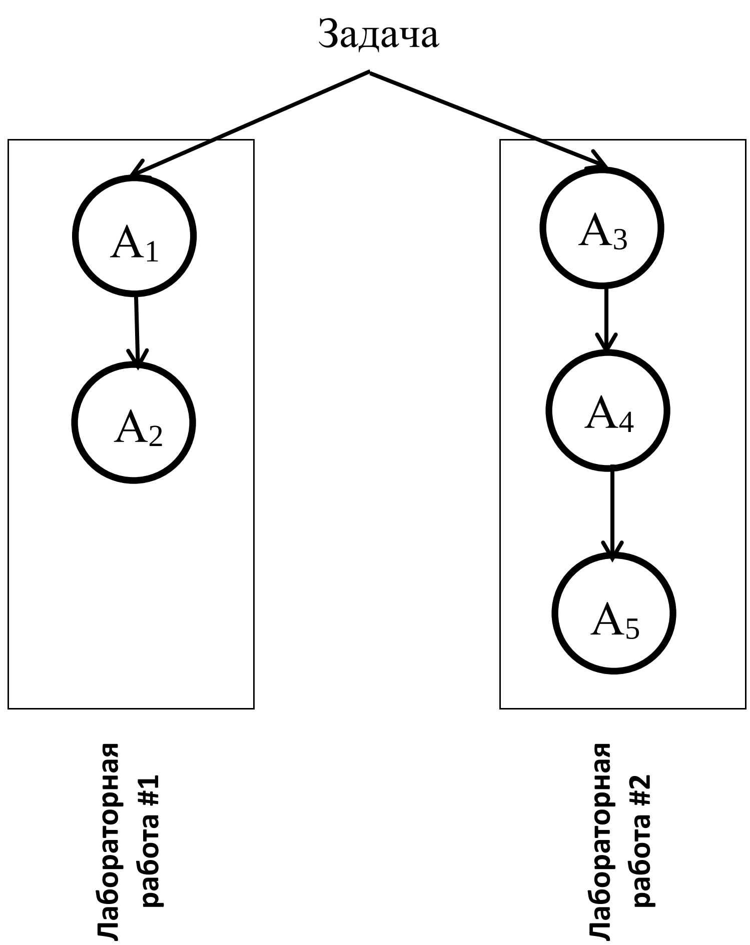
ЗАДАНИЕ ПО ЛАБОРАТОРНОЙ РАБОТЕ

1. Выбрать задачу (используя репозитории машинного обучения).
2. Написать программу на основе общей схемы метрических алгоритмов. В каждой программе необходимо реализовать **5** алгоритмов, решающих задачу из п.1.
3. Схема сдачи лабораторной работы с демонстрацией соответствующих результатов имеет следующий вид



4. Выбор алгоритмов A_i (*i=1,…,5*) зависит от выбора задачи. Если задача поставлена на множестве действительных чисел $\mathbb{R}^n$, то соответствие алгоритмов следующее:

$$A_1 = A_{\mathbf{I.1}}, A_2 = A_{\mathbf{I.3}}, A_3 = A_{\mathbf{II.1}}, A_4 = A_{\mathbf{II.2}}, A_5 = A_{\mathbf{II.3}}.$$

   В случае, когда задача поставлена в булевом пространстве $\mathbb{B}_2^n$, то соответствие алгоритмов следующее:

$$A_1 = A_{\mathbf{I.2}}, A_2 = A_{\mathbf{I.3}}, A_3 = A_{\mathbf{II}\mathit{I}\mathbf{.1}}, A_4 = A_{\mathbf{III.2}}, A_5 = A_{\mathbf{III.3}}.$$

5. График сдачи лабораторных работ согласовывается дополнительно.

ОБЩАЯ СХЕМА МЕТРИЧЕСКИХ АЛГОРИТМОВ РАСПОЗНАВАНИЯ

ЗАДАНЫ

– некоторое множество объектов X, разбитое на подмножества (классы) $X_1, \dots, X_l$ ($l \in \mathbb{N}$). Причем, классы не пересекаются $X_i \cap X_j = \emptyset \ \forall i \neq j$ ($i, j \in \{1, \dots, l\}$).
– выборка объектов $X^0 \subset X$, которая удовлетворяет условию $X^0 \cap X_i \neq \emptyset$ ($\forall i \in \{1, \dots, l\}$). Кроме того, для каждого объекта $x \in X^0$ известна (определена) информация о принадлежности к классам $X_1, \dots, X_l$. Эта информация задается в виде **информационного вектора** $P(x) = (P_1(x), \dots, P_l(x))$, компоненты которого определяются следующим образом

$$P_i(x) = \begin{cases} 1, \text{если } x \in X_i, \\ 0, \text{иначе.} \end{cases}$$

Шаг 0 (предварительный)

Разбиение выборки X^0 на две части

Шаг может повторяться некоторое число раз. Каждый раз выборка X^0 разбивается на две части $X^0_{\text{обуч}}$ и $X^0_{\text{контр}}$ (называются соответственно **обучающей** и **контрольной** выборками). Выборки должны удовлетворять следующим условиям:

$$X^0_{i\,(\text{о})} \stackrel{\text{def}}{=} X^0_{\text{обуч}} \cap X^0_i \neq \emptyset \ \ \forall i \in \{1, \dots, l\},\ \bigcup_{i=1}^{l} X^0_{i\,(\text{о})} = X^0_{\text{обуч}};$$

$$X^0_{i\,(\text{к})} \stackrel{\text{def}}{=} X^0_{\text{контр}} \cap X^0_i \neq \emptyset \ \ \forall i \in \{1, \dots, l\},\ \bigcup_{i=1}^{l} X^0_{i\,(\text{к})} = X^0_{\text{контр}}.$$

Обозначим: $\left|X^0_{i\,(\text{о})}\right| = m_i$, $\left|X^0_{i\,(\text{к})}\right| = t_i$, $\sum_{i=1}^{l} m_i = \left|X^0_{\text{обуч}}\right| = m$, $\sum_{i=1}^{l} t_i = \left|X^0_{\text{контр}}\right| = t$.

Предположим, что в результате шага 0 получено некоторое конечное число разбиений

$$\{X^0_{\text{обуч}}, X^0_{\text{контр}}\}^1, \{X^0_{\text{обуч}}, X^0_{\text{контр}}\}^2, \dots, \{X^0_{\text{обуч}}, X^0_{\text{контр}}\}^k$$

Ограничение:

Число разбиений, полученных на шаге 0должно быть не менее трех ($k \geq 3$). Все выборки $\{X^0_{\text{обуч}}, X^0_{\text{контр}}\}^j$ должны удовлетворять следующему дополнительному условию: $t_i \,/\, m_i \geq 0.2$. После построения всех выборок $\{X^0_{\text{обуч}}, X^0_{\text{контр}}\}^j$, они фиксируются для последующего применения на каждой из них всего набора алгоритмов A_i

---

Далее алгоритм $A_i, i = 1,2,3,4,5$ реализуется как последовательность шагов 1-3.

Шаг 1. Определение функции попарного сравнения объектов из X:

$$s : X \times X \to \mathbb{R} \tag{1}$$

Каждый объект $x \in X$ можно сравнить с объектами из выборки $X^0_{\text{обуч}}$. В результате объекту x можно сопоставить вектор $(s(x, x_1), \dots, s(x, x_m))$ m_1 первых компонент являются результатом сравнения x с объектами из $X^0_{1(\text{о})}$, m_2 следующих являются результатом сравнения x с объектами из $X^0_{2(\text{о})}$ и т.д.

Шаг 2. Определение функции сравнения объектов из $x \in X$ с объектами из обучающей выборки $X^0_{i\,(\text{о})}$:

$$f_i : \mathbb{R}^m \to \mathbb{R},\ i \in \{1, \dots, l\} \tag{2}$$

С помощью функций (2) каждому вектору $(s(x,x_1),\dots,s(x,x_m))\in\mathbb{R}^m$ можно сопоставить вектор $(f_1(x),\dots,f_l(x))\in\mathbb{R}^l$.

Шаг 3. Определение решающего правила P^A в виде:

$$P^A\colon\mathbb{R}^l\to\mathbb{B}_2^l,\ \mathbb{B}_2=\{0,1\} \tag{3}$$

Вектор $P^A(x)=(P_1^A(x),\dots,P_l^A(x))$ в отличие от вектора $P(x)$ называют обычно **классификационным**, тк его значения интерпретируются следующим образом. Считается, что объект $x\in X$ заносится (или не заносится) алгоритмом A в класс X_i, если $P_i^A(x)=1\ (=0)$.

Шаг 4. Тестирование определенного на шагах 1-3 алгоритма распознавания A_i на фиксированной выборке $\{X^0_{\text{обуч}},X^0_{\text{контр}}\}^j, j=1,2,\dots,k$.

Инициализация: полагаем $t^0(X^0_{\text{контр}})$=0.

Шаг 4.1. Последовательно перебираем все объекты $x\in X^0_{\text{контр}}$ и для каждого из них вычисляем:

a) вектор $(s(x,x_1),\dots,s(x,x_m))$ для всех $x\in X^0_{\text{обуч}}$;
b) вектор $(f_1(x),\dots,f_l(x))$ для всех $i\in\{1,\dots,l\}$;
c) вектор $P^A(x)=(P_1^A(x),\dots,P_l^A(x))$
d) если $P^A(x)=\ P(x)$, то $t^0\left(X^0_{\text{контр}}\right)=t^0\left(X^0_{\text{контр}}\right)+1$ и переходим к пункту e). В противном случае $t^0\left(X^0_{\text{контр}}\right)$ не меняется и переходим к пункту e).
e) если не все объекты $x\in X^0_{\text{контр}}$ исчерпаны, то выполняем шаг 4.1 для следующего объекта контрольной выборки. В противном случае переходим к шагу 4.2.

Шаг 4.2. Вычисляем

$$\Phi_A\left(X^0_{\text{контр}}\right)=\frac{t^0\left(X^0_{\text{контр}}\right)}{t} \tag{4}$$

величину **функционала качества** $\Phi_A\left(X^0_{\text{контр}}\right)\in[0,1]$ и заносим ее в таблицу:

Разбиение / Алгоритм	$\{X^0_{\text{обуч}},X^0_{\text{контр}}\}^1$	$\{X^0_{\text{обуч}},X^0_{\text{контр}}\}^2$	……	$\{X^0_{\text{обуч}},X^0_{\text{контр}}\}^k$
A_1				
A_2				
A_3				
A_4				
A_5				

Если все выборки $\{X^0_{\text{обуч}},X^0_{\text{контр}}\}^j$ исчерпаны, то переходим к выбору следующего алгоритма A_i. В противном случае выбираем новую выборку $\{X^0_{\text{обуч}},X^0_{\text{контр}}\}^j$ и возвращаемся на шаг 4.

ВАРИАНТЫ ВЫБОРА ФУНКЦИЙ ДЛЯ ШАГОВ 1-3 СХЕМЫ МЕТРИЧЕСКИХ АЛГОРИТМОВ

ВАРИАНТ I. *Метрическая близость*.

Алгоритм I.1 - $A_{\mathbf{I.1}}$ **.** (для пространства $X \subseteq \mathbb{R}^n$)

Выбор функции (1).

– метрика Евклида

$$s(x_1, x_2) = \left(\sum_{i=1}^{n} (x_{1i} - x_{2i})^2\right)^{1/2}$$

– метрика Минковского ($p \in \mathbb{N}$)

$$s(x_1, x_2) = \left(\sum_{i=1}^{n} (x_{1i} - x_{2i})^p\right)^{1/p}$$

Выбор функции (2).

– среднее (м.б. взвешенное) расстояние до класса X_i

$$f_i(x) = (m_i)^{-1} \sum_{x_j \in X_{i\,(\text{o})}^0} s(x, x_j)$$

– k ближайших соседей

пусть для класса X_i получен набор $s\left(X_{i\,(\text{o})}^0\right) = \{s(x, x_1), \dots, s(x, x_{m_i})\}$;

переупорядочим набор $s\left(X_{i\,(\text{o})}^0\right)$ по **возрастанию** элементов и поставим ему в соответствие новый набор $\bar{X}_{i\,(\text{o})}^0$, в котором содержится k первых элементов $x_j \in X_{i\,(\text{o})}^0$ из полученного в результате переупорядочения набора;

посчитаем среднее расстояние до класса по новому набору $\bar{X}_{i\,(\text{o})}^0$

$$f_i(x) = (k)^{-1} \sum_{x_j \in \bar{X}_{i\,(\text{o})}^0} s(x, x_j)$$

– по минимальному расстоянию до объектов класса X_i

$$f_i(x) = \min_{x_j \in X_{i\,(\text{o})}^0} s(x, x_j)$$

Выбор решающего правила (3).

– по минимуму оценки до класса X_i

$$P_i^A(x) = \begin{cases} 1, \text{если } f_i(x) = \min\limits_{i \in \{1,\dots,l\}} \{f_1(x), \dots, f_l(x)\}; \\ 0, \text{в противном случае} \end{cases}$$

Алгоритм I.2 - $A_{\mathbf{I.2}}$ **.** (для пространства $X \subseteq \mathbb{B}^n$)

Выбор функции (1).

– метрика Хэмминга ()

$$s(x_1, x_2) = \sum_{i=1}^{n} |x_{1i} - x_{2i}|$$

Выбор функции (2).

– среднее (м.б. взвешенное) расстояние до класса X_i

$$f_i(x) = (m_i)^{-1} \sum_{x_j \in X^0_{i\,(\mathrm{o})}} s(x, x_j)$$

– k ближайших соседей

пусть для класса X_i получен набор $s(X^0_{i\,(\mathrm{o})}) = \{s(x, x_1), \dots, s(x, x_{m_i})\}$;

переупорядочим набор $s(X^0_{i\,(\mathrm{o})})$ по **возрастанию** элементов и поставим ему в соответствие новый набор $\bar{X}^0_{i\,(\mathrm{o})}$, в котором содержится k первых элементов $x_j \in X^0_{i\,(\mathrm{o})}$ из полученного в результате переупорядочения набора;

посчитаем среднее расстояние до класса по новому набору $\bar{X}^0_{i\,(\mathrm{o})}$

$$f_i(x) = (k)^{-1} \sum_{x_j \in \bar{X}^0_{i\,(\mathrm{o})}} s(x, x_j)$$

– по минимальному расстоянию до объектов класса X_i

$$f_i(x) = \min_{x_j \in X^0_{i\,(\mathrm{o})}} s(x, x_j)$$

Выбор решающего правила (3).

– по минимуму оценки до класса X_i

$$P_i^A(x) = \begin{cases} 1, \text{если } f_i(x) = \min\limits_{i \in \{1,\dots,l\}} \{f_1(x), \dots, f_l(x)\}; \\ 0, \text{в противном случае} \end{cases}$$

Алгоритм I.3 - $A_{\mathbf{I.3}}$ **.** *метрическое сходство* (для пространств $X \subseteq \mathbb{R}^n$ и $X \subseteq \mathbb{B}^n$)

Предварительный шаг. Для каждого подмножества

$$X_i^0 = X^0 \cap X_i \ (|X_i^0| = m_i)$$

и выбранной в результате построения алгоритмов $A_{\mathbf{I.1}}$ или $A_{\mathbf{I.2}}$ функции $s(x_1, x_2)$ (для упрощения дальнейшего изложения переобозначим ее через $\bar{s}(x_1, x_2)$) находим диаметр множества X_i^0 по следующей формуле

$$d_i = \max_{(x_u, x_v) \in X_i^0 \times X_i^0} \bar{s}(x_u, x_v)$$

В результате получим следующий набор – $(d_1, d_2, \dots, d_l)$, для которого вычисляем величину

$$d = max(d_1, d_2, \dots, d_l)$$

Замечание: величина d никак не зависит от разбиений, т.к. определяется на подмножествах X_i^0 $(i \in \{1, \dots, l\})$.

Выбор функции (1).

$$s(x_1, x_2) = max\{0, 1 - \frac{\bar{s}(x_1, x_2)}{d}\}$$

Выбор функции (2).

– среднее (м.б. взвешенное) расстояние до класса X_i

$$f_i(x) = (m_i)^{-1} \sum_{x_j \in X^0_{i\,(\mathrm{o})}} s(x, x_j)$$

– k ближайших соседей

пусть для класса X_i получен набор $s(X^0_{i\,(\mathrm{o})}) = \{s(x, x_1), \dots, s(x, x_{m_i})\}$;

переупорядочим набор $s\left(X^0_{i\,(\text{o})}\right)$ по **убыванию** элементов и поставим ему в соответствие новый набор $\bar{X}^0_{i\,(\text{o})}$, в котором содержится k первых элементов $x_j \in X^0_{i\,(\text{o})}$ из полученного в результате переупорядочения набора;

посчитаем среднее расстояние до класса по новому набору $\bar{X}^0_{i\,(\text{o})}$

$$f_i(x) = (k)^{-1} \sum_{x_j \in \bar{X}^0_{i\,(\text{o})}} s\left(x, x_j\right)$$

– по максимальному расстоянию до объектов класса X_i

$$f_i(x) = \max_{x_j \in X^0_{i\,(\text{o})}} s\left(x, x_j\right)$$

Выбор решающего правила (3).

– по максимуму оценки до класса X_i

$$P_i^A(x) = \begin{cases} 1, \text{если } f_i(x) = \max\limits_{i \in \{1,\dots,l\}} \{f_1(x), \dots, f_l(x)\}; \\ 0, \text{в противном случае} \end{cases}$$

ВАРИАНТ II. *Метрическое подобие* (для пространства $X \subseteq \mathbb{R}^n_+$, $\mathbb{R}_+ \stackrel{\text{def}}{=} \{x | x \in \mathbb{R}, x > 0\}$).

Алгоритм II.1 - $A_{\mathbf{II.1}}$. *Метрическое подобие*

Выбор функции (1).

– аддитивное подобие

$$s(x_1, x_2) = \left(\sum_{i=1}^n a_i \cdot x_{1i}\right) \cdot \left(\sum_{i=1}^n a_i \cdot x_{1i}\right)^{-1}, \left(a_i \in \mathbb{R}_+, \sum_{i=1}^n a_i = 1\right)$$

– мультипликативное подобие

$$s(x_1, x_2) = \left(\left(\prod_{i=1}^n x_{1i}\right) \cdot \left(\prod_{i=1}^n x_{2i}\right)^{-1}\right)^{1/n}$$

Выбор функции (2).

– среднее геометрическое (м.б. взвешенное) подобие для объектов класса X_i

$$f_i(x) = \Big(\prod_{x_j \in X^0_{i\,(\text{o})}} s\left(x, x_j\right)\Big)^{(m_i)^{-1}}$$

Выбор решающего правила (3).

– по минимуму оценки до класса X_i

$$P_i^A(x) = \begin{cases} 1, \text{если } f_i(x) = \max\limits_{i \in \{1,\dots,l\}} \{\min\{f_1(x), f_1^{-1}(x)\}, \dots, \min\{f_l(x), f_l^{-1}(x)\}\}; \\ 0, \text{в противном случае} \end{cases}$$

Алгоритм II.2 - $A_{\mathbf{II.2}}$. *Метрическая близость на основе подобия*

Выбор функции (1).

– (пусть $\bar{s}(x_1, x_2)$ функция подобия, выбранная при построении алгоритма $A_{\mathbf{II.1}}$

$$s(x_1, x_2) = 1 - \min\{\bar{s}(x_1, x_2), \bar{s}(x_2, x_1)\}$$

Выбор функции (2).

– среднее (м.б. взвешенное) расстояние до класса X_i

$$f_i(x) = (m_i)^{-1} \sum_{x_j \in X^0_{i\,(o)}} s(x, x_j)$$

– k ближайших соседей

пусть для класса X_i получен набор $s\left(X^0_{i\,(o)}\right) = \{s(x, x_1), \dots, s\left(x, x_{m_i}\right)\}$;

переупорядочим набор $s\left(X^0_{i\,(o)}\right)$ по **возрастанию** элементов и поставим ему в соответствие новый набор $\bar{X}^0_{i\,(o)}$, в котором содержится k первых элементов $x_j \in X^0_{i\,(o)}$ из полученного в результате переупорядочения набора;

посчитаем среднее расстояние до класса по новому набору $\bar{X}^0_{i\,(o)}$

$$f_i(x) = (k)^{-1} \sum_{x_j \in \bar{X}^0_{i\,(o)}} s(x, x_j)$$

– по минимальному расстоянию до объектов класса X_i

$$f_i(x) = \min_{x_j \in X^0_{i\,(o)}} s(x, x_j)$$

Выбор решающего правила (3).

– по минимуму оценки до класса X_i

$$P_i^A(x) = \begin{cases} 1, \text{если } f_i(x) = \min\limits_{i \in \{1,\dots,l\}} \{f_1(x), \dots, f_l(x)\}; \\ 0, \text{в противном случае} \end{cases}$$

Алгоритм II.3 - $A_{\mathbf{II.3}}$. *Метрическое сходство на основе подобия*

Предварительный шаг. Для каждого подмножества

$$X_i^0 = X^0 \cap X_i \;(|X_i^0| = m_i)$$

и выбранной в результате построения алгоритма $\mathrm{A}_{\mathbf{II.2}}$ функции $s(x_1, x_2)$ (для упрощения дальнейшего изложения переобозначим ее через $\bar{s}(x_1, x_2)$) находим диаметр множества X_i^0 по следующей формуле

$$d_i = \max_{(x_u, x_v) \in X_i^0 \times X_i^0} \bar{s}(x_u, x_v)$$

В результате получим следующий набор – $(d_1, d_2, \dots, d_l)$, для которого вычисляем величину

$$d = max(d_1, d_2, \dots, d_l)$$

Выбор функции (1).

$$s(x_1, x_2) = max\{0{,}1 - \frac{\bar{s}(x_1, x_2)}{d}\}$$

Выбор функции (2).

– среднее (м.б. взвешенное) расстояние до класса X_i

$$f_i(x) = (m_i)^{-1} \sum_{x_j \in X^0_{i\,(o)}} s(x, x_j)$$

– k ближайших соседей

пусть для класса X_i получен набор $s\left(X^0_{i\,(o)}\right) = \{s(x, x_1), \dots, s\left(x, x_{m_i}\right)\}$;

переупорядочим набор $s\left(X_{i\,(\text{o})}^0\right)$ по **убыванию** элементов и поставим ему в соответствие новый набор $\bar{X}_{i\,(\text{o})}^0$, в котором содержится k первых элементов $x_j \in X_{i\,(\text{o})}^0$ из полученного в результате переупорядочения набора;

посчитаем среднее расстояние до класса по новому набору $\bar{X}_{i\,(\text{o})}^0$

$$f_i(x) = (k)^{-1} \sum_{x_j \in \bar{X}_{i\,(\text{o})}^0} s(x, x_j)$$

– по максимальному расстоянию до объектов класса X_i

$$f_i(x) = \max_{x_j \in X_{i\,(\text{o})}^0} s(x, x_j)$$

Выбор решающего правила (3).

– по максимуму оценки до класса X_i

$$P_i^A(x) = \begin{cases} 1, \text{если } f_i(x) = \max\limits_{i \in \{1,\dots,l\}} \{f_1(x), \dots, f_l(x)\}; \\ 0, \text{в противном случае} \end{cases}$$

ВАРИАНТ III. *Метрическая прецедентность*. (для пространства $\subseteq \mathbb{B}_2^n$, $\mathbb{B}_2 = \{0,1\}$)

Алгоритм III.1 - $A_{\mathbf{III.1}}$. *Метрическая прецедентность*

Подготовительный этап: (подсчет параметров алгоритма)

Шаг 1. Фиксируем номер класса $i \in \{1, \dots, l\}$ и переходим к шагу 2.

Шаг 2. Для всех признаков $j \in \{1, \dots, n\}$ вычисляем:

$$b_{ij} = (m_i)^{-1} \Big(\sum_{x_t \in X_i^{0,\text{обуч}}} x_{tj} \Big)$$

где x_{tj} - значение признака j в векторе $x_t \in X_{i\,(\text{o})}^0$ $(t \in \{1, \dots, m_i\})$

Шаг 3. Шаги 1&2 выполняем до тех пор, пока все номера классов и все признаки в каждом классе не будут исчерпаны. Затем переходим к шагу 4.

Шаг 4. Для всех признаков $j \in \{1, \dots, n\}$ и классов $i \in \{1, \dots, l\}$ вычисляем:

$$b_j = (l)^{-1} \Big(\sum_{i=1}^{l} b_{ij} \Big)$$

$$a_{ij} = \left| b_{ij} - b_j \right| \tag{5}$$

Выбор функции (1).

Функция $s(x_1, x_2)$ зависит от параметров (5). При описании этой зависимости в формуле ниже предполагаем, что x_1 - произвольный объект из X , $x_2 \in X_{i\,(\text{o})}^0$ и для всех таких x_2 вычисляем

$$s(x_1, x_2) = \Big(\sum_{j=1}^{n} a_{ij} \Big)^{-1} \times \Big(\sum_{j=1}^{n} (-1)^u \times a_{ij} \Big)$$

где

$$u = \begin{cases} 1, \text{если } x_{1j} \neq x_{2j} \\ 2, \text{иначе} \end{cases}$$

Выбор функции (2).

Для каждого $i \in \{1, \dots, l\}$ вычисляем

$$f_i(x) = \max_{x_t \in X^0_{i\,(\text{o})}} \{s(x, x_t)\}$$

Выбор решающего правила (3).

– по максимуму оценки до класса X_i

$$P_i^A(x) = \begin{cases} 1, \text{если } f_i(x) = \max\limits_{i \in \{1,\dots,l\}} \{f_1(x), \dots, f_l(x)\}; \\ 0, \text{в противном случае} \end{cases}$$

– пороговое решающее правило ($c_0, c_1 \in [0,1], c_0 \leq c_1$).

Алгоритм III.2 - $A_{\mathbf{III.2}}$. *Метрическая близость на основе прецедентности*

Выбор функции (1).

– пусть $\bar{s}(x_1, x_2)$ функция прецедентности, выбранная при построении алгоритма $\text{А}_{\mathbf{III.1}}$

$$s(x_1, x_2) = 1 - \bar{s}(x_1, x_2)$$

Выбор функции (2).

– среднее (м.б. взвешенное) расстояние до класса X_i

$$f_i(x) = (m_i)^{-1} \sum_{x_j \in X^0_{i\,(\text{o})}} s(x, x_j)$$

– k ближайших соседей

пусть для класса X_i получен набор $s\left(X^0_{i\,(\text{o})}\right) = \{s(x, x_1), \dots, s\left(x, x_{m_i}\right)\}$;

переупорядочим набор $s\left(X^0_{i\,(\text{o})}\right)$ по **возрастанию** элементов и поставим ему в соответствие новый набор $\bar{X}^0_{i\,(\text{o})}$, в котором содержится k первых элементов $x_j \in X^0_{i\,(\text{o})}$ из полученного в результате переупорядочения набора;

посчитаем среднее расстояние до класса по новому набору $\bar{X}^0_{i\,(\text{o})}$

$$f_i(x) = (k)^{-1} \sum_{x_j \in \bar{X}^0_{i\,(\text{o})}} s(x, x_j)$$

– по минимальному расстоянию до объектов класса X_i

$$f_i(x) = \min_{x_j \in X^0_{i\,(\text{o})}} s(x, x_j)$$

Выбор решающего правила (3).

– по минимуму оценки до класса X_i

$$P_i^A(x) = \begin{cases} 1, \text{если } f_i(x) = \min\limits_{i \in \{1,\dots,l\}} \{f_1(x), \dots, f_l(x)\}; \\ 0, \text{в противном случае} \end{cases}$$

Алгоритм III.3 - $A_{\mathbf{III.3}}$. *Метрическое сходство на основе прецедентности*

Предварительный шаг. Для каждого подмножества

$$X_i^0 = X^0 \cap X_i \ (|X_i^0| = m_i)$$

и выбранной в результате построения алгоритма $\text{А}_{\mathbf{III.2}}$ функции $s(x_1, x_2)$ (для упрощения дальнейшего изложения переобозначим ее через $\bar{s}(x_1, x_2)$) находим диаметр множества X_i^0 по следующей формуле

$$d_i = \max_{(x_u,x_v)\in X_i^0\times X_i^0} \bar{s}(x_u, x_v)$$

В результате получим следующий набор – $(d_1, d_2, \dots, d_l)$, для которого вычисляем величину

$$d = max(d_1, d_2, \dots, d_l)$$

Выбор функции (1).

$$s(x_1, x_2) = max\{0{,}1 - \frac{\bar{s}(x_1, x_2)}{d}\}$$

Выбор функции (2).

– среднее (м.б. взвешенное) расстояние до класса X_i

$$f_i(x) = (m_i)^{-1} \sum_{x_j\in X_{i\,(\text{o})}^0} s(x, x_j)$$

– k ближайших соседей

пусть для класса X_i получен набор $s\left(X_{i\,(\text{o})}^0\right) = \{s(x, x_1), \dots, s\left(x, x_{m_i}\right)\}$;

переупорядочим набор $s\left(X_{i\,(\text{o})}^0\right)$ по **убыванию** элементов и поставим ему в соответствие новый набор $\bar{X}_{i\,(\text{o})}^0$, в котором содержится k первых элементов $x_j \in X_{i\,(\text{o})}^0$ из полученного в результате переупорядочения набора;

посчитаем среднее расстояние до класса по новому набору $\bar{X}_{i\,(\text{o})}^0$

$$f_i(x) = (k)^{-1} \sum_{x_j\in \bar{X}_{i\,(\text{o})}^0} s(x, x_j)$$

– по максимальному расстоянию до объектов класса X_i

$$f_i(x) = \max_{x_j\in X_{i\,(\text{o})}^0} s(x, x_j)$$

Выбор решающего правила (3).

– по максимуму оценки до класса X_i

$$P_i^A(x) = \begin{cases} 1, \text{если } f_i(x) = \max\limits_{i\in\{1,\dots,l\}} \{f_1(x), \dots, f_l(x)\}; \\ 0, \text{в противном случае} \end{cases}$$