



ИПМ им.М.В.Келдыша РАН • Электронная библиотека

Препринты ИПМ • Препринт № 255 за 2018 г.



ISSN 2071-2898 (Print)
ISSN 2071-2901 (Online)

Гавриков Б.М., [Гавриков М.Б.](#),
Пестрякова Н.В., Ставицкий Р.В.

Структура базы обучения
статистического
классификатора состояний
систем организма человека

Рекомендуемая форма библиографической ссылки: Структура базы обучения статистического классификатора состояний систем организма человека / Б.М.Гавриков [и др.] // Препринты ИПМ им. М.В.Келдыша. 2018. № 255. 40 с. doi:[10.20948/prepr-2018-255](https://doi.org/10.20948/prepr-2018-255)
URL: <http://library.keldysh.ru/preprint.asp?id=2018-255>

**Ордена Ленина
ИНСТИТУТ ПРИКЛАДНОЙ МАТЕМАТИКИ
имени М.В.Келдыша
Российской академии наук**

**Б.М. Гавриков, М.Б. Гавриков, Н.В. Пестрякова,
Р.В. Ставицкий**

**Структура базы обучения
статистического классификатора
состояний систем организма человека**

Москва — 2018

Б.М. Гавриков, М.Б. Гавриков, Н.В. Пестрякова, Р.В. Ставицкий

Структура базы обучения статистического классификатора состояний систем организма человека

Исследуются свойства классификатора, основанного на методе полиномиальной регрессии, используемого для оценивания состояния здоровья человека по данным периферической крови. Проведен анализ структуры обучающей базы семи систем организма для мужчин и девяти для женщин. По каждому из четырех классов здоровья рассмотрены особенности взаимного расположения своих элементов, пар свой–чужой, а также своих и чужих элементов относительно центра масс. Показана связь между адекватной сложностью классификатора и структурой обучающего множества.

Ключевые слова: состояние здоровья человека, система организма, периферическая кровь, классификация, полиномиальная регрессия

Boris Mikhailovich Gavrikov, Mikhail Borisovich Gavrikov, Nadejda Vladimirovna Pestryakova, Roman Vladimirovich Stavitskii

The structure of the training base of the statistical classifier of the states of the human body systems

The properties of the classifier based on the polynomial regression method used to assess human health according to peripheral blood data are investigated. The structure of the training base of seven body systems for men and nine for women was analyzed. For each of the four health classes, the features of the mutual disposition of their elements are considered, their pairs are alien, as well as their own and alien elements relative to the center of mass. The relationship between the adequate complexity of the classifier and the structure of the training set is shown.

Key words: human health, body system, peripheral blood, classification, polynomial regression

Работа выполнена при поддержке Российского фонда фундаментальных исследований, проект 16-07-00742-а.

Введение

Рассматривается задача создания методов предварительной диагностики состояния здоровья человека (СЗЧ) по результатам анализа крови с использованием данных, полученных в процессе медицинского обследования значительного количества людей. Актуальность данной проблематики обусловлена необходимостью разработки математических методов предварительной диагностики в доказательной медицине. Применение именно статистического подхода определяется тем обстоятельством, что он является наиболее адекватным, учитывая наличие для каждой системы организма значительного объема данных, устанавливающих соответствие между классами здоровья и наборами показателей крови.

Для решения этой задачи был построен классификатор, позволяющий по параметрам периферической крови оценивать состояние каждой из систем организма (СО), используя четырехбалльную систему. Точность классификации на обучающих множествах составляет 95–100%.

В рамках исследования свойств разработанного классификатора необходимо, в частности, провести анализ обучающей базы. Проблема заключается в том, что до сих пор отсутствует стандартная методология и соответствующий инструментарий для изучения и описания конечных множеств объектов, используемых для обучения.

Актуальность этих работ связана с тем, что в публикациях по методам классификации (распознавания) наличие базы обучения рассматривается обычно как некая данность, а ее качество не подлежит сомнению. Однако вполне очевидно, что проблематично говорить о методе, его достоинстве и недостатках в отрыве от описания множества, на котором проходило обучение. Оно не только не является идеальным, но и может быть вообще непригодным для проведения этой процедуры. Результат обучения существенно зависит от его структуры. В то же время канонические принципы формирования и анализа таких множеств отсутствуют.

Постановка задачи классификации

Разработан классификатор, при помощи которого можно оценивать состояние здоровья различных СО человека по результатам лабораторного анализа периферической крови (из пальца). В основе лежит статистический

метод, использующий полиномиально-регрессионный подход и имеющий вероятностные оценки.

По данным анализа крови производится деление на четыре класса, которые соответствуют различным стадиям поражения СО человека:

- 1 класс – здоровые;
- 2 класс – начальные отклонения состояния здоровья;
- 3 класс – выраженное отклонение состояния здоровья;
- 4 класс – тяжелое заболевание.

Для каждой СО строится свой классификатор с использованием отдельного обучающего множества ([1, 5]). Наборы показателей крови практически здоровых людей одинаковы у всех СО. Базы данных для мужчин и женщин различаются.

Приведем общепринятые обозначения и размерность показателей крови:

- RBC [L⁻¹] – эритроциты,
- HGB [g L⁻¹] – гемоглобин,
- PLT [L⁻¹] – тромбоциты,
- WBC [L⁻¹] – лейкоциты,
- LIMPH [L⁻¹], [%] – лимфоциты,
- GRAN [L⁻¹], [%] – гранулоциты
(GRAN = NEUT + EOS + BASO).

Для фиксированной СО и пола определим, какой из четырех градаций СЗЧ соответствует результат анализа периферической крови. Перечень градаций СЗЧ есть множество с $K = 4$ элементами. Введем вектор $\mathbf{v} \in \mathbf{R}^N$, i -я компонента которого – отнормированная на отрезок [0,1] величина i -го показателя крови, причем $N = 8$.

Отождествляем k -й элемент множества градаций СЗЧ с базисным вектором $\mathbf{e}_k = (0 \dots 1 \dots 0)$ (здесь 1 находится на k -м месте, причем $1 \leq k \leq K$) из \mathbf{R}^K . Обозначаем $Y = \{\mathbf{e}_1, \dots, \mathbf{e}_K\}$.

Пусть можно найти $p_k(\mathbf{v})$ – вероятность того, что набор отнормированных показателей крови соответствует k -му элементу СЗЧ, где $1 \leq k \leq K$. Искомый элемент СЗЧ будет иметь порядковый номер, получивший максимальное значение вероятности:

$$p_r(\mathbf{v}) = \max_k \{p_k(\mathbf{v})\}, \quad 1 \leq k \leq K. \quad (1)$$

Приближенные значения $p_1(\mathbf{v}), \dots, p_K(\mathbf{v})$ представляются в виде конечных многочленов от координат $\mathbf{v} = (v_1, \dots, v_N)$ и определяются выбором базисных мономов:

$$p_k(\mathbf{v}) \cong c_0^{(k)} + \sum_{i=1}^N c_i^{(k)} v_i + \sum_{i,j=1}^N c_{i,j}^{(k)} v_i v_j + \dots, \quad 1 \leq k \leq K. \quad (2)$$

Представим упорядоченные базисные мономы из (2) в виде вектора размерности L :

$$\mathbf{x}(\mathbf{v}) = (1, v_1, \dots, v_N, \dots)^T.$$

Тогда (2) можно записать в векторном виде:

$$\mathbf{p}(\mathbf{v}) = (p_1(\mathbf{v}), \dots, p_K(\mathbf{v}))^T \cong A^T \mathbf{x}(\mathbf{v}), \quad (3)$$

где A – матрица размера $L \times K$, столбцами которой являются векторы $\mathbf{a}^{(1)}, \dots, \mathbf{a}^{(K)}$. Каждый такой вектор составлен из коэффициентов при мономах соответствующей строки (2) (с совпадающим верхним индексом), упорядоченных так же, как в векторе $\mathbf{x}(\mathbf{v})$.

Значение A вычисляется приближенно в процессе обучения, используя базу данных: $[\mathbf{v}^{(1)}, \mathbf{y}^{(1)}], \dots, [\mathbf{v}^{(J)}, \mathbf{y}^{(J)}]$. Здесь $\mathbf{v}^{(j)}$ – набор параметров крови, соответствующий элементу СЗЧ с номером k ($1 \leq k \leq K$), $\mathbf{y}^{(j)} = (0 \dots 1 \dots 0)$ – его базисный вектор, где 1 стоит на k -м месте, $1 \leq j \leq J$:

$$A \cong \left(\frac{1}{J} \sum_{j=1}^J \mathbf{x}^{(j)} (\mathbf{x}^{(j)})^T \right)^{-1} \left(\frac{1}{J} \sum_{j=1}^J \mathbf{x}^{(j)} (\mathbf{y}^{(j)})^T \right). \quad (4)$$

При получении правой части (4) используется рекуррентная процедура [2-5].

Рассматривались различные модификации вектора $\mathbf{x}(\mathbf{v})$. Указаны СО и пол, для которых при $\mathbf{x}(\mathbf{v})$ данного вида были получены наилучшие результаты, и более сложные модификации не использовались. Точность классификации приведена в случаях, когда она меньше 100%.

1) Длина полинома 33.

$$\mathbf{x} = (1, \{v_i\}, \{v_i^2\}, \{v_i^3\}, \{v_i^4\}, 1 \leq i \leq 8). \quad (5)$$

Имеются мономы степенного вида первого, второго, третьего и четвертого порядка. Перекрестные произведения отсутствуют.

2) Длина полинома 61. ЦНС, органы чувствительности (мужчины). Печень и желчевыводящие пути (мужчины).

$$\mathbf{x} = (1, \{v_i\}, \{v_i^3\}, \{v_i^4\}, \{v_i v_j\}), 1 \leq i \leq 8, i \leq j \leq 8. \quad (6)$$

Имеются мономы степенного вида первого, второго, третьего и четвертого порядка. Перекрестные произведения используются в качестве мономов второго порядка, а для более высоких порядков отсутствуют.

3) Длина полинома 69. Органы дыхания (женщины). ЦНС, органы чувствительности (женщины). Гинекологическая система (женщины).

$$\mathbf{x} = (1, \{v_i\}, \{v_i^3\}, \{v_i^4\}, \{v_i^5\}, \{v_i v_j\}), 1 \leq i \leq 8, i \leq j \leq 8. \quad (7)$$

Имеются мономы степенного вида первого, второго, третьего, четвертого и пятого порядка. Перекрестные произведения используются в качестве мономов второго порядка, а для более высоких порядков отсутствуют.

4) Длина полинома 77. *Опорно-двигательный аппарат (мужчины).*

$$\mathbf{x}=(1,\{v_i\},\{v_i^3\},\{v_i^4\},\{v_i^5\},\{v_i^6\},\{v_i v_j\}), 1\leq i\leq 8, i\leq j\leq 8. \quad (8)$$

Имеются мономы степенного вида первого, второго, третьего, четвертого, пятого и шестого порядка. Перекрестные произведения используются в качестве мономов второго порядка, а для более высоких порядков отсутствуют.

5) Длина полинома 85. *Печень и желчевыводящие пути (женщины).*

$$\mathbf{x}=(1,\{v_i\},\{v_i^3\},\{v_i^4\},\{v_i^5\},\{v_i^6\},\{v_i^7\},\{v_i v_j\}), 1\leq i\leq 8, i\leq j\leq 8. \quad (9)$$

Имеются мономы степенного вида первого, второго, третьего, четвертого, пятого, шестого и седьмого порядка. Перекрестные произведения используются в качестве мономов второго порядка, а для более высоких порядков отсутствуют.

6) Длина полинома 165. *Пищеварительная система (женщины). Органы дыхания (мужчины). Опорно-двигательный аппарат (женщины). Эндокринная система (женщины). Грудные железы (женщины).*

$$\mathbf{x}=(1,\{v_i\},\{v_i v_j\},\{v_i v_j v_k\}), 1\leq i\leq 8, i\leq j\leq 8, j\leq k\leq 8. \quad (10)$$

Имеются мономы первого, второго и третьего порядка. Перекрестные произведения используются для мономов второго и третьего порядка.

7) Длина полинома 495.

$$\mathbf{x}=(1,\{v_i\},\{v_i v_j\},\{v_i v_j v_k\},\{v_i v_j v_k v_l\}), 1\leq i\leq 8, i\leq j\leq 8, j\leq k\leq 8, k\leq l\leq 8. \quad (11)$$

Имеются мономы первого, второго, третьего и четвертого порядка. Перекрестные произведения используются для мономов второго, третьего и четвертого порядка.

8) Длина полинома 1287. *Пищеварительная система (мужчины – 98,2%). Урологическая система (женщины – 99,2%, мужчины – 97,3%). Эндокринная система (мужчины – 95%).*

$$\mathbf{x}=(1,\{v_i\},\{v_i v_j\},\{v_i v_j v_k\},\{v_i v_j v_k v_l\},\{v_i v_j v_k v_l v_m\}), \\ 1\leq i\leq 8, i\leq j\leq 8, j\leq k\leq 8, k\leq l\leq 8, l\leq m\leq 8. \quad (12)$$

Имеются мономы первого, второго, третьего, четвертого и пятого порядка. Перекрестные произведения используются для мономов второго, третьего, четвертого и пятого порядка.

В (5–12) выражения в фигурных скобках соответствуют цепочкам элементов вектора, вычисляемым по всем показателям крови из имеющегося набора.

Точность классификации на обучающих множествах для различных систем организма находится в диапазоне 95–100 %. Элементы класса «1» распознаются для всех СО при использовании полиномов наиболее простой структуры и минимальной длины.

Параметры структуры базы

В данной работе предлагается сравнительный анализ результатов, полученных для всех СО и обоих полов. Значимость этих исследований определяется тем, что наборы данных по каждому из описанных случаев для классов здоровья «2», «3» и «4» являются уникальными и не повторяются для других СО. Исключением являются множества параметров крови здоровых людей. Имеется только незначительное различие в количестве используемых наборов. В исходных базах этих данных намного больше, чем по остальным классам. Конкретное число элементов «1» класса, используемых в обучающем множестве, определяется из соображений приблизительного выравнивания по объему с ближним по заполненности классом.

Имеется еще одна существенная причина для детального изучения обучающих множеств всего набора СО для обоих полов. А именно, представляется важным выявить структуру данных по каждому из четырех имеющихся классов. Если для визуально наблюдаемых объектов, например, для изображений символов, наличие некой структурированности кажется понятным (при достаточно большой вариативности написания мы осознаем отличие изображений одного символа от другого), то для наборов показателей крови органы чувств человека оказываются бессильны. Мы делаем попытку «визуализировать» структуру, которую сумел «найти» метод статистической классификации. То, что он успешно справился с этой задачей, подтверждается высокой точностью классификации.

Пусть имеется обучающее множество некоторой СО и заданного пола.

Для каждого из четырех классов здоровья в отдельности найдем минимальное, максимальное и среднее расстояние между своими векторами (принадлежащими данному классу). Для множества векторов k -го класса определяем их следующим образом.

Минимальное расстояние:

$$U_{k_{\min}} = \min_{V^k} \{ \|v^k - u^k\| \}, v^k \in V^k, u^k \in V^k, v^k \neq u^k \quad (13)$$

Максимальное расстояние:

$$U_{k_{\max}} = \max_{V^k} \{ \|\mathbf{v}^k - \mathbf{u}^k\| \}, \mathbf{v}^k \in V^k, \mathbf{u}^k \in V^k \quad (14)$$

где \mathbf{v}^k и \mathbf{u}^k – пары различных векторов, принадлежащих множеству элементов k -го класса V^k .

Среднее расстояние определим с приведением алгоритма нахождения этой величины:

$$U_{k_{\text{cp}}} = \sum_{j=1}^{J_k} \sum_{j_1=j+1}^{J_k} \|\mathbf{w}^{k,j} - \mathbf{w}^{k,j_1}\| / (J_k (J_k - 1) / 2), \mathbf{w}^{k,j} \in V^k, j = 1, \dots, J_k, \quad (15)$$

где $\{\mathbf{w}^{k,j}, j = 1, \dots, J_k\} = V^k$ – представление совокупности элементов k -го класса в виде множества перенумерованных векторов.

Аналогично получим соответствующие значения для пар свой–чужой по каждому из классов. Чужой вектор – не принадлежащий рассматриваемому классу.

Минимальное расстояние:

$$U_{kz_{\min}} = \min_V \{ \|\mathbf{v}^k - \mathbf{u}^{-k}\| \}, \mathbf{v}^k \in V^k, \mathbf{u}^{-k} \in V^{-k} \quad (16)$$

Максимальное расстояние:

$$U_{kz_{\max}} = \max_V \{ \|\mathbf{v}^k - \mathbf{u}^{-k}\| \}, \mathbf{v}^k \in V^k, \mathbf{u}^{-k} \in V^{-k} \quad (17)$$

где \mathbf{v}^k и \mathbf{u}^{-k} – пары векторов, из которых \mathbf{v}^k принадлежит множеству элементов k -го класса V^k , а \mathbf{u}^{-k} принадлежит множеству чужих элементов V^{-k} классов, отличных от k -го: $V^{-k} = V \setminus V^k$.

Среднее расстояние:

$$U_{kz_{\text{cp}}} = \sum_{j=1}^{J_k} \sum_{j_1=1}^{J_{-k}} \|\mathbf{w}^{k,j} - \mathbf{w}^{-k,j_1}\| / (J_k J_{-k}), \mathbf{w}^{k,j} \in V^k, j = 1, \dots, J_k, \quad (18)$$

$$\mathbf{w}^{-k,j_1} \in V^{-k}, j_1 = 1, \dots, J_{-k}$$

где $\{\mathbf{w}^{k,j}, j = 1, \dots, J_k\} = V^k$ – представление совокупности своих элементов k -го класса в виде множества перенумерованных векторов, аналогично для множества чужих элементов классов, отличных от k -го: $\{\mathbf{w}^{-k,j_1}, j_1 = 1, \dots, J_{-k}\} = V^{-k}$, $V^{-k} = V \setminus V^k$.

Для каждого из четырех классов здоровья в отдельности получим среднестатистический вектор длины 8, принадлежащий исходному векторному пространству \mathbf{R}^8 . Иногда такой вектор называют центром масс.

Для центра масс k -го класса СЗЧ значение i -го параметра крови равно среднему арифметическому значений i -х параметров крови по всем J_k имеющимся в базе наборам показателей крови, относящихся к данному классу:

$$v_i^{k,cp} = \left(\sum_{j=1}^{J_k} v_i^{k,j} \right) / J_k, \quad (19)$$

где $\mathbf{v}^{k,j}$ – перенумерованные элементы k -го класса: $\{\mathbf{v}^{k,j}=(v^{k,j}_1, \dots, v^{k,j}_N), j = 1, \dots, J_k\} = V^k$. Рассматриваемое обучающее множество, содержащее элементы всех четырех классов, суть $V = \{V^1 \cup V^2 \cup V^3 \cup V^4\}$.

Для каждого из четырех классов здоровья в отдельности найдем минимальное, максимальное и среднее расстояние между центром масс и своими векторами.

Указанные величины для множества векторов k -го класса определяем следующим образом. Минимальное расстояние:

$$D_{k_{\min}} = \min_{V^k} \{ \|\mathbf{v}^{k,cp} - \mathbf{u}^k\| \}, \mathbf{u}^k \in V^k \quad (20)$$

Максимальное расстояние:

$$D_{k_{\max}} = \max_{V^k} \{ \|\mathbf{v}^{k,cp} - \mathbf{u}^k\| \}, \mathbf{u}^k \in V^k \quad (21)$$

где \mathbf{u}^k – вектор, принадлежащий множеству элементов k -го класса V^k , $\mathbf{v}^{k,cp}$ – среднестатистический вектор этого класса.

Среднее расстояние определим более детально с приведением алгоритма нахождения этой величины:

$$D_{k_{cp}} = \sum_{j=1}^{J_k} \|\mathbf{w}^{k,j} - \mathbf{v}^{k,cp}\| / J_k, \mathbf{w}^{k,j} \in V^k, j = 1, \dots, J_k \quad (22)$$

где $\{\mathbf{w}^{k,j}, j = 1, \dots, J_k\} = V^k$ – представление совокупности элементов k -го класса в виде множества перенумерованных векторов.

Аналогично получим соответствующие значения по каждому из классов между центром масс и чужими векторами. Минимальное расстояние:

$$D_{kZ_{\min}} = \min_{V^{-k}} \{ \|\mathbf{v}^{k,cp} - \mathbf{u}^{-k}\| \}, \mathbf{u}^{-k} \in V^{-k} \quad (23)$$

Максимальное расстояние:

$$D_{kz_{\max}} = \max_{V^{-k}} \{ \| \mathbf{v}^{k,cp} - \mathbf{u}^{-k} \| \}, \mathbf{u}^{-k} \in V^{-k} \quad (24)$$

где \mathbf{u}^{-k} – вектор, принадлежащий множеству чужих элементов V^{-k} классов, отличных от k -го: $V^{-k} = V \setminus V^k$, $\mathbf{v}^{k,cp}$ – среднестатистический вектор k -го класса.

Среднее расстояние:

$$D_{kz_{cp}} = \sum_{j=1}^{J-k} \| \mathbf{v}^{k,cp} - \mathbf{w}^{-k,j} \| / J_{-k}, \mathbf{w}^{-k,j} \in V^{-k}, j = 1, \dots, J-k \quad (25)$$

где $\{ \mathbf{w}^{-k,j}, j = 1, \dots, J-k \} = V^{-k}$, $V^{-k} = V \setminus V^k$ – представление совокупности чужих элементов классов, отличных от k -го, в виде множества перенумерованных векторов.

Диапазон расстояний между центром масс k -го класса СЗЧ и векторами этого же класса («своими», $\mathbf{v}^k \in V^k$) по рассматриваемой базе, согласно формулам (20), (21), находится на отрезке $[D_{k_{\min}}, D_{k_{\max}}]$. Диапазон расстояний между центром масс k -го класса СЗЧ и векторами всех других классов («чужими», $\mathbf{z}^k \in \{V \setminus V^k\}$), согласно формулам (23), (24), – на отрезке $[D_{kz_{\min}}, D_{kz_{\max}}]$. Пусть

$$\begin{aligned} Dk_{\min} &= \min(D_{k_{\min}}, D_{kz_{\min}}), \\ Dk_{\max} &= \max(D_{k_{\max}}, D_{kz_{\max}}). \end{aligned} \quad (28)$$

Делим отрезок $[Dk_{\min}, Dk_{\max}]$ на десять равных по длине частей – один отрезок и девять полуинтервалов: $[Dk_{\min}, Dk_{\min} + d]$, $(Dk_{\min} + d, Dk_{\min} + 2d]$, \dots , $(Dk_{\min} + 9d, Dk_{\min} + 10d]$, где $d = (Dk_{\max} - Dk_{\min})/10$. Определим, какое количество своих векторов попало в каждый такой участок (аналогично для чужих векторов). Затем рассмотрим распределение числа своих (чужих) векторов на отрезке $[Dk_{\min}, Dk_{\max}]$.

Везде далее расстояние между векторами определяем в метрике L_2 .

На рисунках 1–16 с поднумерацией а), б), в), г) соответственно для классов «1», «2», «3», «4» представлено минимальное, среднее и максимальное расстояние (значения ординат для точек 1, 2, 3 по оси абсцисс) между своими векторами (Ряд 1), аналогичные величины для пар свой–чужой (Ряд 3), между центром масс и своими векторами (Ряд 2), аналогично между парами центр масс–чужой вектор (Ряд 4).

На рисунках 1–16 с поднумерацией аа), бб), вв), гг) (классы «1», «2», «3», «4») изображено распределение числа своих (Ряд 1) и чужих (Ряд 2) элементов на отрезке $[Dk_{\min}, Dk_{\max}]$ (ось абсцисс).

Пищеварительная система

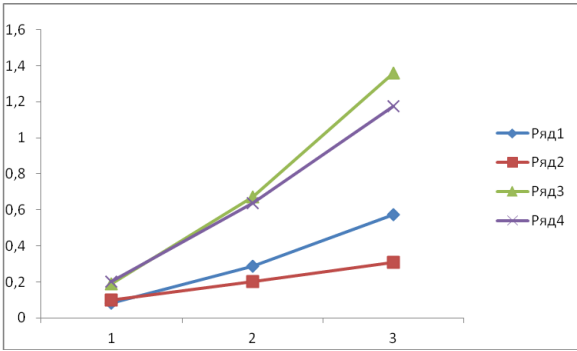
Рассмотрим пищеварительную систему для женщин (рис. 1). Обучающее множество в этом случае имеет 77 элементов. Классы «1» и «4» содержат по 24 набора крови, а классы «2» и «3» включают соответственно 9 и 20 элементов.

Как видно на рисунке 1а), минимальное, среднее и максимальное расстояние между своими векторами (Ряд 1) класса «1» меньше соответствующих расстояний между парами свой–чужой (Ряд 3). Из рисунков 1б), в), г) следует, что этот порядок или существенно нарушается для классов «2», «3», «4», или соответствующие величины сближаются, разница их значений уменьшается. Например, наименьшее расстояние между своими векторами класса «2» больше минимума расстояния между своими и чужими векторами (рис. 1б)); такое же соотношение имеется по классу «4» и для среднего, и для максимального расстояния (рис. 1г)). По классам «2» и «3» указанные средние величины примерно одинаковы (рис. 1б), в)); аналогично для минимальных величин класса «3» и класса «4» (рис. 1в), г)).

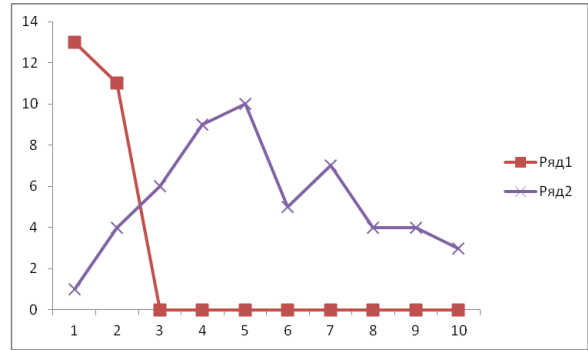
Минимальное, среднее и максимальное расстояние между центром масс и своими векторами (Ряд 2) класса «1» (рис. 1а)) меньше соответствующих расстояний между центром масс и чужими векторами (Ряд 4). Это соотношение для классов «2», «3», «4» (рис. 1б), в), г)) либо в значительной степени нарушается, либо различие между указанными величинами становится меньше. Так, минимальное расстояние между центром масс и своими векторами больше наименьшего расстояния между центром масс и чужими векторами класса «2» (рис. 1б)), аналогично и для среднего, и для максимального расстояния по классу «4» (рис. 1г)). Указанные средние величины для класса «2» примерно одинаковы (рис. 1б)), как и минимальные для классов «3» и «4» (рис. 1в), г)).

Вблизи центра масс класса «1» имеется небольшая окрестность, в которой находятся все его элементы (рис. 1аа), причем их число убывает при удалении от центра масс (Ряд 1). В то же время в этой окрестности чужих элементов мало, а подавляющее их количество располагается вне нее (Ряд 2); соответствующая функция распределения сначала нарастает при удалении от центра масс, а затем имеет тенденцию к убыванию, и ее максимум находится на удалении от этой окрестности, т.е. там, где отсутствуют элементы класса «1».

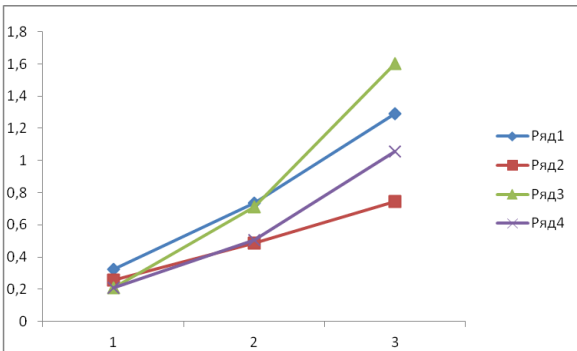
Для классов «2», «3», «4» (рис. 1бб), вв), гг)) функция распределения чужих элементов (Ряд 2) схожа с рис. 1аа), но ее максимум ближе к центру масс. Свои элементы не локализованы вблизи центра масс: они имеются до конца (рис. 1гг)) или почти до конца (рис. 1бб), вв)) отрезка $[Dk_{\min}, Dk_{\max}]$, а максимум у соответствующей функции распределения (Ряд 1) отсутствует.



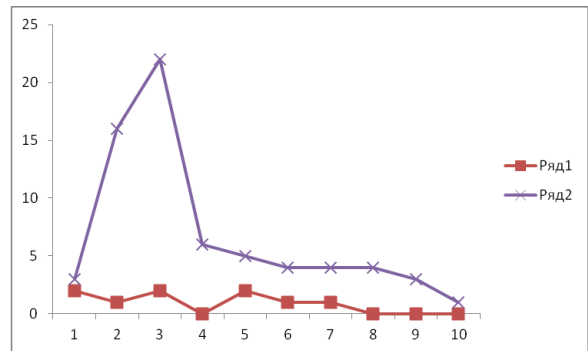
а) «1» min, middle, max



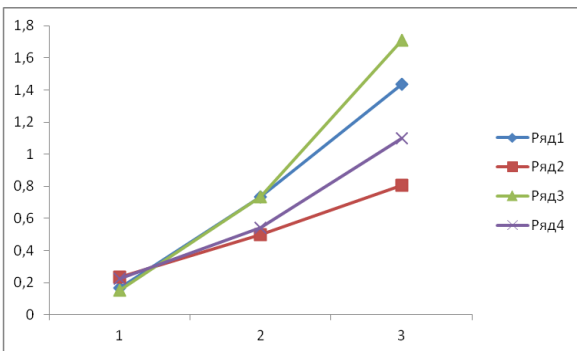
аа) «1» распределение своих/чужих



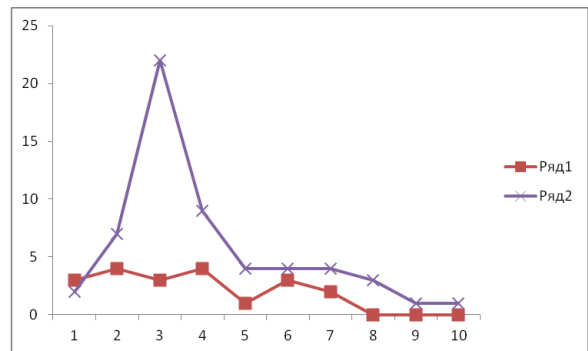
б) «2» min, middle, max



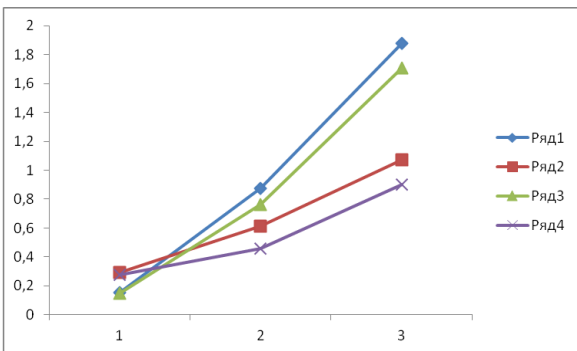
бб) «2» распределение своих/чужих



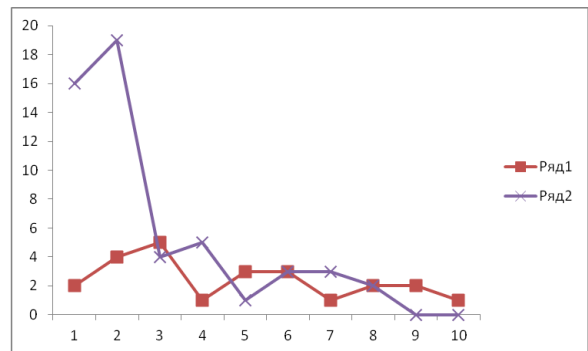
в) «3» min, middle, max



вв) «3» распределение своих/чужих



г) «4» min, middle, max



гг) «4» распределение своих/чужих

Рис. 1. Пищеварительная система (женщины)

Отмеченные здесь закономерности взаимного расположения своих и чужих элементов, а также характер распределения количества своих и чужих элементов по классам «1», «2», «3», «4» вполне соответствуют и могут служить объяснением того факта, что для распознавания элементов класса «1» достаточно использовать полиномы, имеющие малую длину. Классы «2», «3», «4» являются более проблемными в этом отношении. Для их распознавания требуется усложнение структуры и, соответственно, увеличение длины полинома. Эти зависимости имеют место для всех СО (рис. 1–16).

Наиболее сложным объектом, для распознавания которого потребовался многочлен из 165 мономов, оказался элемент класса «2», расположенный вблизи минимума расстояний между своими векторами, который (рис. 1б)) превышает наименьшее расстояние между своими и чужими векторами.

В случае пищеварительной системы для мужчин (рис. 2) обучающее множество имеет 109 элементов. Классы «1» и «4» содержат по 33 набора крови, а классы «2» и «3» включают соответственно 17 и 26 элементов.

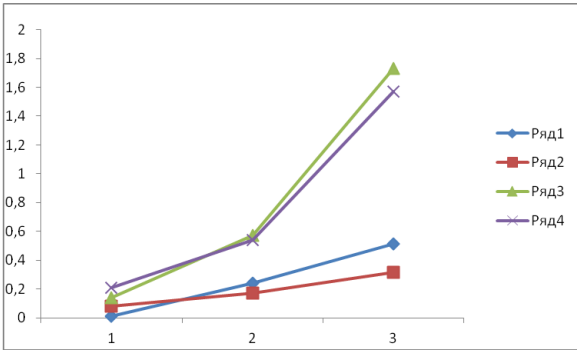
Для класса «1» взаиморасположение Ряда 1 и Ряда 3 на рис. 2а) аналогично рис. 1а). То же имеет место в отношении Ряда 2 и Ряда 4. Для других СО класс «1» (рисунки с литерой а)) рассматривать не будем ввиду однотипности. Заметим, что эти соотношения определяются как мало меняющимися наборами своих элементов, так и множеством чужих, уникальных по каждой СО.

Порядок, характерный для класса «1» (Ряд 1 и Ряд 3), не выполняется для классов «2», «3», «4» (рис. 2б), в), г)). Например, для этих трех классов минимальное (а для класса «4» и среднее) расстояние между своими векторами превышает наименьшее расстояние между своими и чужими векторами. Указанные средние величины примерно одинаковы для класса «2» (рис. 2б)).

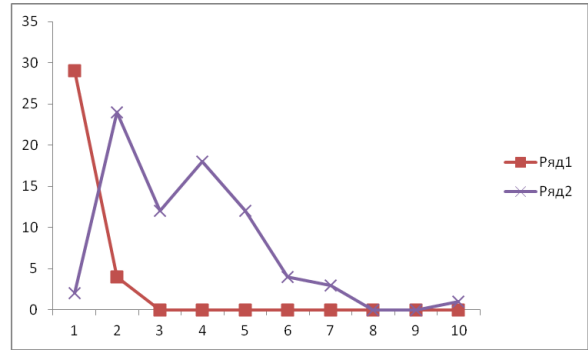
Взаимоположение Ряда 2 и Ряда 4, имеющее место в классе «1», либо в значительной степени нарушается для других классов (рис. 2б), в), г)), либо различие между указанными величинами становится меньше. Из рисунка 2г) (класс «4») следует, что максимальное расстояние между центром масс и своими векторами превышает наибольшее расстояние между центром масс и чужими векторами, а указанные средние величины примерно одинаковы; последнее имеет место как для минимальных, так и для средних расстояний по классу «2» (рис. 2б)).

Взаиморасположение Ряда 1 и Ряда 2 класса «1» на рис. 2аа) аналогично рис. 1аа), как и для прочих СО, – поэтому рис. 2аа)–1баа) обсуждать не будем.

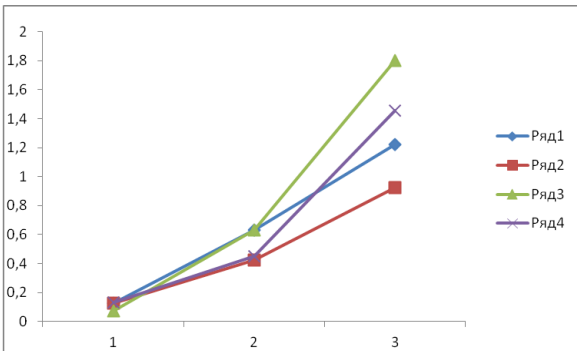
В других классах свои элементы не локализованы около центра масс: они имеются до конца (рис. 2гг)) или почти до конца (рис. 2бб), вв)) отрезка $[Dk_{\min}, Dk_{\max}]$, а максимум соответствующей функции распределения (Ряд 1) смещен вправо от его начальной точки. Там же находится максимум Ряда 2 классов «2», «3» (рис. 2бб), вв)). Максимум функции распределения своих элементов (Ряд 1) класса «4» (рис. 2гг)) дальше от центра масс, чем для чужих элементов (Ряд 2).



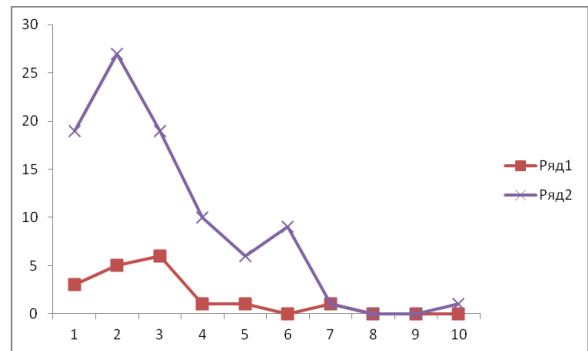
а) «1» min, middle, max



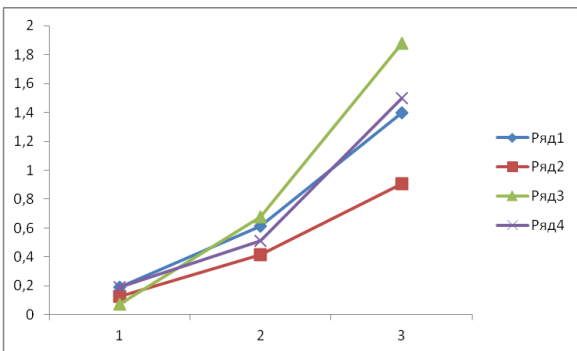
аа) «1» распределение своих/чужих



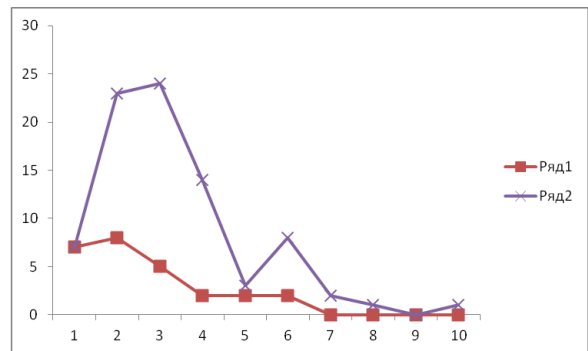
б) «2» min, middle, max



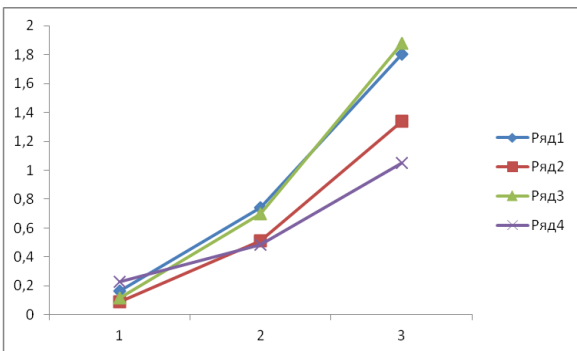
бб) «2» распределение своих/чужих



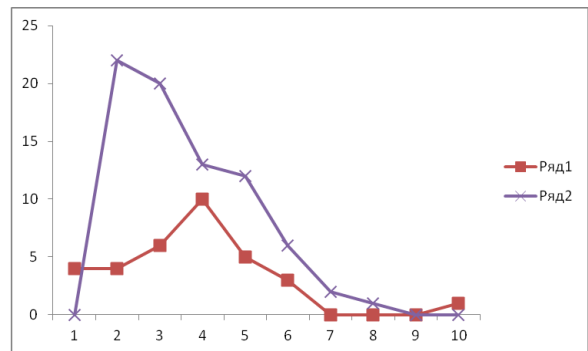
в) «3» min, middle, max



вв) «3» распределение своих/чужих



г) «4» min, middle, max



гг) «4» распределение своих/чужих

Рис. 2. Пищеварительная система (мужчины)

По одному элементу класса «2» и «3» остались нераспознанными даже при использовании полиномов длины 1287. Они расположены вблизи минимума расстояний между своими векторами, который (рис. 2б), в)) превышает наименьшее расстояние между своими и чужими векторами.

Это же имеет место и для класса «4» (рис. 2г)), но в отличие от него у классов «2» и «3» максимумы распределений своих и чужих элементов при удалении от центра масс не разнесены (рис. 2бб), вв)).

Органы дыхания

Рассмотрим органы дыхания для женщин (рис. 3). Обучающее множество имеет 62 элемента. Классы «1» и «4» содержат по 24 набора крови, а классы «2» и «3» включают соответственно 4 и 10 элементов.

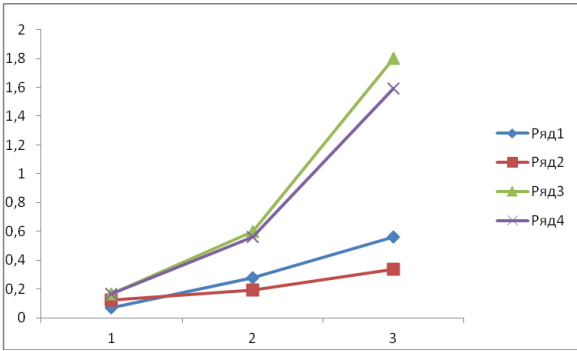
По рисункам 3б), в) для классов «2» и «3» видно (Ряд 1 и Ряд 3), что минимальное расстояние между своими векторами превышает наименьшее расстояние между своими и чужими векторами; аналогичное соотношение наблюдается для среднего расстояния по классам «2» и «4» (рис. 3б), г)), а также максимального по классу «4» (рис. 3г)). Указанные средние величины примерно одинаковы для класса «3» (рис. 3в)).

Из рисунка 3б) для класса «2» следует, что минимальное расстояние между центром масс и своими векторами превышает наименьшее расстояние между центром масс и чужими векторами; аналогичное соотношение имеется для среднего и максимального расстояния по классу «4» (рис. 3г)). Для класса «3», как и для класса «4», соответствующие минимальные величины приблизительно одинаковы (рис. 3в), г)).

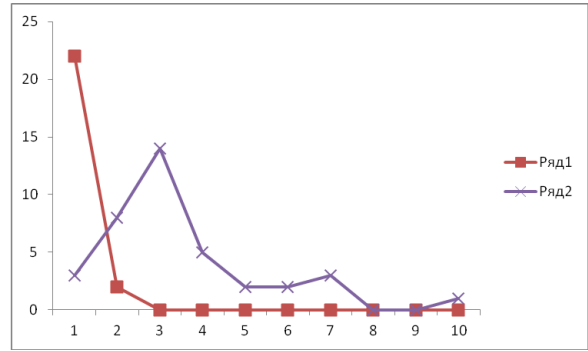
Четыре своих элемента класса «2» (Ряд 1) располагаются вблизи центра масс (рис. 3бб)), но ввиду малочисленности их набор неинформативен для анализа. Свои элементы классов «2» и «3» не локализованы вблизи центра масс: для класса «4» (рис. 3гг)) – находятся до конца, а для класса «3» (рис. 3вв)) почти до конца отрезка $[Dk_{\min}, Dk_{\max}]$. Максимум соответствующей функции распределения (Ряд 1) класса «4» дальше от центра масс, чем аналог для чужих элементов (Ряд 2); для класса «3» – наоборот, подобно классу «1». Поведение Ряда 2 на рис. 3вв), гг) схоже с рис. 3аа), а на рис. 3бб) максимум этой функции – в начале отрезка $[Dk_{\min}, Dk_{\max}]$.

Классификация достигается на коротком полиноме (длина 69).

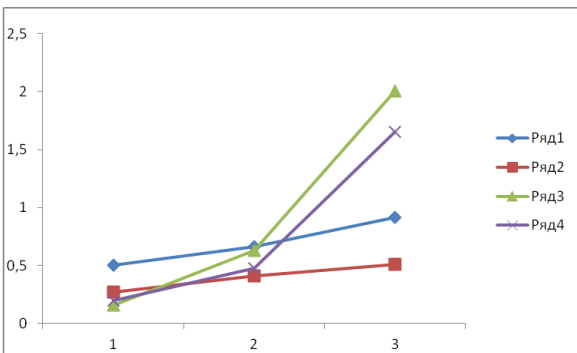
Для классов «2» и «3» минимум расстояний между своими векторами превышает наименьшее расстояние между своими и чужими векторами (рис. 3б), в)), но объем обоих классов маленький.



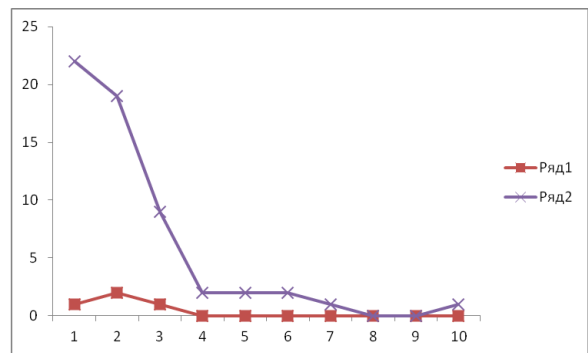
а) «1» min, middle, max



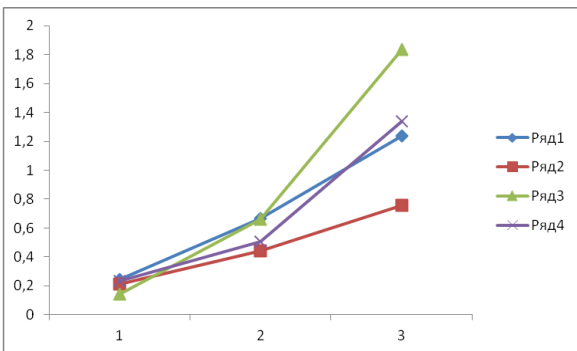
aa) «1» распределение своих/чужих



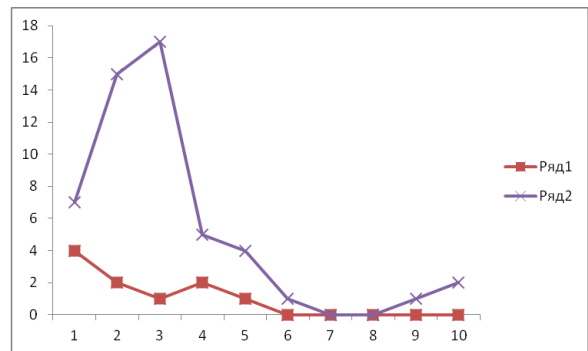
б) «2» min, middle, max



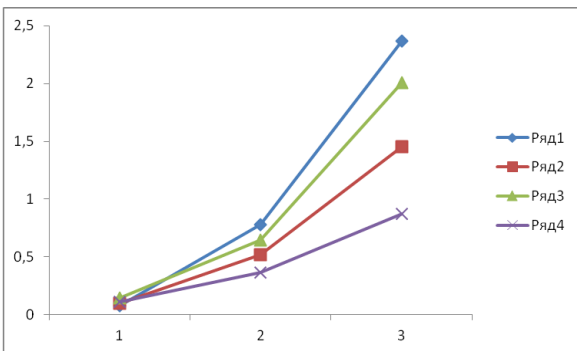
бб) «2» распределение своих/чужих



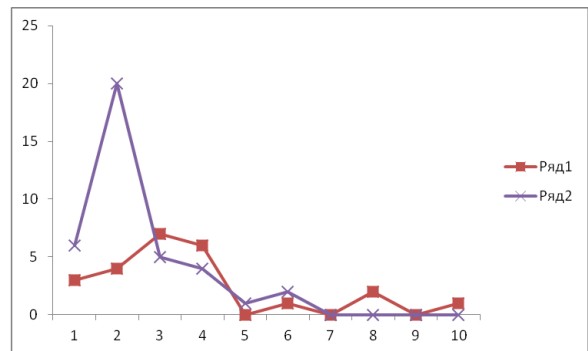
в) «3» min, middle, max



вв) «3» распределение своих/чужих

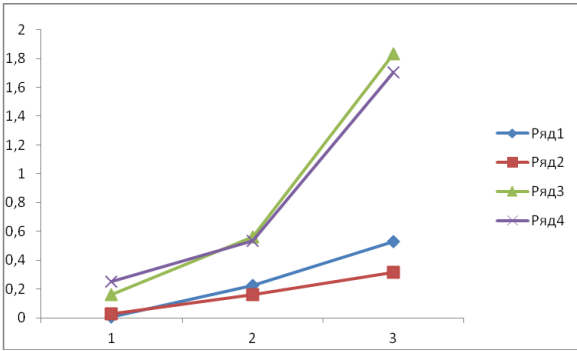


г) «4» min, middle, max

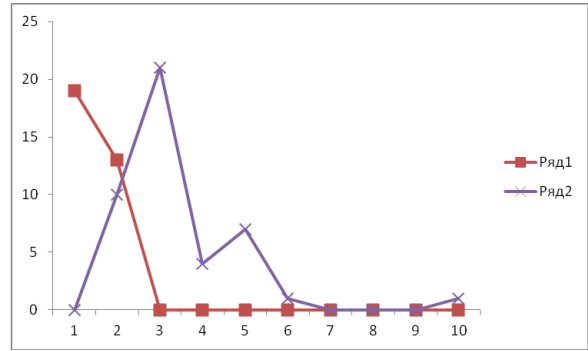


гг) «4» распределение своих/чужих

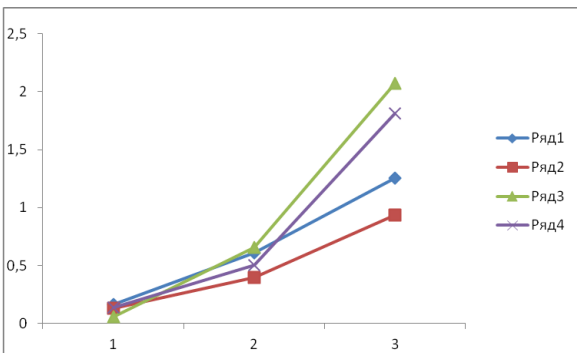
Рис. 3. Органы дыхания (женщины)



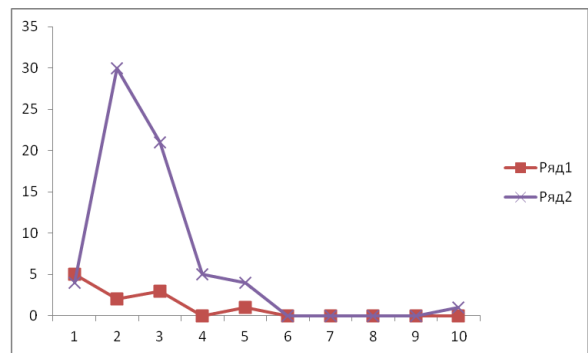
а) «1» min, middle, max



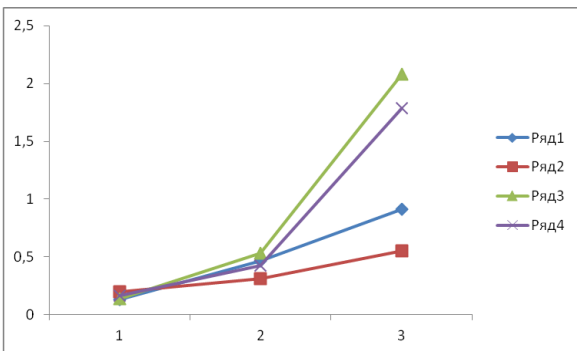
aa) «1» распределение своих/чужих



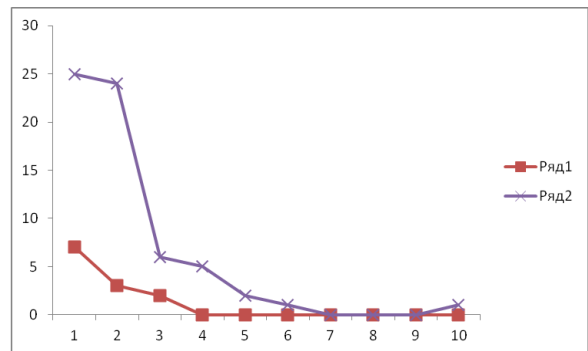
б) «2» min, middle, max



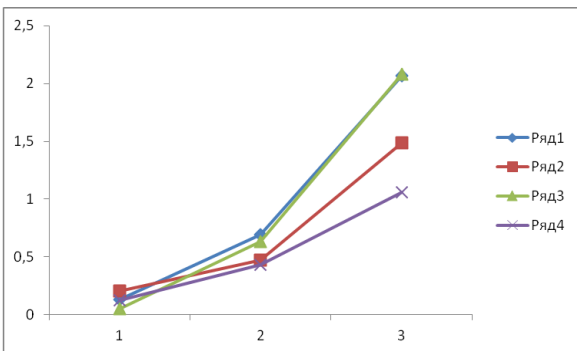
бб) «2» распределение своих/чужих



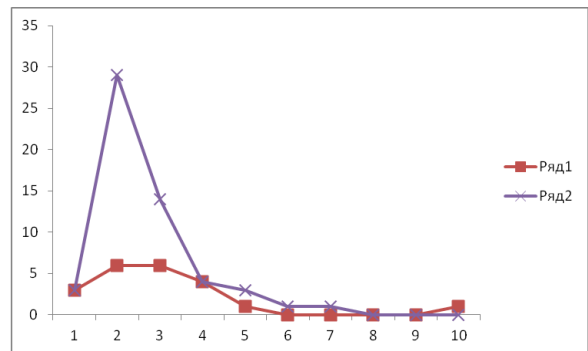
в) «3» min, middle, max



вв) «3» распределение своих/чужих



г) «4» min, middle, max



гг) «4» распределение своих/чужих

Рис. 4. Органы дыхания (мужчины)

В случае классификатора по органам дыхания для мужчин (рис. 4) обучающее множество имеет 76 элементов. Классы «1», «2», «3» и «4» содержат соответственно 32, 11, 12 и 21 набор крови.

По рисункам 4б), г) для классов «2» и «4» видно, что минимальное расстояние между своими векторами превышает наименьшее расстояние между своими и чужими векторами; аналогично и для средних величин по классу «4». Указанные минимальные величины примерно одинаковы для класса «3» (рис. 4в)), как и средние для класса «2» (рис. 4б)), а также максимальные для класса «4» (рис. 4г)).

Из рисунка 4г) для класса «4» следует, что минимальное, среднее и максимальное расстояние между центром масс и своими векторами превышает наибольшее расстояние между центром масс и чужими векторами, причем для среднего значения превышение незначительное. Для класса «2», как и для класса «3», соответствующие минимальные величины приблизительно одинаковы (рис. 4б), в)).

Свои элементы класса «3» относительно локализованы вблизи центра масс (рис. 4вв), а для классов «4» и «2» они имеются до конца (рис. 4гг) или до середины (рис. 4бб)) отрезка $[Dk_{\min}, Dk_{\max}]$; максимум соответствующей функции распределения (Ряд 1) для классов «2» и «3» располагается в его начальной точке, а для класса «4» смещен от нее вправо. Поведение функции распределения чужих элементов (Ряд 2) для классов «2» и «4» схоже с рис. 4аа), но максимум этой функции ближе к центру масс, а по классам «3» и «4» на одинаковом удалении с его аналогом для своих элементов.

Классификация достигнута с многочленом из 165 мономов. Наиболее сложный для распознавания – элемент класса «4», расположенный вблизи минимума расстояний между своими векторами, который (рис. 4г)) превышает наименьшее расстояние между своими и чужими векторами.

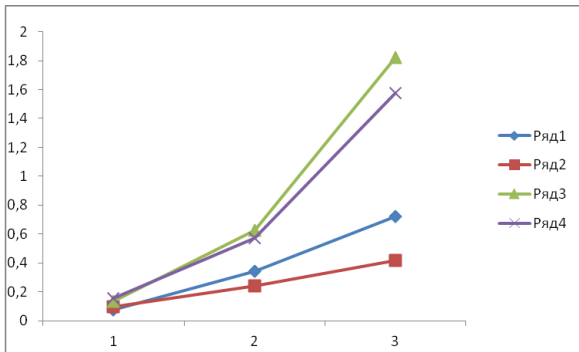
Несмотря на то что это соотношение выполняется и для класса «2» (рис. 4б)), его элементы распознаются с использованием более коротких полиномов, потому что этот класс меньше по объему (11 против 21), а также максимумы распределений своих и чужих элементов при удалении от центра масс разнесены (рис. 4бб)), в отличие от класса «4» (рис. 4гг)).

Опорно-двигательный аппарат

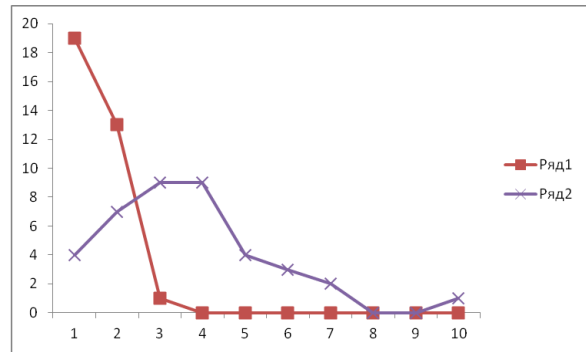
Рассмотрим опорно-двигательный аппарат для женщин (рис. 5). Обучающее множество в этом случае состоит из 72 элементов. Классы «1» и «4» включают по 33 набора крови, а класс «3» имеет 6 элементов. Класс «2» отсутствует. Используем поднумерацию а), в), г).

По рисункам 5в), г) видно, что для класса «3» минимальное расстояние, а для класса «4» как среднее, так и максимальное расстояние между своими векторами больше аналогичного расстояния между своими и чужими векторами. Указанные средние величины для класса «3» примерно одинаковы (рис. 5в)), как и минимальные значения для класса «4» (рис. 5г)).

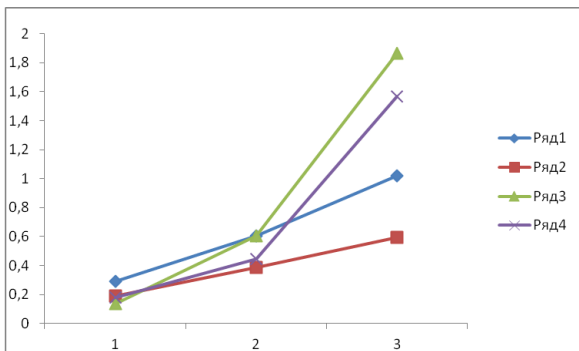
Из рисунка 5г) для класса «4» следует, что среднее и максимальное расстояние между центром масс и своими векторами больше соответствующего расстояния между центром масс и чужими векторами. Для класса «3» соответствующие минимальные величины приблизительно одинаковы (рис. 5в)).



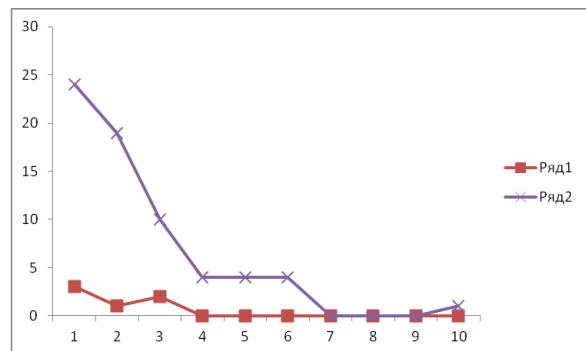
а) «1» min, middle, max



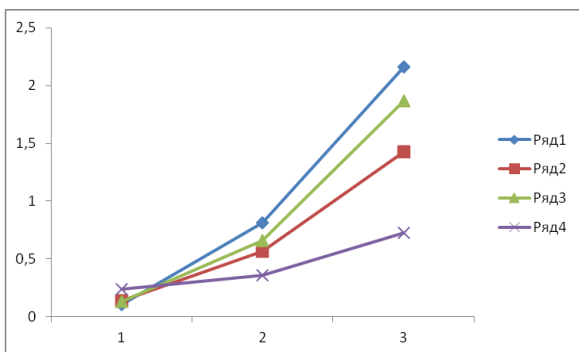
аа) «1» распределение своих/чужих



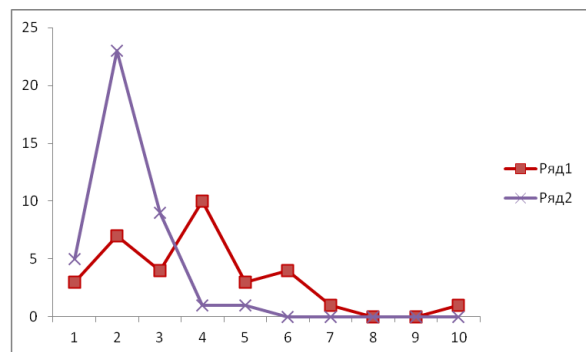
в) «3» min, middle, max



вв) «3» распределение своих/чужих

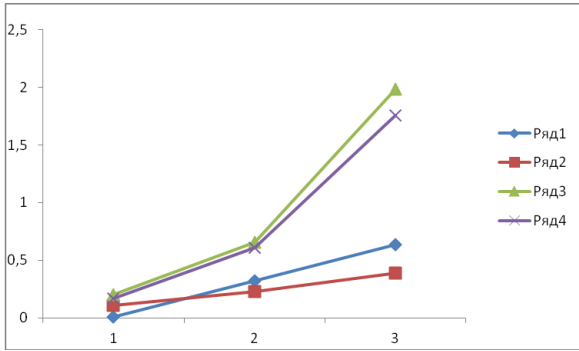


г) «4» min, middle, max

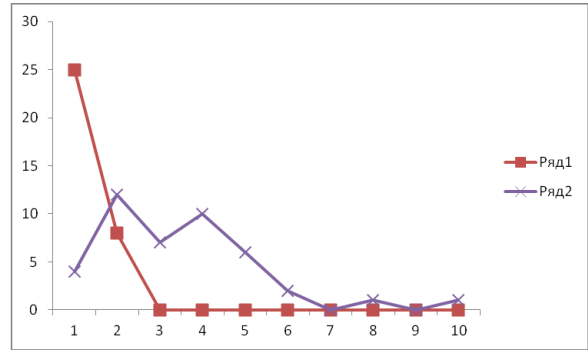


гг) «4» распределение своих/чужих

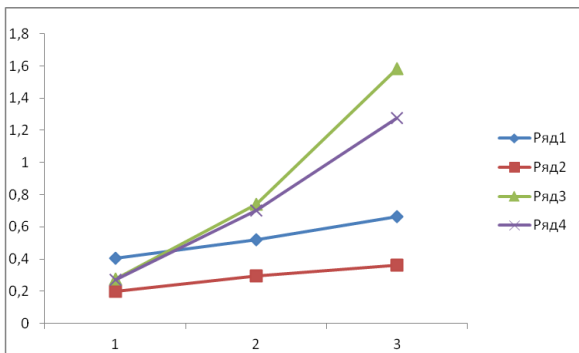
Рис. 5. Опорно-двигательный аппарат (женщины)



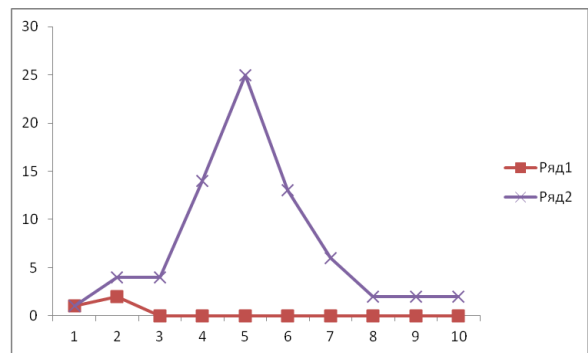
a) «1» min, middle, max



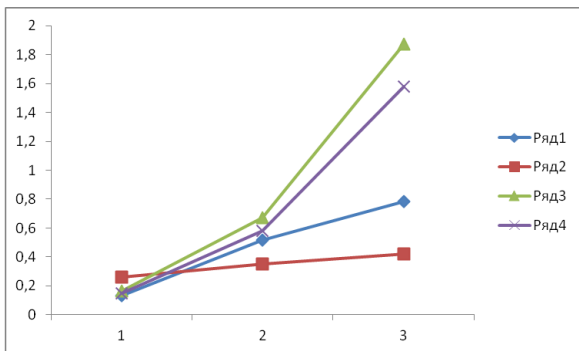
aa) «1» распределение своих/чужих



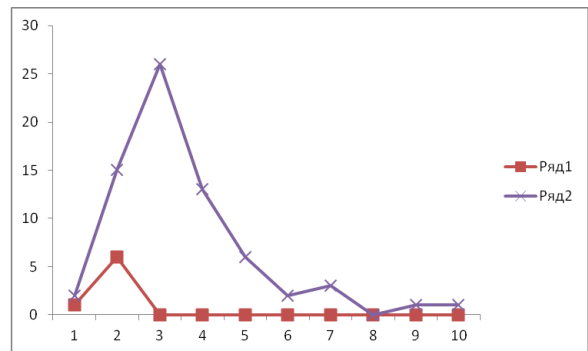
б) «2» min, middle, max



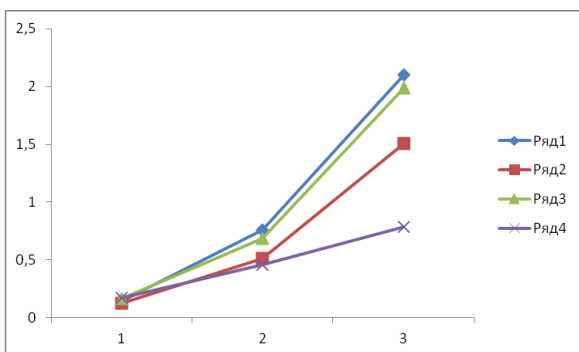
бб) «2» распределение своих/чужих



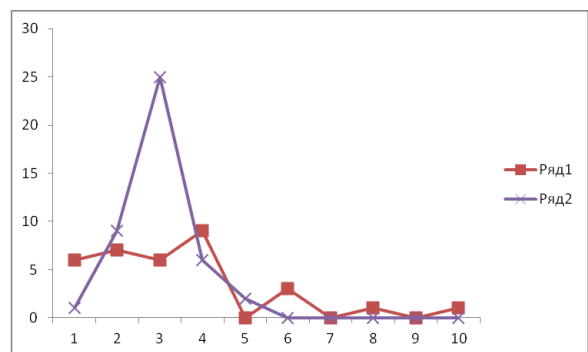
в) «3» min, middle, max



вв) «3» распределение своих/чужих



г) «4» min, middle, max



гг) «4» распределение своих/чужих

Рис. 6. Опорно-двигательный аппарат (мужчины)

Шесть своих элементов (Ряд 1) класса «3» (рис. 5вв)) находятся вблизи центра масс, но ввиду их малого числа неинтересны для анализа. Свои элементы класса «4» (рис. 5гг)) не локализованы вблизи центра масс: располагаются до конца отрезка $[Dk_{\min}, Dk_{\max}]$. Максимум соответствующей функции распределения (Ряд 1) класса «4» дальше от центра масс, чем наибольшая величина для чужих элементов (Ряд 2). Поведение функции распределения чужих элементов (Ряд 2) схоже с рис. 5аа), однако максимум этой функции ближе к центру масс (рис. 5гг)) и особенно (рис. 5вв)).

Самым проблемным объектом, для распознавания которого потребовался многочлен из 165 мономов, оказался элемент класса «3», расположенный вблизи минимума расстояний между своими векторами, который (рис. 5в)) превышает наименьшее расстояние между своими и чужими векторами.

В случае опорно-двигательного аппарата для мужчин (рис. 6) обучающее множество имеет 76 элементов. Классы «1» и «4» содержат по 33 набора крови, а классы «2» и «3» включают соответственно 3 и 7 элементов.

По рисункам 6б), г) видно, что для класса «2» минимальное расстояние, а для класса «4» как среднее, так и максимальное расстояние между своими векторами больше аналогичной величины для расстояния между своими и чужими векторами. Указанные минимальные величины для класса «3» примерно одинаковы (рис. 6в)), как и для класса «4» (рис. 6г)).

Из рисунка 6г) для класса «4» следует, что среднее и максимальное, а для класса «3» (рис. 6в)) минимальное расстояние между центром масс и своими векторами больше соответствующего расстояния между центром масс и чужими векторами, причем для среднего значения (класс «4») превышение незначительное.

Три и семь своих элементов (Ряд 1) классов «2» и «3» (рис. 6бб), вв)) находятся вблизи центра масс, но ввиду их малого числа неинтересны для анализа. Свои элементы класса «4» (рис. 6гг)) не локализованы вблизи центра масс: располагаются до конца отрезка $[Dk_{\min}, Dk_{\max}]$. Максимум соответствующей функции распределения (Ряд 1) для класса «4» удален от центра масс приблизительно так же, как и наибольшая величина для чужих элементов (Ряд 2). Поведение функции распределения чужих элементов (Ряд 2) классов «2», «3» и «4» схоже с рис. 6аа).

Классификация достигается на коротком полиноме из 77 мономов.

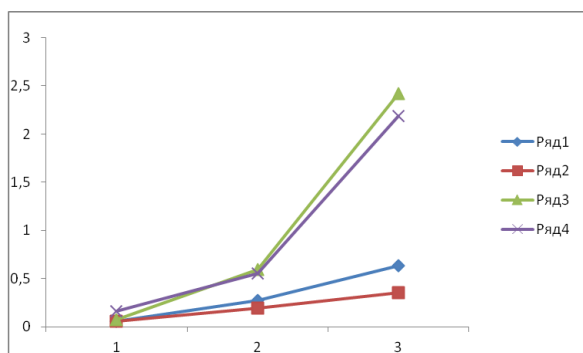
Для класса «2» минимум расстояний между своими векторами превышает наименьшее расстояние между своими и чужими векторами (рис. 6б)), но проблем не возникает, поскольку у класса маленький объем.

Урологическая система

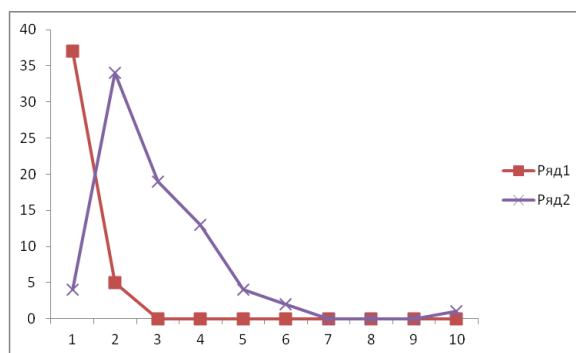
Рассмотрим классификатор урологической системы для женщин (рис. 7). Обучающее множество имеет 119 элементов. Классы «1» и «3» содержат по 42 набора крови, а класс «4» включает 35 элементов. Класс «2» не заполнен.

Из рисунка 7г) следует, что для класса «4» минимальное, среднее и максимальное расстояние между своими векторами больше аналогичной величины для расстояния между своими и чужими векторами. Указанные минимальные величины примерно одинаковы для класса «3» (рис. 7в)). Последнее имеет место и для класса «1» как малозначительное исключение из общего для этого класса правила.

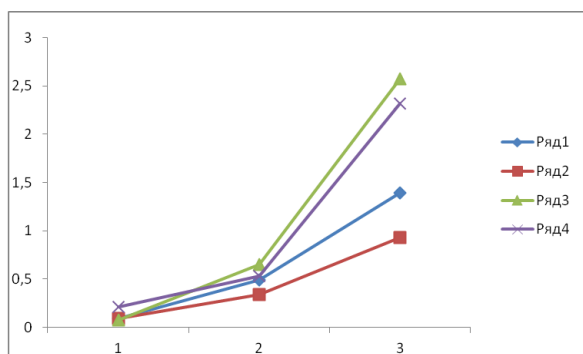
Из рисунка 7г) для класса «4» следует, что среднее и максимальное расстояние между центром масс и своими векторами больше аналогичной величины для расстояния между центром масс и чужими векторами, а соответствующие минимальные величины приблизительно одинаковы.



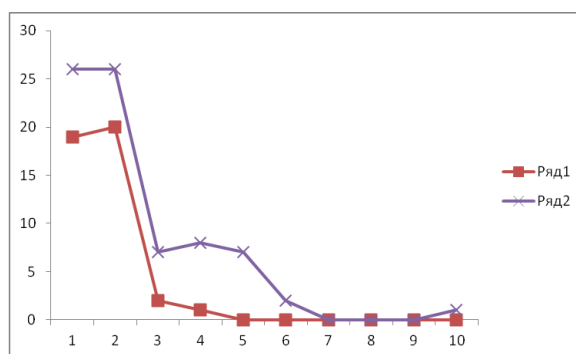
а) «1» min, middle, max



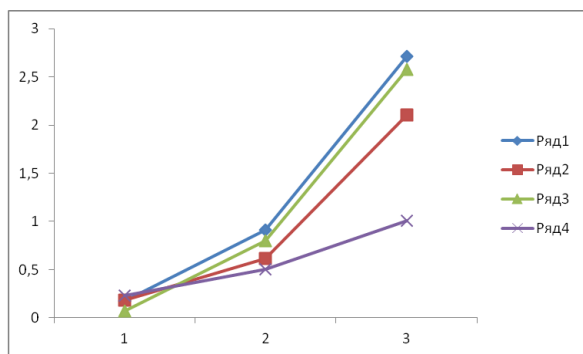
аа) «1» распределение своих/чужих



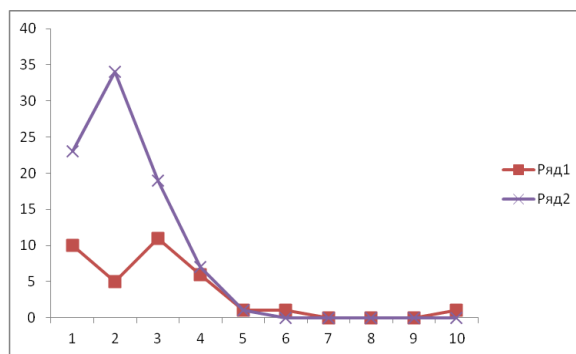
в) «3» min, middle, max



вв) «3» распределение своих/чужих

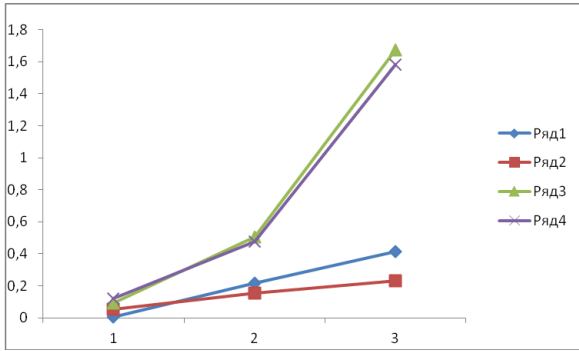


г) «4» min, middle, max

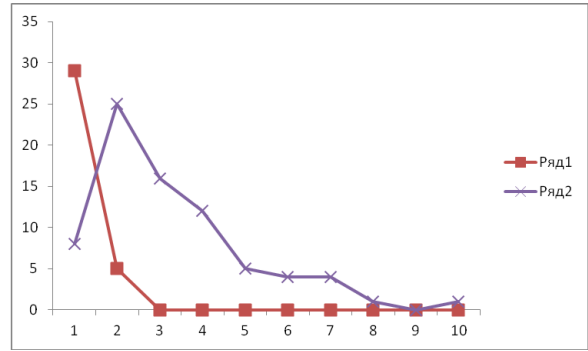


гг) «4» распределение своих/чужих

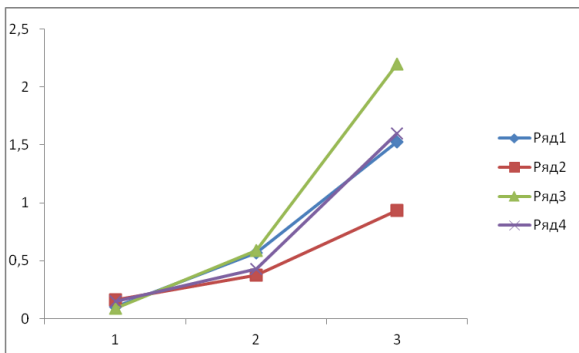
Рис. 7. Урологическая система (женщины)



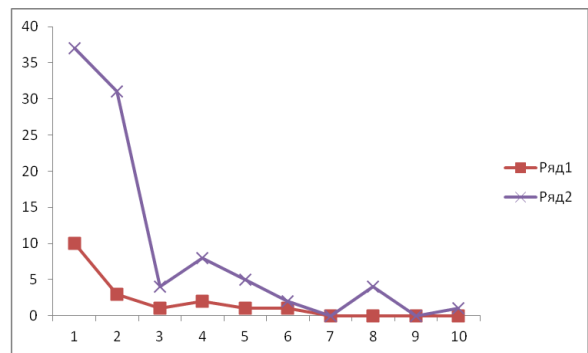
а) «1» min, middle, max



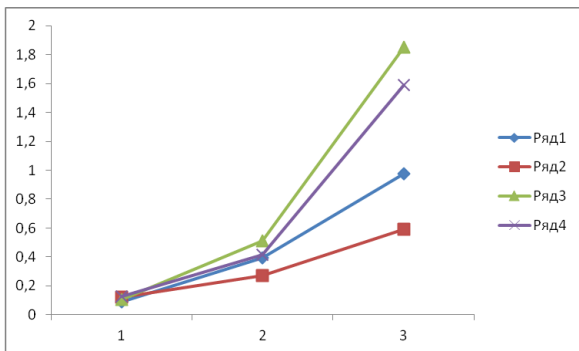
aa) «1» распределение своих/чужих



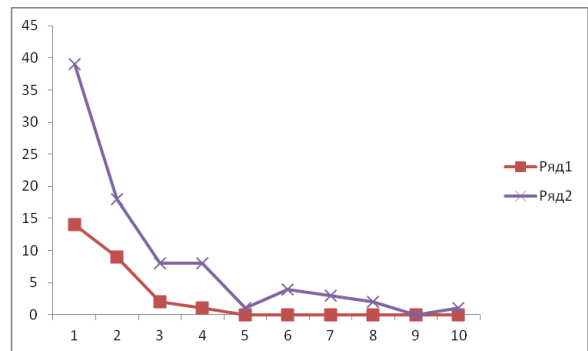
б) «2» min, middle, max



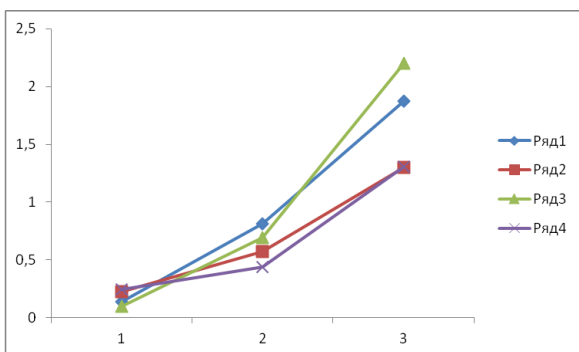
бб) «2» распределение своих/чужих



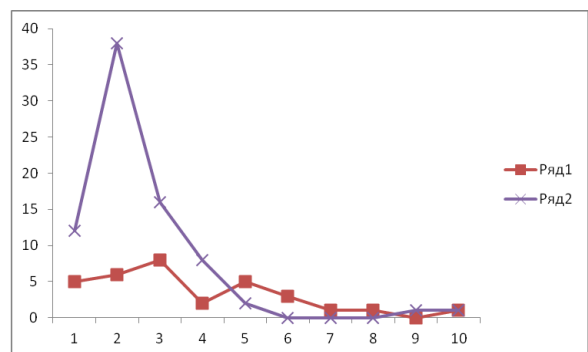
в) «3» min, middle, max



вв) «3» распределение своих/чужих



г) «4» min, middle, max



гг) «4» распределение своих/чужих

Рис. 8. Урологическая система (мужчины)

Свои элементы класса «3» относительно локализованы вблизи центра масс (рис.7вв)). Для класса «4» они имеются до конца (рис.7гг)) отрезка $[Dk_{\min}, Dk_{\max}]$; максимум соответствующей функции распределения (Ряд 1) для классов «3» и «4» располагается вблизи начальной точки. Поведение функции распределения чужих элементов (Ряд 2) по классам «3» и «4» схоже с рис.7аа)), но максимум этой функции ближе к центру масс, приблизительно на одинаковом удалении с его аналогом для своих элементов.

Самым сложным объектом, для распознавания которого потребовался многочлен из 1287 мономов, оказался элемент класса «4», расположенный вблизи минимума расстояний между своими векторами, который (рис. 7г)) превышает наименьшее расстояние между своими и чужими векторами.

В случае урологической системы для мужчин (рис. 8) обучающее множество содержит 110 элементов. Классы «1» и «4» содержат по 33 набора крови, а классы «2» и «3» включают соответственно 18 и 26 элементов.

По рисунку 8г) видно, что для класса «4» минимальное и среднее расстояние между своими векторами больше аналогичного расстояния между своими и чужими векторами. Указанные минимальное и среднее расстояние для класса «2» (рис.8б)), а также минимальное расстояние для класса «3» (рис.8в)) приблизительно одинаковы.

Из рисунка 8г) для класса «4» следует, что среднее расстояние между центром масс и своими векторами больше соответствующего расстояния между центром масс и чужими векторами. Для класса «4» (рис.8г)) соответствующие минимальные и максимальные, для класса «2» (рис.8б)), как и для класса «3» (рис. 8в)), минимальные величины приблизительно одинаковы.

Свои элементы (Ряд 1) класса «3» относительно локализованы вблизи центра масс (рис.8вв)). Для класса «4» они имеются до конца (рис.8гг)), а для класса «2» (рис.8бб)) далее середины отрезка $[Dk_{\min}, Dk_{\max}]$. Поведение функции распределения чужих элементов (Ряд 2) по классу «4» схоже с рис. 8аа), причем ее максимум ближе к центру масс, чем у аналога для своих элементов. По классам «2» и «3» обе функции (Ряд 1, Ряд 2) убывают при удалении от центра масс.

Лишь с использованием полиномов длины 1287 удалось распознать один элемент класса «2», расположенный вблизи минимума расстояний между своими векторами, который чуть превышает наименьшее расстояние между своими и чужими векторами (рис. 8б)).

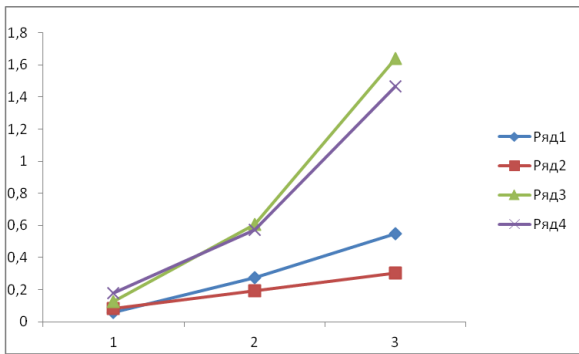
Эта же особенность больше выражена для класса «4» (рис. 8г)). Однако, в отличие от него, у класса «2» максимумы распределений своих и чужих элементов при удалении от центра масс не разнесены (рис. 8бб), гг)).

Эндокринная система

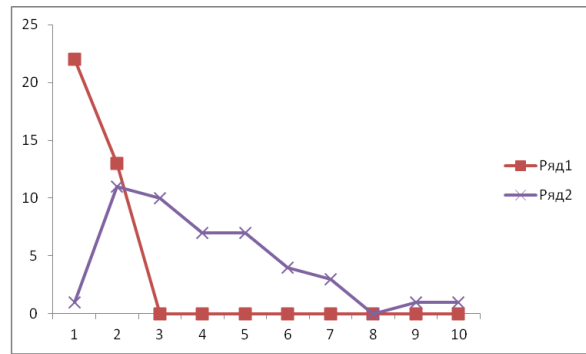
Рассмотрим эндокринную систему для женщин (рис. 9). Обучающее множество в этом случае состоит из 80 элементов. Классы «1» и «4» включают по 35 наборов крови, а класс «2» имеет 10 элементов. Класс «3» отсутствует.

По рисункам 9б), г) видно, что для классов «2» и «4» минимальное и среднее расстояние, а для класса «4» также и максимальное расстояние между своими векторами больше аналогичной величины для расстояния между своими и чужими векторами.

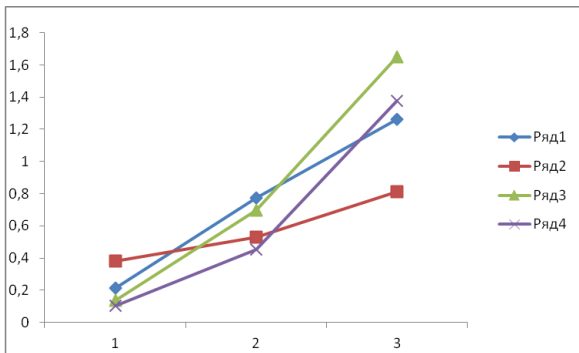
По рисунку 9б) видно, что для класса «2» минимальное и среднее, а для класса «4» (рис. 9г)) среднее и максимальное расстояние между центром масс и своими векторами больше соответствующего расстояния между центром масс и чужими векторами.



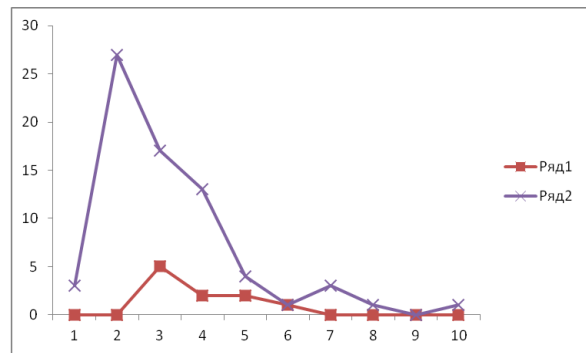
а) «1» min, middle, max



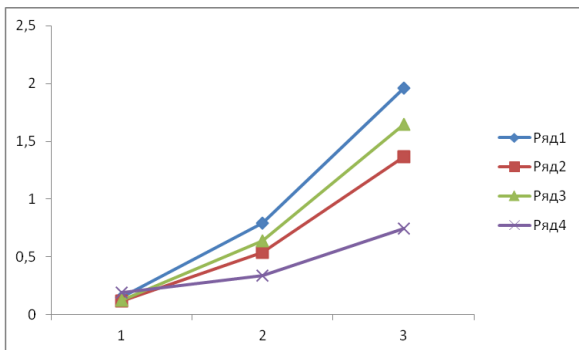
aa) «1» распределение своих/чужих



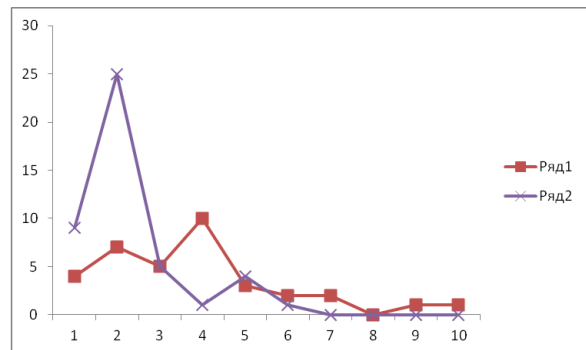
б) «2» min, middle, max



бб) «2» распределение своих/чужих

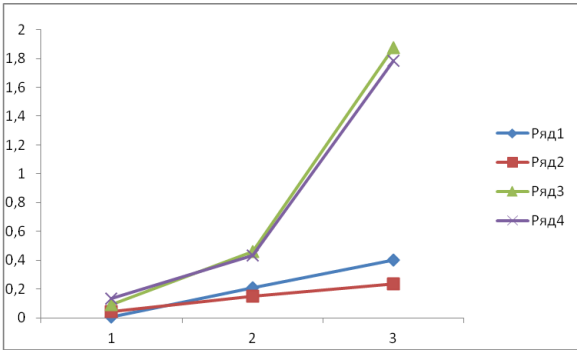


г) «4» min, middle, max

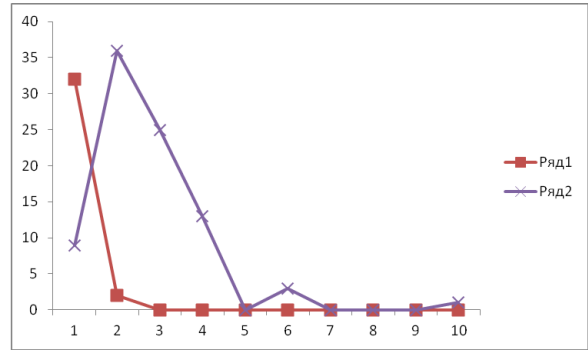


гг) «4» распределение своих/чужих

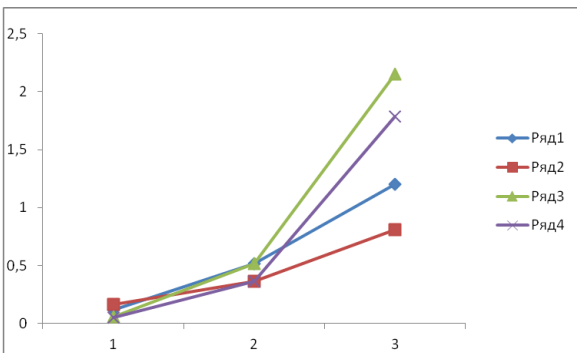
Рис. 9. Эндокринная система (женщины)



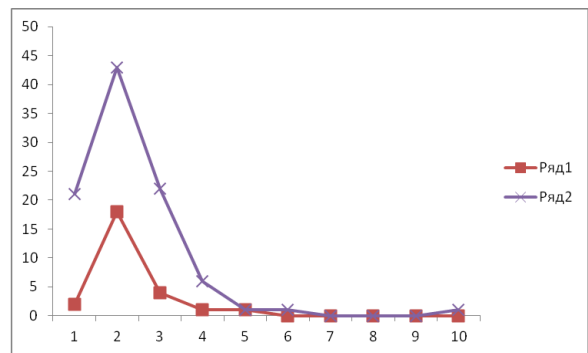
а) «1» min, middle, max



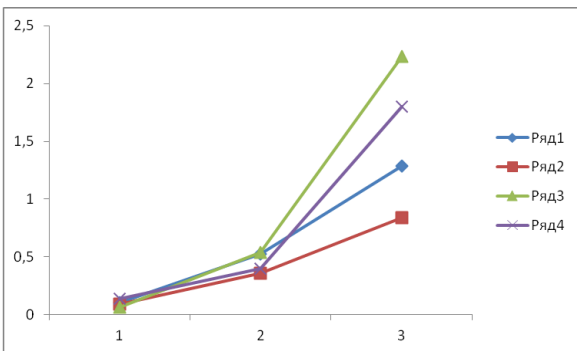
аа) «1» распределение своих/чужих



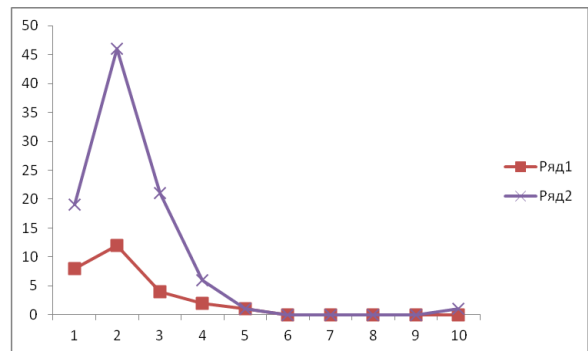
б) «2» min, middle, max



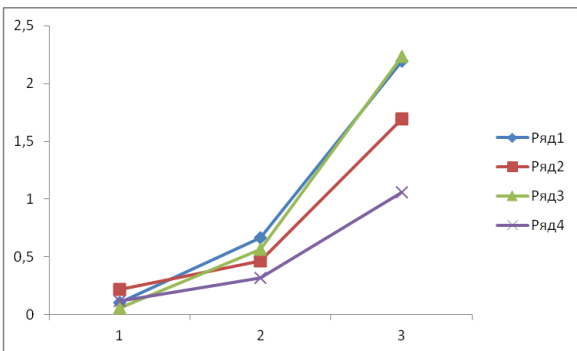
бб) «2» распределение своих/чужих



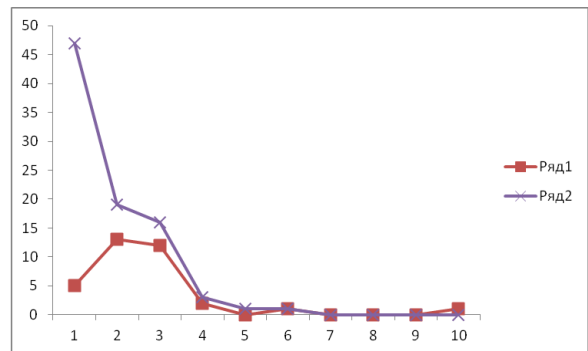
в) «3» min, middle, max



вв) «3» распределение своих/чужих



г) «4» min, middle, max



гг) «4» распределение своих/чужих

Рис. 10. Эндокринная система (мужчины)

Свои элементы классов «2», «4» (рис. 9бб), гг)) не локализованы вблизи центра масс: имеются до конца (рис.9гг)) или удалены от начала и занимают более половины (рис.9бб)) отрезка $[Dk_{\min}, Dk_{\max}]$, а максимум соответствующей функции распределения (Ряд 1) правее его начальной точки. Поведение функции распределения чужих элементов (Ряд 2) схоже с рис. 9аа), причем максимум этой функции ближе к центру масс, чем у своих элементов – незначительно для класса «2» и существенно для класса «4».

Классификация достигнута с многочленом из 165 мономов. Наиболее сложные для распознавания – два элемента класса «2», расположенные вблизи минимума расстояний между своими векторами, который (рис. 9г)) превышает наименьшее расстояние между своими и чужими векторами.

Это соотношение выполняется и для класса «4» (рис. 9г)), но его элементы распознаются с использованием более коротких полиномов, потому что максимумы распределений своих и чужих элементов при удалении от центра масс разнесены (рис.9гг)), в отличие от класса «2» (рис. 9бб)).

В случае эндокринной системы для мужчин (рис. 10) обучающее множество содержит 121 элемент. Классы «1» и «4» содержат по 34 набора крови, а классы «2» и «3» включают соответственно 26 и 27 элементов.

По рисункам 10б),г) видно, что для классов «2» и «4» минимальное расстояние, а для класса «4» еще и среднее расстояние между своими векторами больше аналогичного расстояния между своими и чужими векторами. Указанные средние расстояния для классов «2» и «3» (рис. 10б),в)), минимальные расстояния для класса «3» (рис. 10в)), а также максимальные расстояния для класса «4» (рис.10г)) приблизительно одинаковы.

Из рисунка 10г) для класса «4» следует, что минимальное, среднее и максимальное расстояние между центром масс и своими векторами больше соответствующего расстояния между центром масс и чужими векторами; аналогично минимальное расстояние для класса «2» (рис. 10б)). Для класса «2» соответствующие средние величины приблизительно одинаковы (рис. 10б)).

Свои элементы классов «2», «3», «4» (рис. 10бб), вв), гг)) не локализованы вблизи центра масс: они имеются до конца (рис. 10гг) или до середины (рис. 10бб), вв)) отрезка $[Dk_{\min}, Dk_{\max}]$, а максимум соответствующей функции распределения (Ряд 1) смещен вправо от его начальной точки. Функция распределения чужих элементов (Ряд 2) для классов «2», «3» схожа с рис. 10аа), но ее максимум удален от центра масс так же, как у аналога по своим элементам (Ряд 1); для класса «4» ее максимум в начальной точке этого отрезка (левее, чем у Ряда 1).

Лишь с использованием полиномов длины 1287 удалось распознать четыре элемента класса «2» и два элемента класса «4», расположенные вблизи минимума расстояний между своими векторами, который превышает наименьшее расстояние между своими и чужими векторами (рис. 10б), г)).

ЦНС и органы чувствительности

Обучающее множество ЦНС и органов чувствительности для женщин (рис. 11) имеет 79 элементов. Классы «1» и «4» содержат по 31 набора крови, а классы «2» и «3» включают соответственно 13 и 4 элемента.

Из рисунков 11в), г) следует, что для класса «3» и «4» минимальное и среднее расстояние между своими векторами больше аналогичной величины для расстояния между своими и чужими векторами. Такие же величины примерно одинаковы для класса «2» (рис. 11б)).

Из рисунка 11г) для класса «4» следует, что среднее и максимальное, а для класса «3» (рис. 11в)) минимальное расстояние между центром масс и своими векторами больше соответствующего расстояния между центром масс и чужими векторами. Для класса «2» (рис. 11б)) указанные минимальные и средние величины, а для класса «4» (рис. 11г)) минимальные, приблизительно одинаковы.

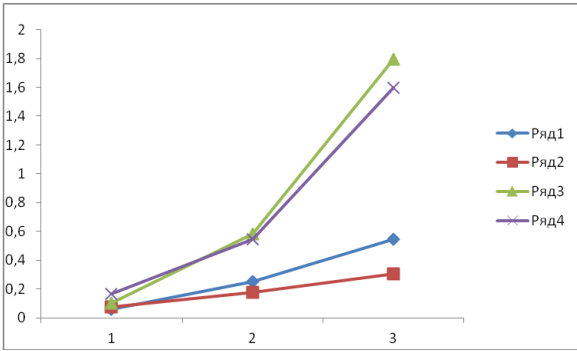
Четыре своих элемента (Ряд 1) класса «3» (рис. 11вв)) находятся вблизи центра масс, но ввиду их малого числа неинтересны для анализа. Свои элементы классов «2» и «4» (рис. 11бб), гг)) не локализованы вблизи центра масс: они имеются до конца (рис. 11гг) или до середины (рис. 11бб)) отрезка $[Dk_{\min}, Dk_{\max}]$, а максимум соответствующей функции распределения (Ряд 1) смещен вправо от его начальной точки. Функция распределения чужих элементов (Ряд 2) классов «2», «3» схожа с рис. 11аа), но ее максимум удален от центра масс относительно аналога по своим элементам (Ряд 1): для класса «2» – так же, для класса «3» – немного левее; для класса «4» (максимум в начальной точке этого отрезка) – левее.

Классификация достигается на коротком полиноме (длина 69).

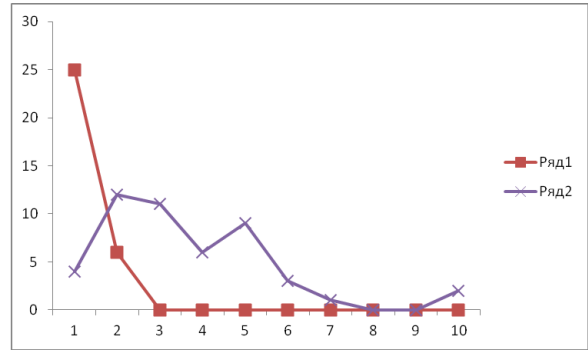
По классу «2» минимальное расстояние между своими векторами равно наименьшему расстоянию между своими и чужими векторами (рис. 11б)).

По классам «3» и «4» – первая упомянутая выше величина больше второй (рис. 11в), г)). Однако класс «3» маленького объема. Для класса «4» максимумы распределений своих и чужих элементов при удалении от центра масс разнесены (рис. 11гг)).

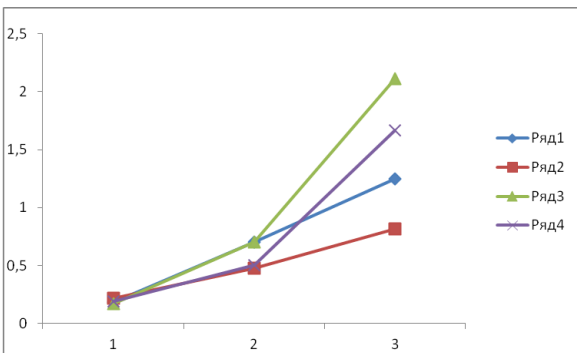
В случае классификатора по ЦНС и органам чувствительности для мужчин (рис. 12) обучающее множество имеет 61 элемент. Классы «1» и «4» содержат по 26, а классы «2» и «3» соответственно 2 и 7 наборов крови.



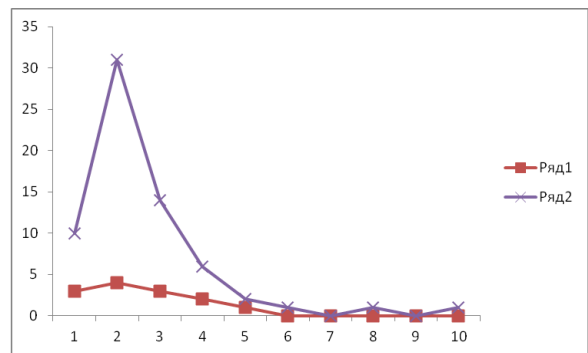
а) «1» min, middle, max



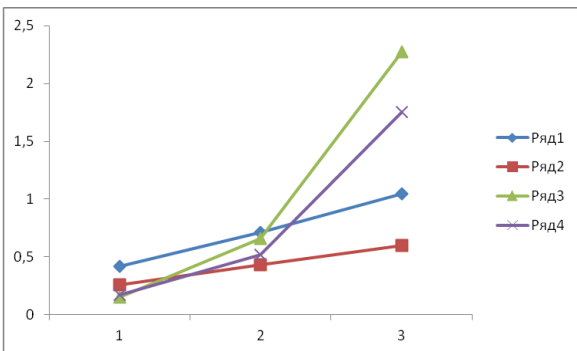
aa) «1» распределение своих/чужих



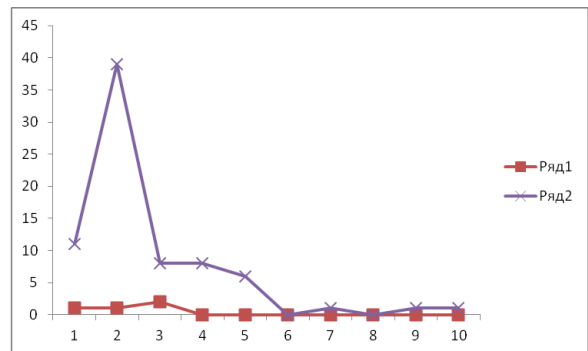
б) «2» min, middle, max



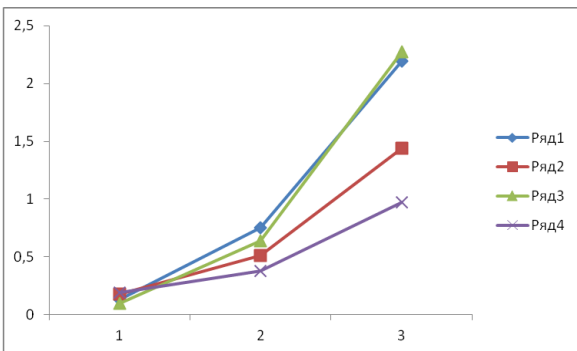
бб) «2» распределение своих/чужих



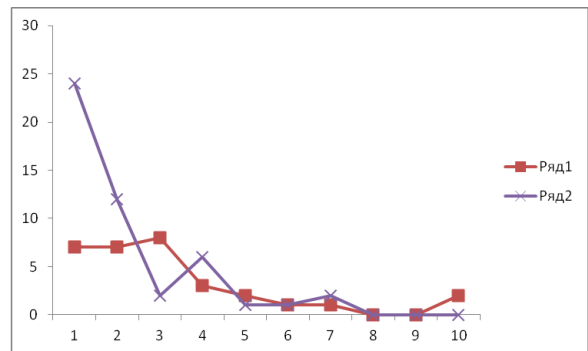
в) «3» min, middle, max



вв) «3» распределение своих/чужих

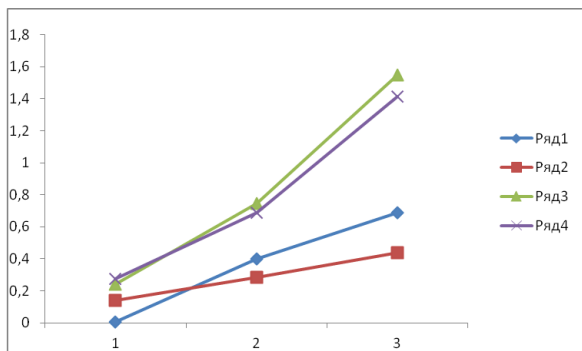


г) «4» min, middle, max

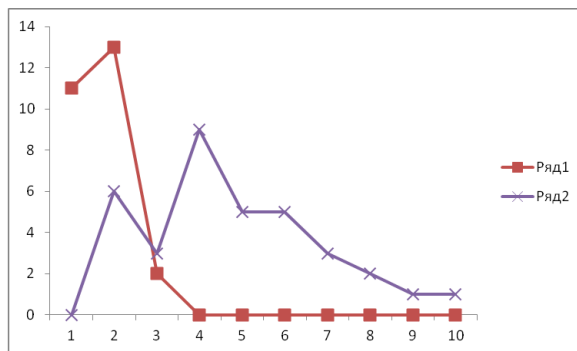


гг) «4» распределение своих/чужих

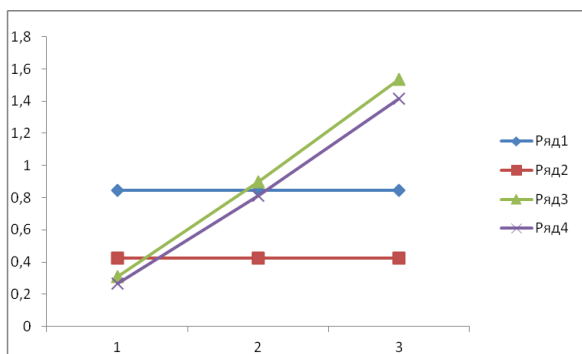
Рис. 11. ЦНС и органы чувствительности (женщины)



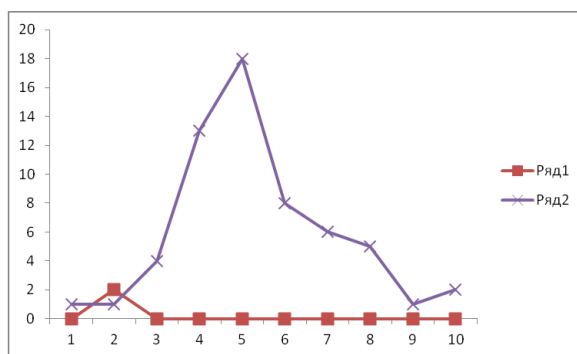
a) «1» min, middle, max



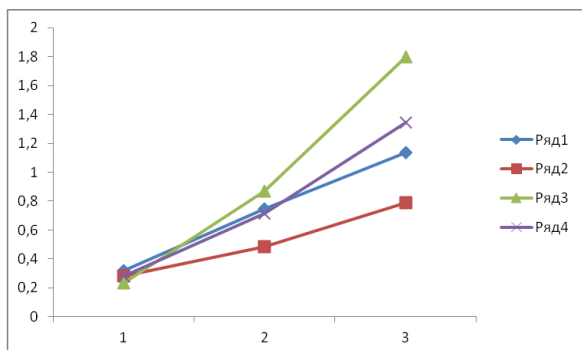
aa) «1» распределение своих/чужих



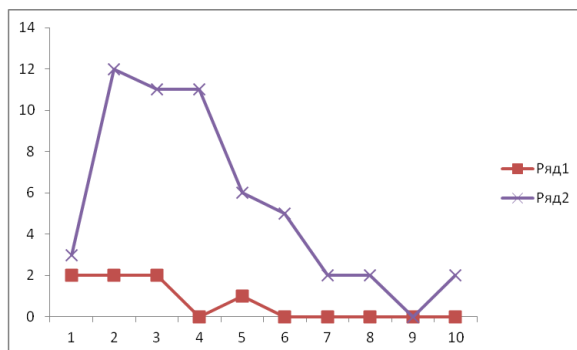
б) «2» min, middle, max



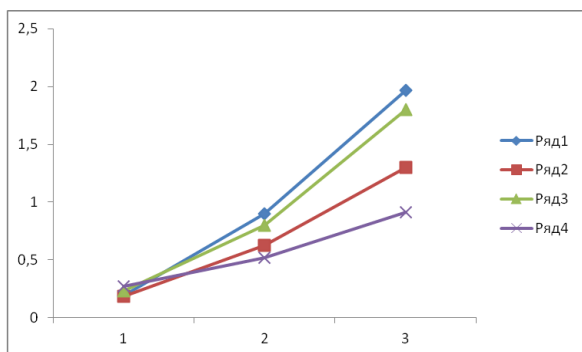
бб) «2» распределение своих/чужих



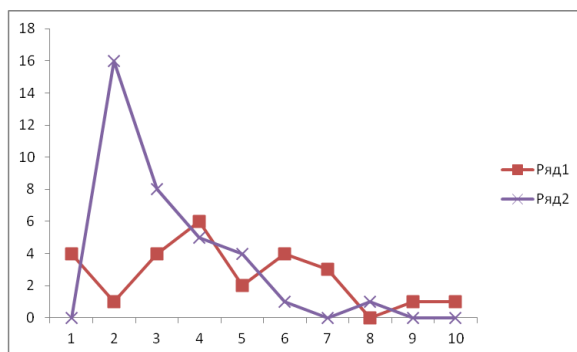
в) «3» min, middle, max



вв) «3» распределение своих/чужих



г) «4» min, middle, max



гг) «4» распределение своих/чужих

Рис. 12. ЦНС и органы чувствительности (мужчины)

По рисункам 12б), в) для классов «2» и «3» видно, что минимальное расстояние между своими векторами превышает наименьшее расстояние между своими и чужими векторами; аналогичное соотношение имеет место для средних и максимальных величин по классу «4» (рис. 12г)). Указанные минимальные величины примерно одинаковы для класса «4» (рис. 12г)).

Из рисунка 12г) для класса «4» следует, что среднее и максимальное, а для класса «2» (рис. 12б)) минимальное расстояние между центром масс и своими векторами больше соответствующего расстояния между центром масс и чужими векторами. Для класса «3» (рис. 12в)) указанные минимальные величины приблизительно одинаковы.

Набор элементов класса «1» уменьшили их числа с целью выравнивания по объему с класса «4». Это привело к незначительному смещению максимума функции распределения своих элементов (Ряд 1) от центра масс, однако вблизи него также имеется небольшая окрестность, в которой все они находятся (рис. 12аа), причем справа от максимума их число убывает при удалении от центра масс.

Два своих элемента (Ряд 1) класса «2» (рис. 12бб)) находятся вблизи центра масс, но ввиду их малого числа неинформативны для анализа. Свои элементы классов «3» и «4» (рис. 12вв), гг)) не локализованы вблизи центра масс: они имеются до конца (рис. 12гг) или до середины (рис. 12вв)) отрезка $[Dk_{\min}, Dk_{\max}]$, а максимум соответствующей функции распределения (Ряд 1) смещен вправо от его начала. Функция распределения чужих элементов (Ряд 2) классов «2», «3», «4» схожа с рис. 12аа), но ее максимум удален от центра масс относительно аналога по своим элементам (Ряд 1): для класса «3» – приблизительно так же, для класса «2» – правее, для класса «4» – левее.

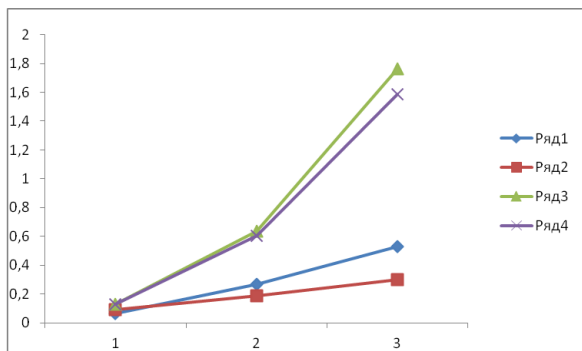
Классификация достигается на коротком полиноме (длина 61).

Для классов «2» и «3» минимальное расстояние между своими векторами превышает наименьшее расстояние между своими и чужими векторами, но они маленького объема (рис. 12б), в)). Для класса «4» первая упомянутая выше величина меньше второй (рис. 12г)).

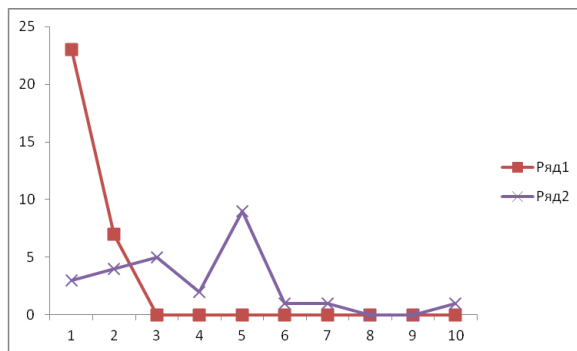
Гинекологическая система

Гинекологическая система присуща только женщинам (рис. 13). Обучение проводилось на множестве из 56 элементов. Классы «1», «2», «3» и «4» включают соответственно 30, 6, 5 и 15 наборов крови.

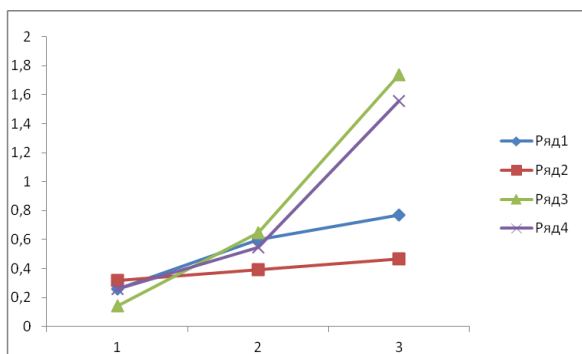
По рисункам 13б), в), г) видно, что для классов «2», «3», «4» минимальное расстояние, для классов «3», «4» также среднее, а для класса «4» и максимальное расстояние между своими векторами больше аналогичной величины для расстояния между своими и чужими векторами.



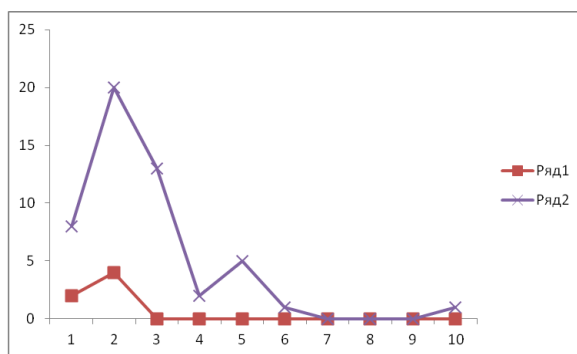
a) «1» min, middle, max



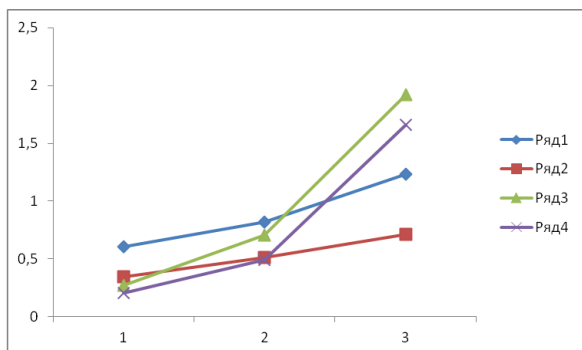
aa) «1» распределение своих/чужих



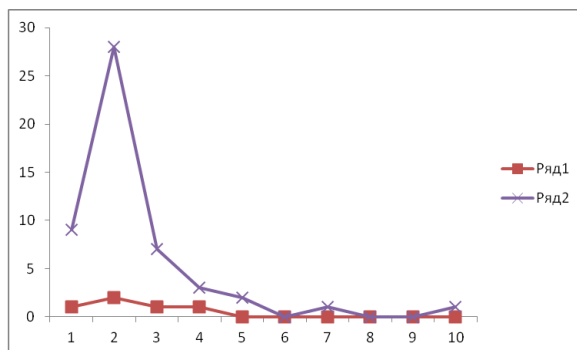
б) «2» min, middle, max



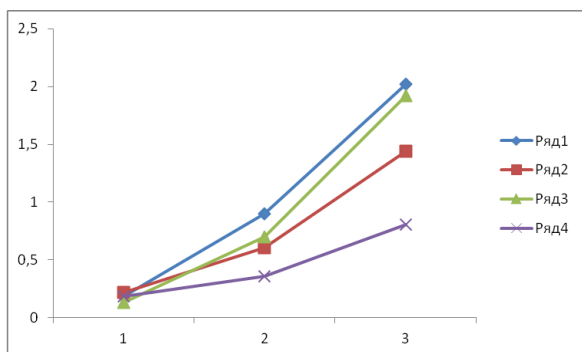
бб) «2» распределение своих/чужих



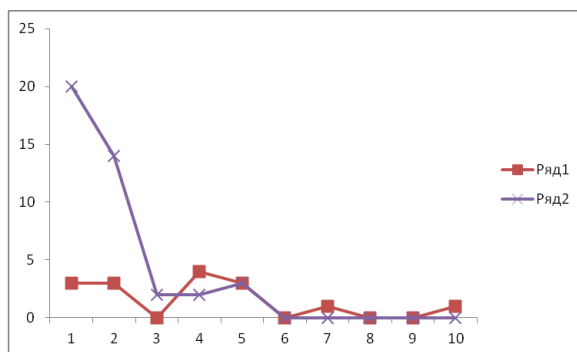
в) «3» min, middle, max



вв) «3» распределение своих/чужих



г) «4» min, middle, max



гг) «4» распределение своих/чужих

Рис. 13. Гинекологическая система (женщины)

Из рисунка 13г) для класса «4» следует, что минимальное (незначительно), среднее и максимальное, а для класса «3» (рис. 13в)) и в некоторой степени для класса «2» (рис. 13б)) минимальное расстояние между центром масс и своими векторами больше соответствующего расстояния между центром масс и чужими векторами. Для класса «3» (рис. 13в)) указанные минимальные величины приблизительно одинаковы.

Шесть и пять своих элементов (Ряд 1) соответственно классов «2» и «3» (рис. 13бб, вв)) находятся вблизи центра масс, но ввиду их малого числа неинформативны для анализа. Свои элементы класса «4» (рис. 13гг)) не локализованы вблизи центра масс: они имеются до конца отрезка $[Dk_{\min}, Dk_{\max}]$, а максимум соответствующей функции распределения (Ряд 1) смазан вправо от его начала. Функция распределения чужих элементов (Ряд 2) классов «2», «3» и «4» схожа с рис. 13аа), но ее максимум удален от центра масс относительно аналога по своим элементам (Ряд 1): для классов «2», «3» – приблизительно так же, для класса «4» (максимум в начальной точке этого отрезка) – левее.

Классификация достигается на коротком полиноме (длина 69).

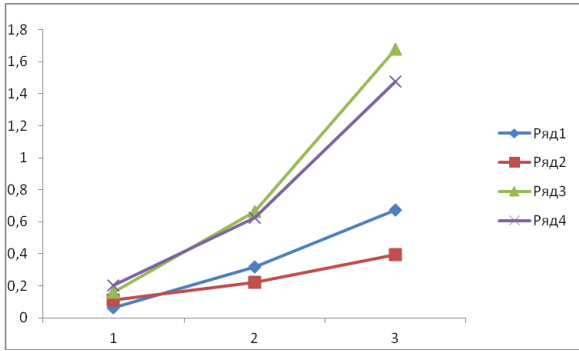
Для классов «2» и «3» минимальное расстояние между своими векторами превышает наименьшее расстояние между своими и чужими векторами, но они маленького объема (рис. 13б), в)). Для класса «4» первая упомянутая выше величина меньше второй (рис. 13г)).

Печень и желчевыводящие пути

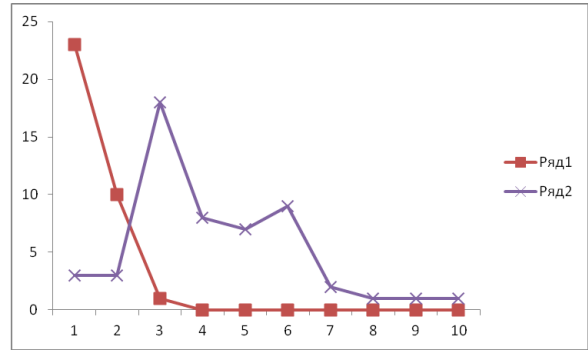
Рассмотрим заболевания печени и желчевыводящих путей у женщин (рис. 14). Обучающее множество состоит из 87 элементов. Классы «1» и «4» включают по 34 набора крови. Класс «2» и «3» имеют 15 и 4 элемента.

По рисункам 14в), г) видно, что для классов «3» и «4» минимальное расстояние, а для класса «4» также среднее и максимальное расстояние между своими векторами больше аналогичной величины для расстояния между своими и чужими векторами. Для класса «3» средние величины приблизительно одинаковы. Для класса «2» (рис. 14б)) сохраняются закономерности, характерные для класса «1», однако с незначительной разницей соответствующих величин.

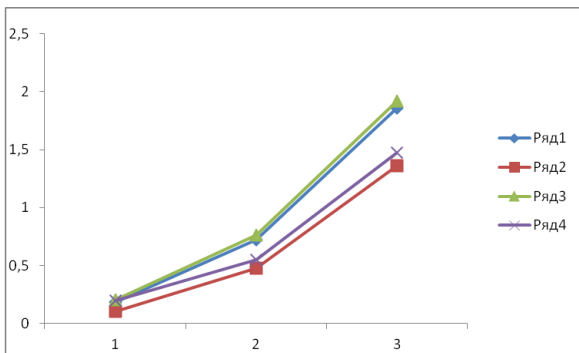
Из рисунка 14г) для класса «4» следует, что среднее и максимальное, а для класса «3» (рис. 14в)) минимальное расстояние между центром масс и своими векторами больше соответствующего расстояния между центром масс и чужими векторами. Для класса «2» (рис. 14б)) сохраняются закономерности, характерные для класса «1», однако с незначительной разницей соответствующих величин.



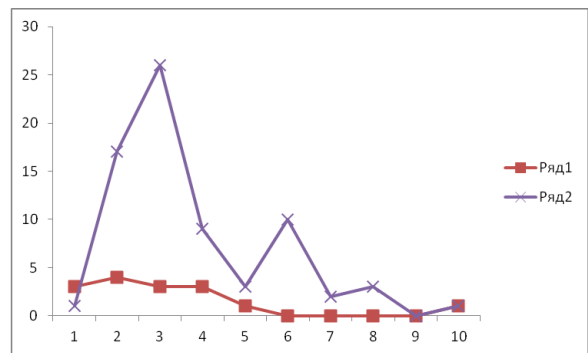
а) «1» min, middle, max



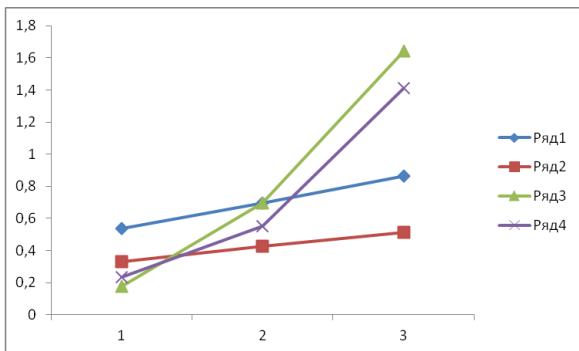
аа) «1» распределение своих/чужих



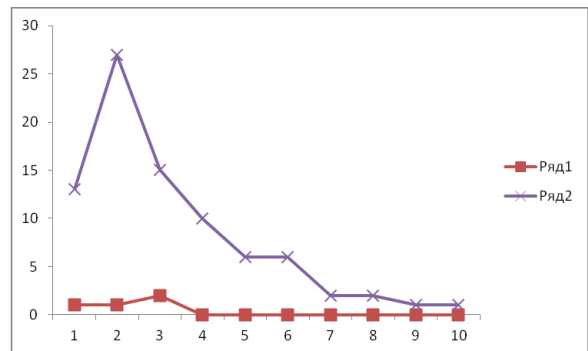
б) «2» min, middle, max



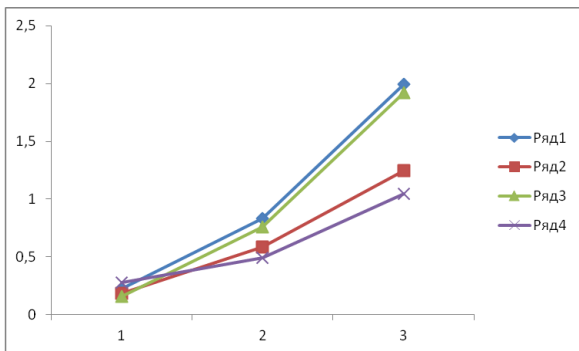
бб) «2» распределение своих/чужих



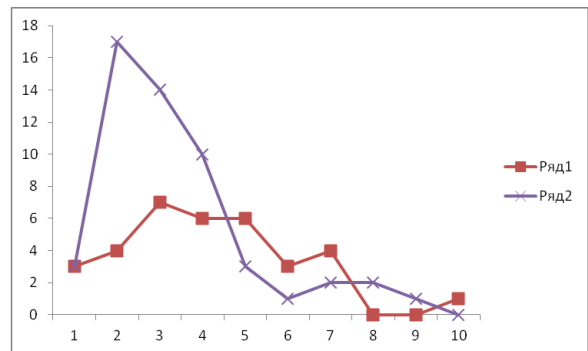
в) «3» min, middle, max



вв) «3» распределение своих/чужих



г) «4» min, middle, max



гг) «4» распределение своих/чужих

Рис. 14. Печень и желчевыводящие пути (женщины)

Четыре своих элемента (Ряд 1) класса «3» (рис. 14вв)) находятся вблизи центра масс, но ввиду их малого числа неинформативны для анализа. Свои элементы классов «2» и «4» (рис. 14бб), гг)) не локализованы вблизи центра масс: они имеются до середины (рис. 14бб)) или до конца (рис. 14гг) отрезка $[Dk_{\min}, Dk_{\max}]$, а максимум соответствующей функции распределения (Ряд 1) смещен вправо от его начала. Функция распределения чужих элементов (Ряд 2) классов «2», «3», «4» схожа с рис. 14аа), но ее максимум удален от центра масс относительно аналога по своим элементам (Ряд 1): для класса «2» – приблизительно так же, для классов «3» и «4» – левее.

Классификация достигается на полиноме средней длины 85.

По классу «2» минимальное расстояние между своими векторами меньше аналогичной величины для расстояния между своими и чужими векторами (рис. 14бб)). По классу «3» первая величина существенно больше второй (рис. 14в)), но этот класс маленького объема. По классу «4» первая величина больше второй (рис. 14г)), но максимумы распределений своих и чужих элементов при удалении от центра масс разнесены (рис. 14гг)).

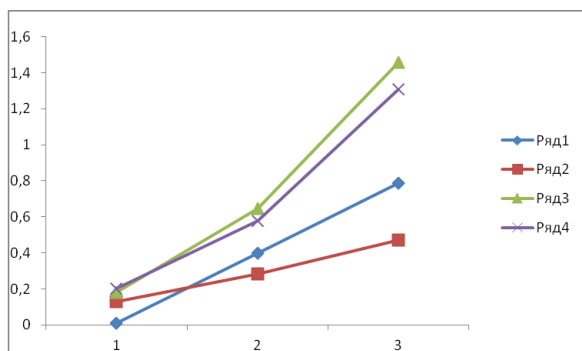
Обучение классификатора по печени и желчевыводящим путям у мужчин (рис. 15) выполнено на множестве из 74 элементов. Классы «1» и «4» включают по 31 набору крови. Класс «2» имеет 12 элементов. Класс «3» отсутствует.

По рисунку 15 г) видно, что для класса «4» среднее расстояние между своими векторами незначительно больше аналогичной величины для расстояния между своими и чужими векторами. Для класса «2» (рис. 15б)) минимальные и средние величины приблизительно одинаковы, как и минимальные и максимальные для класса «4» (рис. 15г)).

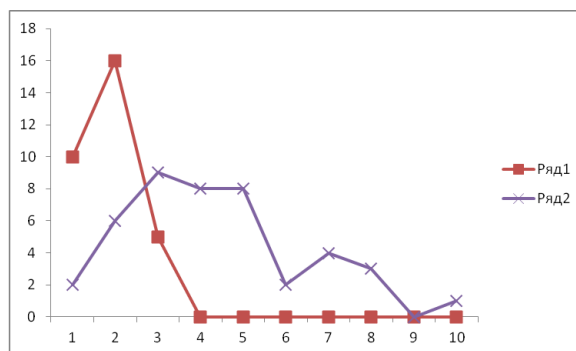
Из рисунка 15г) для класса «4» следует, что среднее и максимальное, а для класса «2» (рис. 15б)) минимальное расстояние между центром масс и своими векторами больше соответствующего расстояния между центром масс и чужими векторами. Для класса «4» (рис. 15г)) указанные минимальные величины приблизительно одинаковы.

Набор элементов класса «1» уменьшили их числа с целью выравнивания по объему с класса «4». Это привело к смещению максимума функции распределения своих элементов (Ряд 1) от центра масс, однако вблизи него так же имеется небольшая окрестность, в которой все они находятся (рис. 15аа), причем справа от максимума их число убывает при удалении от центра масс.

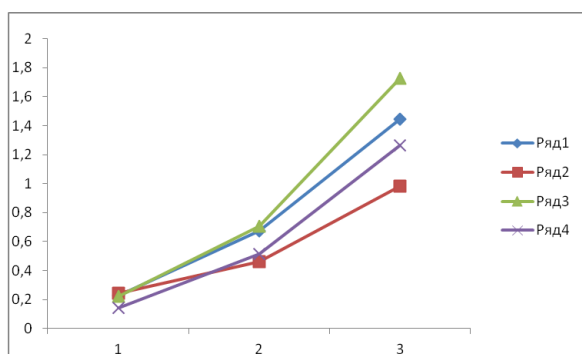
В других классах свои элементы не локализованы около центра масс: они имеются до конца (рис. 15гг) или почти до конца (рис. 15бб)) отрезка $[Dk_{\min}, Dk_{\max}]$, а максимум соответствующей функции распределения (Ряд 1) смещен вправо от его начальной точки. Там же находится максимум Ряда 2 класса «2» (рис. 15бб)). Для класса «4» (рис. 15гг)) он расположен в левой части размазанного максимума функции распределения своих элементов (Ряд 1).



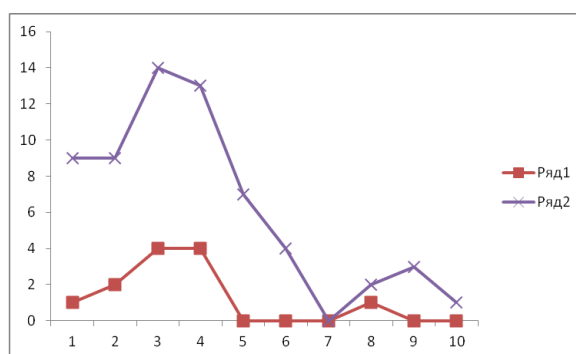
а) «1» min, middle, max



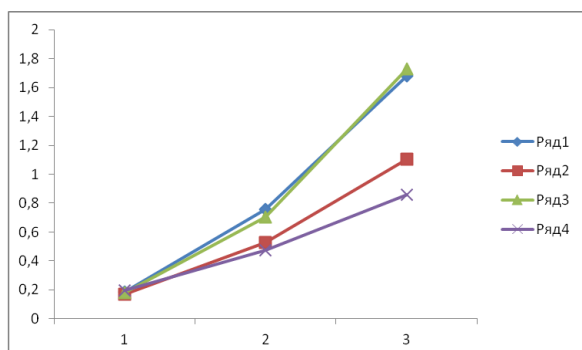
aa) «1» распределение своих/чужих



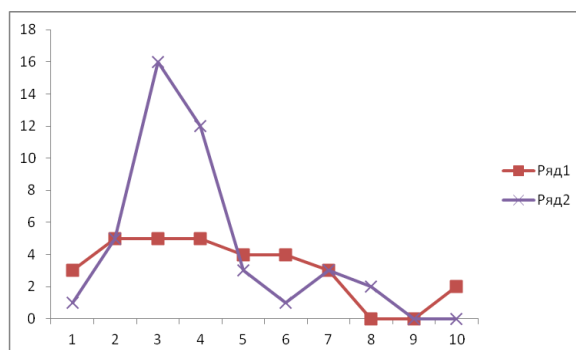
б) «2» min, middle, max



бб) «2» распределение своих/чужих



г) «4» min, middle, max



гг) «4» распределение своих/чужих

Рис. 15. Печень и желчевыводящие пути (мужчины)

Классификация достигается на очень коротком полиноме (длина 61).

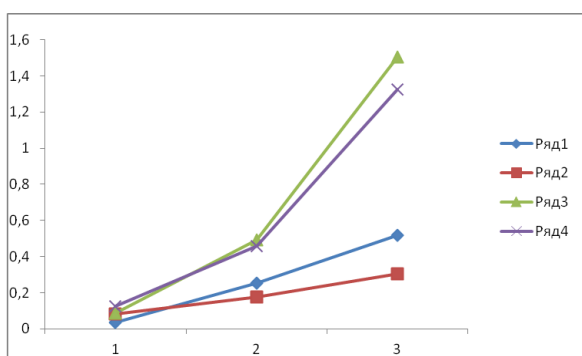
Для класса «2» минимальное расстояние между своими векторами чуть меньше аналогичной величины для расстояния между своими и чужими векторами (рис. 15б)). Для класса «4» эти величины равны (рис. 15г)).

Грудные железы

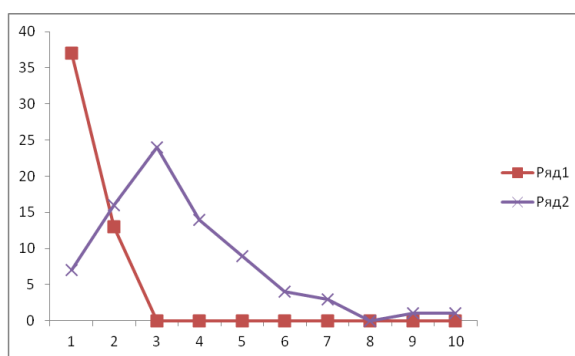
Заболевания грудных желез рассмотрены только для женщин (рис. 16). Обучение проводилось на множестве из 129 элементов. Классы «1», «3» и «4» включают соответственно 50, 29 и 50 наборов крови. Класс «2» отсутствует.

Из рисунка 16г) следует, что для класса «4» среднее расстояние между своими векторами больше аналогичной величины для расстояния между своими и чужими векторами. Указанные минимальные величины для классов «3» и «4» и максимальное для класса «4» примерно одинаковы (рис. 16в),г)).

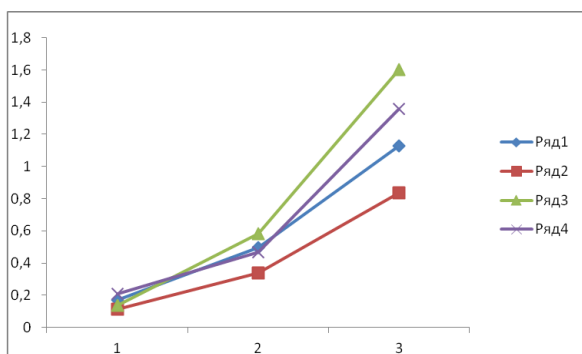
Из рисунка 16г) для класса «4» следует, что среднее и максимальное расстояние между центром масс и своими векторами больше соответствующего расстояния между центром масс и чужими векторами. Для класса «3» (рис. 16в)) сохраняются закономерности, характерные для класса «1».



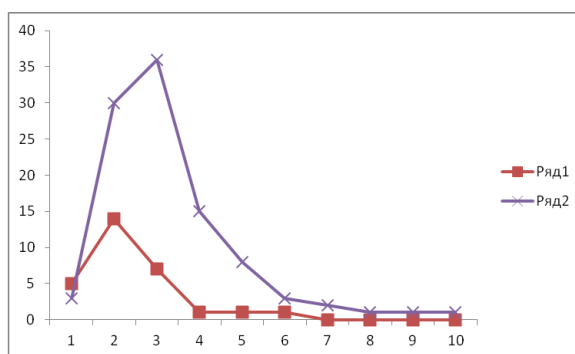
а) «1» min, middle, max



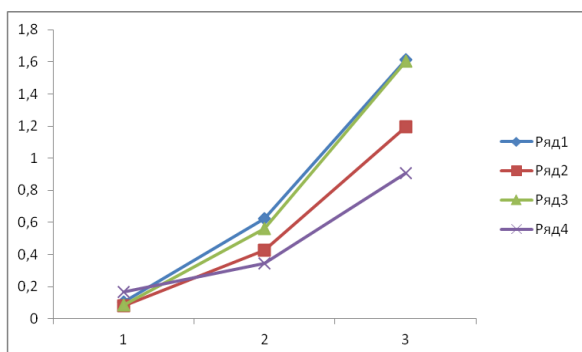
aa) «1» распределение своих/чужих



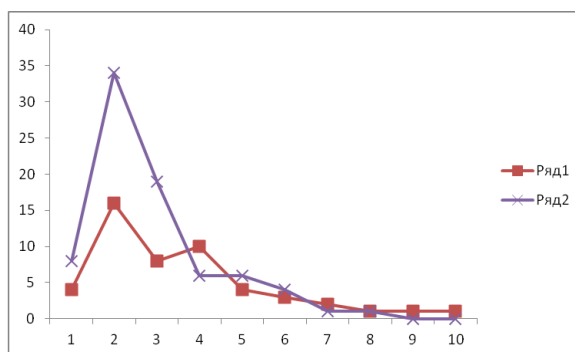
в) «3» min, middle, max



вв) «3» распределение своих/чужих



г) «4» min, middle, max



гг) «4» распределение своих/чужих

Рис. 16. Грудные железы (женщины)

Свои элементы классов «3», «4» (рис. 16вв), гг)) не локализованы вблизи центра масс: они имеются до конца (рис. 16гг) или дальше середины (рис. 10вв)) отрезка $[Dk_{\min}, Dk_{\max}]$, а максимум соответствующей функции распределения (Ряд 1) смещен вправо от его начальной точки. Функция распределения чужих элементов (Ряд 2) для классов «3», «4» схожа с рис. 16аа), а ее максимум удален от центра масс относительно аналога по своим элементам (Ряд 1): для класса «4» – так же, для класса «3» – правее.

Четыре элемента класса «3» удалось классифицировать только с помощью многочлена из 165 мономов. Они расположены вблизи минимального расстояния между своими векторами, которое превышает наименьшее расстояние между своими и чужими векторами (рис. 16в)).

Заключение

Проведено исследование статистического классификатора, основанного на методе полиномиальной регрессии и имеющего вероятностные оценки. Он предназначен для оценивания по четырехбалльной шкале состояния девяти СО у женщин и семи – у мужчин, основываясь на данных лабораторного анализа периферической крови (из пальца). Точность классификации на обучающих множествах для различных СО находится в диапазоне 95–100 %.

Рассматриваемая проблематика является актуальной ввиду необходимости разработки статистических методов предварительной диагностики в доказательной медицине.

Поскольку классификатор, полученный в результате реализации метода, зависит от характера обучения, то множества, на которых оно проходило, стали объектом нашего исследования.

Для каждой СО и конкретного класса здоровья вычислялись величины, характеризующие особенности взаимного расположения пар элементов обучающих множеств. В одном варианте оба вектора принадлежат рассматриваемому классу (свой–свой), а в другом – один из них к нему не относится (свой–чужой). Для каждого класса найдены минимальные, средние и максимальные расстояния между векторами. Показано, как эти значения, полученные для пар различных типов, соотносятся между собой.

Для ряда СО получено, что в случаях когда минимальное расстояние между своим векторами превышает наименьшее расстояние в парах (свой–чужой), возникает необходимость усложнять структуру полиномов и увеличивать количество мономов. Интуитивно это понятно, поскольку в такой ситуации для рассматриваемого проблемного элемента ближайшими являются векторы другого класса.

Введено понятие центра масс по данному классу. Аналогично вышеописанной процедуре рассмотрены пары (центр масс–свой) и (центр масс–чужой). По каждому классу найдены минимальные, средние и максимальные

расстояния между векторами для пар различных типов. Проведено сопоставление этих значений между собой.

В ряде случаев показано наличие структуры в расположении своих и чужих векторов относительно центра масс. А именно, минимальное, среднее и максимальное расстояние между центром масс и своими векторами больше (или, наоборот, меньше) трех аналогичных величин для пар (центр масс–чужой).

Класс «1», универсальный для всех СО и незначительно различающийся лишь количественно, обладает особыми свойствами.

Минимальное, среднее и максимальное расстояние между своими векторами этого класса меньше соответствующих расстояний между парами свой–чужой. Аналогичные соотношения выполняются для пар (центр масс–свой) и (центр масс–чужой). Этот порядок или существенно нарушается для классов «2», «3», «4», или соответствующие величины сближаются, разница их значений уменьшается.

Вблизи центра масс класса «1» имеется небольшая окрестность, в которой находятся все его элементы, причем их число убывает при удалении от центра масс. В то же время в этой окрестности чужих элементов мало, а подавляющее их количество располагается вне нее; соответствующая функция распределения сначала нарастает при удалении от центра масс, а затем имеет тенденцию к убыванию, и ее максимум находится на удалении от этой окрестности, т.е. там, где отсутствуют элементы класса «1».

Отмеченные здесь закономерности взаимного расположения своих и чужих элементов, а также характер распределения количества своих и чужих элементов по классам «1», «2», «3», «4» вполне соответствуют и могут служить объяснением того факта, что для распознавания элементов класса «1» достаточно использовать полиномы более простой структуры, имеющие малую длину. Классы «2», «3», «4» являются более проблемными в этом отношении. Для их классификации требуется усложнение структуры, добавочное включение мономов более высокого порядка и, соответственно, увеличение длины полинома. Эти зависимости имеют место для всех СО.

Библиографический список

1. Количественная оценка гомеостатической активности здоровых и больных людей / Ставицкий Р.В. [и др.] // М.: ГАРТ. 2013. 131 с.

2. Гавриков М.Б., Пестрякова Н.В. Метод полиномиальной регрессии в задачах распознавания печатных и рукопечатных символов // Препринты ИПМ им.М.В.Келдыша. 2004. № 22. 12 с.

3. Об одном методе распознавания символов, основанном на полиномиальной регрессии / Гавриков М.Б. [и др.] // Автоматика и Телемеханика. 2006. № 2. С. 119-134.

4. Пестрякова Н.В. Метод распознавания символов, основанный на полиномиальной регрессии. // М.: УРСС. 2011. 141 с.

5. Гавриков Б.М., Пестрякова Н.В. О построении признакового пространства в задаче обучения // Информационные технологии и вычислительные системы. 2018. №1. С. 22-29.

Оглавление

Введение	3
Постановка задачи классификации	3
Параметры структуры базы.....	7
Пищеварительная система.....	11
Органы дыхания	15
Опорно-двигательный аппарат	18
Урологическая система.....	21
Эндокринная система.....	24
ЦНС и органы чувствительности	28
Гинекологическая система	31
Печень и желчевыводящие пути.....	33
Грудные железы.....	36
Заключение.....	38
Библиографический список.....	39