объектов классов $B \subset E_0$ по заданной метрике $\rho(x, y)$, которое определяется как

$$B = \left\{ S \in E_0 \middle| \rho(S_i, S) = \min_{S_i \in K_j, S_d \in CK_j} \rho(S_i, S_d) \right\}.$$

Объект $S \in B \cap K_j$, $j=1,\dots,l$ принадлежит множеству шумовых объектов D_j класса K_j, если

$$\left| \left\{ S_i \in E_0 \middle| \rho(S_i, S) = \min_{S_i \in CK_j, S_d \in K_j} \rho(S_i, S_d) \right\} \right| > \left| \left\{ S_i \in K_j \middle| \rho(S_i, S) < \min_{S_i \in K_j, S_d \in CK_j} \rho(S_i, S_d) \right\} \right|.$$

Требуется определить все множества шумовых объектов D_i для каждого класса K_i по метрике Журавлёва.

3.15. (*Ob'ektlar bog'langanligi bo'yicha klaster tahlil*).

Berilgan $E_0=(S_1,\dots,S_m)$ tanlanma ob'ektlari $X(n)=(x_1,\dots,x_n)$, $n>2$ – miqdoriy alomatlar bilan tavsiflangan. Tanlanmaning S ob'ekti asosiy deyiladi, agar hech bo'lmaganda k ta ob'ekt undan ε masofada joylashgan bo'lsa (ε – S ob'ektga qo'shnichilik radiusining maksimali), o'zini ham hisobga olgan holda. Ular S ob'ektdan erishish mumkin ob'ektlar hisoblanadi. S^d ob'ekt S ob'ektdan erishish mumkin deyiladi, agar $S_1,\dots,S_r$ yo'l mavjud bo'lib, $S_1=S$ va $S_r=S^d$. Bu erda har bir S_{i+1} ob'ektga S_i ob'ektdan to'g'ridan-to'g'ri erishiladi (yo'ldagi barcha ob'ektlar asosiy hisoblanadi, S^d ob'ektdan tashqari). Agar S ob'ekt asosiy bo'lsa, u undan erishish mumkin bo'lgan barcha ob'ektlar (asosiy va asosiy bo'lmagan) bilan klaster tashkil qiladi.

Talab qilinadi:

1. Berilgan k va ε bo'yicha barcha klasterlar aniqlansin;
2. Klaster soni va ularning tarkibi Evklid va Chebishev metrikalari bo'yicha taqqoslansin.

3.15. (*Кластерный анализ по отношению связанности объектов*). Дано описание множества объектов $E_0=(S_1,\dots,S_m)$ с помощью набора количественных признаков $X(n)=(x_1,\dots,x_n)$. Объект S является основным, если по крайней мере k точек находятся на расстоянии ε (ε является максимальным радиусом соседства от S) включая сам объект. Считается, что эти объекты достижимы из S.

Объект S^d достижим из S, если имеется путь $S_1,\ldots,S_r$ с $S_1=S$ и $S_r=S^d$, где каждый объект S_{i+1} прямо достижим из S_i (все объекты на пути являются основными за исключением S^d). Если объект S является основным, то он формирует кластер со всеми объектами (основными и не основными) достижимыми из этого объекта.

Требуется:

1. по заданным k и ε определить все кластеры;
2. сравнить количество и состав кластеров по метрикам Евклида и Чебышева.

3.16 (*Devis-Bolduin indexsi*). Berilgan $E_0=(S_1,\ldots,S_m)$ – obyektlar to'plami *k means* algoritmi bilan o'zaro kesishmaydigan G_1 ,…, G_k, $k\geq 2$ guruhlarga bo'lingan. Har bir $S\in E_0$ obyekt n ta miqdoriy alomatlar $X(n)$ bilan tavsiflangan. G_i guruh obyektlari va uning M_i markazi o'rtasidag masofa

$$R_i = \sum_{S_r\in G_i} \rho\left(S_r, M_i\right)$$

ko'rinishida, guruhlar markazlari o'rtasida masofa esa

$$D = \sum_{i=1}^{k-1} \sum_{j=i+1}^{k} \rho\left(M_i, M_j\right)$$

formula orqali hisoblanadi.

Klasterlash sifatini baholash talab qilinadi:

$$BAHO = \sum_{i=1}^{k} R_i \,/\, D.$$

3.16 (*Индекс Девиса-Болдуина*)

Множество объектов $E_0=(S_1,\ldots,S_m)$ алгоритмом *k means* разделено на непересекающиеся группы G_1 ,…, G_k, $k\geq 2$. Каждый объект $S\in E_0$ описывается набором из n количественных признаков $X(n)$. Расстояние между объектами группы G_i и её центром M_i вычисляется как

$$R_i = \sum_{S_r\in G_i} \rho\left(S_r, M_i\right)$$

и расстояние между центрами групп

$$D = \sum_{i=1}^{k-1} \sum_{j=i+1}^{k} \rho\left(M_i, M_j\right).$$

Требуется оценку качества кластеризации как

$$BAHO = \sum_{i=1}^{k} R_i \,/\, D.$$

4. Informativ alomatlarni saralash va tanlash (Отбор и выбор информативных признаков)

4.1. O‘zaro kesishmaydigan l ta $K_1, \dots, K_l$ sinflarga bo‘lingan $E_0 = \{S_1, \dots, S_m\}$ o‘rgatuvchi tanlanmada yaqinlik matritsasini nominal alomatlar gradatsiyasi bo‘yicha hisoblash algoritmi dasturi tuzilsin. Har bir $S \in E_0$ – obyekt n ta nominal alomat bilan tavsiflanadi. Berilgan $\{S_a = (x_{a1}, \dots, x_{an}),\ \ S_b = (x_{b1}, \dots, x_{bn})\}$ obyektlar juftligi to‘plamida quyidagi funksiyalar aniqlanadi:

$$g(a,b,i,j) = \begin{cases} 2, \ x_{ai} \neq x_{bi} \quad \text{va} \quad x_{bj} \neq x_{aj}, \\ 1, \ x_{ai} = x_{bi} \quad \text{yoki} \quad x_{aj} = x_{bj}, \\ 0, \ x_{ai} = x_{bi} \quad \text{va} \quad x_{aj} = x_{bj}; \end{cases}$$

$$\alpha(a,b) = \begin{cases} 0, \quad S_a, S_b \in K_i, i = \overline{1,l} \\ 1, \quad S_a \in K_i, S_b \in K_j, i \neq j. \end{cases}$$

Berilgan E_0 tanlanmada x_i, x_j nominal alomatlar juftligi orasidagi yaqinlik o‘lchovi

$$b_{ij} = \begin{cases} \dfrac{\sum_{a=1}^{m} \sum_{b=1}^{m} \alpha(a,b) g(a,b,i,j)}{2 \sum_{p=1}^{l} |K_p| \left(m - |K_p|\right)}, i \neq j \\ 0, \qquad\qquad i = j. \end{cases}$$

formula orqali ifodalanadi.

Qurilgan $\{b_{ij}\}_{n \cdot n}$ matritsa bo’yicha (x_i, x_j) alomatlar juftliklarining o’smaydigan holda tartiblangan ketma-ketligi shakllantirilsin.

4.1. По заданной обучающей выборке $E_0 = \{S_1, \dots, S_m\}$, содержащей представителей l непересекающихся классов объектов $K_1, \dots, K_l$, произвести вычисление матрицы близости по градациям номинальных признаков. Считается, что каждый объект $S_i \in E_0$ описывается n номинальными признаками. На множестве пар объектов $(S_a = (x_{a1}, \dots, x_{an}),\ S_b = (x_{b1}, \dots, x_{bn}))$ определяются функции:

$$g(a,b,i,j) = \begin{cases} 2, \quad x_{ai} \neq x_{bi} \quad \text{и} \quad x_{bj} \neq x_{aj} \\ 1, \quad x_{ai} = x_{bi} \quad \text{либо} \quad x_{aj} = x_{bj} \\ 0, \quad x_{ai} = x_{bi} \quad \text{и} \quad x_{aj} = x_{bj}; \end{cases}$$

$$\alpha(a,b)=\begin{cases}0, & S_a, S_b \in K_i, i=\overline{1,l}\\ 1, & S_a \in K_i, S_b \in K_j, i \neq j.\end{cases}$$

Мера близости между парой номинальных признаков x_i, x_j на E_0 вычисляется как

$$b_{ij}=\begin{cases}\dfrac{\sum_{a=1}^{m}\sum_{b=1}^{m}\alpha(a,b)g(a,b,i,j)}{2\sum_{p=1}^{l}|K_p|(m-|K_p|)}, i \neq j\\ 0, i=j.\end{cases}$$

По матрице $\{b_{ij}\}_{n \cdot n}$ сформировать последовательность из упорядоченных по не возрастанию непересекающихся пар признаков (x_i, x_j).

4.2. O‘zaro kesishmaydigan l ta $K_1, \dots, K_l$ sinflarga bo‘lingan $E_0=\{S_1, \dots, S_m\}$ o‘rgatuvchi tanlanmada sinf ichidagi o‘xshashlikni nominal alomatlar gradatsiyasi bo‘yicha hisoblash algoritmni amalga oshiruvchi dastur tuzilsin. Har bir $S \in E_0$ – obyekt n ta nominal alomat bilan ifodalanadi. C-chi alomatning sinf ichidagi o‘xshashligi F_c ni hisoblash formulasi:

$$F_c=\frac{\gamma_c}{\gamma_{\max}},$$

bu erda, $\gamma_{\max}=\sum_{i=1}^{l}|K_i|(|K_i|-1)$, $\gamma_c=\sum_{i=1}^{l}\sum_{t=1}^{p}g_{ic}^{t}(g_{ic}^{t}-1)$, p– c-alomatning gradatsiyalar soni, g_{ic}^{t} – K_i sinf obyektining c-chi alomatining t gradatsiyasi soni$(1 \leq t \leq p)$.

4.2. По заданной обучающей выборке $E_0=\{S_1, \dots, S_m\}$, содержащей представителей l непересекающихся классов объектов $K_1, \dots, K_l$, произвести вычисление значений внутриклассового сходства по градациям номинальных признаков. Считается, что каждый объект $S_i \in E_0$ описывается n номинальными признаками.

Для вычисления значения внутриклассового сходства F_c по c-му признаку использовать формулу

$$F_c=\frac{\gamma_c}{\gamma_{max}},$$

где $\gamma_{max} = \sum_{i=1}^{l}|K_i|(|K_i| - 1)$, $\gamma_c = \sum_{i=1}^{l}\sum_{t=1}^{p} g_{ic}^t\,(g_{ic}^t\, - 1)$, p – число градаций c-го признака, g_{ic}^t – количество значений t –ой $(1 \le t \le p)$ градации c-го признака в описании объектов класса K_i.

4.3. O'zaro kesishmaydigan l ta $K_1, \dots, K_l$ sinflar obyektlarini o'z ichiga olgan $E_0 = \{S_1, \dots, S_m\}$ tanlanmada nominal alomatlar gradatsiyalari bo'yicha sinflararo farqlanish qiymatlari hisoblansin. Har bir $S_i \in E_0$ obyekt nominal alomatlar bilan tavsiflangan. Nominal c alomat bo'yicha sinflararo farqlanish qiymati R_c

$$R_c = \frac{\beta_c}{\beta_{max}},$$

formula bilan hisoblanadi. Bu erda

$$\beta_c = \sum_{i=1}^{l}\sum_{t=1}^{p}\begin{cases} g_{ic}^t\,(|CK_i| - b_{ic}^t\,), g_{ic}^t\, \ne 0, \\ b_{ic}^t\,|K_i|, \qquad\qquad g_{ic}^t\, = 0, \end{cases}$$

p – c alomatning gradatsiyalari soni, $g_{ic}^t\, - K_i$ sinf obyektlari tavsifidagi c alomatninf t- gradatsiyasi ($1 \le t \le p$) qiymatlarining soni, $b_{ic}^t\, - CK_i$ sinf to'ldiruvchisi obyektlari tavsifidagi c alomatninf t- gradatsiyasi ($1 \le t \le p$) qiymatlarining soni, $\beta_{max} = \sum_{i=1}^{l}|K_i|(m - |K_i|)$.

4.3. По заданной обучающей выборке $E_0 = \{S_1, \dots, S_m\}$, содержащей представителей l непересекающихся классов объектов $K_1, \dots, K_l$, произвести вычисление значений межклассового различия по градациям номинальных признаков. Считается, что каждый объект $S_i \in E_0$ описывается n номинальными признаками. Для вычисления значения межклассового различия R_c по c-му признаку использовать формулу

$$R_c = \frac{\beta_c}{\beta_{max}}, \text{ где } \beta_c = \sum_{i=1}^{l}\sum_{t=1}^{p}\begin{cases} g_{ic}^t\,(|CK_i| - b_{ic}^t\,), g_{ic}^t\, \ne 0, \\ b_{ic}^t\,|K_i|, \qquad\qquad g_{ic}^t\, = 0 \end{cases},$$

где p – число градаций c-го признака, g_{ic}^t – количество значений t-ой ($1 \le t \le p$) градации c-го признака в описании объектов класса K_i, b_{ic}^t – количество значений t-ой градации c-го признака в CK_i – дополнении класса K_i, $\beta_{max} = \sum_{i=1}^{l}|K_i|(m - |K_i|)$.

4.4. Minimal konfiguratsiyali sun'iy neyron to'rlari yordamida klassifikatsiya masalasini echish uchun nominal alomatlar vaznlarini quyidagi formula bo'yicha hisoblovchi dastur tuzing:

$$w_c = \left(\frac{\gamma_c}{\gamma_{max}}\right)\left(\frac{\beta_c}{\beta_{max}}\right).$$

Bu erda $\left(\frac{\gamma_c}{\gamma_{max}}\right)$ – sinf ichidagi o‘xshashlik va $\left(\frac{\beta_c}{\beta_{max}}\right)$ – sinflararo o’xshashliklar (4.2 va 4.3 – masalalarga qaralsin).

4.4. Реализовать вычисление весов номинальных признаков для решения задач классификации с помощью искусственных нейронных сетей с минимальной конфигурацией по формуле

$$w_c = \left(\frac{\gamma_c}{\gamma_{max}}\right)\left(\frac{\beta_c}{\beta_{max}}\right)$$

Для решения использовать постановку задачи и формулы для вычисления внутриклассового сходства $\left(\frac{\gamma_c}{\gamma_{max}}\right)$ и $\left(\frac{\beta_c}{\beta_{max}}\right)$ межклассового различия соответственно из задач 4.2 и 4.3.

4.5. O‘zaro kesishmaydigan l ta $K_1, \dots, K_l$ sinflarga bo‘lingan $E_0 = \{S_1, \dots, S_m\}$ o‘rgatuvchi tanlanmada uchun nominal alomatlar hissalarini hisoblash algoritmining dasturi tuzilsin. Har bir $S \in E_0$ – obyekt n ta nominal alomat bilan tavsiflanadi va p $(1 \le p \le n)$ – alomatning sinflarni ajratishdagi hissasi

$$\lambda_p = \frac{\sum_{i=1}^{l}\sum_{j=1}^{u_p} z_{pj}^i \left(z_{pj}^i - 1\right)}{\sum_{i=1}^{l}|K_i|\left(|K_i| - 1\right)} - \frac{\sum_{i=1}^{l}\sum_{j=1}^{u_p} z_{pj}^i \overline{z_{pj}^i}}{\sum_{i=1}^{l}|K_i||CK_i|}$$

formula bilan hisoblanadi. Bu erda $z_{pj}^i, \overline{z_{pj}^i}$ – mos ravishda K_i va uning to‘ldiruvchisi $CK_i = E_0 \setminus K_i$ sinflarda p-chi alomatning j – gradatsiyasining soni, u_p – p alomatning gradatsiyalar soni.

4.5. Произвести вычисление вкладов номинальных признаков по заданной обучающей выборке $E_0 = \{S_1, \dots, S_m\}$, содержащей представителей l непересекающихся классов объектов $K_1, \dots, K_l$, в процесс принятия решения при классификatsiи объектов. Считается, что каждый объект $S_i \in E_0$ описывается n номинальными признаками и вклад p $(1 \le p \le n)$ признака в разделение классов вычисляется по формуле

$$\lambda_p = \frac{\sum_{i=1}^{l}\sum_{j=1}^{u_p} z_{pj}^i \left(z_{pj}^i - 1\right)}{\sum_{i=1}^{l}|K_i|\left(|K_i| - 1\right)} - \frac{\sum_{i=1}^{l}\sum_{j=1}^{u_p} z_{pj}^i \overline{z_{pj}^i}}{\sum_{i=1}^{l}|K_i||CK_i|},$$

где $z_{pj}^{i}, \overline{z_{pj}^{i}}$ - количество значений j –ой градаций p-го признака, соответственно, класса K_i и его дополнения $CK_i = E_0 \setminus K_i$, u_p –число градаций p-го признака.

4.6. O‘zaro kesishmaydigan K_1, K_2 sinflarga bo‘lingan $E_0 = \{S_1, \ldots, S_m\}$ o‘rgatuvchi tanlanma berilgan. Har bir $S \in E_0$ – obyekt n ta nominal alomat bilan tavsiflangan. $\{x_i, x_j\}_{i,j\in\{1,\ldots n\}}$ juftliklar to’plami bo’yicha sinflar obyektalari o’rtasidagi maksimal farqlanishni aniqlash talab qilinadi.

Agar $x_i, x_j \in X(n), i \neq j$ alomatlar mos ravishda p_i va p_j gradatsiyalarga ega bo’lsa, ularni bitta $y_c = x_i \otimes x_j$ latent alomatga birlashtirishda gradatsiyalar $\mu=p_ip_j$ soni bilan cheklanadi. Latent y_c alomat bo’yicha sinflar farqlanishi

$$\beta_c = 1 - \frac{\sum_{t=1}^{2}\sum_{d=1}^{\mu} g_t^d g_{3-t}^d}{2|K_1||K_2|},$$

ko’rinishida hisoblanadi. Bu erda $g_t^d\left(g_{3-t}^d\right)$ – $K_t\left(K_{3-t}\right)$ sinfdagi d gradatsiyali obyektlar soni.

Izoh. Alomatlarning barcha gradatsiyalari 1,2,… sonlari bilan identifikatsiya qilinadi.

4.6. Дана выборка объектов $E_0 = \{S_1, \ldots, S_m\}$, разделенная на два непересекающихся класса K_1, K_2. Каждый объект выборки описывается набором $X(n) = (x_1, \ldots, x_n)$ номинальных признаков. Требуется определить максимальное различие между объектами классов по множеству пар $\{x_i, x_j\}_{i,j\in\{1,\ldots n\}}$.

Если признаки $x_i, x_j \in X(n), i \neq j$ имеют соответственно p_i и p_j градаций, то при их объединении в один латентный признак $y_c = x_i \otimes x_j$ количество градаций ограничено числом $\mu=p_ip_j$. Различие между классами по y_c вычисляется как

$$\beta_c = 1 - \frac{\sum_{t=1}^{2}\sum_{d=1}^{\mu} g_t^d g_{3-t}^d}{2|K_1||K_2|},$$

где $g_t^d\left(g_{3-t}^d\right)$ – количество объектов с градацией d в классе $K_t\left(K_{3-t}\right)$.

Примечание. Считается, что все градации признаков идентифицируются числами 1,2,…

4.7. Berilgan $E_0 = \{S_1, \dots, S_m\}$ tanlanmadan foydalangan holda alomatlar fazosi o'lchamini n dan k $(k < n)$ gacha qisqartirish amalga oshirilsin. Tanlanma o'zaro kesishmaydigan K_1, K_2 sinflarga bo'lingan va har bir $S_i \in E_0$ obyekt n ta miqdoriy alomatlar to'plami $H(n) = (x_1, \dots, x_n)$ bilan tavsiflangan. $x_j \in H(r), r \le n$ alomatni o'chirishda

$$\frac{\theta_j}{\gamma_j} = max_{H(r)}$$

mezonidan foydalanilsin. Bu erda

$$\theta_j = \sum_{i=1}^{2} \sum_{S_t, S_l \in K_i} |x_{tj} - x_{lj}|, \gamma_j = \sum_{i=1}^{2} \sum_{S_t \in K_i, S_l \in K_{3-i}} |x_{tj} - x_{lj}|.$$

4.7. Произвести сокращение размерности признакового пространства с n до k $(k < n)$ используя выборку объектов $E_0 = \{S_1, \dots, S_m\}$. Выборка разделена на два непересекающихся класса K_1, K_2 и каждый объект $S_i \in E_0$ описывается n количественными признаками из набора $H(n) = (x_1, \dots, x_n)$. Для удаления признака $x_j \in H(r), r \le n$ использовать критерий

$$\frac{\theta_j}{\gamma_j} = max_{H(r)} ,$$

где $\theta_j = \sum_{i=1}^{2} \sum_{S_t, S_l \in K_i} |x_{tj} - x_{lj}|, \gamma_j = \sum_{i=1}^{2} \sum_{S_t \in K_i, S_l \in K_{3-i}} |x_{tj} - x_{lj}|$.

4.8. O'zaro kesishmaydigan K_1, K_2 sinflarga bo'lingan $E_0 = \{S_1, \dots, S_m\}$ o'rgatuvchi tanlanma berilgan. Har bir $S \in E_0$ – obyekt n ta nominal alomat bilan tavsiflangan. Nominal alomatlar gradatsiyalari bo'yicha yaqinlik matritsasi hisoblansin.

Obyektlar juftliklari ($S_a=(x_{a1},\dots,x_{an})$, $S_b=(x_{b1},\dots,x_{bn})$) to'plamida

$$g(a,b,i,j) = \begin{cases} 2, & x_{ai} \ne x_{bi} \quad va \quad x_{bj} \ne x_{aj}; \\ 1, & x_{ai} = x_{bi} \quad yoki \quad x_{aj} = x_{bj}; \\ 0, & x_{ai} = x_{bi} \quad va \quad x_{aj} = x_{bj}, \end{cases}$$

funksiya aniqlangan bo'lsin.

(x_i,x_j), $i,j \in \{1,\dots,n\}$ nominal alomatlar juftligi bo'yicha $S_u \in K_t$, $t=1,2$ bilan K_{3-t} obyektlari o'rtasidagi yaqinlik (farqlanish) o'lchami

$$b_{ij}(S_u) = \begin{cases} \dfrac{\sum\limits_{S_v \in K_{3-t}} g(u,v,i,j)(l_i(x_{ui}) + l_j(x_{uj}))}{2|K_1||K_2|}, i \ne j, \\ 0, i = j, \end{cases}$$

ko'rinishida beriladi. Bu erda $l_i(x_{ui})$ $(l_j(x_{uj}))$ – E_0 tanlanmadagi gradatsiyasi qiymati x_{ui} (x_{uj}) bo'lgan obyektlar soni.

$B(S_u)=\{b_{ij}(S_u)\}_{n*n}$ matritsa elementlari bo'yicha S_u obyektning XI $=(x_1,\ldots,x_r)$, $r<n$ individual informativ alomatlar to'plamini tanlash uchun alomatlarning

$$x_{u_1}, x_{u_2}, \ldots, x_{u_n}$$

tartiblangan ketma-ketligini quruvchi rekursiv protsedurasinin amalga oshirish talab qilinadi.

$B(S_u)$ matritsadan eng katta b_{ij} qiymatiga ega (x_i,x_j) juftlik ajratib olinadi va ketma-ketlikka (chapdan o'ngga) qo'shiladi. (x_i,x_j) juftlikdagi joylashuv tartibi $l_i(x_{ui}) \geq l_j(x_{uj})$ sharti bo'yicha aniqlanadi. Keyingi har juftlik xuddi shu tamoyilda $B(S_u)$ jadvaldan i va j nomerli satr va ustunlarni o'chirishdan keyin anialanadi.

4.8. По заданной обучающей выборке $E_0=\{S_1,\ldots,S_m\}$, содержащей представителей 2-х непересекающихся классов объектов K_1, K_2, произвести вычисление матрицы близости по градациям номинальных признаков. Считается, что каждый объект $S_i \in E_0$ описывается n номинальными признаками. На множестве пар объектов $(S_a=(x_{a1},\ldots,x_{an}),\ S_b=(x_{b1},\ldots,x_{bn}))$ определяются функции:

$$g(a,b,i,j)=\begin{cases} 2, & x_{ai}\neq x_{bi} \quad \text{и} \quad x_{bj}\neq x_{aj} \\ 1, & x_{ai}=x_{bi} \quad \text{либо} \quad x_{aj}=x_{bj} \\ 0, & x_{ai}=x_{bi} \quad \text{и} \quad x_{aj}=x_{bj}; \end{cases}$$

Мера близости (различия) между $S_u \in K_t$, t=1,2 и объектами из K_{3-t} по паре номинальных признаков (x_i,x_j), $i,j \in \{1,\ldots,n\}$ задаётся как

$$b_{ij}(S_u)=\begin{cases} \dfrac{\sum\limits_{S_v\in K_{3-t}} g(u,v,i,j)(l_i(x_{ui})+l_j(x_{uj}))}{2|K_1||K_2|}, i\neq j, \\ 0, i=j, \end{cases}$$

где $l_i(x_{ui})$ $(l_j(x_{uj}))$ – число объектов из E_0 с градацией равной x_{ui} (x_{uj}).

Для выбора по элементам матрицы $B(S_u)=\{b_{ij}(S_u)\}_{n*n}$ индивидуального информативного набора X® $=(x_1,\ldots,x_r)$, $r<n$ объекта S_u требуется реализовать рекурсивную процедуру построения упорядоченной последовательности признаков

$$x_{u_1}, x_{u_2}, \ldots, x_{u_n}.$$

Из матрицы $B(S_u)$ выделяется пара (x_i,x_j) с наибольшим значением b_{ij} и включается (слева-направо) в последовательность. Порядок следования в (x_i,x_j) выбирается по условию $l_i(x_{ui}) \geq l_j(x_{uj})$. Каждая следующая пара признаков по аналогичному принципу определяется из $B(S_u)$ после удаления в ней строк и столбцов с номерами i и j.

4.9. O‘zaro kesishmaydigan K_1, K_2 sinflarga bo‘lingan $E_0 = \{S_1, \ldots, S_m\}$ o‘rgatuvchi tanlanma berilgan. Har bir $S \in E_0$ obyekt n ta $X(n)=(x_1,\ldots,x_n)$ turli toifadagi (miqdoriy va nominal) alomatlar bilan tavsiflangan. $S_i,S_j \in E_0$ obyektlari o’rtasidagi yaqinlik masofisi o’lchovi $\rho(S_i,S_j)$ sifatida Juravlev metrikasidan qaralsin.

Aytaylik, $S_i \in K_p$ obyektdan $\rho(S_i,S_j)$ masofa bo’yicha kamaymaydigan holda tartiblangan

$$S^1,S^2,\ldots,S^u,\ldots,S^m$$

ketma-ketlik aniqlangan bo’lsin. Yaqinlik matritsasi bo’yicha $(x_a,x_b) \in X(n)$ juftlik bahosi

$$R(x_a, x_b) = \sum_{i=1}^{m} \max_u \left(\frac{d_{ip}(u)/|K_p|}{d_{ip}(u)/|K_p| + d_{i,3-p}(u)/|K_{3-p}|} \right) \left(\frac{d_{ip}(u)}{|K_p|} \right)$$

ko’rinishida hisoblansin va informativ alomatlar izlashda birinchi qadamida $(x_\alpha, x_\beta) = \arg \max_{\{(x_a, x_b)\}} R(x_a, x_b)$ shartidan foydalansin.

4.9. Выборка объектов $E_0=\{S_1,\ldots,S_m\}$ разделено на два непересекающихся класса K_1 и K_2. Каждый объект $S \in E_0$ описываются набором разнотипных (количественных и номинальных) признаков $X(n)=(x_1,\ldots,x_n)$. В качестве меры расстояния $\rho(S_i,S_j)$ между объектами $S_i,S_j \in E_0$ используется метрика Журавлёва. Пусть

$$S^1,S^2,\ldots,S^u,\ldots,S^m$$

последовательность объектов, упорядоченная по не убыванию расстояния $\rho(S_i,S_j)$ от объекта $S_i \in K_p$, $d_{ip}(u)$, $d_{i,3-p}(u)$ – количество объектов класса K_p, K_{3-p} из (1) до S^u включительно. Требуется по мере расстояния вычислить оценку пар $(x_a,x_b) \in X(n)$ как

$$R(x_a, x_b) = \sum_{i=1}^{m} \max_u \left(\frac{d_{ip}(u)/|K_p|}{d_{ip}(u)/|K_p| + d_{i,3-p}(u)/|K_{3-p}|} \right) \left(\frac{d_{ip}(u)}{|K_p|} \right)$$

и определить $(x_\alpha, x_\beta) = \arg \max_{\{(x_a, x_b)\}} R(x_a, x_b)$ в качестве первого шага для поиска информативных наборов признаков.

4.10. $E_0 = \{S_1, \ldots, S_m\}$ o’rgatuvchi tanlanma ikkita o’zaro kesishmaydigan K_1, K_2 sinf vakillarini o’z ichiga oladi. Tanlanmaning har bir obekti binary alomatlarning $X(n)=(x_1,\ldots,x_n)$ to'plami bilan tavsiflanadi.

Quyidagi keltirilgan algoritm bo'yicha obektlarning umumlashgan baholari hisiblansin.

Belgilash kiritaylik, g_{1c}^{j}, g_{2c}^{j} orqali mos ravishda K_1 va K_2 sinflar obektlari tavsifidagi $x_c \in X(n)$ alomat j∈{1,2} gradatsiya qiymatlari soni.

X_c alomat bo'yicha sinflar ora farqlanish

$$\lambda_c = 1 - \frac{\sum_{j=1}^{2} g_{1c}^{j} g_{2c}^{j}}{|K_1||K_2|}.$$

kattalik bilan aniqlanadi.

Alomatning K_1, K_2 sinflar bo'yicha garadatsiya qiymatlariga asoslangan birjinslik darajasi (sinf ichidagi o'xshashlik) βc quyidagi formula bilan hisoblanadi:

$$\beta_c = \frac{\sum_{j=1}^{2} g_{1c}^{j}\left(g_{1c}^{j}-1\right) + g_{2c}^{j}\left(g_{2c}^{j}-1\right)}{|K_1|(|K_1|-1)+|K_2|(|K_2|-1)}.$$

$x_c \in X(n)$ alomatning vazni $w_c=\beta_c\lambda_c$ ko'rinishida hisoblanadi, j∈{1,2} gradatsiya hissasi esa

$$\eta_c(j) = \omega_c\left(\frac{\alpha_{cj}^{1}}{|K_1|} - \frac{\alpha_{cj}^{2}}{|K_2|}\right),$$

bilan aniqlanadi. Bu erda $\alpha_{cj}^1, \alpha_{cj}^2$ – mos ravishda K_1 и K_2 sinflar obektlari tavsifidagi j∈{1,2} gradatsiya qiymatlari soni.

Mumkin bo'lgan $S_r \in E_0$, $S_r=\{x_{ri}\}$ obyektning $X(d) \subset X(n)$ tavsifi bo'yicha umumlashgan bahosi

$$R(S_r) = \sum_{x_i \in X(d)} \mu_i(x_{ri})$$

ko'rinishida hisoblanadi.

Algoritmning qadamba-qadam amalga oshirilshi quyidagicha:

1-qadam. P={1,…,n}.

2-qadam. Hisoblansin crit=10; $u = \arg\max_{j \in P} \omega_j$ P=P/{u}; t ∈{1,…,m} bo'yicha takrorlansin R(St) = ηu(xtu). Takrorlash tugashi; cr1=10;

3-qadam. U ∈P Takrorlansin. T ∈{1,…,m} bo'yicha takrorlansin bt= R(St) + ηu(xtu). Takrorlash tugashi; $M_1 = \sum_{S_t \in K_1} b_t$, $M_2 = \sum_{S_t \in K_2} b_t$. M1= M1 /|K1|, M2= M2 /|K2|. Θ=0. Г=0.

T∈{1,…,m} bo'yicha takrorlansin: Agar St ∈ K1 u holda θ= θ +|bt – M1|, γ= γ + |bt – M2|. Aks holda θ= θ +|bt – M2|, γ= γ + |bt – M1|. Takrorlash tugashi;

Agar θ/ γ < crit u holda crit= θ/ γ, q=u. Takrorlash tugashi;

Agar cr1> crit u holda cr1=crit. P=P/{q}.

t ∈{1,…,m} bo'yicha takrorlansin: $R(S_t) = R(S_t) + \eta_q(x_{tq})$. Takrorlash tugashi. 3-qadamga o'tilsin. Aks holda $\{R(S_t)\}$ chop etilsin;

4-qadam. Agar |P|>1 u holda 2-qadamga o'tilsin;

5-qadam. Tamom.

4.10. В обучающей выборке $E_0 = \{S_1, \ldots, S_m\}$ содержаться представители двух непересекающихся классов K_1, K_2. Каждый объект выборки описывается набором $X(n)$=(x1,…,xn) из n бинарных признаков. Требуется вычислить обобщённые оценки объектов по приведённому ниже алгоритму.

Обозначим через g_{1c}^j, g_{2c}^j – количество значений градации j∈{1,2} признака $x_c \in X(n)$ в описании объектов соответственно класса K_1 и K_2. Межклассовое различие по признаку xc определяется как величина

$$\lambda_c = 1 - \frac{\sum_{j=1}^{2} g_{1c}^j g_{2c}^j}{|K_1||K_2|}.$$

Степень однородности (мера внутриклассового сходства) βc значений градаций признака по классам K_1, K_2 вычисляется по формуле:

$$\beta_c = \frac{\sum_{j=1}^{2} g_{1c}^j\left(g_{1c}^j - 1\right) + g_{2c}^j\left(g_{2c}^j - 1\right)}{|K_1|\left(|K_1| - 1\right) + |K_2|\left(|K_2| - 1\right)}.$$

Вес признаку $x_c \in X(n)$ определяется как $\omega_c = \beta_c \lambda_c$. Вклад градации j∈{1,2} как

$$\eta_c(j) = \omega_c\left(\frac{\alpha_{cj}^1}{|K_1|} - \frac{\alpha_{cj}^2}{|K_2|}\right),$$

где $\alpha_{cj}^{1}, \alpha_{cj}^{2}$ – количество значений градации j признака xc соответственно в классах K_1 и K_2. Обобщённая оценка объекта $Sr \in E_0$, $S_r=\{x_{ri}\}$ по описанию на наборе $X(d) \subset X(n)$ вычисляется как $R(S_r) = \sum_{x_i \in X(d)} \eta_i(x_{ri})$.

Реализация алгоритма по шагам такая:

Шаг 1. P={1,…,n}.

Шаг 2. Вычислить crit=10; $u = \arg \max_{j \in P} \omega_j$. P=P/{u};

Цикл по t ∈{1,…,m} R(St) = ηu(xtu). Конец цикла; cr1=10;

Шаг 3. Цикл по u ∈P. Цикл по t ∈{1,…,m}

bt= R(St) + ηu(xtu). Конец цикла; $M_1 = \sum_{S_t \in K_1} b_t$, $M_2 = \sum_{S_t \in K_2} b_t$.

M1= M1 /|K1|, M2= M2 /|K2|. Θ=0. Γ=0. Цикл по t∈{1,…,m} Если St ∈ K1 то θ= θ +|bt – M1|, γ= γ + |bt – M2|. Иначе θ= θ +|bt – M2|, γ= γ + |bt – M1|. Конец цикла;

Если θ/ γ < crit то crit= θ/ γ, q=u. Конец цикла;

Если cr1> crit то cr1=crit. P=P/{q}.

Цикл по t ∈{1,…,m} R(St)= R(St) + ηq(xtq). Конец цикла. Идти 3.

Иначе вывод {R(St)}

Шаг 4. Если |P|>1 то идти 2;

Шаг 5. Конец.

5. Neyron to’rlari (Нейронные сети)

5.1. Tekislikdagi tasvir $A=\{a_{ij}\}_{m*m}$, $m \geq 4$ matritsa ko‘rinishida berilgan bo‘lib, matritsaning elementlari $a_{ij} \in \{0,\ldots,5\}$ piksellarning qiymatlaridan iborat. Tasvirni zichlashtirish uchun 2×2 o‘lchamli matritsa ko‘rinishidagi tugub olish yadrosidan foydalaniladi. Yadro tasvir bo‘yicha sirpanadi (*A* matritsasi bo‘yicha ko‘chadi) (1-rasmga qarang). Natijada zichlashgan tasvirni ifodalovchi $B=\{b_{ij}\}_{(m-1)*(m-1)}$ matritsa hosil bo‘ladi. Matritsa elementining qiymati yadro elementlarining *A* matritsa minoriga ko‘paytmalarining yig‘indisi ko‘rinishida hisoblandi. Masalan, $b_{11} = 1 \cdot 0 + 4 \cdot 2 + 3 \cdot 1 + 0 \cdot 1 = 11$ (2-rasmga qarang).

1^0	4^2	4	3
3^1	0^1	0	2

1	1	3	5
0	2	4	2

11	×	×
×	×	×
×	×	×

2-rasm. Tasvirni 2×2 o‘lchamdagi tugub olish yadrosi bo‘yicha zichlashtirish
(Рис.2. Сжатие изображения по ядру свёртки размера 2×2)

Berilgan *A* tavsirni $C=\begin{pmatrix}0&2\\1&1\end{pmatrix}$ va $D=\begin{pmatrix}1&0\\2&2\end{pmatrix}$ yadrolardan foydalangan holda $m*m$ dan $(m\text{-}2)*(m\text{-}2)$ o‘lchamgacha zichlashtiring.

5.1. Изображение на плоскости задано в виде матрицы $A=\{a_{ij}\}_{m*m}$, $m\ \geq 4$. Элементами матрицы *A* являются значения пикселей $a_{ij}\in\{0,\ldots,5\}$. Для сжатия изображения используются ядра свёртки в виде матриц размера 2×2. Каждое ядро скользит по изображению (перемещается по матрице *A*) как показано на рис. 2. Значение нового элемента (сжатого изображения) матрицы $B=\{b_{ij}\}_{(m\text{-}1)*\ (m\text{-}1)}$ получается как сумма произведений элементов ядра на элементы минора матрицы *A*. Например (см. рис. 2), b_{11} = 1·0 +4·2 + 3·1 + 0·1=11.

Получить свёртку (сжатие) изображения *A* с размера $m*m$ до $(m\text{-}2)*(m\text{-}2)$ путём последовательного использования ядер $C=\begin{pmatrix}0&2\\1&1\end{pmatrix}$ и $D=\begin{pmatrix}1&0\\2&2\end{pmatrix}$.

5.2. Tekislikdagi tasvir $A=\{a_{ij}\}_{m*m}$, $m\ \geq 4$ matritsa ko‘rinishida berilgan bo‘lib, matritsaning elementlari $a_{ij}\in\{0,\ldots,5\}$ piksellarning qiymatlaridan iborat. *A* tavsirning 4 ta chetlariga nol qiymatli qalbaki piksellar qo‘shilgan (“*nolli to‘ldirma*”). Nolli to‘ldirmalar hisobiga yangi *A** matritsa o‘lchami *A* matritsaga nisbatan ikkita satr va ikkita ustunga oshadi. Tasvirni zichlashtirish uchun 2×2 o‘lchamli matritsa ko‘rinishidagi tugub olish yadrosidan foydalaniladi (2-rasmga qarang). Nolli to‘ldirmalar tugub olishdan keyin chiquvchi matritsada tasvir chegarasini ajratib olish imkonini beradi. Berilgan *A* tavsirni $C=\begin{pmatrix}0&2\\1&1\end{pmatrix}$ yadrodan yordamida ikki marta ketma-ket zichlashtirish orqali *A** tasvirni $(m+2)*(m+2)$ dan $m*m$ o‘lchamgacha zichlashtiring.

5.2. Изображение на плоскости задано в виде входной матрицы $A=\{a_{ij}\}_{m*m}$, $m\ \geq 4$. Элементами матрицы *A* являются значения пикселей a_{ij}∈{0,…,5}. К 4–м краям изображения A добавлены поддельные (fake) пиксели нулевого значения (“*нулевое дополнение*”). С учётом нулевого дополнения размеры нового изображения A^* увеличиваются на две строки и два столбца. Для сжатия изображения A^* используется ядро свёртки в виде матрицы размера 2×2. Принцип свертки показан на рис. 5.1. Нулевое дополнение позволяет при одноразовой свёртке (см. рис. 2) по выходной матрице выделять границы изображений. С помощью заданного ядра $C = \begin{pmatrix} 0 & 2 \\ 1 & 1 \end{pmatrix}$ дважды провести последовательную свёртку изображения A^* с размера $(m+2)\cdot(m+2)$ до $m\cdot m$.

5.3. Tekislikdagi tasvir $A=\{a_{ij}\}_{m*m}$, $m \geq 4$ matritsa ko‘rinishida berilgan bo‘lib, matritsaning elementlari a_{ij}∈{0,…,5} piksellarning qiymatlaridan iborat. Tasvirni zichlashtirish uchun 2×2 o‘lchamli matritsa ko‘rinishidagi tugub olish yadrosidan foydalaniladi. Yadroning 4 elementlarining har birining mumkin bo‘lgan qiymatlari {0,1,2} to‘plamidan olinadi. Yadro tasvir bo‘yicha ayrim sohalarni o‘tkazib yuborgan holda sirpanadi (A matritsasi bo‘yicha ko‘chadi).

Ikki qadam bilan yangi qiymatni shakllantirish tamoyili 3-rasmda ko‘rsatilgan. Ikki qadam shuni anglatadiki, siljish ustun va satr bo‘yicha bittadan piksellarni sakrab o‘tish orqali amalga oshiriladi, natijada sirpanish jarayoni ikki barobarga kamayadi. Yangi element (zichlashgan tasvir) yadro elementlarining A matritsaning ajratilgan matritsa osti elementlariga ko‘paytmalarning yig‘indisi sifatida aniqlanadi.

$$\begin{array}{|c|c|c|c|} \hline 1^0 & 4 & 4^2 & 3 \\ \hline 3 & 0 & 0 & 2 \\ \hline 1^1 & 1 & 3^1 & 5 \\ \hline 0 & 2 & 4 & 2 \\ \hline \end{array} \dashrightarrow \begin{array}{|c|c|} \hline 12 & \times \\ \hline \times & \times \\ \hline \end{array}$$

3-rasm. Ikki qadam bilan tasvirni zichlashtirish
(Рис. 3. Сжатие изображения с пролётом через шаг 2)

Berilgan 2×2 o‘lchamdagi tugub olish yadrosi bo‘yicha A tasvirni ikki qadam bilan zichlashtiring.

5.3. Изображение задано в виде матрицы $A=\{a_{ij}\}_{m*m}$, $m \geq 4$. Элементы матрицы (пиксели изображения) a_{ij}∈{0,…,5}. Для сжатия изображения используется ядро свёртки в виде матрицы размера 2×2.

Множеством допустимых значений каждого из 4 элементов ядра является {0,1,2}. Ядро скользит по изображению (перемещается по матрице A), пропуская некоторые области. Принцип формирования нового значения с шагом 2 показан на рис. 3. Шаг 2 означает, что пролеты совершаются через каждые два пикселя, пропуская все другие пролеты в процессе и уменьшая их количество примерно в 2 раза. Новый элемент (сжатого изображения) получается как сумма произведений элементов ядра на элементы выделенной подматрицы матрицы A.

По заданному ядру свёртки размера 2×2 получить сжатие изображения A с шагом 2.

5.4. Tekislikdagi tasvir $A=\{a_{ij}\}_{m*m}$, $3\leq m\leq 5$ matritsa ko‘rinishida beriladi. Matritsaning elementlari $a_{ij}\in\{0,\ldots,5\}$ piksellarning qiymatlaridan iborat. Tasvirni o‘girish uchun $C(k)=\{c_{ij}\}_{k*k}$, $c_{ij}\in\{0,1,3\}$, $1<k<m$ matritsa ko‘rinishidagi tugub olish yadrosidan foydalaniladi. K=2 qiymati uchun tugub olish tamoyili 2-rasmda keltirilgan. Hosil bo‘luvchi matritsa elementining qiymati, $C(k)$ yadro elementlarini A matritsa minori elementlariga ko‘paytmalarining yig‘indisi sifatida aniqlanadi. Tugub olish natijasida tasvir chegaralarini ajratib olish imkoni bo‘lishi uchun A matritsaga qalbaki (*fake*) nol qiymatlar (“*nolli to‘ldirma*”). Bunda boshlang‘ich tasvir 2 ta ustun va 2 ta satrga kengayadi. Nolli to‘ldirmani hisobga olgan holda $\{C(k)\}$, $k=2,\ldots,m-1$ yadrolari to‘plami yordamida tasvirning barcha zichlashgan ko‘rinishlari hosil qilinsin. Masalani yechishda $\{C(k)\}$ yadrolar to‘plami tugub olish yadrolari kutubxonasidan olinadi. Kutubxona tasvirlarni o‘girish uchun tajriba yo‘li bilan shakllantirilgan, satr va ustunlari 5 dan katta bo‘lmagan yadrolardan iborat deb hisoblanadi.

5.4. Изображение на плоскости задано в виде входной матрицы $A=\{a_{ij}\}_{m*m}$, $3\leq m\leq 5$. Элементы матрицы A представлены значениями $a_{ij}\in\{0,\ldots,5\}$ пикселей изображения. Для преобразования изображения используется ядра свёрток в виде матриц $C(k)=\{c_{ij}\}_{k*k}$, $c_{ij}\in\{0,1,3\}$, $1<k<m$. Принцип свёртки при k =2 показан на рис. 2. Значение нового элемента выходной матрицы получается как сумма произведения элементов ядра $C(k)$ на минор матрицы A. Для того чтобы по результатам свёртки иметь возможность анализировать границы изображения к матрице A добавляются поддельные (fake) пиксели нулевого значения (“*нулевое дополнение*”). При этом размеры исходного изображения расширяются на 2 строки и 2 столбца. С учётом нулевого дополнения требуется получить все

представления сжатого изображения по набору ядер $\{C(k)\}$, $k = 2,\ldots,m-1$. Считается, что набор $\{C(k)\}$ извлечён из библиотеки ядер свёрток. Библиотека сформирована экспериментальным путём для преобразования изображений с числом строк и столбцов не больше 5.

5.5. Tasvirni $m*m$, $m \geq 4$ o‘lchamdagi matritsa bo‘yicha subdiskretlash uchun matritsa $h \geq 1$ qadam bilan $d*d$, $2 \leq d < m$ o‘lchamdagi darchalarga bo‘lib chiqiladi. Har bir darchadagi maksimal qiymat yangi (zichlashgan) matritsa elmentining qiymati sifatida olinadi. Boshlang‘ich matritsaning mumkin bo‘lgan qiymatlari to‘plami $\{0,\ldots,5\}$ qiymatlardan iborat hisoblandi. 3-rasmda 2*2 o‘lchamli darcha va 1 va 2 qadam bilan tasvirni subdiskretlashga misol keltirilgan.

1	4	4	3
3	0	0	2
1	1	3	5
0	2	4	2

4	4	4
3	3	5
2	4	5

a)

4	4
2	5

b)

4-rasm.Tasvirni subdiskretlash: a) 1 qadam bilan; b) 2 qadam bilan.
(Рис. 4. Субдискретизация изображений: а) с шагом 1; б) с шагом 2)

Берилган иккита $C=\{c_{ij}\}_{m*m}$ ва $D=\{d_{ij}\}_{m*m}$ матрицалар учун 2*2 ўлчамдаги дарча орқали 1 қадам билан тасвирларни субдискретлаш амалга оширилсин. C ва D матрицалар бўйича зичлаш натижаларини мос равишда $A=\{a_{ij}\}_{(m-1)*(m-1)}$ ва $B=\{b_{ij}\}_{(m-1)*(m-1)}$ матрицаларда ҳосил қилинсин. A ва B матрицалар ўзаро мос келмайдиган элементлари сони – $N=\left|\{(i,j) | a_{ij} \neq b_{ij}\}\right|$ аниқлансин.

5.5. Для субдискретизации (сжатия) изображения по матрице размера $m*m$, $m \geq 4$ производится разбиение его на окна размера $d*d$, $2 \leq d < m$ с шагом $h \geq 1$. Максимальное значение в каждом окне выбирается в качестве значения элемента новой (сжатой) матрицы. Множеством допустимых значений элементов исходной матрицы является $\{0,\ldots,5\}$. Пример субдискретизatsiи с окном размера 2*2 и шагом 1 и 2 показан на рис. 4.

Для двух заданных матриц $C=\{c_{ij}\}_{m*m}$ и $D=\{d_{ij}\}_{m*m}$ произвести субдискретизatsiя изображений через окно размера 2*2 с шагом 1. Результаты сжатия по матрице C и D представить соответственно как $A=\{a_{ij}\}_{(m-1)*(m-1)}$ и $B=\{b_{ij}\}_{(m-1)*(m-1)}$. Определить число несовпадающих элементов матриц A и B как $N=\left|\{(i,j)|\ a_{ij} \neq b_{ij}\}\right|$.

5.6. Tasvirni $m*m$, $m\geq4$ o‘lchamdagi matritsa bo‘yicha subdiskretlash uchun matritsa $h\geq1$ qadam bilan $d*d$, $2\leq d<m$ o‘lchamdagi darchalarga bo‘lib chiqiladi. Har bir darchadagi o‘rtacha qiymat yangi (zichlashgan) matritsa elementining qiymati sifatida olinadi. Boshlang‘ich matritsaning mumkin bo‘lgan qiymatlari to‘plami $\{0,\ldots,5\}$ qiymatlardan iborat hisoblandi. 5-rasmda 2*2 o‘lchamli darcha va 1 va 2 qadam bilan tasvirni subdiskretlashga misol keltirilgan.

1	4	4	3
3	0	0	2
1	1	3	5
0	2	4	2

2	2	2.25
1.25	1	2.5
1	2.5	3.5

a)

2	2.25
1	3.5

b)

5-rasm. Tasvirni subdiskretlash: a) 1 qadam bilan; b) 2 qadam bilan
(Рис. 5. Субдискретизация изображений: а) с шагом 1;б) с шагом 2).

Berilgan $A=\{a_{ij}\}_{m*m}$ matritsani dxd o‘lchamli darcha orqali $h=d$ qadam bilan bo‘lishlar aniqlansin. Agar $p=mod(m,d)>0$ bo‘lsa, A matritsaga nol qiymatli qalbaki (*fake*) piksellar $d-p$ ustunlar $m+d-p$ indekslar bilan va $d-p$ satrlar $m+d-p$ indekslar bilan qo‘shilsin.

5.6. Для субдискретизации (сжатия) изображения по матрице размера $m*m$, $m\geq4$ производится разбиение его на окна размера $d*d$, $2\leq d<m$ с шагом $h\geq1$. Максимальное значение в каждом окне выбирается в качестве значения элемента новой (сжатой) матрицы. Множеством допустимых значений элементов исходной матрицы является $\{0,\ldots,5\}$. Пример субдискретизации с окном размера 2*2 и шагом 1 и 2 показан на рис. 5.

Для заданной матрицы $A=\{a_{ij}\}_{m*m}$ определить разбиение на окна размера $d*d$ с шагом $h=d$. Если p=mod(m,d)>0, то к матрице A добавить поддельные (fake) пиксели нулевого значения (“нулевое дополнение”) в виде $d-p$ столбцов справа с индексами $m+d-p$ и $d-p$ строк снизу с индексами $m+d-p$.

5.7. Tasvirni $m*m$, $m\geq4$ o‘lchamdagi matritsa bo‘yicha subdiskretlash uchun matritsa $h\geq1$ qadam bilan $d*d$, $2\leq d<m$ o‘lchamdagi darchalarga bo‘lib chiqiladi. Har bir darchadagi o‘rtacha qiymat yangi (zichlashgan) matritsa elementining qiymati sifatida olinadi.

Boshlang‘ich matritsaning mumkin bo‘lgan qiymatlari to‘plami {0,…,5} qiymatlardan iborat hisoblandi. 4-rasmda 2x2 o‘lchamli darcha va 1 va 2 qadam bilan tasvirni subdiskretlashga misol keltirilgan.

Berilgan ikkita $C=\{c_{ij}\}_{m*m}$ va $D=\{d_{ij}\}_{m*m}$ matritsalar uchun 2*2 o‘lchamdagi darcha orqali 1 qadam bilan tasvirlarni subdiskretlash amalga oshirilsin. S va D matritsalar bo‘yicha zichlash natijalarini mos ravishda $A=\{a_{ij}\}_{(m-1)*(m-1)}$ va $B=\{b_{ij}\}_{(m-1)*(m-1)}$ matritsalarda hosil qilinsin. A va V matritsalar o‘zaro mos kelmaydigan elementlari soni – $N=\left|\{(i,j) | a_{ij} \neq b_{ij}\}\right|$ aniqlansin.

1	4	4	3
3	0	0	2
1	1	3	5
0	2	4	2

2	2	2.25
1.25	1	2.5
1	2.5	3.5

a)

2	2.25
1	3.5

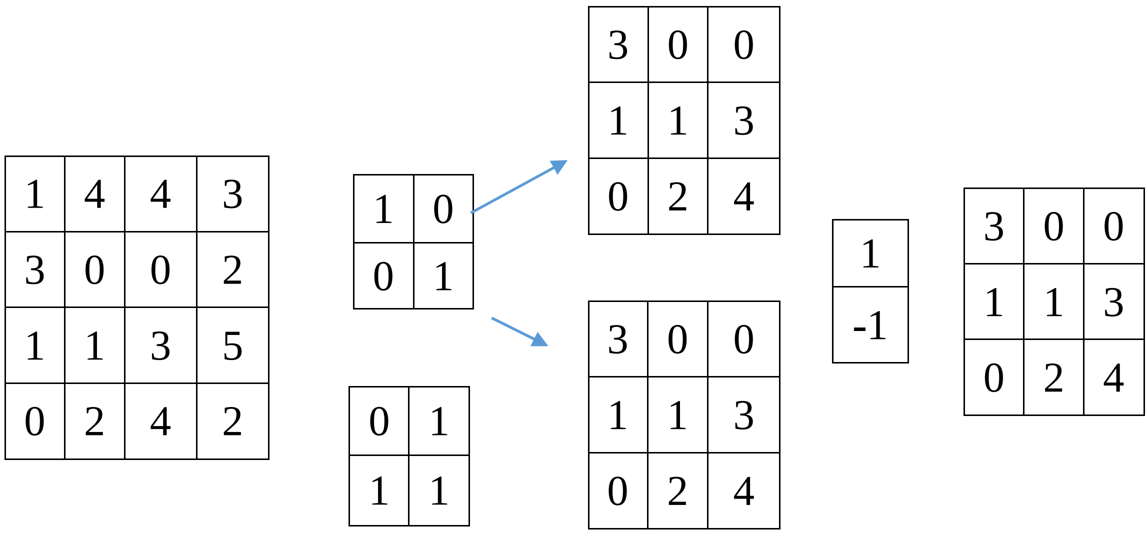
b)

6-rasm. Tasvirni subdiskretlash: a) 1 qadam bilan; b) 2 qadam bilan.
(Рис. 6. Субдискретизация изображений: а) с шагом 1;б) с шагом 2)

5.7. Для субдискретизatsiи (сжатия) изображения по матрице размера $m*m$, $m \geq 4$ производится разбиение его на окна размера $d*d$, $2 \leq d < m$ с шагом $h \geq 1$. Среднее значение в каждом окне выбирается в качестве значения элемента новой (сжатой) матрицы. Множеством допустимых значений элементов исходной матрицы является {0,…,5}. Пример субдискретизации с окном размера 2*2 и шагом 1 и 2 показан на рис. 6.

Для двух заданных матриц $C=\{c_{ij}\}_{m*m}$ и $D=\{d_{ij}\}_{m*m}$ произвести субдискретизatsiя изображений через окно размера 2*2 с шагом 1. Результаты сжатия по матрице C и D представить соответственно как $A=\{a_{ij}\}_{(m-1)*(m-1)}$ и $B=\{b_{ij}\}_{(m-1)*(m-1)}$. Определить число несовпадающих элементов матриц A и B как $N=\left|\{(i,j) | a_{ij} \neq b_{ij}\}\right|$.

5.8. Tekislikdagi tasvir $A=\{a_{ij}\}_{m*m}$, $m \geq 4$ matritsa ko‘rinishida berilgan. A matritsaning $a_{ij} \in \{0,\ldots,5\}$ elementlari piksel qiymatlari. Tasvirni o‘girish ikkita qatlamda amalga oshiriladi. Birinchi qatlamda 2x2 o‘lchamli tugub olish yadrolardan foydalaniladi. Yadro elementlari {-1,1} qiymatlari bilan aniqlanadi. Ikkinchi qatlamda, 7-rasmda ko‘rsatilgandek 2x1 o‘lchamdagi yadro bilan matritsalar birlashtiriladi.



7-rasm. Ikki qatlamli to‘r
(Рис. 7. Двухслойная сеть)

Har bir yadro tasvir bo‘yicha sirpanadi (A matritsa bo‘yicha ko‘chadi). 7-rasmda ko‘rsatilgan sxema bo‘yicha tasvirni o‘girish amalga oshirilsin.

5.8. Изображение на плоскости задано в виде матрицы $A=\{a_{ij}\}_{m*m}$, $m \geq 4$. Элементами матрицы A являются значения пикселей $a_{ij} \in \{0,\ldots,5\}$. Преобразование изображения реализуется в два слоя. На первом слое для сжатия используются ядра свёртки в виде матриц размера 2*2. Элементы ядра представляются с помощью значений {-1,1}. На втором слое производится объединение матриц с помощью ядра размера 2*1. Как показано на рис. 7.

Каждое ядро скользит по изображению (перемещается по матрице A) как показано на рис. 7. Реализовать преобразование изображения по схеме показанной на рис.7.

6. Berilganlarni vizuallashtirish (Визуализация данных)

6.1. Berilgan tasoddifiy kattaliklar qiymatlari to’plami $X = (x_1, \ldots, x_n)$ bo’yicha protsentil taqsimot grafigi chizilsin.

6.1. По заданному множеству значений случайной величины $X = (x_1, \ldots, x_n)$ построить график процентильного распределения.

6.2. Berilgan tasoddifiy kattaliklar qiymatlari to'plami $X = (x_1, \ldots, x_n)$ bo'yicha detsil taqsimot grafigi chizilsin.

6.2. По заданному множеству значений случайной величины $X = (x_1, \ldots, x_n)$ построить график децильного распределения.

6.3. Miqdoriy $X(n)=(x_1,\ldots,x_n)$ alomatlar bilan tavsiflangan $E_0 = (S_1,\ldots,S_m)$ tanlanma obyektlari $A=\{a_{ij}\}_{m\times n}$ jadval ko'rinishida berilgan. Har bir $x_i \in X(n)$ alomat bo'yicha mediana qiymati $M(x_i)$ hisoblangan bo'lsin.

Alomatlar juftligi $\{x_i\ x_j\} \in X(n)$ uchun talab qilinadi:

1. Tekislikni $x=M(x_i)$, $y=M(x_j)$ chiziqlar bo'yicha 4 ta qismga bo'lsin va ular chapdan-o'nga, pastdan-yuqoriga 1,2,3,4 nomerlari bilan identifikatsiyalansin.

2. E_0 tanlanma obyektlari tavsiflari tekislikka akslantirilsin va qiymatlari tekislikdagi obyektni o'z ichiga olgan soha nomerlari bo'lgan yangi alomat shakllantirilsin.

Izoh: Obyektning $x_i \in X(n)$ alomati qiymatini $M(x_i)$ mediana bilan ustma-ust tushmasligini ta'minlash uchun hisob paytida $M^*(x_i)= M(x_i) + \varepsilon$ qiymatidan foydalanilsin. Bu erda ε ($\varepsilon>0$) – etarlicha kichik kattalik.

$$\alpha = \frac{\pi_2 + b}{2}$$

6.3. В заданной таблице $A=\{a_{ij}\}_{m\times n}$ содержатся описание объектов $E_0 = (S_1,\ldots,S_m)$ по n количественным признакам $X(n)=(x_1,\ldots,x_n)$. Считается, что по каждому признаку $x_i \in X(n)$ вычислено значение медианы $M(x_i)$.

Требуется по каждой паре $\{x_i\ x_j\} \in X(n)$:

1.Определить разбиение плоскости на 4 части по линиям $x=M(x_i)$, $y=M(x_j)$ и идентифицировать их номерами 1,2,3,4 слева–направо, снизу–вверх.

2. Отобразить описания объектов E_0 на плоскость и сформировать новый признак, значениями которого являются номера области на плоскости, содержащей описания объектов.

Примечание: для того чтобы исключить совпадение значение признака $x_i \in X(n)$ объекта и медианы $M(x_i)$, при расчёте необходимо использовать значение $M^*(x_i)= M(x_i) + \varepsilon$, где ε – бесконечно малая величина, $\varepsilon>0$.

$$\alpha = \frac{\pi_2 + b}{2}$$

6.4. Butun *m* soni va ular o‘rtasidagi masofalar *A*(*m***m*) simmetrik matritsa bilan berilgan. O‘lchamlarni kamaytirish uchun quyidagi yo‘qotish funksiyasidan foydalanilgan holda ob’ektlarni R^2 fazosiga chiziqsiz akslantirish amalga oshiriladi:

$$\varepsilon = \frac{1}{\sum_{i<j} d_{ij}} \sum_{i<j}^{m} \frac{\left(d_{ij} - \delta_{ij}\right)^2}{d_{ij}},$$

bu yerda d_{ij}, δ_{ij} – mos ravishda qo‘p o‘lchamli va ikki o‘lchamli fazodagi *i* va *j* ob’ektlar o‘rtasidagi masofa. Algoritmni ishlashi iterativ xususiyatga ega bo‘lib, u ikki o‘lchamdagi fazo koordinatalari y_{ik} uchun rekkurient munosabat bilan bog‘liqdir:

$$y_{ik}(t+1) = y_{ik}(t) + \frac{2\cdot\alpha}{\sum_{i<j}^{m} d_{ij}} \cdot \sum_{\substack{i=1 \\ j\neq i}}^{m} \frac{d_{ij} - \delta_{ij}}{d_{ij}\delta_{ij}} \cdot \left(y_{ik}(t) - y_{jk}(t)\right), 0 < \alpha \le 1.$$

6.4. Расстояния между *m* объектами заданы в виде симметричной квадратной матрицы *A*(*m***m*). Для снижения размерности производится нелинейное отображение объектов в пространство R^2 с использованием функции потерь

$$\varepsilon = \frac{1}{\sum_{i<j} d_{ij}} \sum_{i<j}^{m} \frac{\left(d_{ij} - \delta_{ij}\right)^2}{d_{ij}},$$

где d_{ij}, δ_{ij} – расстояния между объектами *i* и *j*, соответственно, в многомерном и двумерном пространстве. Работа алгоритма имеет итеративный характер, связанный со следующим рекуррентным соотношением для координат в двумерном пространстве y_{ik}:

$$y_{ik}(t+1) = y_{ik}(t) + \frac{2\cdot\alpha}{\sum_{i<j}^{m} d_{ij}} \cdot \sum_{\substack{i=1 \\ j\neq i}}^{m} \frac{d_{ij} - \delta_{ij}}{d_{ij}\delta_{ij}} \cdot \left(y_{ik}(t) - y_{jk}(t)\right), 0 < \alpha \le 1.$$

7. Matematik lingvistika (Математическая лингвистика)

7.1. (*So’zlarning semantic yaqinligi*). Tabiiy tildagi *T* – matnli hujjat *m* (*m*>1) ta gaplar to’plami $T=\{t_j\}_{j\in\{1,\ldots,m\}}$ ko’rinishida berilgan. Gap tugashini {.,?,!} to’plamga kiruvchi belgilar bildiradi. Matn tahlili *n*(*n*>1) kalit so'zlar lug'ati $A=\{a_i\}_{i\in\{1,\ldots,n\}}$ orqali amalga oshiriladi. Hujjatga dastlabki ishlov berish – gapdagi $a_i, a_j \in t_k$, $i\neq j$, $k=1,\ldots,m$ so'zlar juftligi uchrashi bo'yicha shakllanuvchi $R=\{r_{ij}\}_{n\times n}$ matrisani hosil qilishga keltiriladi. Matritsaning nol bo'lmagan elementlari

$$r_{ij}=|\{\, a_i,a_j \in t_k \mid \mathrm{pos}(t_k, a_i) < \mathrm{pos}(t_k, a_j),\, k=1,\ldots,m\}|$$

ko'rinishida hisoblanadi. Bu erda $\mathrm{pos}(t_k, a_i)$ ($\mathrm{pos}(t_k, a_j)$) – a_i,(a_j) so'zning t_k gapdagi joylashuv o'rni. Agar (a_i,a_j) juftlik uchun $\mathrm{pos}(t_k, a_i) < \mathrm{pos}(t_k, a_j)$ shart bajariluvchi $t_k\in T$ gap mavjud bo'lmasa $r_{ij}=0$ bo'ladi.

Talab qilinadi:

- so'zlarning juftligi uchrashini ko'rsatuvchi R matritsasini faqat nol bo'lmagan $r_{ij}>0$ elementlarini (i,j,r_{ij}) uchlik ko'rinishida yozish orqali ixcham ko'rinishda yozish, bu erda i–satr nomeri, j– ustun nomeri;
- boshqa so'zlar bilan juftlikda uchrash chastotasi $r_{ij} + r_{ji} = \max$ bo'lgan $a_i\in A$ so'z(lar) aniqlansin.

7.1. (*Семантическая близость слов*). Текстовый документ T на естественном языке задан как множество из m (m>1) предложений $T=\{t_j\}_{j\in\{1,\ldots,m\}}$. Концом предложения является символ из множества {.,?,!}. Анализ текста реализуется через словарь из n(n>1) ключевых слов $A=\{a_i\}_{i\in\{1,\ldots,n\}}$. Предобработка документа сводится к формированию матрицы $R=\{r_{ij}\}_{n\times n}$ попарной встречаемости слов $a_i,a_j\in t_k$, $i\neq j$, $k=1,\ldots,m$ в предложениях, ненулевые значения элементов которой вычисляются как

$$r_{ij}=|\{\, a_i,a_j \in t_k \mid \mathrm{pos}(t_k, a_i) < \mathrm{pos}(t_k, a_j),\, k=1,\ldots,m\}|,$$

где $\mathrm{pos}(t_k, a_i)$ ($\mathrm{pos}(t_k, a_j)$) – позиция слова a_i,(a_j) в предложении t_k. Если для пары (a_i,a_j) не существует $t_k\in T$, для которого $\mathrm{pos}(t_k, a_i) < \mathrm{pos}(t_k, a_j)$, то $r_{ij}=0$.

Требуется:

– записать матрицу попарной встречаемости слов R в упакованном виде, сохраняя лишь ненулевые значения элементов $r_{ij}>0$ в виде троек (i,j,r_{ij}), где i–номер строки, j– номер столбца;

– определить слово(а) $a_i\in A$, частота встречаемости которого(ых) в паре с другими словами $r_{ij} + r_{ji} = \max$.

7.2. (*Hujjatlarning semantic bog'langanligi*). Tabiiy tildagi T – matnli hujjat m (m>1) ta gaplar to'plami $T=\{t_j\}_{j\in\{1,\ldots,m\}}$ ko'rinishida berilgan. Gap tugashini {.,?,!} to'plamga kiruvchi belgilar bildiradi. Matn tahlili n(n>1) ta so'zlar juftligi $D=\{a_i,b_i\}_{i\in\{1,\ldots,n\}}$ to'plami orqali amalga oshiriladi. T hujjatdagi $x,y\in D$ so'zlar juftligini uchraydigan gaplar sonini hisoblovchi $count(x,y,T)$ funksiya aniqlangan bo'lsin.

Berilgan T_1 va T_2 hujjatlar uchun talab qilinadi:

– har bir $a_i, b_i \in D$ so'zlar juftligi uchun $r1_i = count(a_i, b_i, T_1)$, $r2_i = count(a_i, b_i, T_2)$ hisoblansin;

– T_1 va T_2 hujjatlar bog'langanligi

$$ZET = \sum_{\{r1_i + r2_i > 0\}} \left(1 - \frac{|r1_i / m1 - r2_i / m2|}{|r1_i / m1 + r2_i / m2|}\right) / \left|\{i | r1_i + r2_i > 0\}\right|$$

ko'rinishida hisoblansin.

7.2. (*Семантическая связанность документов*). Текстовые документы $T1$ и $T2$ на естественном языке заданы множествами соответственно из $m1=|T1|$ и $m2=|T2|$ ($m1>1$, $m2>1$) предложений. Концом предложения является символ из множества {.,?,!}. Анализ документов реализуется через набор из n(n>1) пар слов $D=\{a_i, b_i\}_{i \in \{1,\ldots,n\}}$. Считается, что определена функция $count(x,y,T)$ для подсчета числа предложений документа T, содержащих пару слов $x,y \in D$.

Требуется:

– определить значения $r1_i = count(a_i, b_i, T1)$, $r2_i = count(a_i, b_i, T2)$ по каждой паре слов $a_i, b_i \in D$;

– вычислить связанность документов $T1$ и $T2$ как

$$ZET = \sum_{\{r1_i + r2_i > 0\}} \left(1 - \frac{|r1_i / m1 - r2_i / m2|}{|r1_i / m1 + r2_i / m2|}\right) / \left|\{i | r1_i + r2_i > 0\}\right|.$$

7.3. (*Trigramm uchrashi chastotasi*). $R=\{'a', 'b', 'c', 'd', 'f'\}$ belgilar to'plamidan shakllangan m (m>1) ta, ixtiyoriy uzunlikdagi $P=\{p_i\}_{i \in \{1,\ldots,m\}}$ zanjirlar to'plami berilgan bo'lsin. Elementlari $\{xyz\}$, $x,y,z \in R$ bo'lgan trigrimmlar to'plamida ularni uchrash chastotalarini $|\{ xyz \in p_i \mid x \neq y, x,y \in R \setminus \{z\}, i=1,\ldots,m\}|$ ko'rinishida hisoblash aniqlangan bo'lsin.

Berilgan P zanjirdagi har bir $z \in R$ bo'yicha uchrash chastotasi maksimal bo'lgan xyz ko'rinishidagi trigrimmlar aniqlansin.

7.3. (*Частота встречаемости триграмм*). Задано множество из m (m>1) цепочек $P=\{p_i\}_{i \in \{1,\ldots,m\}}$ произвольной длины, сформированных из набора символов $R=\{a,b,c,d,f\}$. На множестве триграмм вида $\{xyz\}$, $x,y,z \in R$ определено вычисление частоты их встречаемости как

$$|\{ xyz \in p_i \mid x \neq y, x,y \in R \setminus \{z\}, i=1,\ldots,m\}|.$$

Требуется определить триграммы вида xyz, частоты встречаемости которых по каждой $z \in R$ на P максимальны.

7.4. Tabiiy tildagi n ta so'zlarning asoslari (lug'at) $L=(c_1,\ldots,c_n)$ to'plami bilan berilgan. Matn hujjati uchun $a_i = d_i / d_{um}$ – uchrashlar chastotasi bilan $A=(a_1,\dots,a_n)$ vektor-satrni qurish talab qilinadi. Bu yerda d_i – hujjatdagi lug‘atga kiruvchi $c_i \in L$ so‘zning uchrashlari soni, d_{um} – lug‘atdagi hujjatga kiruvchi so‘zlarning umumiy soni.

7.4. Задан набор (словарь) $L=(c_1,\dots,c_n)$ из n основ слов естественного языка. Требуется для текстового документа построить вектор-строку $A=(a_1,\dots,a_n)$ с частотой встречаемости $a_i = d_i/d_{um}$, где d_i – число встречаемости слова $c_i \in L$ из словаря, d_{um} – общее количество слов из словаря в документе.

7.5. Tabiiy tildagi n ta so‘zlarning asoslari (lug‘at) $L=(c_1,\dots,c_n)$ to‘plami bilan berilgan. $R=(D_1,\dots,D_m)$ kolleksiyada berilgan m matn hujjatlari L lug‘atga kiruvchi n so‘zlar uchrashlari chastotalari bilan berilgan. D_i hujjatdagi $c_i \in L$ so‘zning kirishlar sonining L lug‘atdagi so‘zlarning umumiy soniga nisbati $d_{ij} \in D_i$ qiymati ko‘rinishida hisoblanadi.

Matn hujjatlarning vektor tavsifining $\{(D_i,D_j)\}$ juftliklar to‘plami ichida kosinus masofasi bo‘yicha eng kichik qiymatga ega juftlik topilsin:

$$\cos(D_u,D_v)=\frac{\sum_{i=1}^{n} d_{ui}d_{vi}}{\sqrt{\sum_{i=1}^{n}(d_{ui})^2}\sqrt{\sum_{i=1}^{n}(d_{vi})^2}}.$$

7.5. Задан набор (словарь) $L=(c_1,\dots,c_n)$ из n основ слов естественного языка. В коллекции $R=(D_1,\dots,D_m)$ каждый из m текстовых документов описывается через частоту встречаемости n слов из словаря L. Значение $d_{ij} \in D_i$ вычисляется как количество (доля) вхождений слова $c_i \in L$ относительно общего количества слов из словаря L в документе D_i.

Требуется по множеству пар векторных представлений текстовых документов $\{(D_i,D_j)\}$ выбрать пару с наименьшим значением косинусного расстояния

$$\cos(D_u,D_v)=\frac{\sum_{i=1}^{n} d_{ui}d_{vi}}{\sqrt{\sum_{i=1}^{n}(d_{ui})^2}\sqrt{\sum_{i=1}^{n}(d_{vi})^2}}.$$

7.6. Tabiiy tildagi n ta so‘zlarning asoslari (lug‘at) $L=(c_1,\dots,c_n)$ to‘plami bilan berilgan. $R=(D_1,\dots,D_m)$ kolleksiyada berilgan m matn hujjatlari L lug‘atga kiruvchi n so‘zlar uchrashlari chastotalari bilan berilgan. D_i hujjatdagi $c_i \in L$ so‘zning kirishlar sonining L lug‘atdagi so‘zlarning umumiy soniga nisbati $d_{ij} \in D_i$ qiymati ko‘rinishida hisoblanadi.

Matn hujjatlarning vektor tavsifining $\{(D_i, D_j)\}$ juftliklar to‘plami ichida Xelinger masofasi bo‘yicha eng kichik qiymatga ega juftlik topilsin:

$$H(D_u, D_v) = \frac{1}{\sqrt{2}} \sqrt{\sum_{i=1}^{n} \left(\sqrt{d_{ui}} - \sqrt{d_{vi}} \right)}.$$

7.6. Задан набор (словарь) $L=(c_1,\ldots,c_n)$ из n основ слов естественного языка. В коллекции $R=(D_1,\ldots,D_m)$ каждый из m текстовых документов описывается через частоту встречаемости n слов из словаря L. Значение $d_{ij} \in D_i$ вычисляется как количество (доля) вхождений слова $c_i \in L$ относительно общего количества слов из словаря L в документе D_i.

Требуется по парам векторных представлений текстовых документов (Di,Dj) выбрать пару с наименьшим значением расстояния Хелингера

$$H(D_u, D_v) = \frac{1}{\sqrt{2}} \sqrt{\sum_{i=1}^{n} \left(\sqrt{d_{ui}} - \sqrt{d_{vi}} \right)}.$$

7.7. Tabiiy tildagi n ta so‘zlarning asoslari (lug‘at) $L=(c_1,\ldots,c_n)$ to‘plami bilan berilgan. Berilgan m matn hujjatlaridagi L lug‘atga kiruvchi so‘zlar uchrashlari $R=(D_1,\ldots,D_m)$ kolleksiya ko‘rinishida berilgan bo‘lib, unda hujjatning nashr vaqti (kun, oy, yil), mamlakat nomi, nashriyot nomi va adabiyotlarga navolalarni o‘z ichiga oladi. Berilgan mamlakat va nashriyot nomlari bo‘yicha 20XX va 20UU yillar oralig‘ida nashr qilingan hujjatlar aniqlansin.

7.7. Задан набор (словарь) $L=(c_1,\ldots,c_n)$ из n основ слов естественного языка. В коллекции $R=(D_1,\ldots,D_m)$ каждый из m текстовых документов описывается через частоту встречаемости n слов из словаря L. Коллекция R содержит название документа, сведения об времени издания (число, месяц, год), название страны, название издательства, ссылки на литературные источники. Требуется по названию страны и издательства определить все документы, изданные за период с 20ХХ по 20УУ годы.

7.8. Tabiiy tildagi n ta so‘zlarning asoslari (lug‘at) $L=(c_1,\ldots,c_n)$ to‘plami bilan berilgan. Berilgan m matn hujjatlaridagi L lug‘atga kiruvchi so‘zlar uchrashlari $R=(D_1,\ldots,D_m)$ kolleksiya ko‘rinishida berilgan bo‘lib, unda hujjatning nashr vaqti (kun, oy, yil), mamlakat nomi, nashriyot nomi va adabiyotlarga navolalarni o‘z ichiga oladi. Mamlakat nomi berilganda, nashr yili va oyi ustma-ust tushadigan hujjatlar uchun aniqlansin.

7.8. Задан набор (словарь) $L=(c_1,\ldots,c_n)$ из n основ слов естественного языка. В коллекции $R=(D_1,\ldots,D_m)$ каждый из m текстовых документов описывается через частоту встречаемости n слов из словаря L. Коллекция R содержит название документа, сведения об времени издания (число, месяц, год), название страны, название издательства, ссылки на литературные источники.

Требуется определить документы, у которых название страны, год и месяц совпадают.

7.9. Tabiiy tildagi n ta so‘zlarning asoslari (lug‘at) $L=(c_1,\ldots,c_n)$ to‘plami bilan berilgan. Berilgan m matn hujjatlaridagi L lug‘atga kiruvchi so‘zlar uchrashlari $R=(D_1,\ldots,D_m)$ kolleksiya ko‘rinishida hisoblangan bo‘lib, unda hujjatning nashr vaqti (kun, oy, yil), mamlakat nomi, nashriyot nomi va adabiyotlarga navolalarni o‘z ichiga oladi. Berilgan $D_i \in R$ hujjat uchun unga korrekt va nokorrekt navolaga ega bo‘lgan R ro‘yxatga kiruvchi hujjatlar to‘plam ostisi aniqlansin. Hujjatning navolasi nokorrekt hisoblanadi, agar uning nashr yili D_i hujjatnikidan kichik bo‘lsa.

Izoh: R kolleksiyadagi hujjatga navola $D_i \in R$ hujjatning i indeksining qiymati bilan aniqlanadi.

7.9. Задан набор (словарь) $L=(c_1,\ldots,c_n)$ из n основ слов естественного языка. В заданном наборе $R=(D_1,\ldots,D_m)$ каждый из m текстовых документов описывается через частоту встречаемости n слов из словаря L. Коллекция R содержит название документа, сведения об времени издания (число, месяц, год), название страны, название издательства, ссылки на литературные источники. Требуется для заданного документа $D_i \in R$ определить перечень документов из R корректно или некорректно ссылающиеся на D_i. Некорректной считается ссылка у того документа, год издания у которого меньше чем у D_i.

Примечание: Ссылка на документ из R определяется значением индекса i документа $D_i \in R$

7.10. (*Eng umumiy ketma-ketlik ostini (EUKK) topish*)

Ta’rif. Z ketma-ketlik X va Y ketma-ketliklar uchun umumiy ketma-ketlik ostisi bo‘ladi, agar u bir paytda X hamda Y uchun ketma-ketlik ostisi bo‘lsa.

Berilgan X va Y ketma-ketliklar uchun eng katta uzunlikka ega bo‘lgan ketma-ketlik ostisi topilsin.

1-izoh. EUKK bir nechta bo‘lishi mumkin.

2-izoh. Ketma-ketlik satr ostidan farq qiladi. Masalan, agar boshlang‘ich ketma-ketlik «ABCDEF» bo‘lsa, u holda «ACE» ketma-ketlik ostisi bo‘ladi, satr ostisi emas, «ABC» esa bir paytda ketma-ketlik ostisi hamda satr ostisi bo‘ladi.

7.10. (*Поиск наибольшей общей подпоследовательности* (*НОП*)).

Определение. Будем говорить, что последовательность Z является общей подпоследовательностью последовательностей X и Y, если Z является подпоследовательностью как X, так и Y.

Требуется для двух заданных последовательностей X и Y найти все общие подпоследовательности наибольшей длины.

Примечание 1. НОП может быть несколько.

Примечание 2. Подпоследовательность отличается от подстроки. Например, если есть исходная последовательность «ABCDEF», то «ACE» будет подпоследовательностью, но не подстрокой, а «ABC» будет как подпоследовательностью, так и подстрокой.

7.11. Berilgan $T=(D_1,\ldots,D_m)$ matn hujjatlari to‘plamida ikkita o‘zaro kesishmaydigan sinf va l ta guruhlarga (mavzularga) bo‘lingan. $D_i \in T$ hujjat $D_i = (h_i, g_i)$ ko‘rinishida bo‘lib, unda h_i – K_1 yoki K_2 sinf nomeri, $h_i \in \{1,2\}$ va $g_i \in \{1,\ldots,l\}$, $l \geq 2$ – o‘zaro kesishmaydigan $G_1,\ldots,G_l$ guruhlar nomerlari.

Talab qilinadi:

- har bir G_i guruh uchun ob’ektlarning K_1 sinfga tegishligi funksiyasi $\lambda_i(K_1) = \dfrac{v_{i1}}{|G_i|}$ ko‘rinishida aniqlansin. Bu yerda v_{i1} – G_i guruhdagi K_1 sinf ob’ektlari soni;

- T to‘plamdagi hujjatlarni l guruhga bo‘lingan holdagi kontentli autentlik qiymati quyidagicha hisoblansin:

$$F(l) = \frac{1}{m} \sum_{j=1}^{l} \begin{cases} |G_j| \lambda_j(K_1), \lambda_j(K_1) > 0.5; \\ |G_j| (1 - \lambda_j(K_1)), \lambda_j(K_1) < 0.5. \end{cases}$$

7.11. На множестве текстовых документов $T=(D_1,\ldots,D_m)$ задано разбиение на два непересекающихся класса и l групп (тем). Документ $D_i \in T$ представлен как $D_i = (h_i, g_i)$, где h_i – номер класса K_1 или K_2, $h_i \in \{1,2\}$ и $g_i \in \{1,\ldots,l\}$, $l \geq 2$ – номера непересекающихся групп $G_1,\ldots,G_l$. Требуется:

–для каждой группе G_i определить значение функции принадлежности объектов к классу K_1 по G_i как $\lambda_i(K_1)=\frac{v_{i1}}{|G_i|}$, где v_{i1} – число объектов класса K_1 в G_i.

– вычислить значение контентной аутентичности документов из T при разбиении их на l групп как

$$F(l)=\frac{1}{m}\sum_{j=1}^{l}\begin{cases}|G_j|\lambda_j(K_1), \lambda_j(K_1)>0.5;\\ |G_j|(1-\lambda_j(K_1)), \lambda_j(K_1)<0.5.\end{cases}$$

7.12. Berilgan $T=(D_1,\ldots,D_m)$ matn hujjatlari to‘plami ikkita o‘zaro kesishmaydigan K_1 va K_2 to‘plam ostilariga (sinflarga) bo‘lingan. Har bir hujjat $X(n)$ – n ta miqdoriy alomatlar bilan tavsiflangan. Yevklid metrikasi bo‘yicha chegaraviy ob’ektlar to‘plami $T_c\subset T$ quyidagi ko‘rinishda aniqlanadi:

$$T_c=\left\{\mathrm{D}_i\in K_u \middle| \rho(D,D_i)=\min_{D_j\in K_u, D\in K_{3-u}}\rho(D,D_j)\right\},$$

$$r_i=\min_{D_i\in K_u, D\in K_{3-u}}\rho(D,D_i),\ r_j=\min_{D_j\in K_u, D\in K_{3-u}}\rho(D,D_j).$$

Ikkita D_i, $D_j\in K_u$ ob’ektlar $D\in K_u\cap T_c$ bo‘yicha bog‘langan hisoblanadi, agar $\rho(D_i,D)<r_i$ va $\rho(D_j,D)<r_j$ bo‘lsa.

Berilgan T to‘plamdagi bog‘langan ob’ektlardan tashkil topgan minimal sondagi guruhlarni aniqlash talab qilinadi.

7.12. Множество текстовых документов $T=(D_1,\ldots,D_m)$ представляет объединение двух непересекающихся подмножеств (классов) K_1 и K_2. Каждый документ описывается набором из n количественных признаков $X(n)$. Множество граничных по евклидовой метрике объектов T_c на $T, T_c\subset T$ определяется как

$$T_c=\left\{\mathrm{D}_i\in K_u \middle| \rho(D,D_i)=\min_{D_j\in K_u, D\in K_{3-u}}\rho(D,D_j)\right\},$$

$$r_i=\min_{D_i\in K_u, D\in K_{3-u}}\rho(D,D_i),\ r_j=\min_{D_j\in K_u, D\in K_{3-u}}\rho(D,D_j).$$

Объекты D_i, $D_j\in K_u$ связаны по $D\in K_u\cap T_c$, если $\rho(D_i,D)<r_i$ и $\rho(D_j,D)<r_j$.

Требуется определить минимальное количество групп из связанных объектов на T.

7.13. Berilgan $E_0=(S_1,\ldots,S_m)$ ob’ektlar to‘plami ikkita o‘zaro kesishmaydigan K_1 va K_2 to‘plam ostilariga (sinflarga) bo‘lingan. Har bir ob’ekt $X(n)$ – n ta miqdoriy alomatlar bilan tavsiflangan. E_0 to‘plamida ikkita metrika aniqlangan: $\rho_1(x,y)$ – yevklid va $\rho_2(x,y)$ – vaznli yevklid. Vaznli yevklid metrikasi berilgan $W=(w_1,\ldots,w_n)$ bo‘yicha $\rho_2(x,y)=\sqrt{\sum_{i=1}^{n} w_i(x_i-y_i)^2}$ ko‘rinishida hisoblanadi, bu yerda $0<w_i\le 1$. Har bir $S\in K_t\cap E_0$, t=1,2 ob’ekt uchun $\rho_1(x,y)$ va $\rho_2(x,y)$ metrikalar bo‘yicha kamayish bo‘yicha tartiblangan $\Pi_1(S)=(S_1^1,\ldots,S_m^1)$ va $\Pi_2(S)=(S_1^2,\ldots,S_m^2)$ ketma-ketliklar qurilgan bo‘lsin. $P_1(S)$ va $P_2(S)$ ketma-ketlikdagi pastki indekslar ob’ektlar rangini bildiradi.

Talab qilinadi:

– har bir $S\in K_t\cap E_0$ ob’ekt uchun $P_1(S)$ va $P_2(S)$ ketma-ketlikdagi K_t sinf ob’ektlarining o‘rtacha rangi hisoblansin;

– E_0 to‘plamdagi vaznli yevklid metrikasi bo‘yicha rangi yevklid metrikasidan kichik bo‘lgan ob’ektlar aniqlansin.

7.13. Множество объектов $E_0=(S_1,\ldots,S_m)$ разделено на два непересекающихся подмножества (класса) K_1 и K_2. Каждый объект $S\in E_0$ описывается набором из n количественных признаков $X(n)$. На множестве E_0 определены две метрики: евклидова $\rho_1(x,y)$ и взвешенная евклидова $\rho_2(x,y)$. Взвешенная евклидова метрика вычисляется как $\rho_2(x,y)=\sqrt{\sum_{i=1}^{n} w_i(x_i-y_i)^2}$ по заданному вектору весов $W=(w_1,\ldots,w_n)$, где $0<w_i\le 1$.

Для каждого объекта $S\in K_t\cap E_0$, t=1,2 построены упорядоченные по убывания расстояния $\rho_1(x,y)$ и $\rho_2(x,y)$ последовательности $\Pi_1(S)=(S_1^1,\ldots,S_m^1)$ и $\Pi_2(S)=(S_1^2,\ldots,S_m^2)$. Нижний индекс в последовательностях $\Pi_1(S)$ и $\Pi_2(S)$ определяет ранг объекта.

Требуется:

– для каждого $S\in K_t\cap E_0$ вычислить среднее число рангов объектов класса K_t в последовательностях $\Pi_1(S)$ и $\Pi_2(S)$;

– определить объекты E_0, средний ранг по взвешенной евклидовой метрике у которых меньше чем по евклидовой.

7.14. (*Gauss zichligini aniqlash*).

Berilgan $E_0=(S_1,\ldots,S_m)$ to‘plamida ob’ektlar $X(n)=(x_1,\ldots,x_n)$ – miqdoriy alomatlar bilan tavsiflangan. Ixtiyoriy $S_i\in E_0$ ob’ektning S_j ob’ektga nisbatan shartli ehtimolligi

$$P_{j|i}=\frac{\exp\left(-\rho^2\left(S_i,S_j\right)/\sigma_i^2\right)}{\sum_{k\neq i}\exp\left(-\rho^2\left(S_i,S_k\right)/\sigma_i^2\right)}$$

bilan va mos ravishda $S_j \in E_0$ ob'ektning S_i ob'ektga nisbatan shartli ehtimolligi

$$P_{i|j} = \frac{\exp\left(-\rho^2\left(S_i, S_j\right)/\sigma_j^2\right)}{\sum_{k\neq j}\exp\left(-\rho^2\left(S_j, S_k\right)/\sigma_j^2\right)}$$

bilan hisoblanadi. Bu yerda $\rho(S_i,S_j)$ – evklid metrikasi,

$$\sigma_j^2 = \sum_{k\neq j}\rho^2\left(S_j, S_k\right)/(m-1),\quad \sigma_i^2 = \sum_{k\neq i}\rho^2\left(S_i, S_k\right)/(m-1).$$

Markazi S_i ob'ektda bo'lgan gauss zichligi $P_{ij} = \dfrac{P_{i|j} + P_{j|i}}{2}$ ko'rinishida aniqlanadi. Maksimal gauss zichligiga ega ob'ekt (ob'ektlar) aniqlansin.

7.14. (*Вычисление гауссовской плотности*).

Дано описание множества объектов $E_0=(S_1,\ldots,S_m)$ с помощью набора количественных признаков $X(n)=(x_1,\ldots,x_n)$. Условная вероятность объекта $S_i \in E_0$ относительно S_j вычисляется как

$$P_{j|i} = \frac{\exp\left(-\rho^2\left(S_i, S_j\right)/\sigma_i^2\right)}{\sum_{k\neq i}\exp\left(-\rho^2\left(S_i, S_k\right)/\sigma_i^2\right)}.$$

и соответственно $S_j \in E_0$ относительно S_i

$$P_{i|j} = \frac{\exp\left(-\rho^2\left(S_i, S_j\right)/\sigma_j^2\right)}{\sum_{k\neq j}\exp\left(-\rho^2\left(S_j, S_k\right)/\sigma_j^2\right)},$$

где $\rho(S_i,S_j)$ – евклидова метрика, $\sigma_j^2 = \sum_{k\neq j}\rho^2\left(S_j, S_k\right)/(m-1)$,

$$\sigma_i^2 = \sum_{k\neq i}\rho^2\left(S_i, S_k\right)/(m-1).$$

Гауссовская плотность с центром в объекте S_i определяется как $P_{ij} = \dfrac{P_{i|j} + P_{j|i}}{2}$. Требуется определить объект (объекты) с максимальной гауссовской плотностью.

7.15. (*Kulbak-Leybler mertikasi*).

Berilgan $E_0=(S_1,\ldots,S_m)$ to'plamida ob'ektlar $X(n)=(x_1,\ldots,x_n)$ – miqdoriy alomatlar bilan tavsiflangan. Ixtiyoriy $S_i \in E_0$ ob'ektning S_j ob'ektga nisbatan shartli ehtimolligi

$$P_{j|i} = \frac{\exp\left(-\rho^2\left(S_i, S_j\right)/\sigma_i^2\right)}{\sum_{k\neq i}\exp\left(-\rho^2\left(S_i, S_k\right)/\sigma_i^2\right)}$$

bilan va mos ravishda $S_j \in E_0$ ob'ektning S_i ob'ektga nisbatan shartli ehtimolligi

$$P_{i|j} = \frac{\exp\left(-\rho^2\left(S_i, S_j\right)/\sigma_j^2\right)}{\sum_{k \neq j} \exp\left(-\rho^2\left(S_j, S_k\right)/\sigma_j^2\right)}$$

bilan hisoblanadi. Bu yerda $\rho(S_i,S_j)$ – yevklid metrikasi, $\sigma_j^2 = \sum_{k \neq j} \rho^2\left(S_j, S_k\right)/(m-1)$, $\sigma_i^2 = \sum_{k \neq i} \rho^2\left(S_i, S_k\right)/(m-1)$.

Markazi S_i ob'ektda bo'lgan gauss zichligi $P_{ij} = \frac{P_{i|j} + P_{j|i}}{2}$ ko'rinishida aniqlanadi.

Talab qilinadi::

– E_0 tanlanmada $X(n)$ bo'yicha G_n va $X(l)$ bo'yicha G_l zichlik aniqlansin, $n>l$;

– G_n va G_l orasidagi masofa $D_{kl}\left(G_n, G_l\right) = \sum_{i \neq j} p_{ij}^n \log \frac{p_{ij}^n}{p_{ij}^l}$ ko'rinishidagi Kulbak–Leybler metrikasi orqali hisoblansin.

7.15. (*Метрика Кульбака–Лейблера*).

Дано описание множества объектов $E_0=(S_1,\ldots,S_m)$ с помощью набора количественных признаков $X(n)=(x_1,\ldots,x_n)$. Для анализа структуры отношений объектов используется вычисление условной вероятности объекта $S_i \in E_0$ относительно S_j как

$$P_{j|i} = \frac{\exp\left(-\rho^2\left(S_i, S_j\right)/\sigma_i^2\right)}{\sum_{k \neq i} \exp\left(-\rho^2\left(S_i, S_k\right)/\sigma_i^2\right)}$$

и соответственно $S_j \in E_0$ относительно S_i

$$P_{i|j} = \frac{\exp\left(-\rho^2\left(S_i, S_j\right)/\sigma_j^2\right)}{\sum_{k \neq j} \exp\left(-\rho^2\left(S_j, S_k\right)/\sigma_j^2\right)},$$

где $\rho(S_i,S_j)$ – метрика Евклида, $\sigma_j^2 = \sum_{k \neq j} \rho^2\left(S_j, S_k\right)/(m-1)$, $\sigma_i^2 = \sum_{k \neq i} \rho^2\left(S_i, S_k\right)/(m-1)$. Гауссовская плотность с центром в объекте S_i определяется как $p_{ij} = \frac{P_{i|j} + P_{j|i}}{2}$.

Требуется:

– определить на E_0 гауссовскую плотность G_n по $X(n)$ и G_l по $X(l)$, $n>l$;

– вычислить расстояние Кульбака–Лейблера между G_n и G_l как

$$D_{kl}\left(G_n, G_l\right)=\sum_{i\neq j} p_{ij}^n \log\frac{p_{ij}^n}{p_{ij}^l}.$$

7.16 (*Invariantlik*).

Berilgan $E_0=(S_1,\ldots,S_m)$ to‘plamida ob’ektlar $X(n)=(x_1,\ldots,x_n)$, $n>2$ – miqdoriy alomatlar bilan tavsiflangan.

E_0 tanlanmada berilgan $\rho(x,y)$ metrika uchun chiziqli tartib hosil qilinadi $((i,j)\leq(k,l))\equiv(\rho(S_i,S_j)\leq\rho(S_k,S_l))$. E_0 tanlanmadagi chiziqli tartiblashlar to‘plamini $\Delta(\rho)=\{((i,j)\leq(k,l))\equiv(\rho(S_i,S_j)\leq\rho(S_k,S_l))\}$ bilan belgilaymiz.

Talab qilinadi:

– $\Delta(\rho)$ quvvat hisoblansin;

– $\Delta(\rho^*)$ to‘plam bo‘yicha masofa qiymatiining monoton o‘suvchanligiga nisbatan $\Delta(\rho)$ quvvatni invariantligi isbot qilinsin, bu yerda $\rho^*=\rho/(1+\rho)$ va ρ – Xemming metrikasi.

7.16 (*Инвариантность*).

Объекты выборки данных $E_0=(S_1,\ldots,S_m)$ описываются с помощью набора количественных признаков $X(n)=(x_1,\ldots,x_n)$, $n>2$. Для заданной метрики $\rho(x,y)$ на E_0 индуцируется линейный порядок $((i,j)\leq(k,l))\equiv(\rho(S_i,S_j)\leq\rho(S_k,S_l))$. Обозначим через $\Delta(\rho)=\{((i,j)\leq(k,l))\equiv(\rho(S_i,S_j)\leq\rho(S_k,S_l))\}$ – множество линейных порядков на E_0.

Требуется:

– вычислить мощность $\Delta(\rho)$;

–доказать инвариантность $\Delta(\rho)$ монотонно возрастающим преобразованиям значений расстояний по множеству $\Delta(\rho^*)$, где $\rho^*=\rho/(1+\rho)$ и ρ – метрика Хэмминга.

7.17. (*So’zlar o’zagini ajraitish*). Tabiiy tildagi yozilgan T matn hujjati $m=|T|$ dan iborat jumlalar to’plami bilan berilgan. Jumla tugallashi {.,?,!} to’plam belgisi hisoblanadi. So’z o’zagini ajraib olish $n(n>1)$ ta tabiiy tildagi so’z oxiri bo’lgan $Qush=\{a_i\}_{i\in\{1,\ldots,n\}}$ termasi bo’yicha amalga oshiriladi. Masalan, Q*ush*={*lar, ning, ni, man*}.

Talab qilinadi:

– *T* hujjatdan *Qush* termasiga kiruvchi bilan tugaydigan so’zlar ajratib olinsin;

– T hujjatdan kamida ikkita *Qush* termasiga kiruvchi bilan tugaydigan so'zlar ajratib olinsin;
– ajratib olingan so'zlarning uchrash chastotalari aniqlansin.

7.18 (*Выделение основы слов*). Текстовый документ T на естественном языке задан множеством из $m=|T|$ предложений. Концом предложения является символ из множества {.,?,!}. Выделение основы слов производится по набору $Qush=\{a_i\}_{i\in\{1,\ldots,n\}}$ из $n(n>1)$ окончаний естественного языка. Например, *Qush*={*лар*, *нинг*, *ни*, *ман* }.

Требуется:
– выделить из T слова с окончаниями из *Qush*;
– выделить из T слова, имеющих более одного окончания из *Qush*;
– определить частоту встречаемости выделенных слов в документе T.

7.19 (*Jumlalarni juftli o'xhashligi*). Ikkita S_i va S_j, $i,j\in\{1,\ldots,n\}$ jumlalar so'zlari termalarining juftligi bo'yicha o'xshashlik

$$\mathrm{sim}_{ij}=\frac{\left|\left\{w\middle|\left(w\in S_i\right)\wedge\left(w\in S_j\right)\right\}\right|}{\log\left(\left|S_i\right|\right)+\log\left(\left|S_j\right|\right)}$$

ko'rinishida aniqladi. Bu erda w– jumladagi so'z. Barcha $(i,j)\in\{1,\ldots,n\}$ juftliklar bo'yicha o'xshsashlikni maksimal qiymati aniqlansin.

7.19 (*Попарное сходство предложений*) Вычислить попарное сходство наборов слов из предложений S_i и S_j, $i,j\in\{1,\ldots,n\}$ как

$$\mathrm{sim}_{ij}=\frac{\left|\left\{w\middle|\left(w\in S_i\right)\wedge\left(w\in S_j\right)\right\}\right|}{\log\left(\left|S_i\right|\right)+\log\left(\left|S_j\right|\right)},$$

где w– слово в предложении. Определить максимальное значение сходства по всем парам $(i,j)\in\{1,\ldots,n\}$.

7.20 (*TF-IDF bo'yicha jumlalar o'xshsashlik o'lchovi*). $S=\{S_j\}_{j\in\{1,\ldots,n\}}$ jumlalardan tashkil topgan terma va t parametr aniqlangan. Grafning t bo'sag'adan past bo'lgan qirralari nollashtiriladi.

Jumlalar o'xshashligi quyidagi formula bilan hisoblanadi

$$tf_w^i=\frac{\left|\left\{w_k\middle|\left(w_k\in S_i\right)\wedge\left(w_k=w\right)\right\}\right|}{\left|S_i\right|},$$

$$idf_w=\log\left(\frac{|S|}{\left|\left\{S_i\middle|w\in S_i\right\}\right|}\right),$$

$$sim_{ij} = \frac{\sum_{w \in \{S_i \cup S_j\}} tf_w^{\,i} \cdot tf_w^{\,j} \cdot \left(idf_w\right)^2}{\sqrt{\sum_{w \in S_i} \left(tf_w^{\,i} \cdot idf_w\right)^2} \cdot \sqrt{\sum_{w \in S_j} \left(tf_w^{\,j} \cdot idf_w\right)^2}}.$$

Barcha $(i,j)\in\{1,\ldots,n\}$ juftliklar bo'yicha o'xshashlikning maksimal qiymati aniqlansin.

7.20 (*Мера схожести предложений по* TF-IDF). Определён набор из предложений $\boldsymbol{S}=\{S_j\}_{j\in\{1,\ldots,n\}}$ и параметр t для зануления ребер графа ниже определенного порога t.

Схожесть предложений определяется по следующим формулам

$$tf_w^{\,i} = \frac{\left|\left\{w_k \middle| \left(w_k \in S_i\right) \wedge \left(w_k = w\right)\right\}\right|}{\left|S_i\right|},$$

$$idf_w = \log\left(\frac{|S|}{\left|\left\{S_i \middle| w \in S_i\right\}\right|}\right),$$

$$sim_{ij} = \frac{\sum_{w \in \{S_i \cup S_j\}} tf_w^{\,i} \cdot tf_w^{\,j} \cdot \left(idf_w\right)^2}{\sqrt{\sum_{w \in S_i} \left(tf_w^{\,i} \cdot idf_w\right)^2} \cdot \sqrt{\sum_{w \in S_j} \left(tf_w^{\,j} \cdot idf_w\right)^2}}.$$

Определить максимальное значение сходства по всем парам $(i,j)\in\{1,\ldots,n\}$.

7.21 (*TF(term frequency)-IDF(inverse document frequency) o'lchovi*). Alohida hujjat chegarasida t_i so'zning muhimligini baholash

$$tf(t,d) = \frac{n_t}{\sum_k n_k}$$

formulasi bilan hisoblanadi. Bu erda n_t – t so'zning hujjatga kirishlari soni, maxraj esa ushbu hujjatdagi barcha so'zlar miqdori.

$$idf(t,D) = \log\frac{|D|}{\left|\left\{d_i \in D \middle| t \in d_i\right\}\right|},$$

bu erda $|D|$– kolleksiyadagi jujjatlar soni; $|\{d_i\in D|t\in d_i\}|$ – D kolleksiyadagi t uchraydigan hujjatlar soni ($n_t\neq 0$ holati uchun).

TF-IDF o'lchovi ikkita had ko'paytmasi ko'rinishida hisoblanadi:

$$Tf\text{-}idf(t,d,D) = tf(t,d)\cdot idf(t,D).$$

7.21 (*Мера* TF(*term frequency*)-IDF(*inverse document frequency*)) Оценить важность слова t_i в пределах отдельного документа

$$tf(t,d)=\frac{n_t}{\sum_k n_k},$$

где n_t – число вхождений слова t в документ, а в знаменателе – общее число слов в данном документе.

$$idf(t,D)=\log\frac{|D|}{|\{d_i\in D|t\in d_i\}|},$$

где $|D|$– число документов в коллекции; $|\{d_i \in D|t \in d_i\}|$ – число документов из коллекции D, в которых встречается t (когда $n_t \neq 0$).

Мера TF IDF является произведением двух сомножителей:

$$Tf\text{-}idf(t,d,D)= tf(t,d)\cdot idf(t,D).$$

7.22(*Soxta yangiliklar*). Hujjat yaratuvchilarni tahlil qilish uchun hujjatlar mazmunini tasniflashga imkon beruvchi 2 ta bigrammlar termasi mavjud bo'lsin (*aldayapti/aldamayapti*).

1-terma (*aldayapti)*: {*NUJ foydalanuvchilari, Buxoro metropolisi, Qashqadaryo bug'uchilari, Xorazm muzliklari, Grenlandiya mexanizatori, tibbiy aksioma*}.

2-terma (*aldamayapti*): {*samsa pishirish, ilg'or texnologiyalar, malakali xodim, kasalliklarni davolash, postcovid sindromi, transport kollapsi*}.

Alohida bir hujjat chegarasida *t* bigram muhimligi

$$tf(t,d)=\frac{n_t}{\sum_k n_k}$$

ko'rinishda aniqlalandi. Bu erda n_t – T bigrammning hujjatga kirishlari soni, maxraj – ikkita terma bigrammlarining hujjatdagi umumiy soni.

Quyidagi qoida bo'yicha berilgan hujjat uchun tasniflashni amalga oshirish talab qilinadi:

- agar birinchi termadagi qandaydir bitta bigramm muhimligi ikkinchi termadagi ixtiyoriy bigramm muhimligidan kam bo'lmasa hujjatni soxta xabar deb hisoblash mumkin.

7.22(*Фейковые новости*) Для анализа разработчиков документов имеется 2 набора биграмм, используемых для классификатsiи их содержимого(врут/не врут).

Набор 1*(врут)*: {*пользователи НЛО, мегаполис Бухара, оленеводы Кашкадарьи, ледники Хорезма, механизатор Гренландии, медицинская аксиома*}.

Набор 2 (*не врут*):{*выпечка самсы, прорывные технологии, заслуженный работник, лечение болезней, постковидный синдром, транспортный коллапс*}.

Важность биграмма *t* в пределах отдельного документа определяется как

$$tf(t,d)=\frac{n_t}{\sum_k n_k},$$

где n_t – число вхождений биграмма *t* в документ, в знаменателе – общее число биграмм из двух наборов в документе. Требуется определить классификatsiю для заданного документа по следующему правилу:

если существует хотя бы один биграмм из первого набора, важность которого больше любого или равна важности любого биграмма из второго набора, то следует считать документ фейком.

7.23 (*Soxta yangiliklar*). Jujjat yaratuvchilarni tahlil qilish uchun hujjatlar mazmunini tasniflashga imkon beruvchi 2 ta bigrammlar termasi mavjud bo'lsin (*aldayapti/aldamayapti*).

1-terma (*aldayapti)*: {*xo'rozlar poygalari, Taxtakupir aglomeratsiyasi, jirafa chorvadori, Qoratau muzliklari, chukchi paxtakori, tibbiyot kombinatlari*}

2-terma (*aldamaypti)*: {*somsa pishirish, yig'ish texnologiyasi, ishonchsiz arbob, diabetni davolash, post-covid sindromi, transport echimi*}.

Alohida bir hujjat chegarasida *t* bigram muhimligi

$$tf(t,d)=\frac{n_t}{\sum_k n_k}$$

ko'rinishda aniqlalandi. Bu erda n_t – *T* bigrammning hujjatga kirishlari soni, maxraj – ikkita terma bigrammlarining hujjatdagi umumiy soni.

Birinchi va ikkinchi termasidagi bigrammlar muhimligini qiyoslash talab qilinadi.

7.23 (*Фейковые новости*) Для анализа целей разработчиков документов имеется 2 набора биграмм, используемых для классификatsiи их содержимого(врут/не врут).

Набор 1(врут): {*гонки петухов, агломерatsiя Тахтакупира, пастух жирафов, ледники Каратау, чукча хлопкороб, медицинская комбинаторика* }

Набор 2(не врут): $\{$*выпечка самсы, технология сборки, опальный деятель, лечение диабета, постковидный синдром, транспортная развязка*$\}$.

Важность биграмма t в пределах отдельного документа определяется как

$$tf(t,d)=\frac{n_t}{\sum_k n_k},$$

где n_t – число вхождений биграмма t в документ, в знаменателе – общее число биграмм из двух наборов в документе. Требуется определить соотношение важностей биграмм из первого и второго набора.

Foydalanilgan adabiyotlar (Использованная литература)

1. N.A.Ignatьev, Sh.F. Madraximov R.N. Usmonov. Berilganlarning intellektual tahlili. O’quv qo’llanma.