

Экспертная система поддержки принятия решений в
оценке риска генетического бесплодия и спонтанного
прерывания беременности

Пономарева Н.С.

02.12.2015 г.

Задача классификации

формализованная задача, в которой имеется множество объектов (ситуаций), разделённых некоторым образом на классы. Задано конечное множество объектов, для которых известно, к каким классам они относятся. Это множество называется выборкой. Классовая принадлежность остальных объектов неизвестна. Требуется построить алгоритм, способный классифицировать произвольн

Исходные данные

Microsoft Excel window: данные по генам для программы.xls [Режим совместимости] - Microsoft Excel

| | диагноз | IL-10 G1082A | TNF G308A | IL1b T31C | IL-6 C174G | IL-10 C592A | IL-10 C819T | iTGB3 | Leu33Pr | iTGA2 807 C/T | MTHFR | MTR | MTRR | PAI1 -675 5G/4G | FGB 455G- |
|-----|----------|--------------|-----------|-----------|------------|-------------|-------------|-------|---------|---------------|-------|-----|------|-----------------|-----------|
| 1к | контроль | NP | NP | NP | NN | NP | NP | NN | NP | NP | NP | NN | NP | NP | NP |
| 2к | контроль | NP | NP | NN | NP | NN | NN | NN | NP | NP | NP | NN | NP | NP | NN |
| 3к | контроль | PP | NP | NP | NP | NN | NN | NN | NP | NP | NP | PP | PP | NN | NP |
| 1н | нб | NP | NN | NP | PP | NN | NN | NP | NN | NP | NP | NN | NP | NN | NN |
| 4к | контроль | NP | NN | NN | NN | NP | NP | NN | NN | NN | NN | NP | PP | PP | NN |
| 5к | контроль | NP | NP | NP | PP | NN | NN | NN | NP | NN | NN | NN | NP | NP | NN |
| 6к | контроль | NP | NN | NP | NP | NP | NP | NN | NP | NP | NP | NP | NP | PP | NN |
| 2н | нб | NP | NN | NN | NN | NN | NN | NN | NP | NP | NN | NN | NP | NP | NN |
| 3н | нб | NP | NN | NP | NP | NP | NP | NN | NP | NP | NP | PP | PP | NP | NN |
| 7к | контроль | NN | NP | NN | NP | NP | NP | NN | NP | NN | NN | NN | PP | NP | NN |
| 8к | контроль | NN | NP | NP | NP | NP | NP | PP | NN | NN | NN | NP | NN | NP | NP |
| 9к | контроль | NP | NN | NN | NP | NP | NP | NN | NP | NN | NN | NP | NP | NP | NP |
| 10к | контроль | PP | NP | PP | NP | NN | NN | NP | NP | NN | NN | NN | NP | NP | NN |
| 11к | контроль | NP | NN | NP | NP | NP | NP | NN | PP | PP | NN | PP | NP | NP | NN |
| 12к | контроль | NP | NN | NP | NP | NN | NN | NN | PP | NN | NN | NN | NP | PP | NN |
| 13к | контроль | NP | NP | NN | PP | NN | NN | NP | NP | NN | NN | NN | NP | PP | PP |
| 14к | контроль | NN | NN | NN | PP | NP | NP | PP | NN | NN | NN | NN | NP | NN | NP |
| 15к | контроль | PP | NN | PP | PP | NN | NN | NN | PP | NP | NN | PP | PP | PP | NN |
| 16к | контроль | NP | NN | NP | NP | NP | NP | NP | NN | NN | NN | NN | NP | NP | NN |
| 1а | са | NP | NN | NN | NP | NP | NP | NP | NN | NP | NN | NN | NN | NN | NP |
| 17к | контроль | PP | NN | NN | NN | NN | NN | NP | NN | NN | NN | NN | NN | NP | NN |
| 4н | нб | PP | NP | NN | PP | NN | NN | NN | PP | NN | NP | NN | NN | NN | NP |
| 18к | контроль | NP | NN | NN | NP | NN | NP | NP | NN | NP | NN | NN | NN | NP | NP |
| 5н | нб | NP | NN | NP | NP | NN | NN | NN | PP | NN | NN | NN | NP | NP | NP |
| 6н | нб | NP | NN | NP | PP | NN | NN | NN | NN | NN | NN | NP | NP | PP | NP |
| 19к | контроль | NP | NN | NP | NP | NN | NN | NN | NN | NP | NP | NP | NN | NP | NP |

Гены факторов системы гемостаза

| | | | |
|-----------------------------------|-----------------|-----------|-------------------|
| Интегрин В 3 | <i>ITGB3</i> | <i>g1</i> | <i>T1565C</i> |
| Интегрин А2 | <i>ITGA2</i> | <i>g2</i> | <i>C807T</i> |
| β-фибриноген: | <i>FGB</i> | <i>g3</i> | <i>G-455A</i> |
| Ингибитор активатора плазминогена | <i>SERPINE1</i> | <i>g4</i> | <i>-675 5G/4G</i> |

Гены цитокинов

| | | | |
|-------------------------|--------------|-----------|----------------|
| Фактор некроза опухолей | <i>TNFα</i> | <i>c5</i> | <i>-308G-A</i> |
| Интерлейкин 10 | <i>IL-10</i> | <i>c4</i> | <i>-819C-T</i> |
| Интерлейкин 10 | <i>IL-10</i> | <i>c3</i> | <i>-592C-A</i> |
| Интерлейкин 1β | <i>IL-1β</i> | <i>c1</i> | <i>-31C-T</i> |
| Интерлейкин 6 | <i>IL-6</i> | <i>c2</i> | <i>-174G-C</i> |

Гены фолатного цикла

| | | | |
|--------------------------------------|--------------|-----------|----------------|
| 5,10-метилентетрагидрофолатредуктаза | <i>MTHFR</i> | <i>f1</i> | <i>677C-T</i> |
| Метионинсинтаза редуктаза | <i>MTRR</i> | <i>f2</i> | <i>66A-G</i> |
| Метионинсинтаза | <i>MTR</i> | <i>f3</i> | <i>2756A-G</i> |

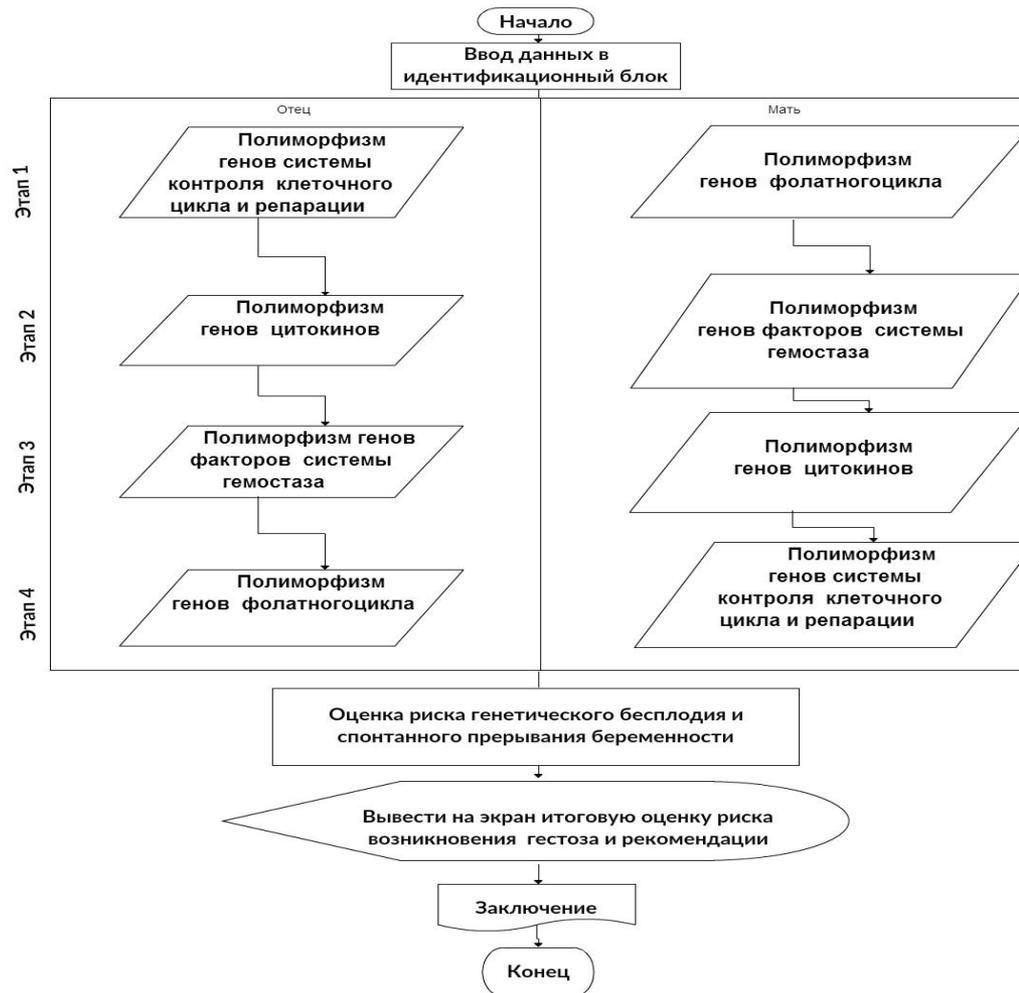
Гены системы контроля клеточного цикла и репарации

| | | | |
|------------------------------------|---------------|-----------|------------------|
| Белок эксцизионной репарации | <i>ERCC 2</i> | <i>k2</i> | <i>Lys751Gln</i> |
| Чекпойнт-киназа 2 | <i>CHEK 2</i> | <i>k3</i> | <i>1100delC</i> |
| Ген белка p53, супрессора опухолей | <i>TP53</i> | <i>k4</i> | <i>Pro72Arg</i> |

Оценка риска является задачей классификации, признаковое пространство генетического бесплодия и самопроизвольного прерывания беременности представляет собой множество данных генетических исследований супругов (мужчины U и женщины X):

а признаковое описание пары принимает вид:

Алгоритм расчета оценки риска генетического бесплодия и самопроизвольного прерывания



Алгоритм расчета оценки риска генетического бесплодия и самопроизвольного прерывания беременности

- 1 этап: $f1(x), \dots, f3(x), k1(y), \dots, k4(y);$
- 2 этап: $g1(x), \dots, g4(x), c1(y), \dots, c5(y);$
- 3 этап: $c1(x), \dots, c5(x), g1(y), \dots, g4(y);$
- 4 этап: $k1(x), \dots, k4(x), f1(y), \dots, f3(y).$

Ирисы Фишера

Ирисы Фишера — это набор данных для задачи классификации, на примере которого Рональд Фишер в 1936 году продемонстрировал работу разработанного им метода дискриминантного анализа. Иногда его также называют ирисами Андерсона, так как данные были собраны американским ботаником Эдгаром Андерсоном. Этот набор данных стал уже классическим, и часто используется в литературе для

Ирисы Фишера

| Длина чашелист ика | Ширина чашелист ика | Длина лепестка | Ширина лепестка | Вид ириса |
|--------------------------|---------------------------|-------------------|--------------------|-------------------|
| 5.1 | 3.5 | 1.4 | 0.2 | <i>setosa</i> |
| 4.9 | 3.0 | 1.4 | 0.2 | <i>setosa</i> |
| 4.7 | 3.2 | 1.3 | 0.2 | <i>setosa</i> |
| 4.6 | 3.1 | 1.5 | 0.2 | <i>setosa</i> |
| 5.0 | 3.6 | 1.4 | 0.2 | <i>setosa</i> |
| 5.4 | 3.9 | 1.7 | 0.4 | <i>setosa</i> |
| 7.0 | 3.2 | 4.7 | 1.4 | <i>versicolor</i> |
| 6.4 | 3.2 | 4.5 | 1.5 | <i>versicolor</i> |
| 6.9 | 3.1 | 4.9 | 1.5 | <i>versicolor</i> |
| 5.5 | 2.3 | 4.0 | 1.3 | <i>versicolor</i> |
| 6.5 | 2.8 | 4.6 | 1.5 | <i>versicolor</i> |
| 6.7 | 3.3 | 5.7 | 2.5 | <i>virginica</i> |
| 6.7 | 3.0 | 5.2 | 2.3 | <i>virginica</i> |
| 6.3 | 2.5 | 5.0 | 1.9 | <i>virginica</i> |
| 6.5 | 3.0 | 5.2 | 2.0 | <i>virginica</i> |
| 6.2 | 3.4 | 5.4 | 2.3 | <i>virginica</i> |
| 5.9 | 3.0 | 5.1 | 1.8 | <i>virginica</i> |

Диаграмма рассеяния ирисов Фишера

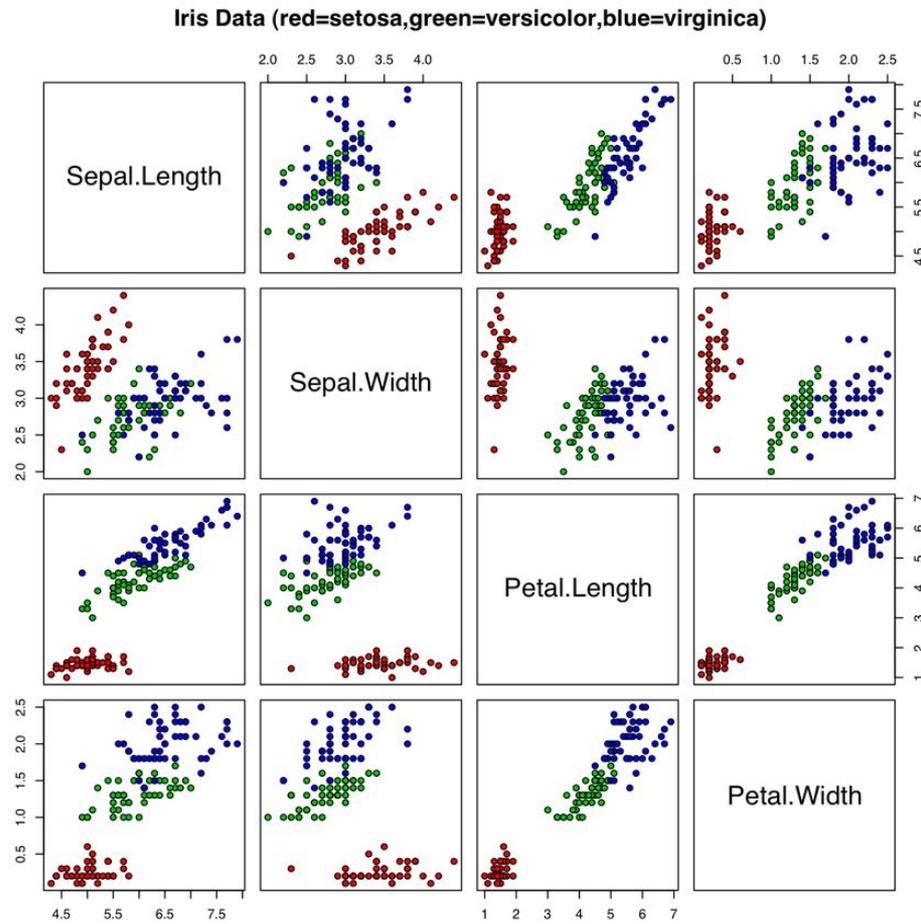


Диаграмма рассеяния генетического бесплодия и спонтанного прерывания беременности



Визуализация многомерных данных с помощью диаграмм

Эндрюса

- Дэвид Эндрюс (Andrews, David F.) в 1972-м году описал удобный способ визуализации многомерных данных.

Суть данного метода такова:

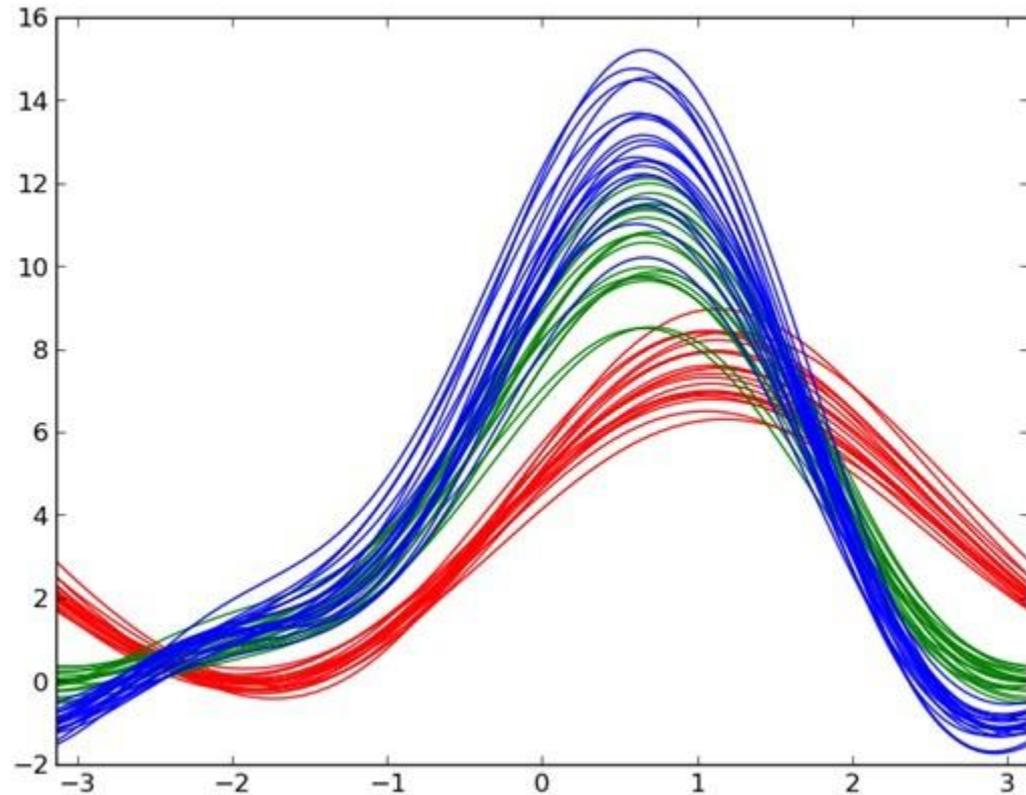
Каждая точка $x = \{x_1, x_2, x_3, \dots, x_n\}$ представляется в

$$f_x(t) = \frac{x_1}{\sqrt{2}} + x_2 \sin(t) + x_3 \cos(t) + x_4 \sin(2t) + x_5 \cos(2t) + x_6 \sin(3t) + \dots$$

$$t = [-\pi, \pi]$$

Получившаяся функция изображается на графике в промежутке

Диаграмма Эндрюса для ирисов Фишер



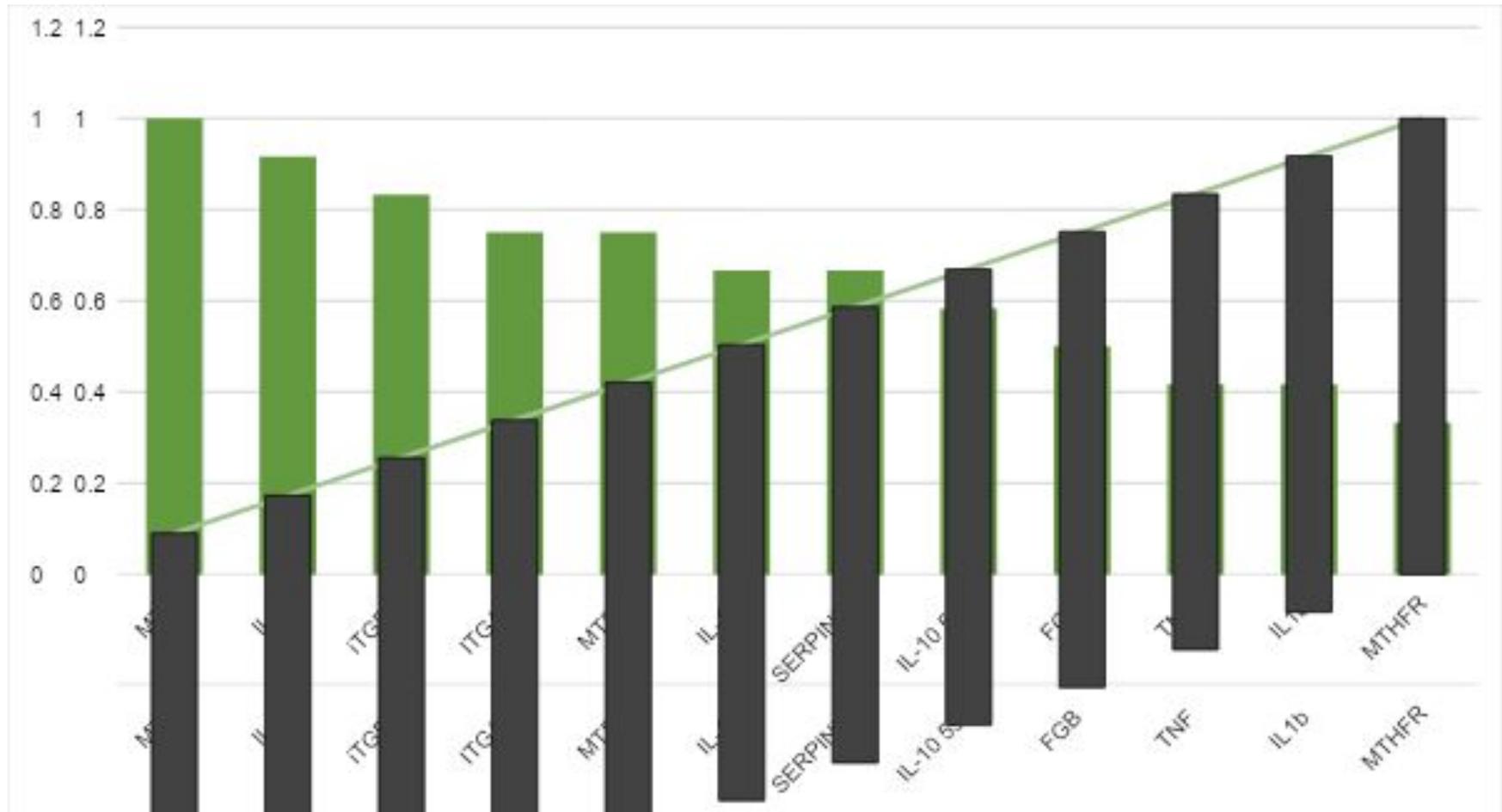
Анализ признаков x_i в осуществляется в порядке убывания их информативности

- $I_{Di}(x_1) > I_{Di}(x_2) > I_{Di}(x_3) > \dots > I_{Di}(x_m)$.

Данные для построения диаграммы Парето

| | | | | |
|-----------|------|------|-------|--------|
| MTR | 1,00 | 1,00 | 12,77 | 12,77 |
| IL-6 | 0,92 | 1,92 | 11,70 | 24,47 |
| iTGB3 | 0,83 | 2,75 | 10,64 | 35,11 |
| ITGA2 | 0,75 | 3,50 | 9,57 | 44,68 |
| MTRR | 0,75 | 4,25 | 9,57 | 54,26 |
| IL-10 | 0,67 | 4,92 | 8,51 | 62,77 |
| SERPINE1 | 0,67 | 5,58 | 8,51 | 71,28 |
| IL-10 592 | 0,58 | 6,17 | 7,45 | 78,72 |
| FGB | 0,50 | 6,67 | 6,38 | 85,11 |
| TNF | 0,42 | 7,08 | 5,32 | 90,43 |
| IL1b | 0,42 | 7,50 | 5,32 | 95,74 |
| MTHFR | 0,33 | 7,83 | 4,26 | 100,00 |

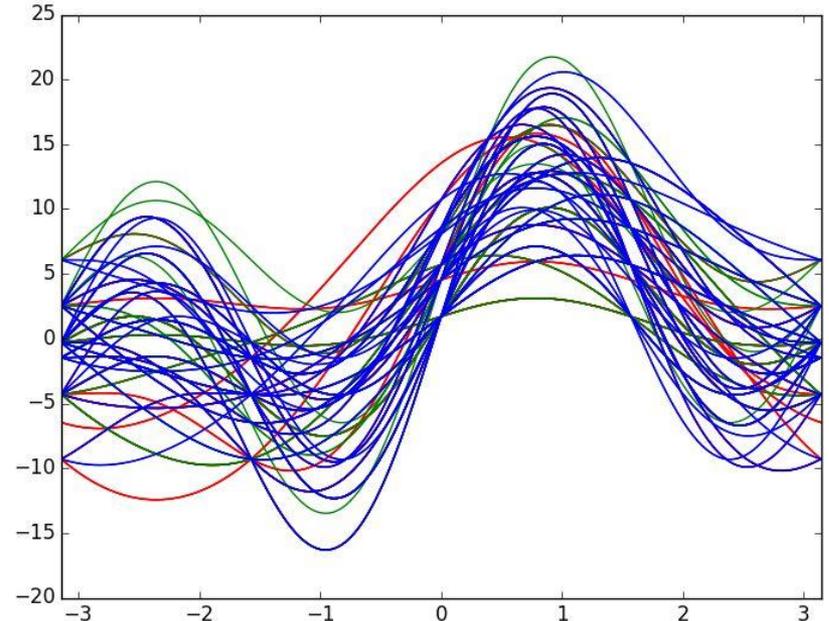
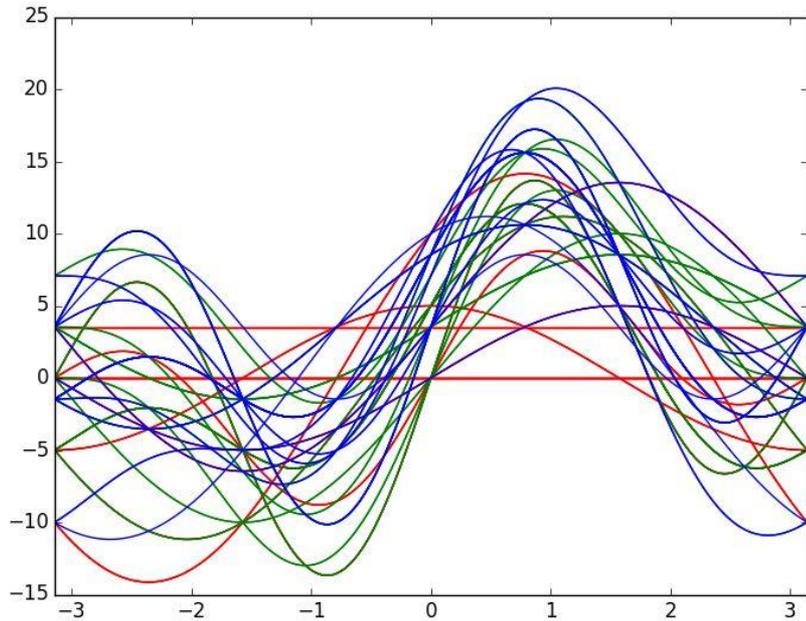
Вклад признаков при репродуктивных потерях



Наличие полиморфизмов в генах MTR; IL-6, iTGB3, ITGA2, MTRR, IL-10

выборка 86

выборка 241



Многослойная искусственная нейронная сеть

Для решения проблемы с линейной неразделимостью образов была предложена двуслойная искусственная нейронная сеть. В качестве архитектуры сети была выбрана сеть с 13 входными нейронами (по количеству признаков), 1 скрытым слоем, содержащим 100 нейронов, и 1 выходным. Классу «0» соответствовал сигнал «0» выходного нейрона, а классу «1» - сигнал «1». В качестве пороговой функции был выбран

Перекрестная проверка

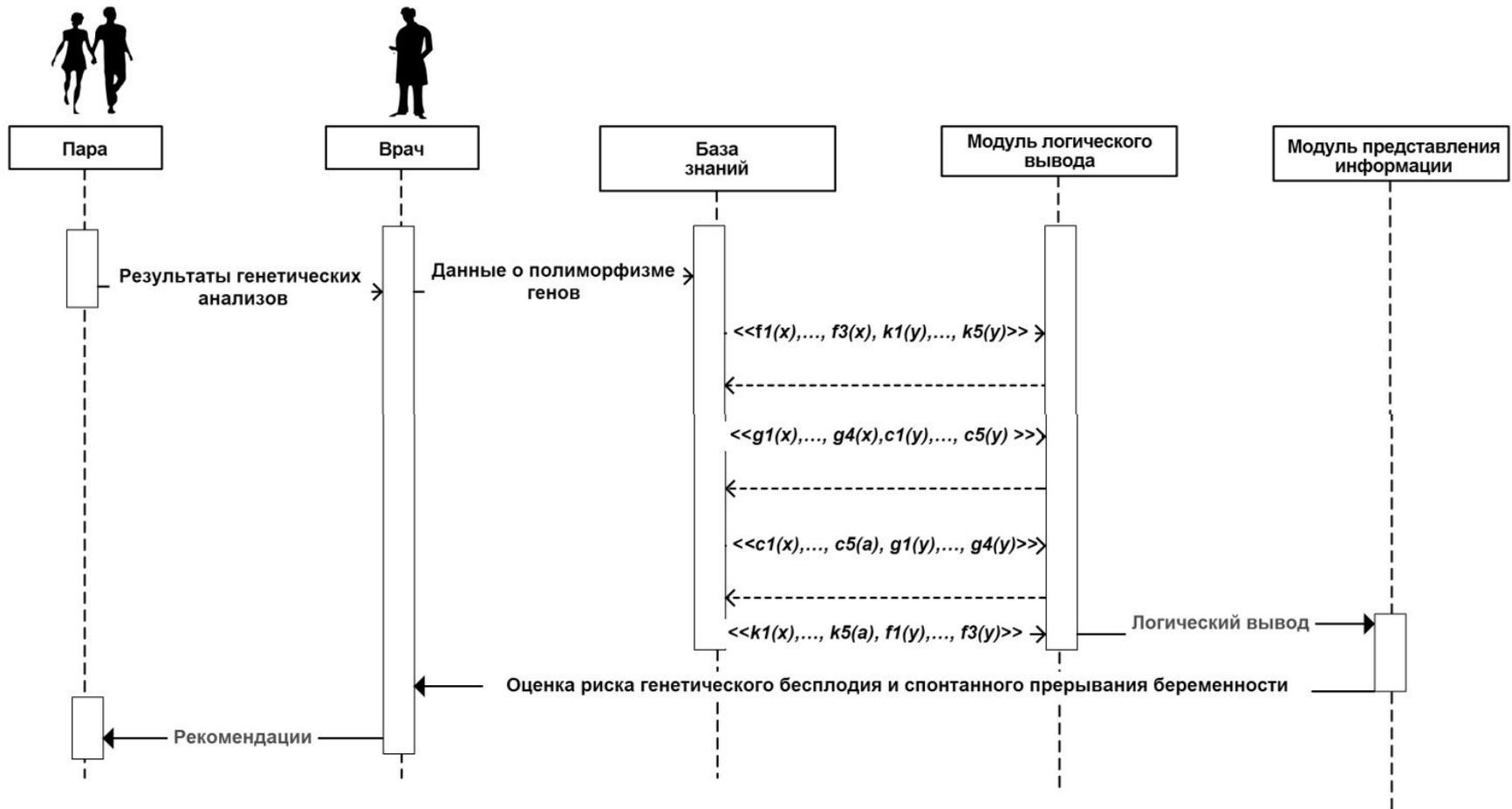
Результаты

перекрестной проверки
приведены.

Исходное обучающее множество было перемешано и разбито на $k=10$ частей равного размера. Затем последовательно выбиралась каждая из 10 частей и служила проверочным множеством, в то время как остальное множество использовалось для обучения.

| | |
|----|------|
| 22 | 0,91 |
| 24 | 1 |
| 23 | 0,65 |
| 21 | 0,88 |
| 24 | 1 |
| 21 | 0,86 |
| 24 | 1 |
| 23 | 0,95 |
| 24 | 1 |
| 18 | 0,75 |

Взаимодействие субъектов и компонентов экспертной системы при расчете оценки риска генетического бесплодия и спонтанного прерывания беременности



Форма ввода данных о полиморфизме генов

Генетическое обследование пациента 0391550001 (мать)

| Символ гена | Ген | Группа генов | Полиморфизм | Генотип |
|-------------|--------------------------------------|---------------------------------|-------------|----------------|
| MTHFR | 5,10-метилентетрагидрофолатредуктаза | Гены фолатного цикла | 677C-T | NP ▾ |
| MTRR | Метионинсинтаза редуктаза | Гены фолатного цикла | 66A-G | NN ▾ |
| MTR | Метионинсинтаза | Гены фолатного цикла | 2756A-G | NN NP PP |
| ITGB3 | Интегрин В3 | Гены факторов системы гемостаза | T1565C | PP NN |
| ITGA2 | Интегрин А2 | Гены факторов системы гемостаза | C807T | NN ▾ |
| FGB | β-фибриноген | Гены факторов системы гемостаза | G455A | NN ▾ |
| SERPINE1 | - | Гены факторов системы гемостаза | 675 5G/4G | NN ▾ |
| IL-1β | Интерлейкин 1β | Гены цитокинов | 31C-T | NN ▾ |
| IL-6 | Интерлейкин 6 | Гены цитокинов | 174G-C | NN ▾ |

Окно результата оценка риска бесплодия



← Я fiav.ru Оценка риска генетического бесплодия и спонтанного прерывания беременности

Оценка риска генетического бесплодия и спонтанного прерывания беременности

Оценка генетического обследования матери по балльной шкале - 10
Оценка генетического обследования отца по балльной шкале - 0
Итоговая оценка - 10
Высокий риск спонтанного прерывания беременности