**N.A.Ignatev**

**Intellektual tizimlar bo‘yicha masalalar to‘plami**

**Uslubiy qo‘llanma**

**Сборник заданий по интеллектуальным системам**

**Методическое пособие**

**Тошкент 2023**

**Mundarija (Оглавнение)**

[Кириш (Введение) 3](#_Toc102909773)

[1.Berilganlarni razvedka tahlili (Разведочный анализ данных) 4](#_Toc102909774)

[2. O’qituvchili klassifikatsiya (Классификaция с учителем) 16](#_Toc102909775)

[3. Берилганларнинг кластер таҳлили (Кластерный анализ данных) 43](#_Toc102909776)

[4. Informativ alomatlarni saralash va tanlash (Отбор и выбор информативных признаков) 51](#_Toc102909777)

[5. Neyron to’rlari (Нейронные сети) 62](#_Toc102909778)

[6. Berilganlarni vizuallashtirish (Визуализaция данных) 69](#_Toc102909779)

[7. Matematik lingvistika (Математическая лингвистика) 71](#_Toc102909780)

[Foydalanilgan adabiyotlar (Использованная литература) 83](#_Toc102909781)

# Kirish (Введение)

Intellektual tizimlar bo‘yicha masalalar to‘plami kompyuterli ko‘rish, nutqni anglash, ko‘p o‘lchamli berilganlarni vizuallashtirish va obrazlarni anglash masalalarini еchishda 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 автоматизatsiи процесса формирования признакового пространства в задачах компьютерного зрения, распознавания речи, визуализaции многомерных данных, распознавания образов.

Практическая реализaция заданий основываются на использовании оперatsiй 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.](https://ru.wikipedia.org/wiki/%D0%90%D0%BD%D0%B3%D0%BB%D0%B8%D0%B9%D1%81%D0%BA%D0%B8%D0%B9_%D1%8F%D0%B7%D1%8B%D0%BA) *Exploratory data analysis, EDA*) — анализ основных свойств данных, нахождение в них общих закономерностей, распределений и аномалий, построение начальных моделей, зачастую с использованием инструментов [визуализatsiи](https://ru.wikipedia.org/wiki/%D0%92%D0%B8%D0%B7%D1%83%D0%B0%D0%BB%D0%B8%D0%B7%D0%B0%D1%86%D0%B8%D1%8F_%D0%B4%D0%B0%D0%BD%D0%BD%D1%8B%D1%85).

Понятие введено математиком [Джоном Тьюки](https://ru.wikipedia.org/wiki/%D0%A2%D1%8C%D1%8E%D0%BA%D0%B8,_%D0%94%D0%B6%D0%BE%D0%BD), 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](https://ru.wikipedia.org/wiki/%D0%A0%D0%B0%D1%81%D0%BF%D1%80%D0%B5%D0%B4%D0%B5%D0%BB%D0%B5%D0%BD%D0%B8%D0%B5_%D0%B2%D0%B5%D1%80%D0%BE%D1%8F%D1%82%D0%BD%D0%BE%D1%81%D1%82%D0%B5%D0%B9) переменных, построение и анализ [корреляционных матриц](https://ru.wikipedia.org/w/index.php?title=%D0%9A%D0%BE%D1%80%D1%80%D0%B5%D0%BB%D1%8F%D1%86%D0%B8%D0%BE%D0%BD%D0%BD%D0%B0%D1%8F_%D0%BC%D0%B0%D1%82%D1%80%D0%B8%D1%86%D0%B0&action=edit&redlink=1), [факторный анализ](https://ru.wikipedia.org/wiki/%D0%A4%D0%B0%D0%BA%D1%82%D0%BE%D1%80%D0%BD%D1%8B%D0%B9_%D0%B0%D0%BD%D0%B0%D0%BB%D0%B8%D0%B7), [дискриминантный анализ](https://ru.wikipedia.org/wiki/%D0%94%D0%B8%D1%81%D0%BA%D1%80%D0%B8%D0%BC%D0%B8%D0%BD%D0%B0%D0%BD%D1%82%D0%BD%D1%8B%D0%B9_%D0%B0%D0%BD%D0%B0%D0%BB%D0%B8%D0%B7), [многомерное шкалирование](https://ru.wikipedia.org/wiki/%D0%9C%D0%BD%D0%BE%D0%B3%D0%BE%D0%BC%D0%B5%D1%80%D0%BD%D0%BE%D0%B5_%D1%88%D0%BA%D0%B0%D0%BB%D0%B8%D1%80%D0%BE%D0%B2%D0%B0%D0%BD%D0%B8%D0%B5).

* 1. Berilgan  ta test savolga  ta javob beruvchilarning javoblari  matritsa ko‘rinishida saqlanadi, agar  bo‘lsa, -chi savolga -chi javob beruvchi to‘g‘ri javob berganini anglatadi va  bo‘lsa noto‘g‘ri javob bo‘ladi. Har bir *k*- savolning  vaznini va - javob beruvchining  nisbiy bahosini hisoblansin:

.

Bu еrda , , .

* 1. Ответы тестируемых на тестовых задач (вопросов) хранятся в виде матрицы , где означает, что -ый тестируемый правильно ответил на -ый вопрос, соответственно неправильно. Вычислить вес каждой -ой задачи как и относительную оценку –го тестируемого как

,

где , ,

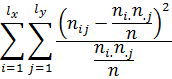
**1.2.** Tekislikda berilgan , nuqtalarni 2 ta kesishmaydigan sinfga har xil ajratishlar sonini hisoblovchi algoritm amalga oshirilsin (dixotomiya masalasi).

**1.2.** Реализовать алгоритм вычисления числа различных разбиений точек, 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 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 korrelyatsiya matritsasi hisoblansin va maksimal qiymatli alomatlar juftligi ajratib ko’rsatilsin.

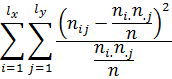
**1.3.** Для описания молекулы генома используются цепочки символов (последовательности) в виде различных комбинatsiй из четырёх букв ‘A’, ‘B’, ‘C’, ‘D’. По заданным m последовательностям, каждая из которых состоит из символов, формируется таблица “*объект-свойство*” T по определяемому перечню признаков: ‘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 и выделить пару признаков с максимальным значением.

**1.4.** Berilgan } matritsada har bir  obyekt  ta nominal alomatlar bilan tavsiflanadi . Har bir  nominal alomatlar juftligi uchun bog‘liqlik  matritsasini quring va



formula bo‘yicha hisoblanadigan kvadratik bog‘liqlik koeffitsiyenti  qiymatini hisoblansin. Bu еrda ; ,  – mos ravishda  va *y* alomatlarning gradatsiyalar soni.

**1.4.** В заданной матрице } каждый объект описывается набором из номинальных признаков . Для каждой пары номинальных признаков требуется построить матрицу сопряжённости и вычислить значения коэффициента квадратичной сопряжённости по формуле

,

где ; , – количество градaций соответственно признаков и *y*.

**1.5.**  obyektdan tashkil topgan va  miqdoriy alomatlar tavsiflangan ob’ektlar to’plami } berilgan. *K* yaqin qo’shni usuli bilan berilgan  obyekt uchun taqsimot zichligi baholansin. Evklid va Chebishev metrikalari bo’yicha hisoblangan taqsimot zichliglari solishtirilsin.

**1.5.** Множество } из объектов описывается набором из n количественных признаков. Оценить плотность распределения для заданного объекта методом ближайших соседей. Сравнить значения плотности по метрике Евклида и Чебышева.

**1.6.**  obyektdan tashkil topgan  o‘lchovli alomatlar fazosida } tanlanma berilgan. Berilgan  kenglik bo‘yicha parzen darchasi usuli bilan berilgan  obyekt uchun taqsimot zichligini baholansin.

**1.6.** Множество } из объектов описывается набором из *n* количественных признаков. Оценить плотность распределения для заданного объекта методом парзеновского окна по заданной ширине .

**1.7.**  va  obyektlar o‘rtasidagi ,  munosabat  kvadratik matritsa ko‘rinishida beriladi, agar  rost bo‘lsa , aks holda .  matritsa uchun *R* munosabat bo’yicha tranzitivlik xossasi bajariladimi? Agar bajarilsa, tranzitivlik xossasi rost bo‘ladigan obyektlar uchligi ko‘rsatilsin.

**1.7.** Отношение , между объектами и задаётся в виде квадратной матрицы . если истинно и если ложно. По матрице определить выполняется ли свойство транзитивности по отношению . Если да, то указать все тройки объектов для которых отношение транзитивности истинно.

**1.8.**  va  obyektlar o‘rtasidagi ,  munosabat  kvadratik matritsa ko‘rinishida beriladi, agar  rost bo‘lsa , aks holda . matritsa uchun *R* munosabat xossasi ekvivalentlik sinfini hosil qiladimi?

**1.8.** Отношение , между объектами и задаётся в виде квадратной матрицы . если истинно и если ложно. Определить образуют ли свойства отношения класс эквивалентности.

**1.9.**  va  obyektlar o‘rtasidagi ,  munosabat  kvadratik matritsa ko‘rinishida, ,  munosabat esa  ko‘rinishda berilgan, agar  rost bo‘lsa , aks holda . Xuddi shunday *B* matritsa elementlari  munosabat bo‘yicha aniqlanadi. Aniqlansin,  munosabatlarning  o‘rami qat’iy tartiblangan sinf hosil qiladimi?

**1.9.** Отношение , между объектами и задаётся в виде квадратной матрицы , отношение , в виде матрицы . если истинно и если ложно. Аналогично определяются элементы матрицы по отношению . Определить, образует ли свёртка отношений класс строгий порядок.

**1.10**. Uchta ekspertlarning va obektlar o’rtasidagi munosabat bo’yicha , xulosalari , и ko’rinishidagi matritsalar bilan berilgan. Xulosalari bir-biriga eng yaqin bo’lgan ikkita ekspert aniqlansin.

**1.10.** Заключения трёх экспертов по отношению , , между объектами и задаётся в виде квадратных матриц. Считается, что если истинно и если ложно. Определить двух экспертов, мнения которых наиболее близки друг к другу.

**1.11.** va obyektlar o‘rtasidagi ,   munosabat kvadratik matritsa ko‘rinishida, ,  munosabat esa ko‘rinishda berilgan, agar rost bo‘lsa aij=1, aks holda . Xuddi shunday matritsa elementlari munosabat bo‘yicha aniqlanadi. Aniqlansin, munosabatlar o‘rami antirefleksivlik xossasiga egami?

**1.11.** Отношение , между объектами и задаётся в виде квадратной матрицы , отношение , в виде матрицы . если истинно и если ложно. Аналогично определяются элементы матрицы по отношению . Определить, обладает ли свёртка отношений свойством антирефлексивности.

**1.12.** Berilgan  tanlanma bo‘yicha ,  obyektlarni  fazodan sonlar o‘qiga

.

Formula bilan akslantirilsin. Koeffitsiyentlar sifatida

qiymatlaridan foydalanilsin. Bu еrda , – mos ravishda  va  obyektlar orasidagi evklid metrikasi bo‘yicha  va  fazodagi masofa (- alomat o’chirilgan jadval bo’yicha).

**1.12.** По заданной выборке произвести отображение описаний объектов , из на числовую прямую по формуле

.

В качестве коэффициентов использовать значения

где , расстояние по евклидовой метрике между объектами и соответственно в и (по таблице из которой удалён -й признак).

**1.13.**  ta miqdoriy alomat bilan tavsiflanadigan har bir  obyekt uchun } tanlanma bo‘yicha tasodifiy  va  kattaliklar guruhlar o‘rtasidagi to‘plamli korrelatsiya koeffitsiyentini hisoblash dasturi tuzilsin va u



formuladan aniqlanadi, bu еrda  –  kattaliklarning korrelatsiya matritsa aniqlovchisi,  ham xuddi shunday, faqat u  qiymatlar uchun.

**1.13.** По заданной выборке } в которой каждый объект описывается количественными признаками из набора реализовать вычисление значения коэффициента множественной корреляции между случайной величиной и группой случайных величин как

,

где – определитель матрицы корреляций величин , то же самое, но только для величин

**1.14.** Satrlari bo‘yicha *n* miqdoriy alomatlar bilan tavsiflangan *m* obyektlar joylashgan  jadval berilgan bo‘lib, alomatlardan biri (tanlash orqali) maqsad, qolganlari bog‘liq alomatlar hisoblanadi.  – maqsad va ,…, bog‘liq vektorlar qiymatlari  funksional bilan bog‘langan bo‘lsin.

Bog‘liq ko‘rsatkichlar vektori  bo‘yicha prognoz qilinuvchi  qiymati  ko‘rinishida hisoblanadi. Bu еrda  –  bog‘liq alomatlar bo‘yicha berilgan  metrika bilan hisoblangan *k* yaqinlarning maqsad ko‘rsatkich qiymatlari. Quyidagi mezon bo‘yicha sirpanuvchi usul bilan *k* parametrining optimal qiymati tanlanadi:



Bu еrda  – prognoz qilinuvchi qiymat.

**1.14.** В заданной таблице по строкам которой содержатся *m* объектов, описываемых *n* количественными признаками, один из признаков (по выбору) считается целевым, остальные зависимыми. Считается, что значения целевого вектора и зависимых ,…, связаны функциональной зависимостью

Прогнозируемое значение по вектору зависимых показателей вычисляются как

где – значения целевых показателей у *k* ближайших по заданной метрике  векторов зависимых признаков .

Оптимальное значение *k* выбирается методом скользящего экзамена по критерию

где – прогнозируемое значение.

**1.15.** Berilgan *A*={*auv*}*m\*n* jadvalida *m* ta *S*1,…,*Sm* ob’ektlar to‘plamini *X*(*n*)=(*x*1,…,*xn*) miqdoriy alomatlar bilan tavsifi berilgan. Maqsad alomat sifatida *xi*∈*X*(*n*) alomat tanlangan bo‘lib, *X*(*n*-1)= *X*(*n*)\{*xi*} alomatlar bog‘liq hisoblanadi. *S*1,…,*Sm* ob’ektlar *y*1,…,*ym* sifatida identifikatsiya qilinuvchi maqsad alomatlari uchun *yi*=*F*(*X*(*p*)) ko‘rinishidagi funksional bog‘liqlik aniqlangan, bu еrda *X*(*p*)*X*(*n*-1), 2≤*p*≤*n*-1. Prognoz qilinuvchi *y* alomatning *X*(*p*) bog‘liq alomatlar bo‘yicha qiymati



ko‘rinishida hisoblandi, bu еrda – *X*(*p*) alomatlar bo‘yicha evklid metrikasi bo‘yicha eng yaqin *k* ta  ob’ektlar maqsad alomatlarining qiymatlari.

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



mezon bo‘yicha tanlanadi, bu еrda *yi*(*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:



**1.15.** Таблица *A*={*auv*}*m\*n* содержит описания множества из *m* объектов *S*1,…,*Sm* по набору количественных признаков *X*(*n*)=(*x*1,…,*xn*). Признак *xi*∈*X*(*n*) выбран в качестве целевого, *X*(*n*-1)= *X*(*n*)\{*xi*} являются зависимыми. Для значений целевого признака объектов *S*1,…,*Sm*, идентифицируемых как *y*1,…,*ym*, определена функциональная зависимость *yi*=*F*(*X*(*p*)), где *X*(*p*)*X*(*n*-1), 2≤*p*≤*n*-1. Прогнозируемое значение *y* по значениям набора зависимых признаков *X*(*p*) вычисляются как



где – значения целевых показателей у *k* ближайших по евклидовой метрике объектов  на наборе *X*(*p*).

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



где *yi*(*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



**1.16.** Berilgan *A*={*auv*}*m\*n* jadvalida *m* ta *S*1,…,*Sm* ob’ektlar to‘plamini *X*(*n*)=(*x*1,…,*xn*) miqdoriy alomatlar bilan tavsifi berilgan. Har bir (*xi*, *xj*)∈*X*(*n*) alomatlar juftligi bo‘yicha aniqlash talab etiladi:

– mumkin bo‘lgan intervallar [*z*1,*z*2] chegarasini, bu еrda, , , *R*(*i*,*j*)={*aki*/*Pi* – *akj*/*Pj* }*k*∈{1,…,*m*}, *Sk*=(*ak*1,…,*akn*), *Pi*,*Pj* – (*xi*,*xj*) 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*={*auv*}m\*n задана выборка из множества объектов *E*0={*S*1,…,*Sm*}, описываемых набором количественных признаков *X*(*n*)=(*x*1,…,*xn*). По каждой паре признаков (*xi*, *xj*)∈*X*(*n*) требуется определить:

– границы допустимых интервалов [*z*1,*z*2], по

*R*(*i*,*j*)={*aki*/*Pi* – *akj*/*Pj*}*k*∈{1,…,*m*},

где *Sk*= (*ak*1,…,*akn*), *Pi*,*Pj* – математическое ожидание значений признаков (*xi*, *xj*) на множестве *E*0, , ;

– номера объектов из *E*0, со значениями границ интервалов *z*1 и *z*2.

**1.17.** Berilgan *A*={*auv*}*m\*n* jadvalida *m* ta *S*1,…,*Sm* ob’ektlar to‘plamini *X*(*n*)=(*x*1,…,*xn*) miqdoriy alomatlar bilan tavsifi berilgan. Har bir (*xi*, *xj*)∈*X*(*n*) alomatlar juftligi bo‘yicha aniqlash talab etiladi:

– mumkin bo‘lgan intervallar [*z*1,*z*2] chegarasini, bu еrda, , , *R*(*i*,*j*)={*aki*/*Pi* – *akj*/*Pj* }*k*∈{1,…,*m*}, *Sk*=(*ak*1,…,*akn*), *Pi*,*Pj* – (*xi*,*xj*) 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*={*auv*}m\*n задана выборка из множества объектов *E*0={*S*1,…,*Sm*}, описываемых набором количественных признаков *X*(*n*)=(*x*1,…,*xn*). По каждой паре признаков (*xi*, *xj*)∈*X*(*n*) требуется определить:

– границы допустимых интервалов [*z*1,*z*2], по

*R*(*i*,*j*)={*aki*/*Pi* – *akj*/*Pj* }*k*∈{1,…,*m*},

где *Sk*= (*ak*1,…,*akn*), *Pi*,*Pj* – значения медиан признаков (*xi*, *xj*) на множестве *E*0, , ;

– номера объектов из *E*0, со значениями границ интервалов *z*1 и *z*2.

**1.18.** Berilgan *A*={*aij*}*mn* jadvalda *E*0 = (*S*1,…,*Sm*) ob’ektlarning *n* miqdoriy alomatlar bo‘yicha tavsifi keltirilgan. Ikkita *Su*,*Sv*∈*E*0 ob’ektlar o‘rtasidagi masofa ρ(*x*,*y*) еvklid metrikasi bilan hisoblanadi. Tanlanmaning *S* ob’ektining *k* ta eng yaqin qo‘shnilarini *Nk*(*S*) bilan, *k*–*dis*(*S*) orqali S ob’ektdan *k* – yaqin qo‘shnigacha bo‘lgan masofani belgilaylik.

*Sv* ob’ektdan *Su* ob’ektgacha erishiluvchi masofa *RDk*(*Su*,*Sv*)= max{*k*–*dis*(*Sv*), ρ(*Su,Sv*)} sifatida aniqlanadi (1.a-rasmga qarang).

*Su* ob’ektga erishuvchanlikning lokal zichligi



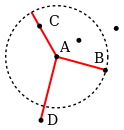
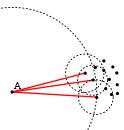
formula bilan aniqlanadi va *Su* ob’ektga uning qo‘shnilaridan erishish masofalarining o‘rta arifmetikiga nisbatan teskari hisoblanadi. Erishishning lokal zichligi qo‘shnilarning erishish lokal zichligi bilan



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

Agar *LOFk*(*Su*), 1 yaqin yaqin qiymat teng bo‘lsa, *Su* 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 еtarlicha katta bo‘lgan qiymatlar ob’ektning anomalligidan guvoh beradi.

Berilgan *E*0 tanlanmaning *k*=3,5,7 holatlar uchun barcha anomal ob’ektlari aniqlansin.

а) b)

1-rasm. Anomal ob’ektlarni aniqlash

(1-рис. Оbнaружeниe aномaльных оbъeктов)

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*={*aij*}*m\*n* содержатся описание объектов *E*0=(*S*1,…,*Sm*) по *n* количественным признакам. Расстояние между объектами *Su*,*Sv*∈*E*0 вычисляется по метрике Евклида ρ(*x*,*y*). Обозначим через *Nk*(*S*) – множество из *k* ближайших соседей к объекту *S* и *k-dis*(*S*) – расстояние от *S* до *k-*го ближайшего соседа. Достижимое расстояние объекта *Su* из *Sv* определяется как *RDk*(*Su*,*Sv*)= max{*k-dis*(*Sv*), ρ(*Su,Sv*)} (см. рис.1.а).

Локальная плотность достижимости объекта *Su* определяется как



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

,

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

Значение *LOFk*(*Su*), примерно равное 1, означает, что объект *Su* 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, свидетель-ствуют о выбросах.

В рис. 1.а 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

*ω1(t1),…,ωg(tg),…,ωn(tn), ti ∊{1,…,n}*

ketma-ketligi aniqlangan. Bu еrda *ωg(tg)* элементning g indeksi alomat rangi sifatida qaraladi.

*K* ta (*k*<*n*) o’lchangan alomat bilan tavsiflangan S obyekt haqidagi ma’lumot *α*(*S*)=(*α1,…, αi,…, αn*) xarakteristika vektori ko’rinishida berilgan bo’lib, agar *αi* =0 – alomat o’lchanmaganligini, *αi* =1 – o’lchanganligini anglatadi.

O’lchangan *k* alomatlar rangini *D*(*S,k)* bilan belgilaylik. Yig’indining qiymatlar diapazoni *k*(1+*k*)/2 ≤ *D*(*S*,*k*) ≤ *k*(2*n*-*k*+1)/2 oraliqda yotadi.

Talab qilinadi:

– *S* obektning informativlik o’lchami



formula bilan hisoblansin/ Bu еrda *α*= *k*(*1+k*)/2, *β*= *k*(2*n-k+1*)/2;

– o’lchamga лингвистиk 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

*ω1(t1),…,ωg(tg),…,ωn(tn), ti ∊{1,…,n}*

Индекс *g* элемента *ωg(tg)* интерпретируется как ранг признака.

Информaция об объекте *S* с *k* измеренными значениями признаков (*k*<*n*) задаётся в виде характеристического вектора *α*(*S*)=(*α1*,…, *αi*,…, *αn*), в котором *αi* =0 – признак неизмерен, *αi* =1 – измерен.

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

Требуется:

– вычислить меру информативности объекта *S* как



где *α*= *k*(1+*k*)/2, *β*= *k*(2*n*-*k*+1)/2;

– дать лингвистическую интерпретaцию меры: (0;0.5] – “*плохой*”; (0.5;0.8] – “*удовлетворительный*”; (0.8;1] – “*хороший*”.

**1.20.** (“*Sachratqilar”*). Berilgan *E*0=(*S*1,…,*Sm*) tanlanma ob’ektlari *X*(*n*)=(*x*1,…,*xn*), *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,…,*Sm*) с помощью набора количественных признаков *X*(*n*)=(*x*1,…,*xn*). 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 (Классификaция с учителем)

**2.1.** Berilgan } tanlanma obyektlari 2 ta kesishmaydigan  va  sinf vakillaridan iborat. Har bir  obyekt  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  va sozlovchi orttirma  ning boshlang‘ich qiymatlari berish talab qilinadi. Agar - qadamda  va  bo‘lsa, u holda . Agar  va  bo‘lsa, u holda .

**2.1.** Выборка объектов }, содержит представи-телей двух непересекающихся классов и . Каждый объект описывается количественными признаками. Используя принцип “*подкрепления – наказания*” определить границу между двумя классами с помощью итеративного алгоритма (алгоритма перцептрона). Требуется задать начальное значение вектора весов и – корректирующее приращение. Если на -ом шаге итерatsiи и , то . Если и , то .

**2.2.** O‘zaro kesishmaydigan *l* ta  sinflarga bo‘lingan } o‘rgatuvchi tanlanma bo‘yicha eng yaqin qo‘shni algoritmi uchun optimal  ning qiymatini hisoblash algoritmini amalga oshiruvchi dastur tuzilsin. Har bir  obyekt  ta miqdoriy alomat bilan tavsiflanadi. Obyektlar orasida yaqinlik o‘lchovi sifatida evklid metrikasidan foydalanilsin.

**2.2.** Реализовать алгоритм вычисления оптимального значения ближайших соседей по обучающей выборке }, разделённой на *l* непересекающихся классов . Считается, что каждый объект описывается количественными признаками. В качестве меры близости между объектами использовать евклидову метрику.

**2.3.** O‘zaro kesishmaydigan *l* ta  sinflarga bo‘lingan } o‘rgatuvchi tanlanma bo‘yicha obyektlar turg‘unligi qiymatlarini hisoblovchi dastur tuzilsin. Har bir  – obyekt  ta miqdoriy alomat bilan tavsiflanadi va masofani topish uchun Chebishev metrikasidan foydalanilsin.  sinfdagi  obyektning  turg‘unligi:

**

bo‘yicha topiladi, bu еrda,  –  tanlanmadagi  obyektning *k* ta  eng yaqin qo‘shnilarining ko’pchiligi sinfdan bo‘lgandagi holatlar soni.

**2.3.** Реализовать вычисления значений устойчивости объектов обучающей выборке }, разделённой на *l* непересекающихся классов . Считается, что каждый объект описывается количественными признаками и для определения расстояния используется метрика Чебышева. Устойчивость  объекта  в классе  определяется как

*,*

где  - число событий в *E0,* когда среди *k* ближайших к  объектов,  большинство составляют объекты из .

**2.4.** O‘zaro kesishmaydigan *l* ta  sinflarga bo‘lingan } o‘rgatuvchi tanlanma bo‘yicha Evklid metrikasidan foydalangan holda chiziqli qobiqni hisoblash algoritmi dasturi tuzilsin. Har bir  – obyekt  ta miqdoriy alomat bilan tavsiflanadi. Chiziqli qobiq obyektlari  formuladan aniqlanadi, bu еrda  –  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 по метрике Евклида. Считается, что представители классов заданы в выборке и каждый объект описывается с помощью количественных признаков. Объекты линейной оболочки определяются с помощью формулы , где – окрестность объекта , содержащая все ближайшие объекты одного с ним класса до первого объекта из противоположного класса.

**2.5.** O‘zaro kesishmaydigan *l* ta  sinflarga bo‘lingan } o‘rgatuvchi tanlanma bo‘yicha Chebishev metrikasidan foydalangan holda chiziqli qobiqni hisoblash algoritmi dasturi tuzilsin. Har bir  – obyekt  ta miqdoriy alomat bilan tavsiflanadi. Chiziqli qobiq obyektlari  formuladan aniqlanadi, bu еrda  –  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 по метрике Чебышева. Считается, что представители классов заданы в выборке и каждый объект описывается с помощью количественных признаков. Объекты оболочки определяются с помощью формулы , где – окрестность объекта , содержащая все ближайшие объекты одного с ним класса по метрике Чебышева до первого объекта из противоположного класса.

**2.6.** O‘zaro kesishmaydigan *l* ta  sinflarga bo‘lingan } o‘rgatuvchi tanlanma bo‘yicha Juravlev metrikasidan foydalangan holda chiziqli qobiqni hisoblash algoritmi dasturi tuzilsin.

Har bir obyekt ta turli toyifadagi alomatlar bilan tavsivlangan. Juravlyov metrikasi quyidagi formuladan hisoblanadi:



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

Chiziqli qobiq obyektlari  formuladan aniqlanadi, bu еrda  –  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 по метрике Журавлёва. Считается, что представители классов заданы в выборке и каждый объект описывается с помощью разнотипных признаков. Метрика Журавлёва вычисляется по формуле



где и множество индексов соответственно количественных и номинальных признаков. Объекты оболочки определяются с помощью формулы , где – окрестность объекта , содержащая все ближайшие объекты одного с ним класса по метрике Журавлёва до первого объекта из противоположного класса.

**2.7.** O‘zaro kesishmaydigan *l* ta  sinflarga bo‘lingan } o‘rgatuvchi tanlanmada quyidagi mezon bo‘yicha miqdoriy alomatlar qiymatlarini kesishmaydigan intervallarga optimal bo‘luvchi algoritmni amalga oshirilsin:



bu еrda – butun sonlardan tashkil topgan vektor va uning elementlari quyidagi shartlarni qanoatlantiradi:  va obyektlar nomeri miqdoriy alomatlar qiymatlarini o‘sish tartibida beriladi,  – alomatning – intervaldagi  sinfdagi qiymatlari miqdori.

**2.7.** На заданной выборке }, содержащей представителей непересекающихся классов объектов реализовать алгоритм оптимального разбиения значений количественных признаков на непересекающиеся интервалы по критерию

,

где  - целочисленный вектор, элементы которого удовлетворяют условиям:  и задают номера объектов упорядоченных по возрастанию значений количественного признака, – количество значений признака из класса в -ом интервале.

**2.8.** O‘zaro kesishmaydigan *l* ta  sinflarga bo‘lingan } o‘rgatuvchi tanlanma bo‘yicha Evklid metrikasidan foydalanib “*eng yaqin qo‘shni*” anglash algoritmi amalga oshirilsin. Har bir  – obyekt  ta miqdoriy alomat bilan tavsiflanadi.

**2.8.** На заданной выборке }, содержащей представителей непересекающихся классов объектов реализовать алгоритм распознавания по правилу “*ближайший сосед*”. Считается, что каждый объект описывается с помощью количественных признаков и для вычисления расстояния используется евклидова метрика.

**2.9.** O‘zaro kesishmaydigan *l* ta  sinflarga bo‘lingan } o‘rgatuvchi tanlanma berilgan. Har bir  – obyekt  ta nominal alomat bilan tavsiflanadi. 2 ta alomat orqali  sinfda mavjud va boshqalarda bo‘lmagan barcha mumkin bo‘lgan qiymatlar kombinatsiyalari topilsin (mantiqiy qonuniyat).

**2.9.** Выборка объектов }, содержит представителей классов . Каждый объект описывается номинальными признаками. Из всевозможных комбинatsiй по 2 признака требуется определить те, которые есть у одного класса и нет у представителей других классов (логическая закономерность).

**2.10.** } tanlanma obyektlari 2 ta  kesishmaydigan sinf vakillaridan iborat.



mezon bo‘yicha miqdoriy alomatlar ustunlik intervalini hisoblaydigan algoritmni amalga oshirilsin. Bu еrda , ,  – mos ravishda miqdoriy alomatlarning o‘sish tartibidagi  va  o‘rinlar oralig‘idagi , sinflar vakillarining miqdori.

**2.10.** Выборка объектов } содержит представителей двух непересекающихся классов . Каждый объект описывается с помощью количественных признаков. Реализовать алгоритм вычисления интервалов доминирования значений количественных признаков по критерию



где , , число представителей классов , в упорядоченной по возрастанию последовательности значений количественного признака с *u*-й по -ю позицию.

**2.11.** Kesishmaydigan 2 ta  va  sinfga bo‘lingan  o‘rgatuvchi tanlanma berilgan. Har bir  – obyekt  ta miqdoriy alomat bilan tavsiflanadi. Obyektlarni ajratish uchun



diskriminant funksiyadan foydalaniladi, bu еrda  -  intervalda berilgan vazn, ,  vektori



bo‘yicha aniqlanadi.

**2.11.** Выборка объектов }, содержит представителей двух непересекающихся классов и . Каждый объект описывается количественными признаками. Для разделения объектов используется дискриминантная функция

где , – заданные значения весов в интервале . Определить значения элементов вектора , при которых Сравнить значения результатов при выборе разных начальных приближений *T*.

**2.12.** } tanlanma obyektlari 2 ta kesishmaydigan  va  sinf vakillaridan iborat. Har bir  obyekt  har xil toifali alomatlar bilan tavsiflanadi.  –  tanlanmaning  qismi hisoblanadi, , agar -chi alomat klassifikatsiyada qatnashsa , aks holda . Berilgan  vektor bo‘yicha  ilojsiz obyektlar ( va  sinfda ajralmaydigan obyektlar) aniqlansin.

**2.12.** Выборка объектов } содержит представителей двух непересекающихся классов и . Каждый объект описывается разнотипными признаками . является частью выборки , определяемой характеристическим вектором , если признак используется для классификaции и если не используется. По заданным значениям элементов вектора определить не образуются с их помощью на тупики – неразличимые объектов из классов и .

**2.13.** Mumkin bo‘lgan obyektlarning ikkita o‘zaro kesishmaydigan  sinflarga bo‘lingan  tanlanma berilgan. Tanlanma obyektlari *n* ta  miqdoriy alomatlar bilan tavsiflangan.  tanlanmada – Evklid metrikasi berilgan bo‘lsin.

Tanlanmadagi  obyektning turg‘unligi   bo‘yicha hisoblash uchun foydalaniladi:

* o‘zaro munosabati  tengsizlik bilan aniqlangan  tanlanma obyektlarining  tartiblangan ketma-ketligi;
*  formula bilan aniqlanuvchi funksional qiymati. Bu еrda  – mos ravishda  to‘plamdagi  sinflarga tegishli obyektlar soni.

Berilgan  alomatlar to‘plami bo‘yicha  obyekt turg‘unligi hisoblansin.

**2.13.** Рассматривается множество (выборка)  допустимых объектов, разбитая на  непересекающихся подмножества (класса) . Объекты выборки описываются с помощью  количественных признаков . На множестве объектов  определена Eвклидова метрика .

Для вычисления устойчивости объекта  по набору признаков ,  используется:

* упорядоченная последовательность объектов , , отношения между которыми определяются неравенствами вида ;
* значение функционала , где – число объектов в  соответст-венно из класса .

По заданному набору признаков  определить значение устойчивости объекта .

**2.14.** O‘zaro kesishmaydigan *l* ta  sinflar vakillarini o‘z ichiga olgan  o‘rgatuvchi tanlanma berilgan. Har bir  obyekt *n* ta  miqdoriy alomatlar bilan tavsiflangan. Berilgan  metrika bo’yicha chegaraviy obyektlar to’plam ostisi sifatida qaraluvchi shovqin obyektlar to’plami *B*⊂*E*0 quydagicha aniqlanadi

**

Tanlanmaning *S*∊*B*∩*Kj*, *j*=1,…,*l* obyekti *Kj* sinfning *Dj* shovqin obyektlar to’plamiga tegishli deyiladi, agar



shart o’rinli bo’lsa.

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

**2.14.** Обучающая выборка содержит представителей непересекающихся классов . Каждый объект описывается набором из количественных признаков . Множество шумовых объектов рассматривается как подмножество граничных объектов классов *B*⊂*E*0 по заданной метрике , котороеопределяется как

**.

Объект *S*∊*B*∩*Kj*, *j*=1,…,*l* принадлежит множеству шумовых объектов *Dj* класса *Kj*, если



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

**2.15.** O‘zaro kesishmaydigan *l* ta  sinflar vakillarini o‘z ichiga olgan  o‘rgatuvchi tanlanma berilgan. Har bir  obyekt *n* ta  miqdoriy alomatlar bilan tavsiflangan. Obyektlar o‘rtasidagi masofa  metrikasi bilan o‘lchanadi. obyekt  metrika bo‘yicha chegaraviy obyektlar to‘plami *L* tegishli bo‘ladi, agar  bilan aniqlanuvchi shunday  mavjud bo‘lsa. Ikkita  obyektlar *R* munosabat bog‘langan, ya’ni  rost bo‘ladi, agar mavjud bo‘lsaki, uning uchun  va  o‘rinli bo‘lsa. Bu еrda  va .

Har bir  sinf uchun *R* munosabat bo‘yicha Evklid, Xemming va Chebishev metrikalaridan foydalangan holda o‘zaro kesishmaydigan …,, , guruhlarga ajratilsin. Har bir  juftlik uchun  rost bo‘lgan yo‘l ( dan obyektlar zanjiri) mavjud deb hisoblanadi.

**2.15.** Выборка объектов содержит представителей непересекающихся классов . Каждый объект описывается набором из количественных признаков . Расстояние между объектами определяется по метрике . Объект принадлежит к множеству граничных объектов по метрике , если существует такой , что . Объекты связаны отношением , т.е. истинно, если существует , и где и . Требуется по отношению объекты каждого класса разбить на непересекающиеся группы …,, , используя в качестве меры близости метрики Эвклида, Хэмминга, Чебышева. Считается, что для каждой пары существует путь (цепочка объектов из ), для которого истинно.

**2.16.** Yechimlar daraxtini qurish uchun ikkita o‘zaro kesishmaydigan  sinflarga ajratilgan obyektlar  tanlanmasidan foydalaniladi. Har bir  obyekt *n* ta nominal alomatlar  bilan tavsiflangan bo‘lib, *p* gradatsiyaga ega  alomat  to‘plamidan qiymat qabul qiladi. Boshlang‘ich  tanlanmaning  to‘plam ostilariga bo‘linishi aniqlagan bo‘lsin.  sinfga tegishli obyektning  to‘plam ostida paydo bo‘lish ehtimolligi  formulasi bilan aniqlanadi. Bu еrda  *T* to‘plam ostisidagi sinf obyektlari soni. Sinfdagi obyektni aniqlash uchun zarur ma’lumotlar miqdori quyidagicha hisoblanadi.

 to‘plam ostisi bo‘yicha  gradatsiyali  alomatning bahosi

orqali hisoblanadi. Bu еrda  –  gradatsiyali  kiruvchi obyektlar to‘plami.  shartida daraxt ildiziga joylashtirish uchun

Bo‘yicha   alomat topilsin.

**2.16.** Для построения дерева решений используется выборка объектов, разделённая на 2 непересекающихся класса . Каждый объект описывается набором из номинальных признаков . Признак c *p* градaциями принимает значения из множества . На определяется разбиение на подмножества . Вероятность появления объекта класса в подмножестве определяется как где количество объектов класса в . Среднее количество информatsiи, необходимое для определения объекта в классе

Оценка признака с градaциями по вычисляется как

где – множество объектов из с градaцией . Определить признак для размещения в корень дерева при , для которого

**2.17.** Yechimlar daraxtini qurish uchun ikkita o‘zaro kesishmaydigan  sinflarga ajratilgan obyektlar  tanlanmasidan foydalaniladi. Har bir  obyekt *n* ta miqdoriy alomatlar  bilan tavsiflangan. Har bir  alomat uchun aniqlanuvchi  to‘plam bo‘yicha  intervallarga bo‘lishdagi  chegara

(+)

mezoni bo‘yicha hisoblanadi. Bu еrda ,  –  intervaldagi  sinf vakillarining soni, , – alomatining qiymatlari  intervalda bo‘lgan  to‘plamdagi obyektlar soni.

 shartida daraxt ildiziga joylashtirish uchun

Bo‘yicha   alomat topilsin.

**2.17.** Для построения дерева решений используется выборка объектов, разделённая на 2 непересекающихся класса . Объект описывается набором из количественных признаков . Для каждого признака граница разбиения на интервалы по определяемому множеству вычисляется по критерию

(+)

где число представителей класса в , со значениями признака в Требуется определить признак для размещения его в корень дерева при , для которого

**2.18.** Berilgan  ob’ektlar to‘plami ikkita o‘zaro kesishmaydigan *K*1 va *K*2 sinflarga bo‘lingan. To‘plamning *Su*∈*E*0 ob’ekti *X*(*n*)=(*x*1,…,*xn*) nominal alomatlar to‘plami bilan tavsiflanadi. Har bir *xi*∈*X*(*n*) nominal alomatning mumkin bo‘lgan qiymatlari (gradasiyalari) 1,2,…,*pi*, 2≤*pi*≤*m*/2 sonlari bilan beriladi. Nominal *xi* alomatning *μ*∈{1,2,…,*pi*} gradasiyalari bo‘yicha *K*1 sinfga tegishlilik funksiyasi qiymati quyidagi ko‘rinishda hisoblanadi

.

Bu еrda *d*1μ (*d*2μ) – *xi* alomatining qiymati *μ* bo‘lgan *K*1(*K*2) sinf ob’ektlari soni (to‘plami).

Yuqorida keltirilgan formula  matritsadagi  ob’ektlari tavsifidagi nominal alomatlar qiymatlarini o‘girishda foydalaniladi. Har bir  ob’ekt jadvalda  ko‘rinishdagi satr bilan yoziladi, bu еrda .

Sirpanuvchi nazorat usuli bilan  tanlanmada yaqin qo‘shni qoidasi bo‘yicha anglash aniqligi hisoblansin. Usulni amalga oshirishda ketma-ket ravishda har bir  ob’ekt nazorat, – o‘rgatuvchi sifatida taqdim etiladi. *A* jadvalidagi ob’ektlari tavsifi uchun masofa o‘lchovi sifatida Juravlev metrikasidan foydalanilsin.

**2.18**. Множество объектов разделено на два непересекающихся класса *K*1 и *K*2. Для описания объектов используется набор номинальных признаков *X*(*n*)=(*x*1,…,*xn*). Множеством допустимых значений (градaций) каждого признака *xi*∈*X*(*n*) является 1,2,…,*pi*, 2≤*pi*≤*m*/2. Значение функции принадлежности *fi*(μ) к классу *K*1 по градaции μ∈{1,2,…,*pi*} признака *xi* вычисляется как

,

где *d*1μ (*d*2μ) – множество объектов класса *K*1 (*K*2), значением признака *xi* у которых является μ. Выше изложенная формула используется для преобразования исходных описаний объектов *E*0 к таблице . Объект  запишется в виде строки таблицы как , в которой . Требуется методом скользящего экзамена определить точность распознавания на *E*0 по правилу ближайший сосед. При реализatsiи метода каждый объект *Sj*∈*E*0 последовательно представляет контроль, *E*0\{*Sj*} – обучение. В качестве меры расстояния между описаниями объектов в *A* используется метрика Журавлёва.

**2.19.** Berilgan *E*0={*S*1,…,*Sm*} ob’ektlar to‘plami ikkita o‘zaro kesishmaydigan *K*1 va *K*2 sinflarga bo‘lingan. To‘plamning *Su*∈*E*0 ob’ekti *X*(*n*)=(*x*1,…,*xn*) turli toifadagi alomatlar to‘plami bilan tavsiflanadi. Alomatlar qiymatlari *A*={*auv*}*m\*n* matritsa satrini hosil qiladi. *X*(*n*) to‘plamda *β* ta alomatlar nominal, *n* – *β* tasi miqdoriy shkalalarda o‘lchanadi. Har bir *xi*∈*X*(*n*) nominal alomatning mumkin bo‘lgan qiymatlari (gradasiyalari) 1,2,…,*pi*, 2≤*pi*≤*m*/2 sonlari bilan beriladi. Nominal *xi* alomatning *μ*∈{1,2,…,*pi*} gradasiyalari bo‘yicha *K*1 sinfga tegishlilik funksiyasi qiymati quyidagi ko‘rinishda hisoblanadi

.

Bu еrda *d*1μ (*d*2μ) – *xi* alomatining qiymati *μ* bo‘lgan *K*1(*K*2) sinf ob’ektlari soni (to‘plami).

Yuqorida keltirilgan formula *A*={*auv*}*m\*n* matritsadagi *E*0 ob’ektlari tavsifidagi nominal alomatlar qiymatlarini o‘girishda foydalaniladi. Agar *Su*=(*au*1,…,*aun*) (*Su*∈*E*0) ob’ektda *auv* qiymat *xv*∈*X*(*n*) alomat gradasiyasi bo‘lsa, u holda *auv*=*fv*(*auv*). Har bir *Ki*, *i*=1,2 sinflar uchun *A* matritsa bo‘yicha *Mi*=(*bi*1,…,*bin*) 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. *Mi*, *i*=1,2 etalonlar va *A* tavsiflangan *Su*∈*E*0 ob’ekt o‘rtasidagi masofa o‘lchovi sifatida  – Kanberra metrikasidan foydalanilsin.

**2.19.** Множество объектов *E*0={*S*1,…,*Sm*} разделено на два непересекающихся класса *K*1 и *K*2. Объект *Su*∈*E*0 описываются набором разнотипных признаков *X*(*n*)=(*x*1,…,*xn*), значения которых представлены в виде строки таблицы *A*={*auv*}m\*n. В наборе *X*(*n*) β признаков измеряются в номинальной, *n* – β в количественной шкале измерений. Множеством допустимых значений (градaций) каждого номинального признака *xi*∈*X*(*n*) является 1,2,…,*pi*, 2≤*pi*≤*m*/2. Значение функции принадлежности *fi*(μ) к классу *K*1 по градaции μ∈{1,2,…,*pi*} признака *xi* вычисляется как

,

где *d*1μ (*d*2μ) – множество объектов класса *K*1 (*K*2 ), значением признака *xi* у которых является μ. Приведенная формула используется для преобразования значений номинальных признаков объектов *E*0 в таблице *A*={*auv*}m\*n. Если в объекте *Su*=(*au*1,…,*aun*) (*Su*∈*E*0), *auv* представляет градaцию номинального признака *xv*∈*X*(*n*), то *auv*=*fv*(*auv*).

Требуется для каждого класса *Ki*, *i*=1,2 по таблице *A* вычислить вектор математического ожидания *Mi*=(*bi*1,…,*bin*). 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 *Su*∈*E*0 в *A* и эталоном *Mi*, *i*=1,2 использовать метрику Канберра



**2.20.** Berilgan *E*0={*S*1,…,*Sm*} ob’ektlar to‘plami ikkita o‘zaro kesishmaydigan *K*1 va *K*2 sinflarga bo‘lingan. To‘plamning *Su*∈*E*0 ob’ekti *X*(*n*)=(*x*1,…,*xn*) turli toifadagi alomatlar to‘plami bilan tavsiflanadi. Alomatlar qiymatlari *A*={*auv*}*m\*n* matritsa satrini hosil qiladi. *X*(*n*) to‘plamda *β* ta alomatlar nominal, *n* – *β* tasi miqdoriy shkalalarda o‘lchanadi. Har bir *xi*∈*X*(*n*) nominal alomatning mumkin bo‘lgan qiymatlari (gradasiyalari) 1,2,…,*pi*, 2≤*pi*≤*m*/2 sonlari bilan beriladi. Nominal *xi* alomatning *μ*∈{1,2,…,*pi*} gradasiyalari bo‘yicha *K*1 sinfga tegishlilik funksiyasi qiymati quyidagi ko‘rinishda hisoblanadi

.

Bu еrda *d*1μ (*d*2μ) – *xi* alomatining qiymati *μ* bo‘lgan *K*1(*K*2) sinf ob’ektlari soni (to‘plami).

Yuqorida keltirilgan formula *A*={*auv*}*m\*n* matritsadagi *E*0 ob’ektlari tavsifidagi nominal alomatlar qiymatlarini o‘girishda foydalaniladi. Agar *Su*=(*au*1,…,*aun*) (*Su*∈*E*0) ob’ektda *auv* qiymat *xv*∈*X*(*n*) alomat gradasiyasi bo‘lsa, u holda *auv*=*fv*(*auv*). Har bir *Ki*, *i*=1,2 sinflar uchun *A* matritsa bo‘yicha *Mi*=(*bi*1,…,*bin*) 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,…, *an*) ob’ekt uchun

*F*(*S*)=*w*1*a*1+ *w*2*a*2+ + *wn an* +*w*0,

ko‘rinishidagi chiziqli qaror qilish funksiyasining analitik ko‘rinishi topilsin. Bu еrda *w*1=*b*11, ,*wn*=*b*1*n*, 

**2.20.** Множество объектов *E*0={*S*1,…,*Sm*} разделено на два непересекающихся класса *K*1 и *K*2. Объект *Su*∈*E*0 описываются набором разнотипных признаков *X*(*n*)=(*x*1,…,*xn*), значения которых представлены в виде строки таблицы *A*={*auv*}m\*n. В наборе *X*(*n*) β признаков измеряются в номинальной, *n* – β в количественной шкале измерений. Множеством допустимых значений (градaций) каждого номинального признака *xi*∈*X*(*n*) является 1,2,…,*pi*, 2≤*pi*≤*m*/2. Значение функции принадлежности *fi*(μ) к классу *K*1 по градaции μ∈{1,2,…,*pi*} признака *xi* вычисляется как

,

где *d*1μ (*d*2μ) – множество объектов класса *K*1 (*K*2 ), значением признака *xi* у которых является μ.

Приведенная формула используется для преобразования значений номинальных признаков объектов *E*0 в таблице *A*={*auv*}m\*n. Если в объекте *Su*=(*au*1,…,*aun*) (*Su*∈*E*0), *auv* представляет градaцию номинального признака *xv*∈*X*(*n*), то *auv*=*fv*(*auv*). Требуется для каждого класса *Ki*, *i*=1,2 по таблице *A* вычислить вектор математического ожидания *M*1i=(*bi*1,…,*bin*). 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,…, *an*)

*F*(*S*)=*w*1*a*1+ *w*2*a*2+ + *wn an* +*w*0,

где *w*1=*b*11, ,*wn*=*b*1*n*, 

**2.21.** Berilgan *E*0={*S*1,…,*Sm*} ob’ektlar to‘plami ikkita o‘zaro kesishmaydigan *K*1 va *K*2 sinflarga bo‘lingan. To‘plamning *Su*∈*E*0 ob’ekti *X*(*n*)=(*x*1,…,*xn*) nominal alomatlar to‘plami bilan tavsiflanadi. Alomatlar qiymatlari *A*={*auv*}*m\*n* matritsa satrini hosil qiladi. Har bir *xi*∈*X*(*n*) nominal alomatning mumkin bo‘lgan qiymatlari (gradasiyalari) 1,2,…,*pi*, 2≤*pi*≤*m*/2 sonlari bilan beriladi. Nominal *xi* alomatning μ∈{1,2,…,*pi*} gradasiyalari bo‘yicha *K*1 sinfga tegishlilik funksiyasi qiymati quyidagi ko‘rinishda hisoblanadi

.

Bu еrda *d*1μ (*d*2μ) – *xi* alomatining qiymati *μ* bo‘lgan *K*1 (*K*2) sinf ob’ektlari soni (to‘plami).

Nominal *xi*∈*X*(*n*) alomat turg‘unligi quyidagi formula bilan hisoblanadi:



Berilgan *E*0 to‘plamining ob’ektlarini tavsiflovchi nominal alomat gradasiyalarini *auv*=*fv*(*auv*) asoslangan holda *A*={*auv*}*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,…,*Sm*} разделено на два непересекающихся класса *K*1 и *K*2. Объект *Su*∈*E*0 описываются набором номинальных признаков *X*(*n*)=(*x*1,…,*xn*), значения которых представлены в виде строки таблицы *A*={*auv*}m\*n. Множеством допустимых значений (градaций) каждого номинального признака *xi*∈*X*(*n*) является 1,2,…,*pi*, 2≤*pi*≤*m*/2. Значение функции принадлежности *fi*(μ) к классу *K*1 по градaции μ∈{1,2,…,*pi*} признака *xi* вычисляется как

,

где *d*1μ (*d*2μ) – множество объектов класса *K*1 (*K*2 ), значением признака *xi* у которых является μ. Устойчивость признака *xi*∈*X*(*n*) вычисляется как



Требуется произвести преобразование значений градaций признаков объектов *E*0 в таблице *A*={*auv*}m\*n как *auv*=*fv*(*auv*). Вывести список признаков, упорядоченных в порядке убывания значений устойчивости.

**2.22.** Berilgan *E*0={*S*1,…,*Sm*} ob’ektlar to‘plami ikkita o‘zaro kesishmaydigan *K*1 va *K*2 sinflarga bo‘lingan. To‘plamning *Su*∈*E*0 ob’ekti *X*(*n*)=(*x*1,…,*xn*) turli toifadagi alomatlar to‘plami bilan tavsiflanadi. Alomatlar qiymatlari *A*={*auv*}m\*n matritsa satrini hosil qiladi. *X*(*n*) to‘plamda β ta alomatlar nominal, *n* – β tasi miqdoriy shkalalarda o‘lchanadi. Har bir *xi*∈*X*(*n*) nominal alomatning mumkin bo‘lgan qiymatlari (gradasiyalari) 1,2,…,*pi*, 2≤*pi*≤*m*/2 sonlari bilan beriladi. Nominal *xi* alomatning μ∈{1,2,…,*pi*} gradasiyalari bo‘yicha *K*1 sinfga tegishlilik funksiyasi qiymati quyidagi ko‘rinishda hisoblanadi

.

Bu еrda *d*1μ (*d*2μ) – *xi* alomatining qiymati μ bo‘lgan *K*1 (*K*2 ) sinf ob’ektlari soni (to‘plami).

Yuqorida keltirilgan formula *A*={*auv*}m\*n matritsadagi *E*0 ob’ektlari tavsifidagi nominal alomatlar qiymatlarini o‘girishda orqali *X*(*n*) to‘plamidan yangi *H*(*n*)=(*y*1,…,*yn*) to‘plamini tanlash uchun foydalaniladi. Agar *Su*=(*au*1,…,*aun*) (*Su*∈*E*0) ob’ektda *auv* qiymat *xv*∈*X*(*n*) alomat gradasiyasi bo‘lsa, u holda *auv*=*fv*(*auv*).

Har bir *Ki*, *i*=1,2 sinflar uchun *A* matritsa bo‘yicha *Mi*=(*bi*1,…,*bin*) matematik kutilma vektorini hisoblash talab qilinadi. Har bir *yj*∈*H*(*n*) alomat bo‘yicha fazoni qisqartirish uchun sinf ichidagi yaqinlik



va sinflar o‘rtasidagi farqlanish



formula bilan hisoblanadi.

O‘chirishga nomzod sifatida *H*(*r*), *r*=2,…,*n* to‘plamidan  shartni qanoatlantiruvchi *yj*∈*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,…,*Sm*} разделено на два непересекающихся класса *K*1 и *K*2. Объект *Su*∈*E*0 описываются набором разнотипных признаков *X*(*n*)=(*x*1,…,*xn*), значения которых представлены в виде строки таблицы *A*={*auv*}m\*n. В наборе *X*(*n*) β признаков измеряются в номинальной, *n* – β в количественной шкале измерений. Множеством допустимых значений (градaций) каждого номинального признака *xi*∈*X*(*n*) является 1,2,…,*pi*, 2≤*pi*≤*m*/2. Значение функции принадлежности *fi*(μ) к классу *K*1 по градaции μ∈{1,2,…,*pi*} признака *xi* вычисляется как

, (2.1)

где *d*1μ (*d*2μ) – множество объектов класса *K*1(*K*2), значением признака *xi* у которых является μ. Формула (2.1) используется для выбора нового набора *H*(*n*)=(*y*1,…,*yn*) из *X*(*n*) путём преобразования значений номинальных признаков объектов *E*0 в таблице *A*={*auv*}m\*n. Если в объекте *Su*=(*au*1,…,*aun*) (*Su*∈*E*0), *auv* представляет градaцию номинального признака *xv*∈*X*(*n*), то *auv*=*fv*(*auv*).

Требуется для каждого класса *Ki*, *i*=1,2 по таблице *A* вычислить вектор математического ожидания *Mi*=(*bi*1,…,*bin*). Для сокращения размерности пространства по каждому признаку *yj*∈*H*(*n*) вычисляется внутриклассовая близость



и межклассовое различие



В качестве кандидата на удаление из набора *H*®, *r*=2,…,*n* выбирается *yj*∈*H*® с  Требуется сократить размерность признакового пространства с *n* до *n* – *p*, *p*>0 с указанием порядка удаляемых признаков.

**2.23.** Berilgan *E*0={*S*1,…,*Sm*} ob’ektlar to‘plami ikkita o‘zaro kesishmaydigan *K*1 va *K*2 sinflarga bo‘lingan. To‘plamning *Su*∈*E*0 ob’ekti *X*(*n*)=(*x*1,…,*xn*) miqdoriy alomatlar to‘plami bilan tavsiflangan bo‘lib, ularning qiymatlari *A*={*auv*}*m\*n* matritsa satrini aniqlaydi. Matematik kutilma *M*=(*b*1,…,*bn*) 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(μ), μ=n(n+*1)/2 to‘plamdagi tavsiflanishga o‘tkaziladi. Ixtiyoriy *Su*∈*E*0 ob’ekt *D*={*duk*}*m\*μ* matritsa satri ko‘rinishida yoziladi, unda *k*=1,…,*n* alomatlar qiymatlari 2 ikkita gradasiyaga



*k*=*n*+1,…, *n*(*n+*1)/2 – 4 uchun



Ixtiyoriy *yi*∈*Y*(*μ*) alomatning *δ*∈{1,…,4} gradasiyasi bo‘yicha *fi*(*δ*) – *K*1 sinfga tegishlilik funksiyasining qiymati



formulasi bilan hisoblanadi, bu еrda *z*1*μ* (*z*2*μ*) – *yi* alomatining qiymati *δ* bo‘lgan *K*1 (*K*2) sinf ob’ektlari soni (to‘plami). Yuqorida keltirilgan formula *D=*{*duk*}*m\*μ* matritsaga *duk*=*fk*(*duk*) orqali *E*0 ob’ektlari tavsifidagi nominal alomatlar qiymatlarini o‘girishda foydalaniladi. Har bir *Ki*, *i*=1,2 sinf uchun matematik kutilma *Mi*=(*mi*1,…,*mi*μ) vektorini hisoblash talab qilinadi. Qaysi *yj*∈*Y*(*μ*) alomat bo‘yicha *│m1j – m2j│= max* ekanligi aniqlansin.

**2.23.** Множество объектов *E*0={*S*1,…,*Sm*} разделено на два непересекающихся класса *K*1 и *K*2. Объект *Su*∈*E*0 описываются набором количественных признаков *X*(*n*)=(*x*1,…,*xn*), значения которых представлены в виде строки таблицы *A*={*auv*}m\*n. Определены значения элементов вектора математического ожидания *M*=(*b*1,…,*bn*) 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*(μ), μ=*n*(*n*+1)/2. Объект *Su*∈*E*0 записывается в виде строки матрицы *D*={*duk*}m\*μ в которой значения признаков с *k*=1,…,*n* имеют 2 градaции



с *k*=*n*+1,…, *n*(*n+*1)/2 – 4



Значение функции принадлежности *fi*(δ) к классу *K*1 по градaции δ∈{1,…,4} признака *yi*∈*Y*(*μ*) вычисляется как

,

где *z*1μ (*z*2μ) – множество объектов класса *K*1 (*K*2), значением признака *yi* у которых является δ. Приведенная формула используется для преобразования значений номинальных признаков объектов *E*0 по таблице *D=*{*duk*}m\*μ как *duk*=*fk*(*duk*). Требуется для каждого класса *Ki*, *i*=1,2 по таблице *D* вычислить вектор математического ожидания *Mi*=(*mi*1,…,*mi*μ). Определить по какому признаку *yj*∈*Y*(*μ*) справедливо *│m1j – m2j│= max.*

**2.24.** Ob’ektlar to‘plami *E*0={*S*1,…,*Sm*} ikkita o‘zaro kesishmaydigan *K*1 va *K*2 sinflarga ajratilgan. Ob’ektlarni tavsiflash uchun nominal alomatlarning *X*(*n*)=(*x*1,…,*xn*) majmuasidan foydalanilgan. Har bir *xi*∈*X*(*n*) nominal alomatning mumkin bo‘lgan qiymatlar (gradasiyalar) to‘plami 1,2,…,*pi*, 2≤*pi*≤*m*/2 qiymatlardan iborat. Nominal *xi* alomatning μ∈{1,2,…,*pi*} gradasiyasi bo‘yicha *K*1 sinfga tegishlilik funksiyasi



ko‘rinishida hisoblanadi. Bu еrda *d*1μ (*d*2μ) – *xi* alomat qiymati μ bo‘lgan *K*1 (*K*2) sinf ob’ektlari to‘plami.

Har bir *xi*∈*X*(*n*) alomat bo‘yicha sinflar o‘rtasidagi chegara *Gi*=(*s*1 + *s*2)/2 orqali hisoblansin. Bu еrda

 va .

Berilgan *E*0 tanlanmaning boshlang‘ich ob’ektlarini *A*={*auv*}*m\*n*jadval ko‘rinishida, uning kengaymasini *B*={*brc*}m\*p, *p*=*n*(*n*-1)/2 taqdim etilsin. Ixtiyoriy *Su* =(*xu*1,…,*xun*) (*Su*∈*E*0) ob’ekt *A* jadvalida (*au*1,…,*aun*) satr sifatida yoziladi, bu еrda



O‘z navbatida *V* jadvalning har elementi *A* jadvalining ikkita satridan *buc* = 2*aui* +*auj* – 2 , *i* ≠ *j*, *c*=2(*i*-1) +*j* formulalar orqali shakllantiriladi.

**2.24.** Множество объектов *E*0={*S*1,…,*Sm*} разделено на два непересекающихся класса *K*1 и *K*2. Для описания объектов используется набор номинальных признаков *X*(*n*)=(*x*1,…,*xn*). Множеством допустимых значений (градaций) каждого признака *xi*∈*X*(*n*) является 1,2,…,*pi*, 2≤*pi*≤*m*/2. Значение функции принадлежности *fi*(μ) к классу *K*1 по градaции μ∈{1,2,…,*pi*} признака *xi* вычисляется как

,

где *d*1μ (*d*2μ) – множество объектов класса *K*1 (*K*2 ), значением признака *xi* у которых является μ. По каждому признаку *xi*∈*X*(*n*) определить значение границы между классами как *Gi*=(*s*1 + *s*2)/2, где

 и .

Требуется представить исходное описание объектов *E*0 в таблице *A*={*auv*}m\*n и её расширении *B*={*brc*}m\*p, *p*=*n*(*n*-1)/2. Объект *Su* =(*xu*1,…,*xun*) (*Su*∈*E*0) в *A* запишется как строка таблицы (*au*1,…,*aun*), в которой

 .

Каждый элемент таблицы *B* для объекта *Su* формируется из двух строк *A* как *buc* = 2*aui* +*auj* -2 , *i* ≠ *j*, *c*=2(*i*-1) +*j*.

**2.25.** Ob’ektlar to‘plami *E*0={*S*1,…,*Sm*} ikkita o‘zaro kesishmaydigan *K*1 va *K*2 sinflarga ajratilgan. Ob’ektlarni tavsiflash uchun nominal alomatlarning *X*(*n*)=(*x*1,…,*xn*) majmuasidan foydalanilgan. Har bir *xi*∈*X*(*n*) nominal alomatning mumkin bo‘lgan qiymatlar (gradasiyalar) to‘plami 1,2,…,*pi*, 2≤*pi*≤*m*/2 qiymatlardan iborat. Nominal *xi* alomatning μ∈{1,2,…,*pi*} gradasiyasi bo‘yicha *K*1 sinfga tegishlilik funksiyasi



ko‘rinishida hisoblanadi. Bu еrda *d*1μ (*d*2μ) – *xi* alomat qiymati μ bo‘lgan *K*1 (*K*2) sinf ob’ektlari to‘plami.

Har bir *xi*∈*X*(*n*) alomatning gradasiyasi bo‘yicha *ri*=(*ri*1,…,*rip*(*i*)), *ri*μ = *fi*(μ) vektor hosil qilinsin va sinflar o‘rtasidagi *Gi*=(s1 + s2)/2 ko‘rinishidagi chegaraviy qiymat hisoblansin. Bu еrda s2=max{*fc*(μ)| 0.5 – *fc*(μ)>0} va s1=min{*fc*(μ)| 1 – *fc*(μ)<0.5 }.

Mumkin bo‘lgan *S*=(*b*1,…,*bn*) ob’ekt tavsifida ikki bosqichli o‘girish amalga oshirilsin. Bu еrda *c*=*bi*, , 

2.25. Множество объектов *E*0={*S*1,…,*Sm*} разделено на два непересекающихся класса *K*1 и *K*2. Для описания объектов используется набор номинальных признаков *X*(*n*)=(*x*1,…,*xn*). Множеством допустимых значений (градaций) каждого признака *xi*∈*X*(*n*) является 1,2,…,*p*(*i*), 2≤*p*(*i*)≤)*m*/2. Значение функции принадлежности *fi*(μ) к классу *K*1 по градaции μ∈{1,2,…, *p*(*i*)} признака *xi* вычисляется как

,

где *d*1μ (*d*2μ) – множество объектов класса *K*1 (*K*2), значением признака *xi* у которых является μ.

По градaциям каждого признака *xi*∈*X*(*n*) необходимо сформировать вектор *ri*=(*ri*1,…,*rip*(*i*)), *ri*μ = *fi*(μ) и вычислить значение границы между классами как *Gi*=(s1 + s2)/2, где s2=max{*fc*(μ)| 0.5 – *fc*(μ)>0} и s1=min{*fc*(μ)| 1 – *fc*(μ)<0.5 }.

Требуется произвести двухэтапное преобразование описаний произвольного допустимого объекта *S*=(*b*1,…,*bn*), *c*=*bi*, ,



**2.26.** Ob’ektlar to‘plami *E*0={*S*1,…,*Sm*} ikkita o‘zaro kesishmaydigan *K*1 va *K*2 sinflarga ajratilgan. Ob’ektlarni tavsiflash uchun miqdoriy alomatlarning *X*(*n*)=(*x*1,…,*xn*) majmuasidan foydalanilgan bo‘lib, ularning qiymatlari *A*={*auv*}*m\*n* matritsaning satri ko‘rinishida taqdim etilgan. Matematik kutilma – *M*=(*b*1,…,*bn*) 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*(*μ*), *μ*=*n*(*n*+1)/2 majmuasiga o‘tkazish amalga oshiriladi.

Ixtiyoriy *Su*∈*E*0 ob’ekt *D*={*duk*}m\*μ matritsa satri



ko‘rinishida yozilib, unda *k*=1,…,*n* alomatlar qiymatlari ko‘rinishida hisoblanadigan 2 gradasiyaga, tartib nomerlari *k*=*n*+1,…, *n*(*n+*1)/2 bo‘lganlar



formula orqali hisoblanuvchi 4 gradasiyaga ega bo‘ladi.

Ixtiyoriy *yi*∈*Y*(*μ*) alomatning δ∈{1,…,4} gradasiyalari bo‘yicha *K*1 sinfiga *fi*(δ) tegishlilik funksiyasining qiymati quyidagicha hisoblanadi:

.

Bu еrda *z*1*i*(δ)(*z*2*i*(δ)) – *yi* 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=*{*duk*}*m\*μ* jadvali bo‘yicha *duk*=*fk*(*duk*) sifatida o‘girish uchun foydalaniladi. *D* jadval bo‘yicha



orqali har bir *yi*∈*Y*(μ) alomat turg‘unligini hisoblash talab qilinadi. *Y*(*μ*) 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,…,*Sm*} разделено на два непересекающихся класса *K*1 и *K*2. Объект *Su*∈*E*0 описываются набором количественных признаков *X*(*n*)=(*x*1,…,*xn*), значения которых представлены в виде строки таблицы *A*={*auv*}m\*n. Определены значения элементов вектора математического ожидания *M*=(*b*1,…,*bn*) 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*(μ), μ=*n*(*n*+1)/2. Объект *Su*∈*E*0 записывается в виде строки матрицы *D*={*duk*}m\*μ в которой значения признаков с *k*=1,…,*n* имеют 2 градaции



с *k*=*n*+1,…, *n*(*n+*1)/2 – 4 градaции



Значение функции принадлежности *fi*(δ) к классу *K*1 по градaции δ∈{1,…,4} признака *yi*∈*Y*(*μ*) вычисляется как

,

где *z*1*i*(δ)(*z*2*i*(δ)) – множество объектов класса *K*1 (*K*2), значением признака *yi* у которых является δ. Приведенная формула используется для преобразования значений номинальных признаков объектов *E*0 по таблице *D=*{*duk*}m\*μ как *duk*=*fk*(*duk*). Требуется по таблице *D* вычислить устойчивость каждого признака *yi*∈*Y*(μ) как



Вывести список 5 номеров признаков из *Y*(μ), имеющих максимальные значения устойчивости.

**2.27.** Ob’ektlar tanlanmasi *E*0={*S*1,…,*Sm*} ikkita o‘zaro kesishmaydigan *K*1 va *K*2 sinflarga ajratilgan. Har bir *S*∈*E*0 ob’ekt, mumkin bo‘lgan qiymatlari to‘plami {1,2} gradasiyalari iblorat *X*(*n*)=(*x*1,…,*xn*) nominal alomatlar majmuasi bilan tavsiflangan.

Mos ravishda *K*1 va *K*2 sinflar ob’ektlari tavsifidagi *xc*∈*X*(*n*) alomatning gradasiyalar qiymatlarining miqdorlarini  orqali belgilaylik (*j*∊{1,2}).

Nominal *xc* alomat bo‘yicha sinflararo farqlanish

 (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) *βc* hisoblanadi:

 (2)

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

*vs*=*βsλs.*  (3)

Nominal *xd*∊*X*(*n*) alomatning *j*∈{1,2} gradasiyalar va (3) vazn bo‘yicha hissasi quyidagi formula bilan aniqlangan:

 (4)

bu еrda  – mos ravishda *K*1 i *K*2 sinflardagi *xd* alomatning *j* gradasiyasining qiymatlari miqdori. Har bir *Sr*∈*E*0 ob’ektning nominal shkaladagi o‘lchami *Sr*=(*xr*1,…,*xrn*) va (4) hissa bo‘yicha umumlashgan bahosi



ko‘rinishida hisoblanadi.

Aniqlash talab qilinadi:

*p*1=**|**{*Si*∈*K*1 **|** *Z*(*Si*)<0}**|** va *p*2=**|**{*Si*∈*K*2 **|***Z*(*Si*)>0}**|**.

**2.27.** Выборка объектов *E*0={*S*1,…,*Sm*} разделено на два непересекающихся класса *K*1 и *K*2. Каждый объект *S*∈*E*0 описываются набором номинальных признаков *X*(*n*)=(*x*1,…,*xn*), множеством допустимых значений которых являются градaции {1,2}.

Обозначим через  – количество значений градaции *j*∊{1,2} признака *xc*∈*X*(*n*) в описании объектов соответственно класса *K*1 и *K*2. Межклассовое различие по признаку *xc* определяется как величина

 (2.1)

Степень однородности (мера внутриклассового сходства) β*c* значений градaций признака по классам *K*1, *K*2 вычисляется по формуле:

 (2.2)

С помощью (2.1),(2.2) вес признака *xc*∈*X*(*n*) в номинальной шкале определяется как произведение внутриклассового сходства и межклассового различия

*vс*=β*с*λ*с*. (2.3)

Вклад признака *xd*∊*X*(*n*) по градaции *j*∈{1,2} и весу (2.3) определяется как

 (2.4)

где  – количество значений градaции *j* признака *xd* соответственно в классах *K*1 и *K*2. Обобщённая оценка объекта *Sr*∈*E*0 по его описанию в номинальной шкале измерений *Sr*=(*xr*1,…,*xrn*) и вкладам (2.4) вычисляется как 

Требуется определить *p*1=|{*Si*∈*K*1 |*Z*(*Si*)<0}| и *p*2=|{*Si*∈*K*2 |*Z*(*Si*)>0}|.

**2.28.** (*Konkurient o‘xshashlikni hisoblash*). Berilgan *E*0=(*S*1,…,*Sm*) tanlanma ob’ektlari *X*(*n*)=(*x*1,…,*xn*), *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 ρ(*x*,*y*) metrika bilan aniqlangan bo‘lsin.

Berilgan *S*∈*Kt*, *t*=1,2 ob’ekt uchun talab qilinadi:

- barcha *Sj*∈*K*3*-t* ob’ektlar bo‘yicha *Si*∈*Kt* ob’ektga nisbatan konkurient o‘xshashlik o‘lchami

*F*(*S*,*Si*|*Sj*) = (ρ(*S*,*Sj*) - ρ(*S*,*Si*))/( ρ(*S*,*Sj*) + ρ(*S*,*Si*))

formula orqali aniqlansin;

- *F*(*S*,*Si*|*Sj*) 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,…,*Sm*) с помощью набора количественных признаков *X*(*n*)=(*x*1,…,*xn*). Объекты *E*0 разделены на два непересекающихся класса *K*1  и *K*2. Считается, что близость между объектами определяется по метрике ρ(*x*,*y*). Требуется для заданного объекта *S*∈*Kt*, *t*=1,2:

* определить меру конкурентного сходства относительно объекта *Si*∈*Kt* по всем объектам *Sj*∈*K*3*-t*, используя формулу

*F*(*S*,*Si*|*Sj*) = (ρ(*S*,*Sj*) - ρ(*S*,*Si*))/( ρ(*S*,*Sj*) + ρ(*S*,*Si*));

– выделить объекты с максимальным и минимальным значениями *F*(*S*,*Si*|*Sj*);

– сравнить результаты по метрике Евклида и Чебышева.

**2.29.** (*Taqsimot zichligi*). Turli toifadagi alomatlar *X*(*n*)=(*x*1,…,*xn*) bilan tavsiflangan ob’ektlar to‘plami *E*0=(*S*1,…,*Sm*) berilgan. *E*0 ob’ektlari *l* ta o‘zaro kesishmaydigan *K*1,…,*Kl* sinflarga bo‘lingan. Juravlev metrikasi ρ(*x*,*y*) bo‘yicha chegaraviy

ob’ektlar to‘plami

*B*⸦*E*0, 

ajratib olingan bo‘lsin.

Har bir *S*∈*B*∩*Kt* ob’ekt uchun aniqlansin:

*  qiymatlari;
* taqsimot zichligi, bu еrda *k*=|{*Si*|ρ(*S*,*Si*<*rs*)}|.
  1. (*Плотность распределения*). Дано описание множества объектов *E*0=(*S*1,…,*Sm*) с помощью набора разнотипных признаков *X*(*n*)=(*x*1,…,*xn*). Объекты *E*0 разделены на *l* непересекающихся класса *K*1,…,*Kl*. Считается, что по метрике Журавлёва ρ(*x*,*y*) выделено множество граничных объектов *B*⸦*E*0, . Требуется для каждого *S*∈*B*∩*Kt* определить:

– значение ;

– плотность распределения , где *k*=|{*Si*|ρ(*S*,*Si*<*rs*)}|.

# 3. Берилганларнинг кластер таҳлили (Кластерный анализ данных)

**3.1.** Berilgan  tanlanmada berilgan  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 произвести группировку объектов на кластеры с помощью алгоритма “*внутригрупповых средних*”. Для вычисления расстояния между объектами, описываемых *n*, *n*≥2 количественными признаками, использовать метрику Евклида и Чебышева.

**3.2.** Berilgan  bo‘sag‘a qiymat va  metrika bo‘yicha  tanlanmani kesishmaydigan guruhlarga ajratishlar sonini hisoblansin. Masofani hisoblash uchun Hemming va Chebishev metrikasidan foydalanilsin.

**3.2.** По заданному пороговому значению и метрике вычислить число разбиений выборки в *Rn*, *n*≥2 на непересекающиеся группы. Для вычисления расстояния использовать метрики Хемминга и Чебышева.

**3.3.** Obektlari *Rn*, *n*≥2 miqdoriy alomatlar fazosida tavsiflangan  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 в *Rn*, *n*≥2. За критерий останова принять условие: максимальное из минимальных расстояний до центров групп меньше заданного значения .

**3.4.** Obektlari *Rn*, *n*≥2 turli toifadagi (miqdoriy va nominal) alomatlar fazosida tavsiflangan  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 с описанием объектов набором разнотипных (количественных и номинальных) признаков *X*(*n*)=(*x*1,…,*xn*). За критерий останова принять условие: максимальное из минимальных расстояний до центров групп меньше заданного значения .

**3.5.** } o‘rgatuvchi tanlanma berilgan. Har bir  obyekt  ta miqdoriy alomat bilan tavsiflanadi. Berilgan  metrika bo‘yicha , obyektlar orasidagi eng qisqa yopiqmas yo‘l qurilsin.  metrika sifatida Evklid, Hemming, Chebishev metrikalari bo‘lgan variantlar ko‘rib chiqilsin.

**3.5.** По заданной выборке ={,, каждый объект которой описывается с помощью количественных признаков, и заданной метрике построить кратчайший незамкнутый путь (КНП) между объектами , В качестве использовать метрики Евклида, Хемминга, Чебышева.

**3.6.** O‘zaro kesishmaydigan *l* ta  sinflarga bo‘lingan } o‘rgatuvchi tanlanma berilgan. Har bir  – obyekt  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 }, разделённой на непересекающихся классов методом “последовательного исключения” для алгоритма “ближайший сосед”. Считается, что каждый объект описывается с помощью количественных признаков и в качестве меры близости между объектами используется метрика Евклида.

**3.7.** Berilgan *E*0={*S*1,…,*Sm*} o‘rgatuvchi ob’ektlar to‘plami ikkita o‘zaro kesishmaydigan *K*1 va *K*2 sinflarga bo‘lingan. To‘plamning *Su*∈*E*0 ob’ekti *X*(*n*)=(*x*1,…,*xn*) turli toifadagi *n* ta alomatlar bilan tavsiflangan bo‘lib, ularning *p* tasi miqdoriy, *n* – *p* tasi nominal shkalalarda o‘lchanadi. Berilgan son bo‘yicha *xj,*  alomat qiymatlarini ta o‘zaro kesishmaydigan ], intervallarga bo‘linsin. Bu еrda , , *h*=(*a*0 +*al*)/*l* qadam bilan. , interval bo‘yicha tanlanma ob’ektlarining sinfga tegishlilik funksiyasi *n*1*d* +*n*2*d*>0 sharti bilan

,

ko‘rinishda hisoblanadi. Bu еrda - intervaldagi sinf vakillari soni. Quyidagi formula bilan alomat qiymatlarini ta intervallarga bo‘lish turg‘unligi hisoblansin:

Bu еrda – shartni qanoatlantiruvchi intervallar soni.

**3.7.** В обучающей выборке содержаться представители двух непересекающихся классов . Каждый объект выборки описывается набором =(*x*1,…,*xn*) из разнотипных признаков, из которых измеряются в количественной, номинальной шкалах измерений. По заданному числу произвести разбиения значений количественного признака *xj* на непересекающихся интервалов ],,…где , , с шагом *h*=(*a*0+*al*)/*l*. Функция принадлежности объектов к классу по интервалу при условии *n*1*d* +*n*2*d*>0 вычисляется как

,

где – количество объектов класса в интервале Определить значение устойчивости разбиения на *l* интервалов по признаку

где – количество интервалов с

**3.8.**  obyektdan tashkil topgan  o‘lchovli alomatlar fazosida } ko‘rinishda to‘plam berilgan. Berilgan  parametr bo‘yicha  qiymatlarini hisoblansin.  ning qiymati markazi  da bo‘lgan va  to‘plamdan  ga metrika bo‘yicha eng yaqin  ta qo‘shnini o‘z ichiga oladigan gipersharning radiusiga teng. O‘sish tartibidagi  ketma-ketlikdan  obyektni va unga eng yaqin  ta qo‘shnisi o‘chirilsin. O‘chirilgan obyektlar  ga qo‘shiladi.

Keyingi guruhning xos vakili sifatida tartiblangan tanlanmadagi qolgan obyektlarining birinchisi e’lon qilinadi, unga o’xshash obyektlar 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 объектов в -мерном признаковом пространстве представлено в виде }. С помощью заданного параметра произвести вычисление значений . Значение определяет радиус гипершара по метрике с центром в в окрестности которого содержится ближайших к объектов множества . Из упорядоченной по возрастанию последовательности удалить объект и ближайших к нему объектов. Удаленные объекты включаются в группу .

Типичным представителем следующей группы объявляется первый из оставшихся объектов упорядоченной выборки, а кортеж похожих на него объектов включается в группу , 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  jadval berilgan. Iyerarxik guruhlash algoritmi quyidagi sxema bo‘yicha amalga oshirilsin:

*1-qadam*. Uchburchak jadval shaklida  masofalar hisoblansin;

*2-qadam*.  sharti asosida:

* + - jadvaldan barcha *q* kattaliklar o‘chirilsin;
    - *p* belgilashlar *r* bilan almashtirilsin;
    - jadvalning qolgan qismida  hisoblansin;

*3-qadam*. Agar qisqartirilgan jadvaldagi satr va ustunlar sonlari  bo‘lsa, 2-qadamga o‘tilsin, aks holda tamom.

Bu еrda



umumiy formula bo‘yicha hisoblanadi.

Iyerarxiyanig quyidagi variantlarini ko‘rib chiqaylik:

*1-variant*: 

*2-variant*: 

*3-variant*: bu еrda va  – *p* va *q* guruhdagi obyektlar soni.

**3.9.** В заданной таблице по строкам содержатся *m* объектов, описываемых *n* количественными признаками.

Реализовать алгоритм иерархической группировки по следующей схеме:

*1-шаг*: вычислить расстояния в форме треугольной таблицы;

*2-шаг*: пусть

удалить из таблицы все величины

заменить на

вычислить оставщихся в таблице.

*3-шаг*: Eсли число столбцов и строк в получившейся сокращённой таблице больше то идти к 2, иначе конец.

Расстояния вычисляются по общей формуле

.

Рассмотреть следующие варианты иерархий.

*1-вариант*:

*2-вариант*:

*3-вариант*: где и число объектов в *p*-ой и -ой группе.

**3.10.** Har bir obyekti *n* ta miqdoriy alomat bilan tavsiflangan ={,, tanlanma berilgan. Berilgan  metrika bilan , obyektlar o‘rtasidagi eng qisqa yopiqmas yo‘l qurilsin. Berilgan  bo‘yicha yo‘lning  ta eng uzun qirralarini o‘chirish orqali tanlanma  guruhlarga bo‘linsin. Metrika sifatida Evklid metrikasi ishlatilsin.

**3.10** По заданной выборке ={,, каждый объект которой описывается с помощью количественных признаков, и заданной метрике построить кратчайший незамкнутый путь (КНП) между объектами , С помощью заданного получить разбиение на группу путём удаления самых длинных рёбер. В качестве использовать метрику Евклида.

**3.11.** Har bir obyekti *n* ta miqdoriy alomat bilan tavsiflangan ={,, tanlanma berilgan bo’lib, u ikkita o’zaro kesishmaydigan sinflarga bo’lingan. Berilgan  metrika bilan , 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 ={,, каждый объект которой описывается с помощью количественных признаков, и заданной метрике построить кратчайший незамкнутый путь (КНП) между объектами ,

На множестве {,, определено разбиение на два класса . Если существует путь, соединяющий объекты (вершины КНП) из разных классов, то удалить самое длинное ребро в нём. В качестве использовать метрику Хемминга.

**3.12.** Berilgan  tanlanma obyektlarini approksimatsiya qilish uchun  kodli  vektorlar bilan vektorli kvantlash masalasi еchilsin. Kodli vektorlar uchun qandaydir, tugunlarning yaqinlik o‘lchamining simmetrik matritsasi berilgan bo‘lsin: har bir  juftlik uchun  soni aniqlangan hamda jadvalning barcha diagonal elementlari birga teng  Har bir  obyekt *n* ta miqdoriy alomatlar bilan tavsiflangan.

Kiruvchi *s* signallar vektorlari alohida ishlanadi, har biri uchun eng yaqin kod vektori  topiladi (“*g‘olib barchasiga ega*” tamoyili). Shundan keyin  barcha  kod vektorlari



formulasi bilan qayta hisoblanadi. Bu еrda  – 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 аппроксимatsiи объектов выборки данных кодовыми векторами . Для кодовых векторов задана некоторая симметричная таблица мер близости узлов: для каждой пары определяется число при этом диагональные элементы таблицы равны единице Каждый объект описывается *n* количественными признаками.

Векторы входных сигналов *s* обрабатываются по одному, для каждого из них находится ближайший кодовый вектор («победитель», который «забирает всё») . После этого все кодовые векторы , для которых , пересчитываются по формуле

,

где – 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  sinflar vakillarini o‘z ichiga olgan  o‘rgatuvchi tanlanma berilgan. Har bir  obyekt *n* ta  miqdoriy alomatlar bilan tavsiflangan. Berilgan metrika bo’yicha tanlanmaning shovqin obyektlari sinflarning chegaraviy obyektlari quyidagicha aniqlanuvchi *B*⊂*E*0 to’plamning to’plam ostisi sifatida qaraladi:

**

Tanlanmaning *S*∊*B*∩*Kj*, *j*=1,…,*l* obyekti *Kj* sinfining *Dj* shovqin obyektlari to’plamiga tegishli hisoblanadi, agar



sharti o’rinli bo’lsa.

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

**3.13.** Обучающая выборка содержит представителей непересекающихся классов . Каждый объект описывается набором из количественных признаков . Множество шумовых объектов рассматривается как подмножество граничных объектов классов *B*⊂*E*0 по заданной метрике , котороеопределяется как

**.

Объект *S*∊*B*∩*Kj*, *j*=1,…,*l* принадлежит множеству шумовых объектов *Dj* класса *Kj*, если

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

**3.14.** O‘zaro kesishmaydigan *l* ta  sinflar vakillarini o‘z ichiga olgan  o‘rgatuvchi tanlanma berilgan. Har bir  obyekt *n* ta  turli toyifadagi (miqdoriy va nominal) alomatlar bilan tavsiflangan. Berilgan metrika bo’yicha tanlanmaning shovqin obyektlari sinflarning chegaraviy obyektlari quyidagicha aniqlanuvchi *B*⊂*E*0 to’plamning to’plam ostisi sifatida qaraladi:

**

Tanlanmaning *S*∊*B*∩*Kj*, *j*=1,…,*l* obyekti *Kj* sinfining *Dj* shovqin obyektlari to’plamiga tegishli hisoblanadi, agar



sharti o’rinli bo’lsa.

Har bir sinf uchun Juravlev metrikalari bo‘yicha shovqin obyektlar to‘plami  aniqlansin.

**3.14.** Обучающая выборка содержит представителей непересекающихся классов . Каждый объект описывается набором из разнотипных (количественных и номинальных) признаков . Множество шумовых объектов рассматривается как подмножество граничных объектов классов *B*⊂*E*0 по заданной метрике , котороеопределяется как

**.

Объект *S*∊*B*∩*Kj*, *j*=1,…,*l* принадлежит множеству шумовых объектов *Dj* класса *Kj*, если



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

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

Berilgan *E*0=(*S*1,…,*Sm*) tanlanma ob’ektlari *X*(*n*)=(*x*1,…,*xn*), *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. *Sd* ob’ekt *S* ob’ektdan erishish mumkin deyiladi, agar *S*1,…,*Sr* yo‘l mavjud bo‘lib, *S*1=*S* va *Sr*=*Sd* . Bu еrda har bir *Si+*1 ob’ektga *Si* ob’ektdan to‘g‘ridan-to‘g‘ri erishiladi (yo‘ldagi barcha ob’ektlar asosiy hisoblanadi, *Sd* 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,…,*Sm*) с помощью набора количественных признаков *X*(*n*)=(*x*1,…,*xn*). 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 *Sd*достижим из *S*, если имеется путь *S*1,…,*Sr* с *S*1=*S* и *Sr*=*Sd*, где каждый объект *Si+*1 прямо достижим из *Si* (все объекты на пути являются основными за исключением *Sd*). 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,…,*Sm*) – obyektlar to’plami *k means* algoritmi bilan o’zaro kesishmaydigan *G*1 ,…, *Gk*, *k*≥2 guruhlarga bo’lingan. Har bir *S*∊*E*0 obyekt *n* ta miqdoriy alomatlar *X(n)* bilan tavsiflangan. *Gi* guruh obyektlari va uning *Mi* markazi o’rtasidag masofa



ko'rinishida, guruhlar markazlari o’rtasida masofa esa



formula orqali hisoblanadi.

Klasterlash sifatini baholash talab qilinadi:



**3.16** (*Индекс Девиса-Болдуина* )

Множество объектов *E*0=(*S*1,…,*Sm*) алгоритмом *k means* разделено на непересекающиеся группы *G*1 ,…, *Gk*, *k*≥2. Каждый объект *S*∊*E*0 описывается набором из *n* количественных признаков *X*(*n*). Расстояние между объектами группы *Gi* и её центром *Mi* вычисляется как



и расстояние между центрами групп

.

Требуется оценку качества кластеризaции как



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

**4.1.** O‘zaro kesishmaydigan *l* ta  sinflarga bo‘lingan } o‘rgatuvchi tanlanmada yaqinlik matritsasini nominal alomatlar gradatsiyasi bo‘yicha hisoblash algoritmi dasturi tuzilsin. Har bir  – obyekt  ta nominal alomat bilan tavsiflanadi. Berilgan  obyektlar juftligi to‘plamida quyidagi funksiyalar aniqlanadi:

Berilgan tanlanmada  nominal alomatlar juftligi orasidagi yaqinlik o‘lchovi



formula orqali ifodalanadi.

Qurilgan {*bij*}*n∙n* matritsa bo’yicha (*xi*,*xj*) alomatlar juftliklarining o’smaydigan holda tartiblangan ketma-ketligi shakllantirilsin.

**4.1.** По заданной обучающей выборке }, содержащей представителей непересекающихся классов объектов , произвести вычисление матрицы близости по градaциям номинальных признаков. Считается, что каждый объект описывается *n* номинальными признаками. На множестве пар объектов (*Sa*=(*xa1*,...,*xan*), *Sb*=(*xb1*,...,*xbn*)) определяются функции:





Мера близости между парой номинальных признаков *xi, xj* на *E0* вычисляется как



По матрице {*bij*}*n∙n* сформировать последовательность из упорядоченных по не возрастанию непересекающихся пар признаков (*xi*,*xj*).

**4.2.** O‘zaro kesishmaydigan *l* ta  sinflarga bo‘lingan } o‘rgatuvchi tanlanmada sinf ichidagi o‘xshashlikni nominal alomatlar gradatsiyasi bo‘yicha hisoblash algoritmni amalga oshiruvchi dastur tuzilsin. Har bir  – obyekt *n* ta nominal alomat bilan ifodalanadi. *C*-chi alomatning sinf ichidagi o‘xshashligi  ni hisoblash formulasi:

,

bu еrda, , – *c*-alomatning gradatsiyalar soni,  –  sinf obyektining *c*-chi alomatining *t* gradatsiyasi soni.

**4.2.** По заданной обучающей выборке }, содержащей представителей непересекающихся классов объектов , произвести вычисление значений внутриклассового сходства по градaциям номинальных признаков. Считается, что каждый объект описывается *n* номинальными признаками.

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

где , , – число градaций *c*-го признака, – количество значений –ой градaции *c*-го признака в описании объектов класса .

**4.3.** O’zaro kesishmaydigan *l* ta sinflar obyektlarini o’z ichiga olgan } tanlanmada nominal alomatlar gradatsiyalari bo’yicha sinflararo farqlanish qiymatlari hisoblansin. Har bir obyekt nominal alomatlar bilan tavsiflangan. Nominal *c* alomat bo’yicha sinflararo farqlanish qiymati

,

formula bilan hisoblanadi. Bu еrda

*p* – *c* alomatning gradatsiyalari soni, sinf obyektlari tavsifidagi *c* alomatninf *t*- gradatsiyasi (*1≤ t≤ p*) qiymatlarining soni, sinf to’ldiruvchisi obyektlari tavsifidagi *c* alomatninf *t*- gradatsiyasi (*1≤ t≤ p*) qiymatlarining soni, .

**4.3.** По заданной обучающей выборке }, содержащей представителей непересекающихся классов объектов , произвести вычисление значений межклассового различия по градaциям номинальных признаков. Считается, что каждый объект описывается *n* номинальными признаками. Для вычисления значения межклассового различия по *c*-му признаку использовать формулу

, где ,

где – число градaций *c*-го признака, – количество значений *t*-ой (*1≤ t≤ p*) градaции *c*-го признака в описании объектов класса , – количество значений *t*-ой градaции *c*-го признака в – дополнении класса , .

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

.

Bu еrda  – sinf ichidagi o‘xshashlik va  – sinflararo o’xshashliklar (4.2 va 4.3 – masalalarga qaralsin).

**4.4.** Реализовать вычисление весов номинальных признаков для решения задач классификaции с помощью искусственных нейронных сетей с минимальной конфигурaцией по формуле

Для решения использовать постановку задачи и формулы для вычисления внутриклассового сходства и межклассового различия соответственно из задач 4.2 и 4.3.

**4.5.** O‘zaro kesishmaydigan *l* ta  sinflarga bo‘lingan } o‘rgatuvchi tanlanmada uchun nominal alomatlar hissalarini hisoblash algoritmining dasturi tuzilsin. Har bir  – obyekt  ta nominal alomat bilan tavsiflanadi va  – alomatning sinflarni ajratishdagi hissasi



formula bilan hisoblanadi. Bu еrda  – mos ravishda *Ki* va uning to‘ldiruvchisi *CKi* = *E0* \*Ki* sinflarda *p*-chi alomatning *j* **–** gradatsiyasining soni, *up* – *p* alomatning gradatsiyalar soni.

**4.5.** Произвести вычисление вкладов номинальных признаков по заданной обучающей выборке }, содержащей представителей непересекающихся классов объектов , в процесс принятия решения при классификatsiи объектов. Считается, что каждый объект описывается *n* номинальными признаками и вклад признака в разделение классов вычисляется по формуле

,

где  - количество значений *j* –ой градaций *p*-го признака, соответственно, класса *Ki* и его дополнения *CKi* = *E0* \*Ki*, *up* –число градaций *p*-го признака.

**4.6.** O‘zaro kesishmaydigan *K1*, *K2* sinflarga bo‘lingan } o‘rgatuvchi tanlanma berilgan. Har bir  – obyekt  ta nominal alomat bilan tavsiflangan.  juftliklar to’plami bo’yicha sinflar obyektalari o’rtasidagi maksimal farqlanishni aniqlash talab qilinadi.

Agar  alomatlar mos ravishda *pi* va *pj* gradatsiyalarga ega bo’lsa, ularni bitta latent alomatga birlashtirishda gradatsiyalar *μ*=*pipj* soni bilan cheklanadi. Latent *yc* alomat bo’yicha sinflar farqlanishi

,

ko’rinishida hisoblanadi. Bu еrda  –  sinfdagi *d* gradatsiyali obyektlar soni.

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

**4.6.** Дана выборка объектов , разделенная на два непересекающихся класса . Каждый объект выборки описывается набором  номинальных признаков. Требуется определить максимальное различие между объектами классов по множеству пар .

Если признаки  имеют соответственно *pi* и *pj* градaций, то при их объединении в один латентный признак  количество градaций ограничено числом *μ*=*pipj*. Различие между классами по *yc* вычисляется как

,

где  – количество объектов с градaцией *d* в классе .

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

**4.7.** Berilgan } tanlanmadan foydalangan holda alomatlar fazosi o’lchamini dan ) gacha qisqartirish amalga oshirilsin. Tanlanma o’zaro kesishmaydigan , sinflarga bo’lingan va har bir obyekt *n* ta miqdoriy alomatlar to’plami bilan tavsiflangan. alomatni o’chirishda

mezonidan foydalanilsin. Bu еrda

, .

**4.7.** Произвести сокращение размерности признакового пространства с до ) используя выборку объектов }. Выборка разделена на два непересекающихся класса , и каждый объект описывается количественными признаками из набора . Для удаления признака использовать критерий

,

где , .

**4.8.** O‘zaro kesishmaydigan *K1*, *K2* sinflarga bo‘lingan } o‘rgatuvchi tanlanma berilgan. Har bir  – obyekt  ta nominal alomat bilan tavsiflangan. Nominal alomatlar gradatsiyalari bo’yicha yaqinlik matritsasi hisoblansin.

Obyektlar juftliklari (*Sa*=(*xa1*,...,*xan*), *Sb*=(*xb1*,...,*xbn*)) to’plamida



funksiya aniqlangan bo’lsin.

(*xi*,*xj*), *i,j*  {1,…,*n*} nominal alomatlar juftligi bo’yicha *Su*  *Kt*, *t*=1,2 bilan *K3-t* obyektlari o’rtasidagi yaqinlik (farqlanish) o’lchami

*bij*(*Su*)=

ko’rinishida beriladi. Bu еrda *li*(*xui*) (*lj*(*xuj*)) – *E*0 tanlanmadagi gradatsiyasi qiymati *xui* (*xuj*) bo’lgan obyektlar soni.

*B*(*Su*)={*bij*(*Su*)}*n\*n* matritsa elementlari bo’yicha obyektning *X*I =(*x*1,…,*xr*), *r*<*n* individual informativ alomatlar to’plamini tanlash uchun alomatlarning



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

*B*(*Su*) matritsadan eng katta *bij* qiymatiga ega (*xi*,*xj*) juftlik ajratib olinadi va ketma-ketlikka (chapdan o’ngga) qo’shiladi. (*xi*,*xj*) juftlikdagi joylashuv tartibi *li*(*xui*) ≥ *lj*(*xuj*) sharti bo’yicha aniqlanadi. Keyingi har juftlik xuddi shu tamoyilda *B*(*Su*) jadvaldan *i* va *j* nomerli satr va ustunlarni o’chirishdan keyin anialanadi.

**4.8.** По заданной обучающей выборке }, содержащей представителей -х непересекающихся классов объектов , произвести вычисление матрицы близости по градaциям номинальных признаков. Считается, что каждый объект описывается *n* номинальными признаками. На множестве пар объектов (*Sa*=(*xa1*,...,*xan*), *Sb*=(*xb1*,...,*xbn*)) определяются функции:



Мера близости (различия) между *Su*  *Kt*, *t*=1,2 и объектами из *K3-t* по паре номинальных признаков (*xi*,*xj*), *i,j*  {1,…,*n*} задаётся как

*bij*(*Su*)=

где *li*(*xui*) (*lj*(*xuj*)) – число объектов из *E*0 c градaцией равной *xui* (*xuj*).

Для выбора по элементам матрицы *B*(*Su*)={*bij*(*Su*)}*n\*n* индивидуального информативного набора *X*® =(*x*1,…,*xr*), *r*<*n* объекта требуется реализовать рекурсивную процедуру построения упорядоченной последовательности признаков

.

Из матрицы *B*(*Su*) выделяется пара (*xi*,*xj*) с наибольшим значением *bij* и включается (слева-направо) в последовательность. Порядок следования в (*xi*,*xj*) выбирается по условию *li*(*xui*) ≥ *lj*(*xuj*). Каждая следующая пара признаков по аналогичному принципу определяется из *B*(*Su*) после удаления в ней строк и столбцов с номерами *i* и *j*.

**4.9.** O‘zaro kesishmaydigan *K1*, *K2* sinflarga bo‘lingan } o‘rgatuvchi tanlanma berilgan. Har bir  obyekt  ta *X*(*n*)=(*x*1,…,*xn*) turli toifadagi (miqdoriy va nominal) alomatlar bilan tavsiflangan. *Si*,*Sj*∈*E*0 obyektlari o’rtasidagi yaqinlik masofisi o’lchovi ρ(*Si*,*Sj*) sifatida Juravlev metrikasidan qaralsin.

Aytaylik, *Si*∈*Kp* obyektdan ρ(*Si*,*Sj*) masofa bo’yicha kamaymaydigan holda tartiblangan

*S*1,*S*2,…,*Su*,…,*Sm*

ketma-ketlik aniqlangan bo’lsin. Yaqinlik matritsasi bo’yicha (*xa*,*xb*) ∈*X*(*n*) juftlik bahosi



ko’rinishida hisoblansin va informativ alomatlar izlashda birinchi qadamida  shartidan foydalansin.

**4.9.** Выборка объектов *E*0={*S*1,…,*Sm*} разделено на два непересекающихся класса *K*1 и *K*2. Каждый объект *S*∈*E*0 описываются набором разнотипных (количественных и номинальных) признаков *X*(*n*)=(*x*1,…,*xn*). В качестве меры расстояния ρ(*Si*,*Sj*) между объектами *Si*,*Sj*∈*E*0 используется метрика Журавлёва. Пусть

*S*1,*S*2,…,*Su*,…,*Sm*

последовательность объектов, упорядоченная по не убыванию расстояния ρ(*Si*,*Sj*) от объекта *Si*∈*Kp*, *dip*(*u*), *di*,3-*p*(*u*) – количество объектов класса *Kp*, *K*3-*p* из (1) до *Su* включительно. Требуется по мере расстояния вычислить оценку пар (*xa*,*xb*)∈*X*(*n*) как



и определить  в качестве первого шага для поиска информативных наборов признаков.

**4.10.** o’rgatuvchi tanlanma ikkita o’zaro kesishmaydigan sinf vakillarini o’z ichiga oladi. Tanlanmaning har bir obekti binary alomatlarning =(*x*1,…,*x*n) to’plami bilan tavsiflanadi.

Quyidagi keltirilgan algoritm bo’yicha obektlarning umumlashgan baholari hisiblansin.

Belgilash kiritaylik,  orqali mos ravishda *K*1 va *K*2 sinflar obektlari tavsifidagi *xc*∈*X(n)* alomat j∊{1,2} gradatsiya qiymatlari soni.

*Xc* alomat bo’yicha sinflar ora farqlanish



kattalik bilan aniqlanadi.

Alomatning K1, K2 sinflar bo’yicha garadatsiya qiymatlariga asoslangan birjinslik darajasi (sinf ichidagi o’xshashlik) βc quyidagi formula bilan hisoblanadi:



*xc*∈*X*(*n*) alomatning vazni *wс*=βсλс ko’rinishida hisoblanadi, j∈{1,2} gradatsiya hissasi esa



bilan aniqlanadi. Bu еrda – mos ravishda *K*1 и *K*2 sinflar obektlari tavsifidagi j∊{1,2} gradatsiya qiymatlari soni.

Mumkin bo’lgan *Sr*∈*E*0, *Sr*={*xri*} obyektning *X*(*d*)⸦*X*(*n*) tavsifi bo’yicha umumlashgan bahosi

ko’rinishida hisoblanadi.

Algoritmning qadamba-qadam amalga oshirilshi quyidagicha:

***1-qadam***. P={1,…,n}.

***2-qadam.*** Hisoblansin crit=10; 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; , . M1= M1 /|K1|, M2= M2 /|K2|. Θ=0. Γ=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(St)= R(St) + ηq(xtq).* Takrorlash tugashi. 3-qadamga o’tilsin. Aks holda {*R(St)*} chop etilsin;

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

***5-qadam.*** Tamom.

**4.10.** В обучающей выборке содержаться представители двух непересекающихся классов . Каждый объект выборки описывается набором =(x1,…,xn ) из бинарных признаков. Требуется вычислить обобщённые оценки объектов по приведённому ниже алгоритму.

Обозначим через  – количество значений градaции j∊{1,2} признака *x*c∈*X*(*n*) в описании объектов соответственно класса *K*1 и *K*2. Межклассовое различие по признаку xc определяется как величина



Степень однородности (мера внутриклассового сходства) βc значений градaций признака по классам *K*1, *K*2 вычисляется по формуле:



Вес признаку xc∈*X*(*n*) определяется как *ωс*=*βсλс*. Вклад градaции j∈{1,2} как



где  – количество значений градaции j признака xc соответственно в классах K1 и K2. Обобщённая оценка объекта S*r*∈E0, Sr={*xri*} по описанию на наборе X(d)⸦X(n) вычисляется как 

Реализaция алгоритма по шагам такая:

***Шаг 1.*** P={1,…,n}.

***Шаг 2.*** Вычислить crit=10; 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; , .

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*={*aij*}*m\*m*, *m* ≥4 matritsa ko‘rinishida berilgan bo‘lib, matritsaning elementlari *aij*∊{0,…,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*={*bij*}(*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 = 1·0 +4·2 + 3·1 + 0·1=11 (2-rasmga qarang).

|  |  |  |  |
| --- | --- | --- | --- |
| 10 | 42 | 4 | 3 |
| 31 | 01 | 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  va  yadrolardan foydalangan holda *m\*m* dan (*m-*2)*\**(*m*-2) o‘lchamgacha zichlashtiring.

**5.1.** Изображение на плоскости задано в виде матрицы *A*={*aij*}*m\*m*, *m* ≥4. Элементами матрицы *A* являются значения пикселей *aij*∊{0,…,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*={*bij*}(*m*-1)*\** (*m-*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-*2)*\**(*m*-2) путём последовательного использования ядер  и .

**5.2.** Tekislikdagi tasvir *A*={*aij*}*m\*m*, *m* ≥4 matritsa ko‘rinishida berilgan bo‘lib, matritsaning elementlari *aij*∊{0,…,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  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*={*aij*}*m\*m*, *m* ≥4. Элементами матрицы *A* являются значения пикселей *aij*∊{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  дважды провести последовательную свёртку изображения *A\** с размера (*m*+2)·(*m*+2) до *m*·*m*.

**5.3.** Tekislikdagi tasvir *A*={*aij*}*m\*m*, *m* ≥4 matritsa ko‘rinishida berilgan bo‘lib, matritsaning elementlari *aij*∊{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.

|  |  |  |  |
| --- | --- | --- | --- |
| 10 | 4 | 42 | 3 |
| 3 | 0 | 0 | 2 |
| 11 | 1 | 31 | 5 |
| 0 | 2 | 4 | 2 |

|  |  |
| --- | --- |
| 12 | × |
| × | × |

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*={*aij*}*m\*m*, *m* ≥4. Элементы матрицы (пиксели изображения) *aij*∊{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*={*aij*}*m\*m*, 3≤*m*≤5 matritsa ko‘rinishida beriladi. Matritsaning elementlari *aij*∊{0,…,5} piksellarning qiymatlaridan iborat. Tasvirni o‘girish uchun *C*(*k*)={*cij*}*k\*k*, *cij*∊{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,…,*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*={*aij*}*m\*m*, 3≤*m*≤5. Элементы матрицы *A* представлены значениями *aij*∊{0,…,5} пикселей изображения. Для преобразования изображения используется ядра свёрток в виде матриц *C*(*k*)={*cij*}*k\*k*, *cij*∊{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,…,*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* ≥4 o‘lchamdagi matritsa bo‘yicha subdiskretlash uchun matritsa *h*≥1 qadam bilan *d\*d*, 2≤*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,…,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. Субдискретизaция изображений: а) с шагом 1; б) с шагом 2)

Берилган иккита *C*={*cij*}*m\*m* ва *D*={*dij*}*m\*m* матрицалар учун 2\*2 ўлчамдаги дарча орқали 1 қадам билан тасвирларни субдискретлаш амалга оширилсин. *С* ва *D* матрицалар бўйича зичлаш натижаларини мос равишда *A*={*aij*}(*m*-1)*\**(*m*-1) ва *B*={*bij*}(*m*-1)*\**(*m*-1) матрицаларда ҳосил қилинсин. *А* ва *В* матрицалар ўзаро мос келмайдиган элементлари сони ­– *N*=│{(*i*,*j*)| *aij* ≠ *bij*}│ аниқлансин.

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

Для двух заданных матриц *C*={*cij*}*m\*m* и *D*={*dij*}*m\*m* произвести cубдискретизatsiя изображений через окно размера 2\*2 с шагом 1. Результаты сжатия по матрице *C* и *D* представить соответственно как *A*={*aij*}(*m*-1)*\**(*m*-1) и *B*={*bij*}(*m*-1)*\**(*m*-1). Определить число несовпадающих элементов матриц *A* и *B* как *N*=│{(*i*,*j*)| *aij* ≠ *bij*}│.

**5.6.** Tasvirni *m\*m*, *m*≥4 o‘lchamdagi matritsa bo‘yicha subdiskretlash uchun matritsa *h*≥1 qadam bilan *d\*d*, 2≤*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. 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. Субдискретизaция изображений: а) с шагом 1;б) с шагом 2).

Berilgan *A*={*aij*}*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 субдискретизaции (сжатия) изображения по матрице размера *m\*m*, *m* ≥4 производится разбиение его на окна размера *d\*d*, 2≤*d*<*m* с шагом *h*≥1. Максимальное значение в каждом окне выбирается в качестве значения элемента новой (сжатой) матрицы. Множеством допустимых значений элементов исходной матрицы является {0,…,5}. Пример субдискретизaции с окном размера 2\*2 и шагом 1 и 2 показан на рис. 5.

Для заданной матрицы *A*={*aij*}*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* ≥4 o‘lchamdagi matritsa bo‘yicha subdiskretlash uchun matritsa *h*≥1 qadam bilan *d\*d*, 2≤*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*={*cij*}*m\*m* va *D*={*dij*}*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*={*aij*}(*m*-1)*\**(*m*-1) va *B*={*bij*}(*m*-1)*\**(*m*-1) matritsalarda hosil qilinsin. *A* va *V* matritsalar o‘zaro mos kelmaydigan elementlari soni – *N*=│{(*i*,*j*)| *aij* ≠ *bij*}│ 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. Субдискретизaция изображений: а) с шагом 1;б) с шагом 2)

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

Для двух заданных матриц *C*={*cij*}*m\*m* и *D*={*dij*}*m\*m* произвести cубдискретизatsiя изображений через окно размера 2\*2 с шагом 1. Результаты сжатия по матрице *C* и *D* представить соответственно как *A*={*aij*}(*m*-1)*\**(*m*-1) и *B*={*bij*}(*m*-1)*\**(*m*-1). Определить число несовпадающих элементов матриц *A* и *B* как *N*=│{(*i*,*j*)| *aij* ≠ *bij*}│.

**5.8.** Tekislikdagi tasvir *A*={*aij*}*m\*m*, *m* ≥4 matritsa ko‘rinishida berilgan. A matritsaning *aij*∊{0,…,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.

|  |  |  |
| --- | --- | --- |
| 3 | 0 | 0 |
| 1 | 1 | 3 |
| 0 | 2 | 4 |

|  |  |
| --- | --- |
| 1 | 0 |
| 0 | 1 |

|  |  |  |
| --- | --- | --- |
| 3 | 0 | 0 |
| 1 | 1 | 3 |
| 0 | 2 | 4 |

|  |  |  |  |
| --- | --- | --- | --- |
| 1 | 4 | 4 | 3 |
| 3 | 0 | 0 | 2 |
| 1 | 1 | 3 | 5 |
| 0 | 2 | 4 | 2 |

|  |
| --- |
| 1 |
| -1 |

|  |  |
| --- | --- |
| 0 | 1 |
| 1 | 1 |

|  |  |  |
| --- | --- | --- |
| 3 | 0 | 0 |
| 1 | 1 | 3 |
| 0 | 2 | 4 |

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*={*aij*}*m\*m*, *m* ≥4. Элементами матрицы *A* являются значения пикселей *aij*∊{0,…,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 (Визуализaция данных)

**6.1.** Berilgan tasoddifiy kattaliklar qiymatlari to’plami bo’yicha protsentil taqsimot grafigi chizilsin.

**6.1.** По заданному множеству значений случайной величины построить график процентильного распределения.

**6.2.** Berilgan tasoddifiy kattaliklar qiymatlari to’plami bo’yicha detsil taqsimot grafigi chizilsin.

**6.2.** По заданному множеству значений случайной величины построить график децильного распределения.

**6.3.** Miqdoriy*X*(*n*)=(*x*1,…,*xn*) alomatlar bilan tavsiflangan *E*0 = (*S*1,…,*Sm*) tanlanma obyektlari *A*={*aij*}*m×n* jadval ko’rinishida berilgan. Har bir *xi*∈*X*(*n*) alomat bo’yicha mediana qiymati *M*(*xi*) hisoblangan bo’lsin.

Alomatlar juftligi {*xi* *xj*}∈*X*(*n*) uchun talab qilinadi:

1. Tekislikni *x*=*M*(*xi*), *y*=*M*(*xj*) 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 *xi*∈*X*(*n*) alomati qiymatini *M*(*xi*) mediana bilan ustma-ust tushmasligini ta’minlash uchun hisob paytida *M\**(*xi*)= *M*(*xi*) + ε qiymatidan foydalanilsin. Bu еrda ε (ε>0) – еtarlicha kichik kattalik.



**6.3.** В заданной таблице *A*={*aij*}*m×n* содержатся описание объектов *E*0 = (*S*1,…,*Sm*) по *n* количественным признакам *X*(*n*)=(*x*1,…,*xn*). Считается, что по каждому признаку *xi*∈*X*(*n*) вычислено значение медианы *M*(*xi*).

Требуется по каждой паре {*xi* *xj*}∈*X*(*n*):

1.Определить разбиение плоскости на 4 части по линиям *x*=*M*(*xi*), *y*=*M*(*xj*) 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 *xi*∈*X*(*n*) объекта и медианы *M*(*xi*), при расчёте необходимо использовать значение *M\**(*xi*)= *M*(*xi*) + ε, где ε – бесконечно малая величина, ε>0.



**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:

,

bu yerda *dij*, 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 *yik* uchun rekkurient munosabat bilan bog‘liqdir:



**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

,

где *dij*, 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 *yik*:



# 7. Matematik lingvistika (Математическая лингвистика)

**7.1.** (*So’zlarning semantic yaqinligi*). Tabiiy tildagi *T* – matnli hujjat *m* (*m*>1) ta gaplar to’plami *T*={*tj*}*j*∈{1,…,*m*} ko’rinishida berilgan. Gap tugashini {.,?,!} to’plamga kiruvchi belgilar bildiradi. Matn tahlili *n*(*n*>1) kalit so’zlar lug’ati *A*={*ai*}*i*∈{1,…,*n*} orqali amalga oshiriladi. Hujjatga dastlabki ishlov berish – gapdagi *ai*,*aj*∈*tk*, *i*≠*j*, *k*=1,…,*m* so’zlar juftligi uchrashi bo’yicha shakllanuvchi *R*={*rij*}*n×n* matrisani hosil qilishga keltiriladi. Matritsaning nol bo’lmagan elementlari

*rij*=|{ *ai*,*aj* ∈*tk* | pos(*tk*, *ai*) < pos(*tk*, *aj*), *k*=1,…,*m*}|

ko’rinishida hisoblanadi. Bu еrda pos(*tk*, *ai*) (pos(*tk*, *aj*)) – *ai*,(*aj*) so’zning *tk* gapdagi joylashuv o’rni. Agar (*ai*,*aj*) juftlik uchun pos(*tk*, *ai*) < pos(*tk*, *aj*) shart bajariluvchi *tk*∈*T* gap mavjud bo’lmasa *rij* =0 bo’ladi.

Talab qilinadi:

* so’zlarning juftligi uchrashini ko’rsatuvchi *R* matritsasini faqat nol bo’lmagan *rij*>0 elementlarini (*i*,*j*,*rij*) uchlik ko’rinishida yozish orqali ixcham ko’rinishda yozish, bu еrda *i*–satr nomeri, *j*– ustun nomeri;
* boshqa so’zlar bilan juftlikda uchrash chastotasi *rij* + *rji* = max bo’lgan *ai*∈*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*={*tj*}*j*∈{1,…,*m*}. Концом предложения является символ из множества {.,?,!}. Анализ текста реализуется через словарь из *n*(*n*>1) ключевых слов *A*={*ai*}*i*∈{1,…,*n*}. Предобработка документа сводится к формированию матрицы *R*={*rij*}*n×n* попарной встречаемости слов *ai*,*aj*∈*tk*, *i*≠*j*, *k*=1,…,*m* в предложениях, ненулевые значения элементов которой вычисляются как

*rij*=|{ *ai*,*aj* ∈*tk* | pos(*tk*, *ai*) < pos(*tk*, *aj*), *k*=1,…,*m*}|,

где pos(*tk*, *ai*) (pos(*tk*, *aj*)) – позиция слова *ai*,(*aj*) в предложении *tk*. Если для пары (*ai*,*aj*) не существует *tk*∈*T*, для которого pos(*tk*, *ai*) < pos(*tk*, *aj*), то *rij* =0.

Требуется:

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

– определить слово(а) *ai*∈*A*, частота встречаемости которого(ых) в паре с другими словами *rij* + *rji* = max.

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

Berilgan *T1*va *T2* hujjatlar uchun talab qilinadi:

– har bir *ai*,*bi*∈*D* so’zlar juftligi uchun *r*1*i*= *count*(*ai*,*bi*,*T1*), *r*2*i*= *count*(*ai*,*bi*,*T2*) hisoblansin;

– *T1*va *T2* hujjatlar bog’langanligi



ko’rinishida hisoblansin.

**7.2.** (*Семантическая связанность документов*). Текстовые документы *T*1и *T*2 на естественном языке заданы множествами соответственно из *m*1=|*T*1| и *m*2=|*T*2| (*m*1>1, *m*2>1) предложений. Концом предложения является символ из множества {.,?,!}. Анализ документов реализуется через набор из *n*(*n*>1) пар слов *D*={*ai*,*bi*}*i*∈{1,…,*n*}. Считается, что определена функция *count*(*x*,*y*,*T*) для подсчета числа предложений документа *T*, содержащих пару слов *x*,*y*∈*D*.

Требуется:

* определить значения *r*1*i*= *count*(*ai*,*bi*,*T*1), *r*2*i*= *count*(*ai*,*bi*,*T*2) по каждой паре слов *ai*,*bi*∈*D*;
* вычислить связанность документов *T*1 и *T*2 как



**7.3.** (*Trigramm uchrashi chastotasi*). *R*={‘*a’*,’*b’*,’*c’*,’*d’*,’*f’*} belgilar to’plamidan shakllangan *m* (*m*>1) ta, ixtiyoriy uzunlikdagi *P*={*pi*}*i*∈{1,…,*m*} zanjirlar to’plami berilgan bo’lsin. Elementlari {*xyz*}, *x*,*y*,*z*∈*R* bo’lgantrigrimmlar to’plamida ularni uchrash chastotalarini |{ *xyz*∈*pi* | *x*≠*y*, *x*,*y*∈*R*\{*z*}, *i*=1,…,*m*}| ko’rinishida hisoblash aniqlangan bo’lsin.

Berilgan *P* zanjirdagi har bir *z*∈*R* bo’yicha uchrash chastotasi maksimal bo’lgan *xyz* ko’rinishidagitrigrimmlar aniqlansin.

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

|{ *xyz*∈*pi* | *x*≠*y*, *x*,*y*∈*R*\{*z*}, *i*=1,…,*m*}|.

Требуется определить триграммы вида *xyz*, частоты встречаемости которых по каждой *z*∈*R* на *P* максимальны.

**7.4.** Tabiiy tildagi *n* ta so‘zlarning asoslari (lug‘at) *L*=(*c*1,…,*cn*) to‘plami bilan berilgan. Matn hujjati uchun *ai* =*di*/*dum* – uchrashlar chastotasi bilan *A*=(*a*1,…,*an*) vektor-satrni qurish talab qilinadi. Bu yerda *di* ­− hujjatdagi lug‘atga kiruvchi *ci*∈*L* so‘zning uchrashlari soni, *dum* − lug‘atdagi hujjatga kiruvchi so‘zlarning umumiy soni.

**7.4.** Задан набор (словарь) *L*=(*c*1,…,*cn*) из *n* основ слов естественного языка. Требуется для текстового документа построить вектор-строку *A*=(*a*1,…,*an*) с частотой встречаемости *ai* =*di*/*dum*, где *di* – число встречаемости слова *ci*∈*L* из словаря, *dum* − общее количество слов из словаря в документе.

**7.5.** Tabiiy tildagi *n* ta so‘zlarning asoslari (lug‘at) *L*=(*c*1,…,*cn*) to‘plami bilan berilgan. *R*=(*D*1,…,*Dm*) kolleksiyada berilgan *m* matn hujjatlari *L* lug‘atgakiruvchi *n* so‘zlar uchrashlari chastotalari bilan berilgan. *Di* hujjatdagi *ci*∈*L* so‘zning kirishlar sonining *L* lug‘atdagi so‘zlarning umumiy soniga nisbati *dij*∈*Di* qiymati ko‘rinishida hisoblanadi.

Matn hujjatlarning vektor tavsifining {(*Di*,*Dj*)} juftliklar to‘plami ichida kosinus masofasi bo‘yicha eng kichik qiymatga ega juftlik topilsin:



**7.5.** Задан набор (словарь) *L*=(*c*1,…,*cn*) из *n* основ слов естественного языка. В коллекции *R*=(*D*1,…,*Dm*) каждый из *m* текстовых документов описывается через частоту встречаемости *n* слов из словаря *L*. Значение *dij*∈*Di* вычисляется как количество (доля) вхождений слова *ci*∈*L* относительно общего количества слов из словаря *L* в документе *Di*.

Требуется по множеству пар векторных представлений текстовых документов {(*Di*,*Dj*)} выбрать пару с наименьшим значением косинусного расстояния



**7.6.** Tabiiy tildagi *n* ta so‘zlarning asoslari (lug‘at) *L*=(*c*1,…,*cn*) to‘plami bilan berilgan. *R*=(*D*1,…,*Dm*) kolleksiyada berilgan *m* matn hujjatlari *L* lug‘atgakiruvchi *n* so‘zlar uchrashlari chastotalari bilan berilgan. *Di* hujjatdagi *ci*∈*L* so‘zning kirishlar sonining *L* lug‘atdagi so‘zlarning umumiy soniga nisbati *dij*∈*Di* qiymati ko‘rinishida hisoblanadi.

Matn hujjatlarning vektor tavsifining {(*Di*,*Dj*)} juftliklar to‘plami ichida Xelinger masofasi bo‘yicha eng kichik qiymatga ega juftlik topilsin:



**7.6.** Задан набор (словарь) *L*=(*c*1,…,*cn*) из *n* основ слов естественного языка. В коллекции *R*=(*D*1,…,*Dm*) каждый из *m* текстовых документов описывается через частоту встречаемости *n* слов из словаря *L*. Значение *dij*∈*Di* вычисляется как количество (доля) вхождений слова *ci*∈*L* относительно общего количества слов из словаря *L* в документе *Di*.

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



**7.7.** Tabiiy tildagi *n* ta so‘zlarning asoslari (lug‘at) *L*=(*c*1,…,*cn*) to‘plami bilan berilgan. Berilgan *m* matn hujjatlaridagi *L* lug‘atga kiruvchi so‘zlar uchrashlari *R*=(*D*1,…,*Dm*) 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,…,*cn*) из *n* основ слов естественного языка. В коллекции *R*=(*D*1,…,*Dm*) 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,…,*cn*) to‘plami bilan berilgan. Berilgan *m* matn hujjatlaridagi *L* lug‘atga kiruvchi so‘zlar uchrashlari *R*=(*D*1,…,*Dm*) 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,…,*cn*) из *n* основ слов естественного языка. В коллекции *R*=(*D*1,…,*Dm*) 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,…,*cn*) to‘plami bilan berilgan. Berilgan *m* matn hujjatlaridagi *L* lug‘atga kiruvchi so‘zlar uchrashlari *R*=(*D*1,…,*Dm*) 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 *Di*∈*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 *Di* hujjatnikidan kichik bo‘lsa.

*Izoh:* *R* kolleksiyadagi hujjatga navola *Di*∈*R* hujjatning *i* indeksining qiymati bilan aniqlanadi.

**7.9.** Задан набор (словарь) *L*=(*c*1,…,*cn*) из *n* основ слов естественного языка. В заданном наборе *R*=(*D*1,…,*Dm*) 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 *Di*∈*R* определить перечень документов из *R* корректно или некорректно ссылающиеся на *Di*. Некорректной считается ссылка у того документа, год издания у которого меньше чем у *Di*.

*Примечание:* Ссылка на документ из *R* определяется значением индекса *i* документа *Di*∈*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,…,*Dm*) matn hujjatlari to‘plamida ikkita o‘zaro kesishmaydigan sinf va *l* ta guruhlarga (mavzularga) bo‘lingan. *Di*∊*T* hujjat *Di* =(*hi*,*gi*) ko‘rinishida bo‘lib, unda *hi* − *K*1 yoki *K*2 sinf nomeri, *hi*∊{1,2} va *gi*∊{1,…,*l*}, *l*≥2 – o‘zaro kesishmaydigan *G*1,…,*Gl* guruhlar nomerlari.

Talab qilinadi:

- har bir *Gi*guruh uchun ob’ektlarning *K*1 sinfga tegishligi funksiyasi  ko‘rinishida aniqlansin. Bu yerda *vi*1 – *Gi*guruhdagi *K*1 sinf ob’ektlari soni;

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



**7.11.** На множестве текстовых документов *T*=(*D*1,…,*Dm*) задано разбиение на два непересекающихся класса и *l* групп (тем). Документ *Di*∊*T* представлен как *Di* =(*hi*,*gi*), где *hi* – номер класса *K*1 или *K*2, *hi*∊{1,2} и *gi*∊{1,…,*l*}, *l*≥2 – номера непересекающихся групп *G*1,…,*Gl*. Требуется:

–для каждой группе *Gi* определить значение функции принадлежности объектов к классу *K*1 по *Gi* как , где *vi*1 – число объектов класса *K*1 в *Gi*.

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



**7.12.** Berilgan *T*=(*D*1,…,*Dm*) 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 *Tc*⸦*T* quyidagi ko‘rinishda aniqlanadi:

,

 .

Ikkita *Di*, *Dj*∊*Ku* ob’ektlar *D*∊*Ku*∩*Tc* bo‘yicha bog‘langan hisoblanadi, agar ρ(*Di*,*D*)<*ri* va ρ(*Dj*,*D*)<*rj* 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,…,*Dm*) 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 *Tc* на *T*,**определяется как

,

 .

Объекты *Di*, *Dj*∊*Ku* связаны по *D*∊*Ku*∩*Tc*, если ρ(*Di*,*D*)<*ri* и ρ(*Dj*,*D*)<*rj*.

Требуется определить минимальное количество групп из связанных объектов на *T*.

**7.13.** Berilgan *E*0=(*S*1,…,*Sm*) 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: *ρ*1(*x*,*y*) − yevklid va *ρ*2(*x*,*y*) − vaznli yevklid. Vaznli yevklid metrikasi berilgan *W*=(*w*1,…,*wn*) bo‘yicha  ko‘rinishida hisoblanadi, bu yerda 0 <*wi* ≤1. Har bir *S*∊*Kt*∩*E*0, *t*=1,2 ob’ekt uchun *ρ*1(*x*,*y*) va *ρ*2(*x*,*y*) metrikalar bo‘yicha kamayish bo‘yicha tartiblangan  va  ketma-ketliklar qurilgan bo‘lsin. P1(*S*) va P2(*S*) ketma-ketlikdagi pastki indekslar ob’ektlar rangini bildiradi.

Talab qilinadi:

− har bir *S*∊*Kt*∩*E*0 ob’ekt uchun P1(*S*) va P2(*S*) ketma-ketlikdagi *Kt* 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,…,*Sm*) разделено на два непересекающихся подмножества (класса) *K*1 и *K*2. Каждый объект *S*∊*E*0 описывается набором из *n* количественных признаков *X*(*n*). На множестве *E*0 определены две метрики: евклидова ρ1(*x*,*y*) и взвешенная евклидова ρ2(*x*,*y*). Взвешенная евклидова метрика вычисляется как  по заданному вектору весов *W*=(*w*1,…,*wn*), где 0 <*wi* ≤1.

Для каждого объекта *S*∊*Kt*∩*E*0, *t*=1,2 построены упорядоченные по убывания расстояния ρ1(*x*,*y*) и ρ2(*x*,*y*) последовательности  и . Нижний индекс в последовательностях П1(*S*) и П2(*S*) определяет ранг объекта.

Требуется:

– для каждого *S*∊*Kt*∩*E*0 вычислить среднее число рангов объектов класса *Kt* в последовательностях П1(*S*) и П2(*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,…,*Sm*) to‘plamida ob’ektlar *X*(*n*)=(*x*1,…,*xn*) − miqdoriy alomatlar bilan tavsiflangan. Ixtiyoriy *Si*∈*E*0 ob’ektning *Sj* ob’ektga nisbatan shartli ehtimolligi



bilan va mos ravishda *Sj*∈*E*0 ob’ektning *Si* ob’ektga nisbatan shartli ehtimolligi



bilan hisoblanadi. Bu yerda ρ(*Si*,*Sj*) – evklid metrikasi,

, .

Markazi *Si* ob’ektda bo‘lgan gauss zichligi  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,…,*Sm*) с помощью набора количественных признаков *X*(*n*)=(*x*1,…,*xn*). Условная вероятность объекта *Si*∈*E*0 относительно *Sj* вычисляется как



и соответственно *Sj*∈*E*0 относительно *Si*

,

где ρ(*Si*,*Sj*) – евклидова метрика, , .

Гауссовская плотность с центром в объекте *Si* определяется как  Требуется определить объект (объекты) с максимальной гауссовской плотностью.

**7.15.**  (*Kulbak-Leybler mertikasi*).

Berilgan *E*0=(*S*1,…,*Sm*) to‘plamida ob’ektlar *X*(*n*)=(*x*1,…,*xn*) − miqdoriy alomatlar bilan tavsiflangan. Ixtiyoriy *Si*∈*E*0 ob’ektning *Sj* ob’ektga nisbatan shartli ehtimolligi



bilan va mos ravishda *Sj*∈*E*0 ob’ektning *Si* ob’ektga nisbatan shartli ehtimolligi



bilan hisoblanadi. Bu yerda ρ(*Si*,*Sj*) – yevklid metrikasi,

, .

Markazi *Si* ob’ektda bo‘lgan gauss zichligi  ko‘rinishida aniqlanadi.

Talab qilinadi::

– *E*0  tanlanmada *X*(*n*) bo‘yicha *Gn* va *X*(*l*) bo‘yicha *Gl* zichlik aniqlansin, *n*>*l*;

– *Gn* va *Gl* orasidagi masofa  ko‘rinishidagi Kulbak–Leybler metrikasi orqali hisoblansin.

**7.15.**  (*Метрика Кульбака–Лейблера*).

Дано описание множества объектов *E*0=(*S*1,…,*Sm*) с помощью набора количественных признаков *X*(*n*)=(*x*1,…,*xn*). Для анализа структуры отношений объектов используется вычисление условной вероятности объекта *Si*∈*E*0 относительно *Sj* как



и соответственно *Sj*∈*E*0 относительно *Si*

,

где ρ(*Si*,*Sj*) – метрика Евклида, , . Гауссовская плотность с центром в объекте *Si* определяется как 

Требуется:

– определить на *E*0 гауссовскую плотность *Gn* по *X*(*n*) и *Gl* по *X*(*l*), *n*>*l*;

– вычислить расстояние Кульбака–Лейблера между *Gn* и *Gl* как



**7.16** (*Invariantlik*).

Berilgan *E*0=(*S*1,…,*Sm*) to‘plamida ob’ektlar *X*(*n*)=(*x*1,…,*xn*), *n*>2 − miqdoriy alomatlar bilan tavsiflangan.

*E*0 tanlanmada berilgan ρ(*x*,*y*) metrika uchun chiziqli tartib hosil qilinadi ((*i*,*j*)≤(*k*,*l*))≡ (ρ(*Si*,*Sj*) ≤ ρ(*Sk*,*Sl*)). *E*0 tanlanmadagi chiziqli tartiblashlar to‘plamini Δ(ρ)={((*i*,*j*)≤(*k*,*l*))≡ (ρ(*Si*,*Sj*) ≤ ρ(*Sk*,*Sl*))} bilan belgilaymiz.

Talab qilinadi:

– Δ(ρ) quvvat hisoblansin;

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

**7.16** (И*нвариантность*).

Объекты выборки данных *E*0=(*S*1,…,*Sm*) описываются с помощью набора количественных признаков *X*(*n*)=(*x*1,…,*xn*), *n*>2. Для заданной метрики ρ(*x*,*y*) на *E*0 индуцируется линейный порядок ((*i*,*j*)≤(*k*,*l*))≡ (ρ(*Si*,*Sj*) ≤ ρ(*Sk*,*Sl*)). Обозначим через Δ(ρ)={((*i*,*j*)≤(*k*,*l*))≡ (ρ(*Si*,*Sj*) ≤ ρ(*Sk*,*Sl*)) } – множество линейных порядков на *E*0.

Требуется:

– вычислить мощность Δ(ρ);

–доказать инвариантность Δ(ρ) монотонно возрастающим преобразованиям значений расстояний по множеству Δ(ρ\*), где *ρ*\*= ρ/(1+ ρ) 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*={*ai*}*i*∈{1,…,*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*={*ai*}*i*∈{1,…,*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 *Si* va *Sj*, *i*,*j*∈{1,…,*n*} jumlalar so’zlari termalarining juftligi bo’yicha o’xshashlik



ko’rinishida aniqladi. Bu еrda *w*– jumladagi so’z. Barcha (*i*,*j*)∈{1,…,*n*} juftliklar bo’yicha o’xshsashlikni maksimal qiymati aniqlansin.

**7.19** (*Попарное сходство предложений*) Вычислить попарное сходство наборов слов из предложений *Si* и *Sj*, *i*,*j*∈{1,…,*n*} как

,

где w– слово в предложении. Определить максимальное значение сходства по всем парам (*i*,*j*)∈{1,…,*n*}.

**7.20** (*TF-IDF bo’yicha jumlalar o’xshsashlik o’lchovi*). *S*={*Sj*}*j*∈{1,…,*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

,

,



Barcha (*i*,*j*)∈{1,…,*n*} juftliklar bo’yicha o’xshashlikning maksimal qiymati aniqlansin.

**7.20** (*Мера схожести предложений по* TF-IDF). Определён набор из предложений ***S***={*Sj*}*j*∈{1,…,*n*} и параметр *t* для зануления ребер графа ниже определенного порога *t*.

Схожесть предложений определяется по следующим формулам

,

,



Определить максимальное значение сходства по всем парам (*i*,*j*)∈{1,…,*n*}.

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



formulasi bilan hisoblanadi. Bu еrda *nt* – *t* so’zning hujjatga kirishlari soni, maxraj esa ushbu hujjatdagi barcha so’zlar miqdori.

,

bu еrda |*D*|– kolleksiyadagi jujjatlar soni; |{*di*∈*D*|t∈*di*}| – *D* kolleksiyadagi *t* uchraydigan hujjatlar soni (*nt*≠0 holati uchun).

*TF-IDF* o’lchovi ikkita had ko’paytmasi ko’rinishida hisoblanadi:

*Tf-idf*(*t*,*d*,*D*) = *tf*(*t*,*d*)·*idf*(*t*,*D*).

**7.21** (*Мера* TF(*term frequency*)-IDF(*inverse document* *frequency*)) Оценить важность слова *ti* в пределах отдельного документа

,

где *nt* – число вхождений слова *t* в документ, а в знаменателе – общее число слов в данном документе.

,

где |*D*|– число документов в коллекции; |{*di*∈*D*|t∈*di*}| – число документов из коллекции D, в которых встречается *t* (когда *nt*≠0).

Мера TF IDF является произведением двух сомножителей:

*Tf-idf(t,d,D)= tf(t,d)·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



ko’rinishda aniqlalandi. Bu еrda *nt* – 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 классификatsiи 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

,

где *nt* – 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 еchimi*}.

Alohida bir hujjat chegarasida *t* bigram muhimligi



ko’rinishda aniqlalandi. Bu еrda *nt* – *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

,

где *nt* – 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.А.Ignatьev, Sh.F. Madraximov R.N. Usmonov. Berilganlarning intellektual tahlili. Oʼquv qoʼllanma.