

Документ подписан простой электронной подписью
Информация о владельце:
ФИО: Локтионова Оксана Геннадьевна
Должность: проректор по учебной работе
Дата подписания: 16.06.2023 13:46:30
Уникальный программный ключ:
0b817ca911e6668abb13a5d426d39e5f1c11eabbf73e943df4a4850511f55d089

МИНОБРАЗОВАНИЯ РОССИИ

Федеральное государственное бюджетное образовательное учреждение
высшего образования

«Юго-Западный государственный университет»
(ЮЗГУ)

Кафедра программной инженерии

УТВЕРЖДАЮ
Проректор по учебной работе
О.Г. Локтионова
« 14 » 12 2021 г.



Информационно-аналитические системы

методические указания к лабораторным занятиям для магистров направления
02.04.03 Математическое обеспечение и администрирование информационных
систем

Курск 2021

УДК 004.94

Составитель: Ю.А. Халин

Рецензент

Кандидат технических наук, с.н.с, доцент А.В. Ткаченко

Информационно-аналитические системы: методические указания к лабораторным занятиям / Юго-Зап. гос. ун-т; сост. Ю.А. Халин. - Курск, 2021. - 29 с. - Библиогр.: с. 29.

В работе рассматриваются информационно-аналитические системы поддержки принятия решений. Изложены краткие теоретические сведения, приведены примеры информационно-аналитических систем, а также задания для самостоятельного решения.

Методические рекомендации предназначены для студентов, обучающихся по направлению подготовки 02.04.03 Математическое обеспечение и администрирование информационных систем.

Текст печатается в авторской редакции.

Подписано в печать 14.12.2021 . Формат 60x84 1/16.
Усл.печ. л. 1,7 п.л . Уч.-изд. л. 1,5. Тираж 100 экз. Заказ 1724. Бесплатно.
Юго-Западный государственный университет.
305040, г. Курск, ул. 50 лет Октября, 94.

Содержание

Изучение методов интеллектуального анализа данных в среде Statgraphics. компонентный анализ. кластерный анализ.	4
Решение задачи кластеризации с помощью нейросетевого моделирования	18
Список литературы:	29

работа №1

Изучение методов интеллектуального анализа данных в среде Statgraphics. компонентный анализ. кластерный анализ.

1 Цель работы

Изучение особенностей применения компонентного и кластерного анализа в среде *StatGraphics* с целью исследования структуры данных и извлечения знаний.

2 Теоретические сведения

2.1 Компонентный анализ

С геометрических позиций алгоритм построения главных компонент (ГК) состоит в следующем.

Производится центрирование исходных данных (рисунок 1а); система координат переносится в центр распределения данных (центроид) (рисунок 1б).

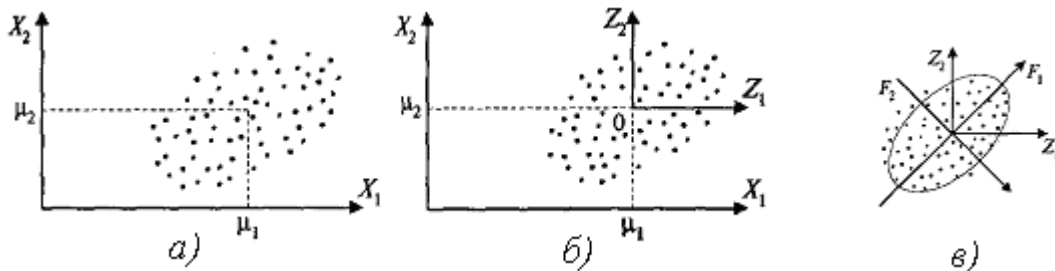


Рисунок 1 – Формирование главных компонент

Затем осуществляется формирование главных компонент F_1, F_2, \dots, F_p (рисунок 1в). Линейные комбинации выбираются таким образом, что среди всех возможных комбинаций первая ГК $F_1(X)$ обладает наибольшей дисперсией. Дисперсия σ стремится к максимуму. Графически это выглядит как ориентация новой координатной оси F_1 вдоль направления наибольшей вытянутости эллипсоида рассеивания объектов в исходном пространстве P признаков.

Вторая ГК $F_2(X)$ перпендикулярна первой и строится исходя из предположений нахождения максимальной дисперсии среди всех оставшихся линейных комбинаций.

Графически это интерпретируется направлением наибольшей вытянутости эллипсоида рассеивания, который перпендикулярен первой главной компоненте.

Достоинства метода.

1. С точки зрения визуализации многомерных данных метод обладает свойствами наименьшего искажения структуры исходного пространства при проецировании в пространство меньшей размерности.

2. Метод применяется успешно в системе с другими методами исследования данных, например в корреляционно-регрессионном анализе.

Недостатки метода.

Возможна ситуация, когда весовые коэффициенты имеют близкие по величине значения. В этом случае результат слабо интерпретируем. Эта проблема решается применением других видов анализа, например факторного, добавлением или исключением переменных из анализа.

2.2 Кластерный анализ

Постановка задачи кластеризации. Дано: множество n объектов, характеризующихся m признаками. Необходимо выполнить разбиение заданного множества объектов на заранее неизвестное или в редких случаях заданное количество групп (кластеров) на основании некоторого математического критерия кластеризации.

Cluster (гроздь, пучок, скопление) – группа элементов, которые характеризуются какими-то общими свойствами. Критерий качества кластеризации в той или иной мере должен отражать следующие неформальные требования:

- 1) внутри групп объекты должны быть тесно связаны между собой;
- 2) объекты разных групп должны быть далеки друг от друга;
- 3) распределение объектов по группам должно быть равномерным.

Методы кластерного анализа позволяют решать следующие задачи:

- 1) проведение классификации объектов с учетом множества признаков с целью углубления знаний о множестве изучаемых признаков;
- 2) проверка выдвигаемых предположений о наличии некоторой структуры в изучаемом множестве объектов;
- 3) построение новых классификаций для слабо изученных явлений, то есть поиск в изучаемом множестве заранее неизвестной структуры.

Все методы кластерного анализа делятся на две группы:

- агломеративные (объединяющие) – построены на основе последовательного объединения объектов в группы.
- дивизионные (разделяющие) – построены на основе расчленения группы на отдельные объекты.

Основные проблемы в кластерном анализе.

1 Определение мер сходства (метрики).

Метрика – мера близости между двумя объектами в m -мерном пространстве.

От выбранной метрики зависит окончательный вариант разбиения.

1) евклидово расстояние:
$$d_{ij} = \sqrt{\sum_{k=1}^m (x_{ik} - x_{jk})^2},$$

2) взвешенное евклидово расстояние:
$$d_{ij} = \sqrt{\sum_{k=1}^m w_k (x_{ik} - x_{jk})^2},$$

3) расстояние *city-block*:
$$d_{ij} = \sum_{k=1}^m |x_{ik} - x_{jk}|,$$

$$4) \text{ расстояние Минковского: } d_{ij} = \left(\sum_{k=1}^m |x_{ik} - x_{jk}|^p \right)^{1/p},$$

где d_{ij} – расстояние между i -м и j -м объектами;

x_{il}, x_{jl} – значения l -й переменной соответственно у i -го и j -го объектов;

w_k – вес, приписываемый k -й переменной.

Обобщенный алгоритм кластерного анализа.

Шаг 1. Задается начальное (искусственное или произвольное) разбиение на кластеры, и определяется некоторый математический критерий качества автоматической классификации.

Шаг 2. Объекты переносятся из кластера в кластер до тех пор, пока значение критерия качества не перестанет улучшаться. При этом возможен либо полный перебор вариантов, либо сокращенный на основании каких-либо эвристик.

3 Ход работы

3.1 Рассмотрим пример, относящийся к сравнительному оцениванию автомобилей. Создадим таблицу с данными: выборочные сведения о фирме-изготовителе, название модели, а также параметры автомобиля: вес (*weight*), число цилиндров (*cylinders*), ускорение (*accel*), объем двигателя (*displace*) и мощность (*horsepower*) (рисунок 2).

	изготовитель	Модель	Вес	кол цилиндро	Ускорение	Объем	Мощность
1	Volkswagen	Rabbit DI	1985	4	21.5	90	48
2	Ford	Fiesta	1800	4	14.4	98	66
3	Mazda	GLC Deluxe	1985	4	19.4	78	52
4	Datsun	B210 GX	2070	4	18.6	85	70
5	Honda	Civic CVCC	1800	4	16.4	91	60
6	Oldsmobile	Cutlass	3365	8	15.5	260	110
7	Dodge	Diplomat	3735	8	13.2	318	140
8	Mercury	Monarch	3570	8	12.8	302	139
9	Pontiac	Phoenix	3535	6	19.2	231	105
10	Chevrolet	Malibu	3155	6	18.2	200	95
11	Ford	Fairmont A	2965	6	15.8	200	85
12	Ford	Fairmont M	2720	4	15.4	140	88
13	Plymouth	Volare	3430	6	17.2	225	100
14	AMC	Concord	3210	6	17.2	232	90
15	Buick	Century	3380	6	15.8	231	105
16	Mercury	Zephyr	3070	6	16.7	200	85
17	Dodge	Aspen	3620	6	18.7	225	110
18	AMC	Concord D1	3410	6	15.1	258	120
19	Chevrolet	MonteCarlo	3425	8	13.2	305	145
20	Buick	Regal Turbo	3445	6	13.4	231	165
21	Ford	Futura	3205	8	11.2	302	139
22	Dodge	Magnum XE	4080	8	13.7	318	140
23	Chevrolet	Chevette	2155	4	16.5	98	68

Рисунок 2 - Таблица с данными

Рассмотрим метод главных компонент. Выберем *Special > Multivariate Methods > Principal Components*. Появляется окно диалога для задания анализируемых переменных, выберем переменные, показанные на рисунке 3.

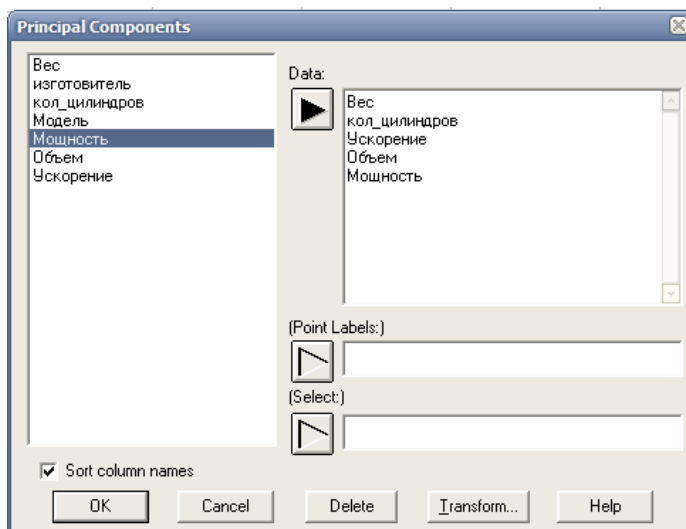


Рисунок 3 – Окно задания переменных для анализа

Получаем исходную сводку анализа метода ГК (рисунок 4), из которой заключаем, что анализу подвергаются переменные *weight (вес)*, *cylinders (кол_цилиндров)*, *accel (ускорение)*, *displace (объем)* и *horspower (мощность)* и число объектов составляет 151.

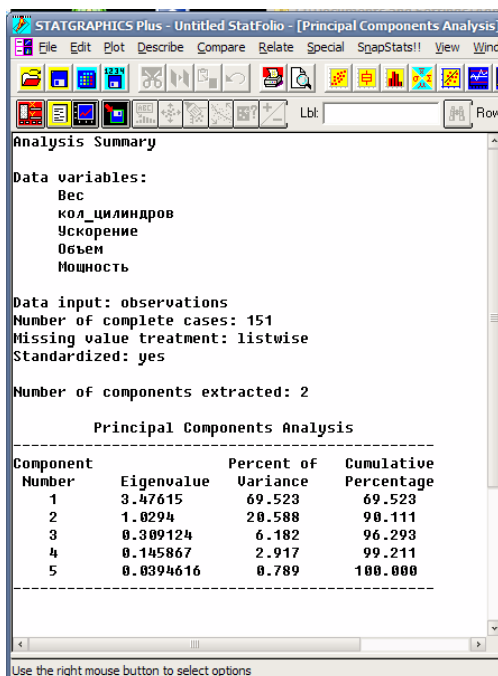


Рисунок 4 – Исходная сводка метода главных компонент

Далее следует информация непосредственно метода ГК: собственные значения ГК, упорядоченные по величине (*Eigenvalue*); процент дисперсии, приходящийся на каждую выделенную ГК (*Percent of Variance*); накопленный

процент дисперсии (*Cumulative Percentage*). Приведенные цифры говорят о том, что уже первые две главные компоненты описывают 90,11 % дисперсии исходных данных.

Нажав на правую клавишу мыши, выберем пункт *Analysis Options* и установим количество компонент, равное трем (рисунок 5).

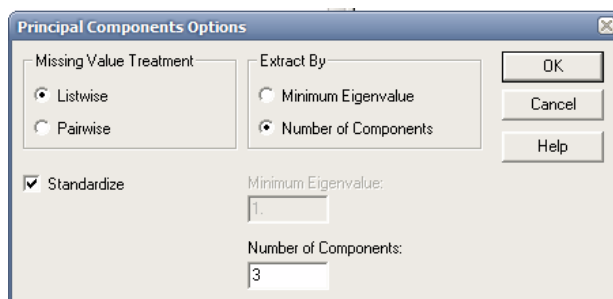


Рисунок 5 – Установка числа компонент

Для более детального анализа нажмем кнопку табличных опций (вторая слева в верхнем ряду) и в соответствующем окне диалога (рисунок 6) установим флажок компонентных весов (*Component Weights*), после чего получим следующую таблицу (рисунок 7).

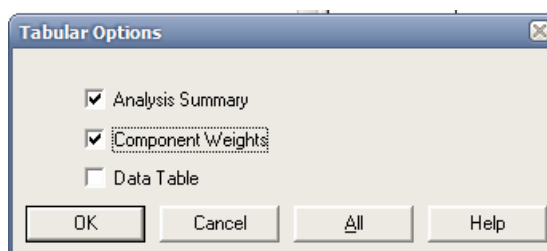


Рисунок 6 – Установка табличных опции

Как следует из полученных цифр, в первой ГК примерно одинаковые по величине положительные коэффициенты имеют вес, количество цилиндров, объем двигателя и мощность. Вместе с тем, во второй ГК превалирует только одна величина: ускорение. А в третьей ГК наблюдается сочетание веса машины и ее мощности (с положительным знаком), которому противопоставляется количество цилиндров (с отрицательным знаком).

Перейдем к рассмотрению диаграммы рассеивания всей совокупности автомашин в пространстве выделенных трех первых ГК. Для этого щелкнем левой кнопкой мыши на кнопке графических опций и инициализируем данное трехмерное отображение (рисунок 8).

Analysis Summary

Data variables:
 Вес
 кол_цилиндров
 Ускорение
 Объем
 Мощность

Data input: observations
 Number of complete cases: 151
 Missing value treatment: listwise
 Standardized: yes

Table of Component Weights

	Component 1	Component 2	Component 3
Вес	0.492492	0.272583	0.0518503
кол_цилиндров	0.501315	0.0928374	-0.421138
Ускорение	-0.161338	0.926091	0.251486
Объем	0.517691	0.125393	-0.292074
Мощность	0.460544	-0.209072	0.819391

Рисунок 7 – Веса признаков в главных компонентах

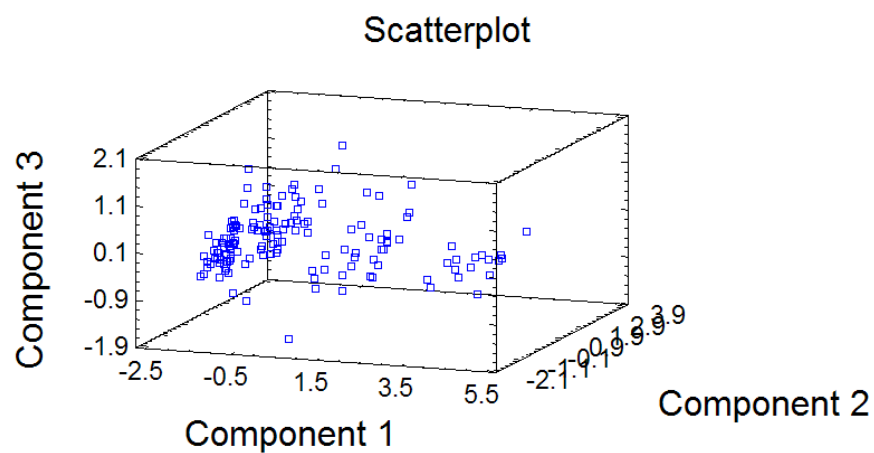


Рисунок 8 – Проекция исследуемых автомобилей

На представленном рисунке хорошо видно, что вся исследуемая совокупность автомашин разделилась на три достаточно четко выраженные группы. Для первой, наиболее многочисленной группы характерны сравнительно небольшие вес, количество цилиндров, мощность и объем двигателя (первая группа слева). Вместе с тем, большая доля автомашин этой группы обладают хорошим ускорением (высокие значения 2-й ГК) и высоким соотношением веса и мощности к количеству цилиндров (3-я ГК).

Рассмотрим применение кластерного анализа.

Выберем Special > Multivariate Methods > Cluster Analysis. Появляется окно диалога для ввода данных в кластерный анализ (рисунок 9).

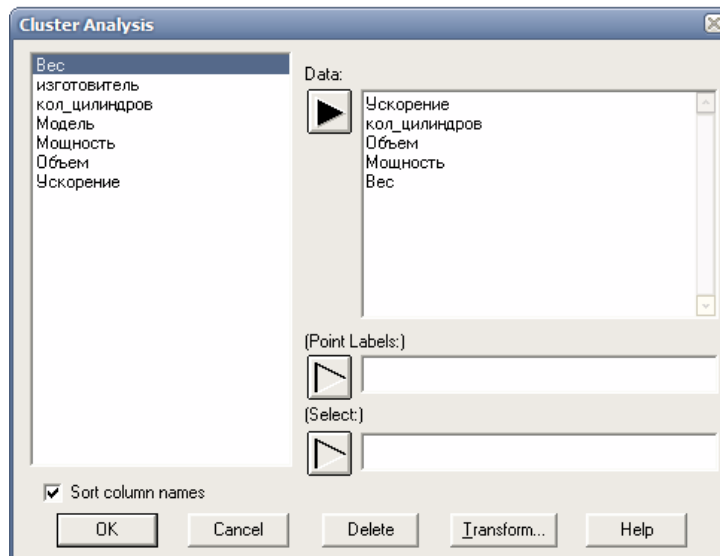


Рисунок 9 - Окно диалога ввода данных для кластерного анализа

Щелчком правой кнопкой мыши — на экране появляется окно диалога для выбора параметров кластерного анализа. Установим флажок Wards, чтобы выделение кластеров происходило по методу Варда (рисунок 10).

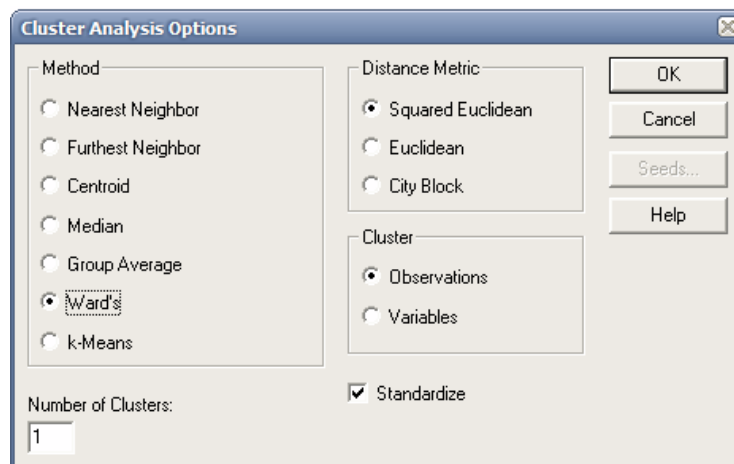


Рисунок 10 – Окно диалога для выбора параметров кластерного анализа

Выберем отображение в виде дендрограммы (Dendrogram) (рисунок 11).

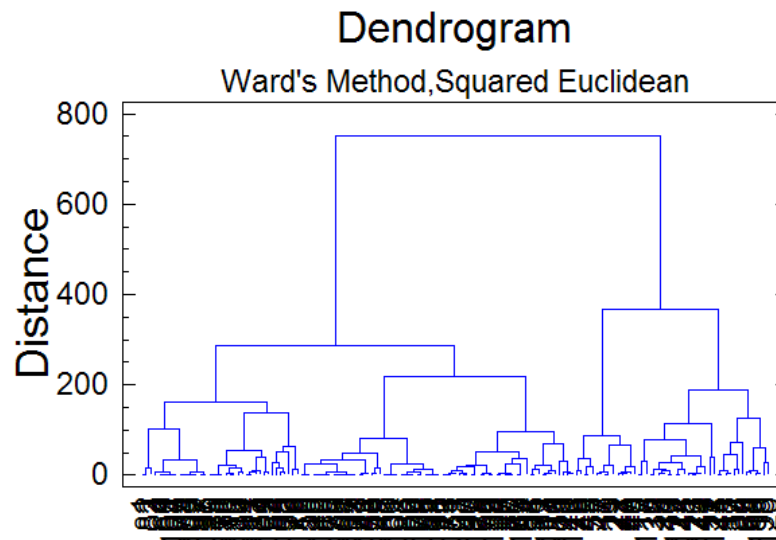


Рисунок 11 - Дендрограмма, полученная методом Варда для одного кластера

Дендрограмма отображает иерархическую структуру группирования объектов. На ней отчетливо видны как минимум три группы – отсюда следует, что для более подробного рассмотрения группировок следует задать их количество равным 3 (рисунок 12). Тогда дендограмма примет вид, изображенный на рисунке 13.

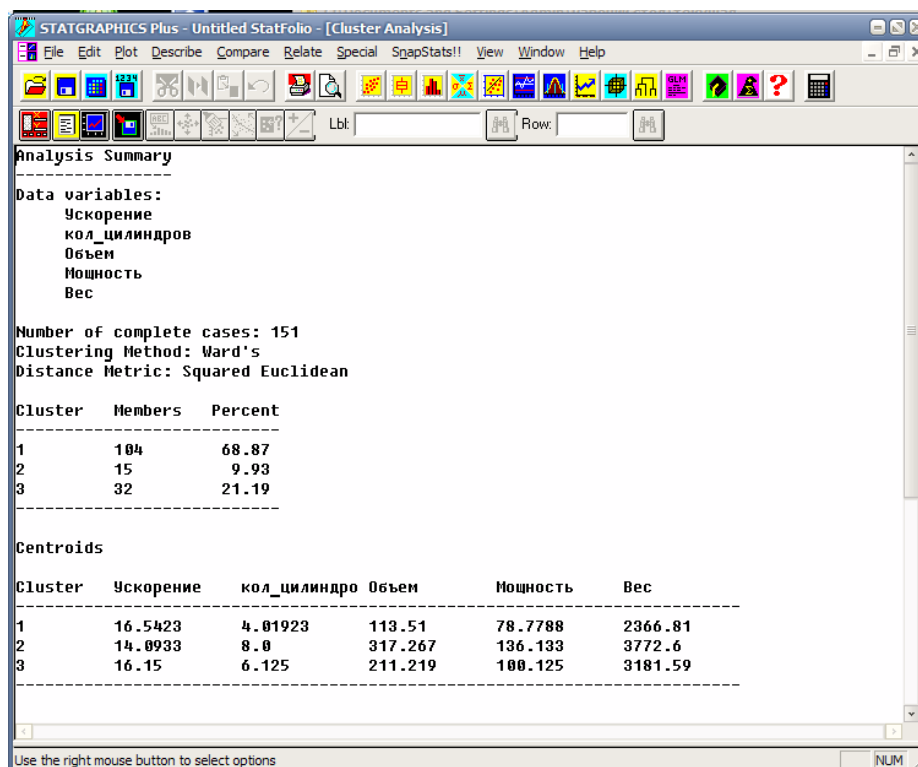


Рисунок 12 – Сводка кластерного анализа

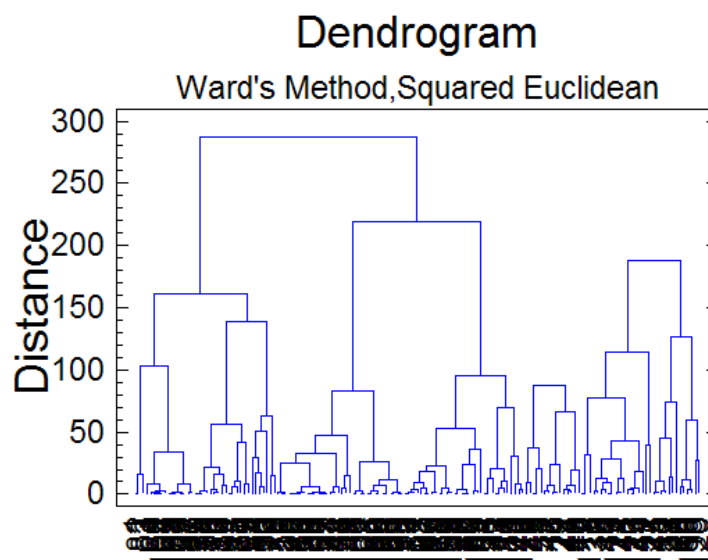


Рисунок 13 - Дендрограмма, полученная методом Варда для трех кластеров

Нажмем кнопку табличных опций. Установим Membership Table (таблица принадлежности наблюдений). В данной таблице описаны выбранные параметры кластерного анализа и дается полный список всех наблюдений, их имена и номера кластеров, в которые входят указанные наблюдения (рисунок 14).

Membership Table		

Clustering Method: Ward's		
Distance Metric: Squared Euclidean		
Row	Label	Cluster

1	Volkswagen	1
2	Ford	1
3	Mazda	1
4	Datsun	1
5	Honda	1
6	Oldsmobile	2
7	Dodge	2
8	Mercury	2
9	Pontiac	3
10	Chevrolet	3
11	Ford	3
12	Ford	1
13	Plymouth	3
14	AMC	3
15	Buick	3
16	Mercury	3
17	Dodge	3
18	AMC	3
19	Chevrolet	2
20	Buick	3
21	Ford	2
22	Dodge	2
23	Chevrolet	1
24	Toyota	1
25	Datsun	1
26	Dodge	1
27	Toyota	1
28	Plymouth	1
29	Oldsmobile	1
30	Datsun	1
31	Audi	1
32	Volvo	3
33	Saab	1
34	Peugeot	3
35	Volkswagen	1

Рисунок 14 – Таблица принадлежности наблюдений кластерам

Создадим двухмерную диаграмму рассеивания (рисунок 15), выбрав по оси X значения веса, по оси Y – мощности.

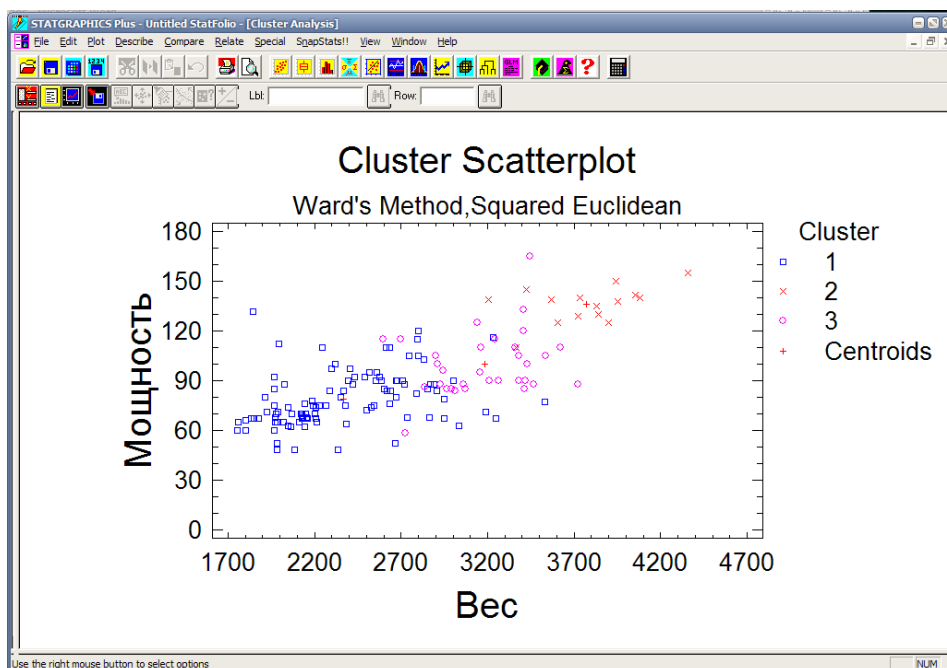


Рисунок 15 – Двухмерная диаграмма рассеивания

3.2 Рассмотрим пример, относящийся к деятельности предприятия. На рисунке 16 показана таблица, содержащая основные показатели:

Y1 - производительность труда;

X5 - удельный вес рабочих в составе промышленно-производственного персонала;

X7 - коэффициент сменности оборудования (смен);

X9 - удельный вес потерь от брака (%);

X10 - фондоотдача активной части основных производственных фондов.

	Y1	X5	X7	X9	X10	C
1	9.26	0.78	1.37	0.23	1.45	
2	12.11	0.68	1.44	0.43	1.37	
3	10.81	0.7	1.42	0.18	1.65	
4	9.35	0.62	1.35	0.15	1.91	
5	9.87	0.76	1.39	0.34	1.68	
6	9.12	0.71	1.27	0.09	1.89	
7	9.37	0.79	1.4	0.21	2.3	
8	10.02	0.76	1.22	0.32	2.62	
9	9.42	0.7	1.2	0.28	2.03	
10	6.61	0.72	1.23	0.48	0.88	
11	4.32	0.68	1.39	0.41	0.62	
12	7.37	0.77	1.38	0.62	1.09	
13	6.64	0.77	1.35	0.5	1.32	
14	5.52	0.72	1.24	1.2	0.68	
15	5.68	0.71	1.28	0.66	1.43	
16	5.22	0.79	1.33	0.74	1.82	
17	6.7	0.79	1.35	0.39	1.24	
18						
19						

Рисунок 16 – Основные показатели деятельности предприятия

Проведем компонентный анализ. Выберем Special > Multivariate Methods > Principal Components. В результате (рисунок 17) получаем исходную сводку

анализа метода, из которой заключаем, что анализу подвергаются переменные Y1(производительность труда, X7 (коэффициент сменности оборудования) (смен), X9 (удельный вес потерь от брака) (%);и что число объектов составляет 17.

Principal Components Analysis

Analysis Summary

Data variables:
X10
X5
X7
X9
Y1

Data input: observations
Number of complete cases: 17
Missing value treatment: listwise
Standardized: yes

Number of components extracted: 3

Principal Components Analysis

Component Number	Eigenvalue	Percent of Variance	Cumulative Percentage
1	2.23567	44.713	44.713
2	1.11917	22.383	67.097
3	1.05293	21.059	88.155
4	0.373657	7.473	95.628
5	0.218579	4.372	100.000

Рисунок 17 – Сводка анализа

С помощью пункта *Analysis Options* установим количество компонент, равное трем. В табличных опциях установим флажок компонентных весов (*Component Weights*), после чего получим следующую таблицу (рисунок 18).

Principal Components Analysis

Table of Component Weights

	Component 1	Component 2	Component 3
X10	-0.514545	0.523025	0.008468
X5	0.0983432	0.486613	0.810687
X7	-0.182441	-0.690191	0.584215
X9	0.582919	0.0892258	0.0263245
Y1	-0.593711	-0.0729885	-0.0267324

Рисунок 18 – Веса признаков в главных компонентах

Как следует из полученных цифр, в первой ГК примерно одинаковые по величине положительные коэффициенты имеют производительность труда (Y1), удельный вес потерь от брака (X9). Вместе с тем, во второй ГК превалирует только одна величина: коэффициент сменности оборудования (X7). А в третьей ГК – удельный вес рабочих в составе персонала (X5).

Перейдем к рассмотрению диаграммы рассеивания. Для этого щелкнем левой кнопкой мыши на кнопке графических опций и инициализируем данное трехмерное отображение (рисунок 19).

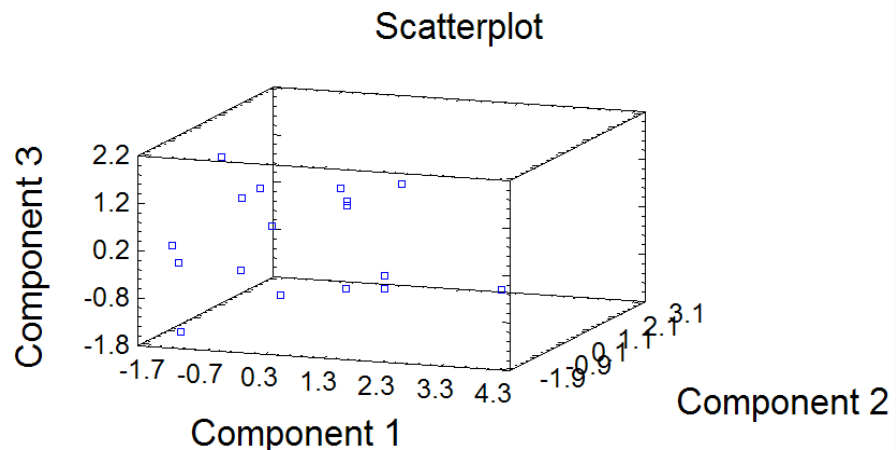


Рисунок 19 – Проекция исследуемых предприятий в пространство трех ГК

Рассмотрим применение кластерного анализа.

Выберем Special > Multivariate Methods > Cluster Analysis. В окне диалога для выбора параметров кластерного анализа установим флажок Wards, чтобы выделение кластеров происходило по методу Варда. Выберем отображение в виде дендрограммы (Dendrogram) (рисунок 20). Затем зададим количество кластеров равным трем, тогда дендрограмма будет иметь вид, показанный на рисунке 21.

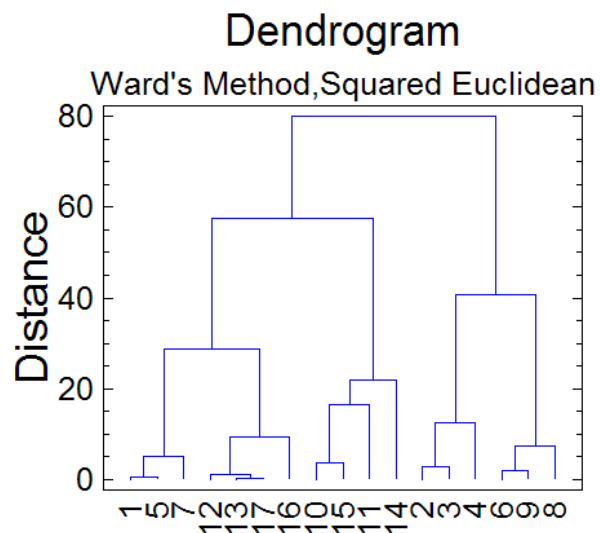


Рисунок 20 - Дендрограмма, полученная методом Варда для одного кластера

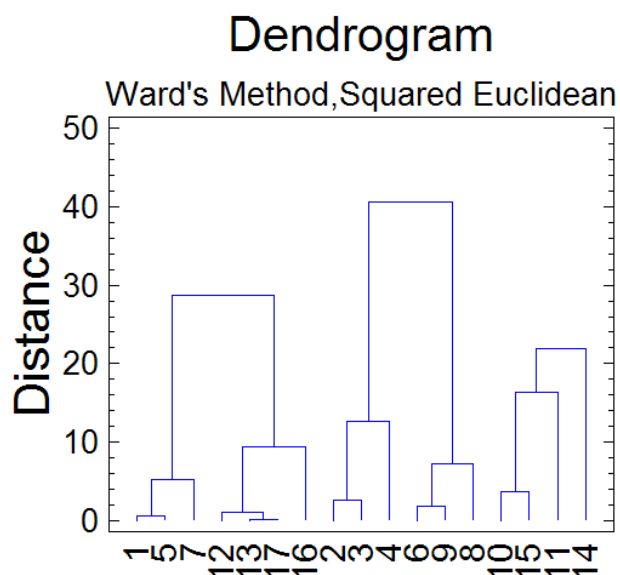


Рисунок 21 - Дендрограмма, полученная методом Варда для трех кластеров

Нажмем кнопку табличных опций. Установим Membership Table (таблица принадлежности наблюдений). В данной таблице описаны выбранные параметры кластерного анализа и дается полный список всех наблюдений, их имена и номера кластеров, в которые входят указанные наблюдения (рисунок 22).

```

Membership Table
-----
Clustering Method: Ward's
Distance Metric: Squared Euclidean

Row      Cluster
-----
1         1
2         2
3         2
4         2
5         1
6         2
7         1
8         2
9         2
10        3
11        3
12        1
13        1
14        3
15        3
16        1
17        1
-----

```

Рисунок 22 – Таблица принадлежности наблюдений кластерам

Создадим двухмерную диаграмму рассеивания (рисунок 23), выбрав по оси X значения коэффициента сменности оборудования, по оси Y – производительность труда.

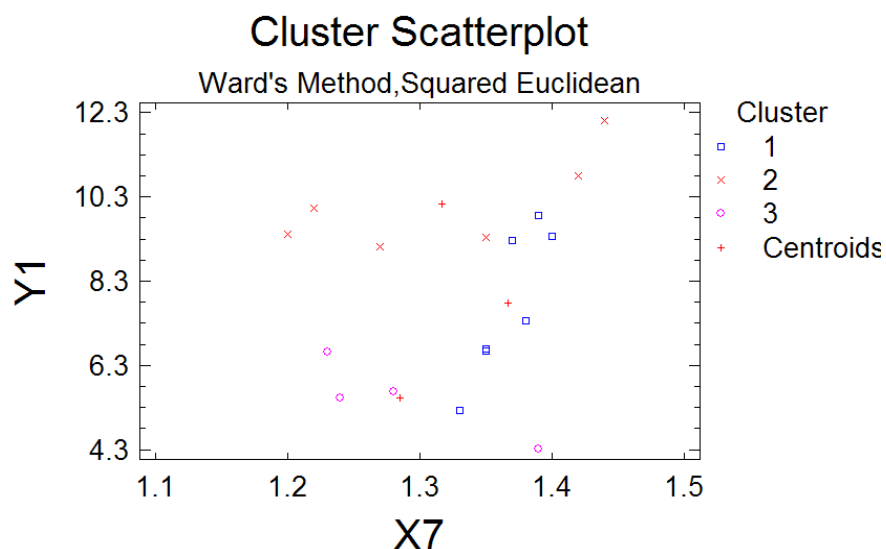


Рисунок 23 – Двухмерная диаграмма рассеивания

Таким образом, в работе были изучены два метода анализа данных – компонентный и кластерный. Рассмотрены две области – рынок автомобилей и деятельность предприятия. Выделены главные компоненты, выявлены наиболее характерные для них признаки, построены диаграммы рассеивания показателей, дендограммы методом Варда.

Контрольные вопросы:

1. Компонентный анализ, достоинства и недостатки
2. Кластерный анализ, достоинства и недостатки
3. Какие неформальные требования должен отражать критерий качества кластеризации
4. На какие группы делятся методы кластерного анализа
5. Какие основные проблемы при проведении кластерного анализа

лабораторная работа №2

Решение задачи кластеризации с помощью нейросетевого моделирования

Цель работы: изучение алгоритма решения задачи кластеризации с помощью нейросетевого моделирования и исследование процесса применения нейросетевого пакета SOMap Analyzer 1.0 для решения этой задачи.

Задание:

1. Сформировать обучающую выборку размером 30 объектов.
2. Произвести обучение сети Кохонена.
3. Построить карты по выделенным признакам.
4. Провести нейросетевой анализ построенных карт. Дать характеристику построенным кластерам.

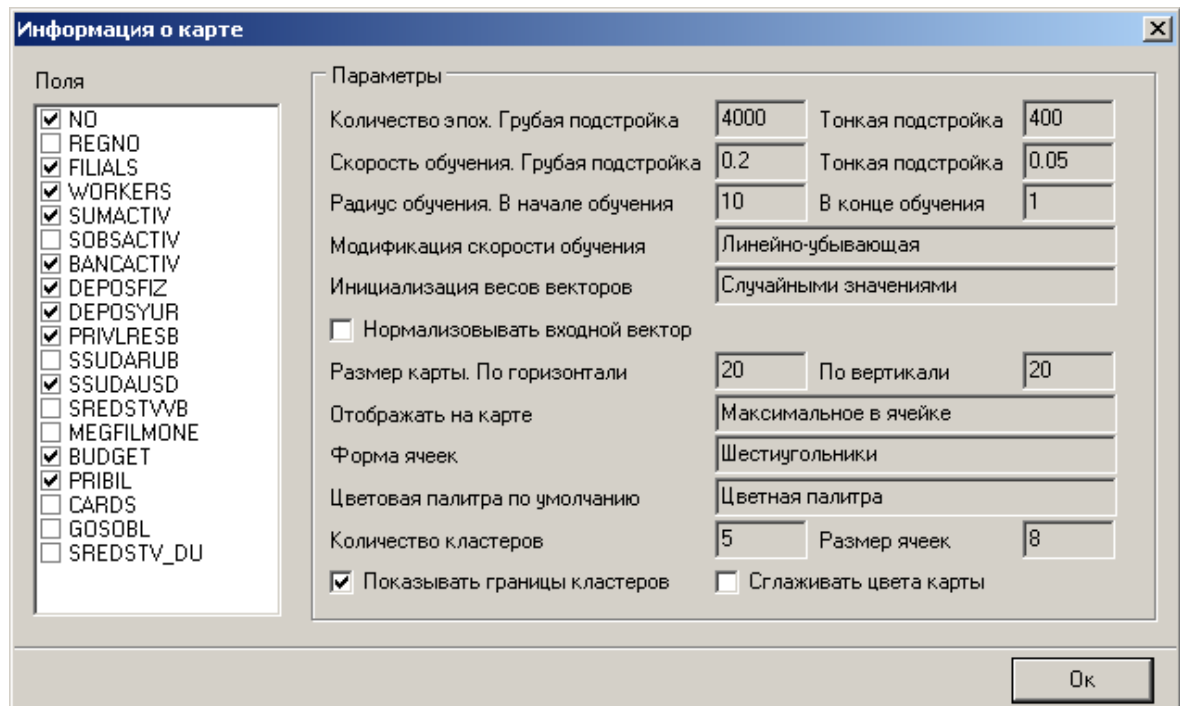
Обучающая выборка:

NO	BANC	REGNO	REUTERS	FILIALS	CITY	WORKERS	SUMACTIV	SOBSACTIV	BANCACTIV	DEPOS
2	Внешторгбанк	1000	-	32	Москва	3297	101660298	23236327	84343558	2086
3	Газпромбанк	354	GZPM	27	Москва	2559	79012789	9255041	74409960	3946
4	ООО "Международный Промышленный банк"	2056	TIBP	4	Москва	459	77888642	26409116	58647197	157
5	Международный Московский Банк	1	IMBX	1	Москва	621	63910966	1176462	62436148	1291
6	ОАО "АЛЬФА-БАНК"	1326	ALFM	17	Москва	2323	57510886	12446938	52348562	4052
7	ОАО "ПСБ"	439	ICSP	44	Санкт-Петербург	3579	49406525	1275859	17091603	1531
8	Банк Москвы	2748	-	34	Москва	2810	31352124	3335734	30287158	2496
9	АКБ "РОСБАНК" (ОАО)	2272	-	13	Москва	989	28105202	4691449	25807591	2583
10	АКБ "ДИБ"	2783	DIBM	0	Москва	377	27350369	2616993	26986210	1231
11	КБ "Ситибанк Т/О"	2557	-	1	Москва	210	26240408	2063168	23291673	327
12	ОАО МАКБ "Возрождение"	1439	VOZM	62	Москва	3439	25446423	1520076	9879932	966
13	ОАО РИКБ "Башкредитбанк"	2275	BKRE	4	Уфа	1838	23304860	2833897	21190313	713
14	ЗАО Банк "МЕНАТЕП СПб"	3279	MTSP	48	Санкт-Петербург	2145	19541965	2268343	9479194	1210
15	БНП-Дрезднер Банк	2455	NDMO	1	Санкт-Петербург	185	17689833	84312	13082069	232
16	АВТОБАНК	30	-	25	Москва	2732	17229288	1530584	15502811	1673
17	ОАО Банк "Петрокоммерц"	1776	PCBM	13	Москва	1006	16690301	727690	10372701	1092
18	АКБ "НРБанк"	2170	NRB/B	0	Москва	320	15791359	5075844	15773889	226
19	АКБ "Еврофинанс"	2402	EFIN	1	Москва	298	15191022	2835944	13398039	1496
20	ООО Райффайзенбанк Австрия	3292	RZBM	0	Москва	212	14892329	593460	11685822	865
21	КБ "ГУТА-БАНК"	1623	GUTA	33	Москва	1906	14799899	2008471	12398573	567
22	ГЛОБЭКСБАНК	1942	GLBX	8	Москва	133	12769906	5016721	12576820	192
23	ОАО "Ханты-Мансийский банк "	1971	-	6	Ханты-Мансийск	473	12722589	323652	2987333	214
24	ОАО "Банк "Петровский "	729	PETR	18	Санкт-Петербург	3395	12589085	439907	5943321	1186
25	АБМ АМРО Банк АО	2594	AABM	1	Москва	175	10566709	1079184	10007563	297
26	Банк ЗЕНИТ	3255	ZENT	3	Москва	456	10407230	1000661	8303786	623
27	НОМОС-БАНК	2209	NMOS	1	Москва	260	8916553	1416318	8799375	161
28	АКБ "Московский Индустриальный банк"	912	MIND	36	Москва	2845	8467224	938868	6145063	456
29	АКБ "Ак-Барс"	2590	AKBA	20	Казань	1667	8401583	2515240	7218500	425
30	ИМПЭКСБАНК (ООО)	2291	IMEX	41	Москва	1731	7963291	1015977	3631085	494
31	АКБ "ЧЕЛИНДБАНК"	485	-	28	Челябинск	1392	7767092	463445	3934709	520
32	ЗАО "Конверсбанк"	122	CONV	10	Москва	957	7618503	1716464	6341373	806
33	ОАО "Банк "Санкт-Петербург"	436	BS PB	7	Санкт-Петербург	1503	6730121	255345	2866942	173

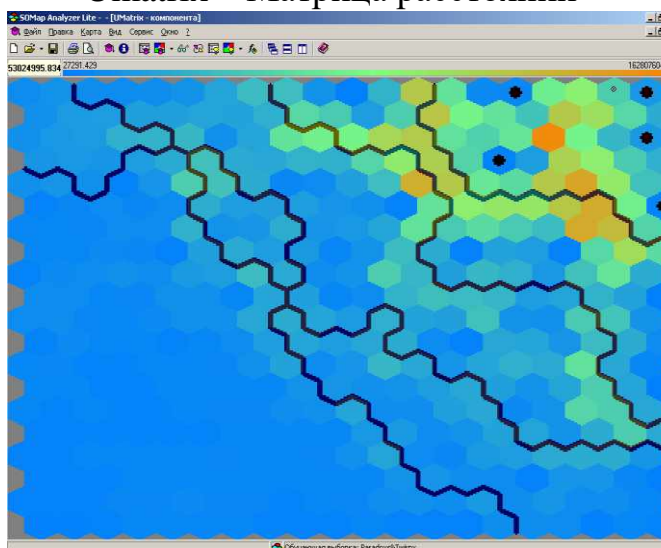
2. Процесс обучения сети состоит из 5 шагов: выбор источника данных; настройка полей; установка параметров обучения; установка параметров визуализации; обучение (построение) карты.

В качестве источника данных используем созданную обучающую выборку. Параметры обучения, параметры визуализации и процесс обучения карты представлены на рисунках.

3. Для построения карты открываем окно с картами по обучающей выборке и выбираем нужные компоненты. Результат построения представлен на рисунке
4. Обучение сети по параметрам:

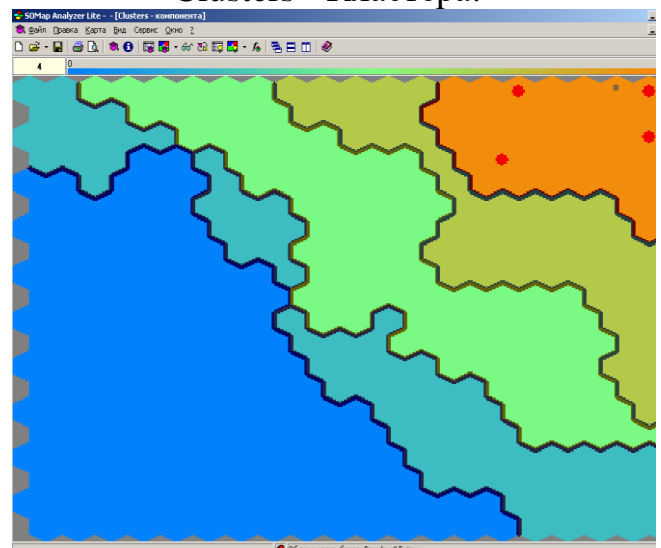


Umatrix – Матрица расстояний

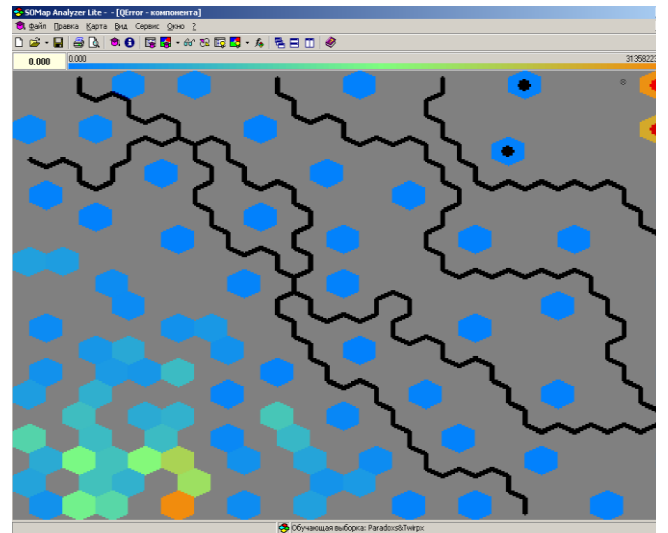
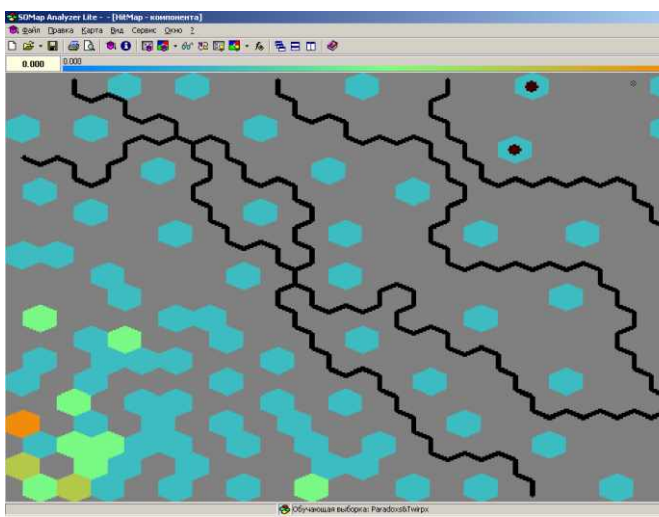


HitMap – количество элементов ячейке

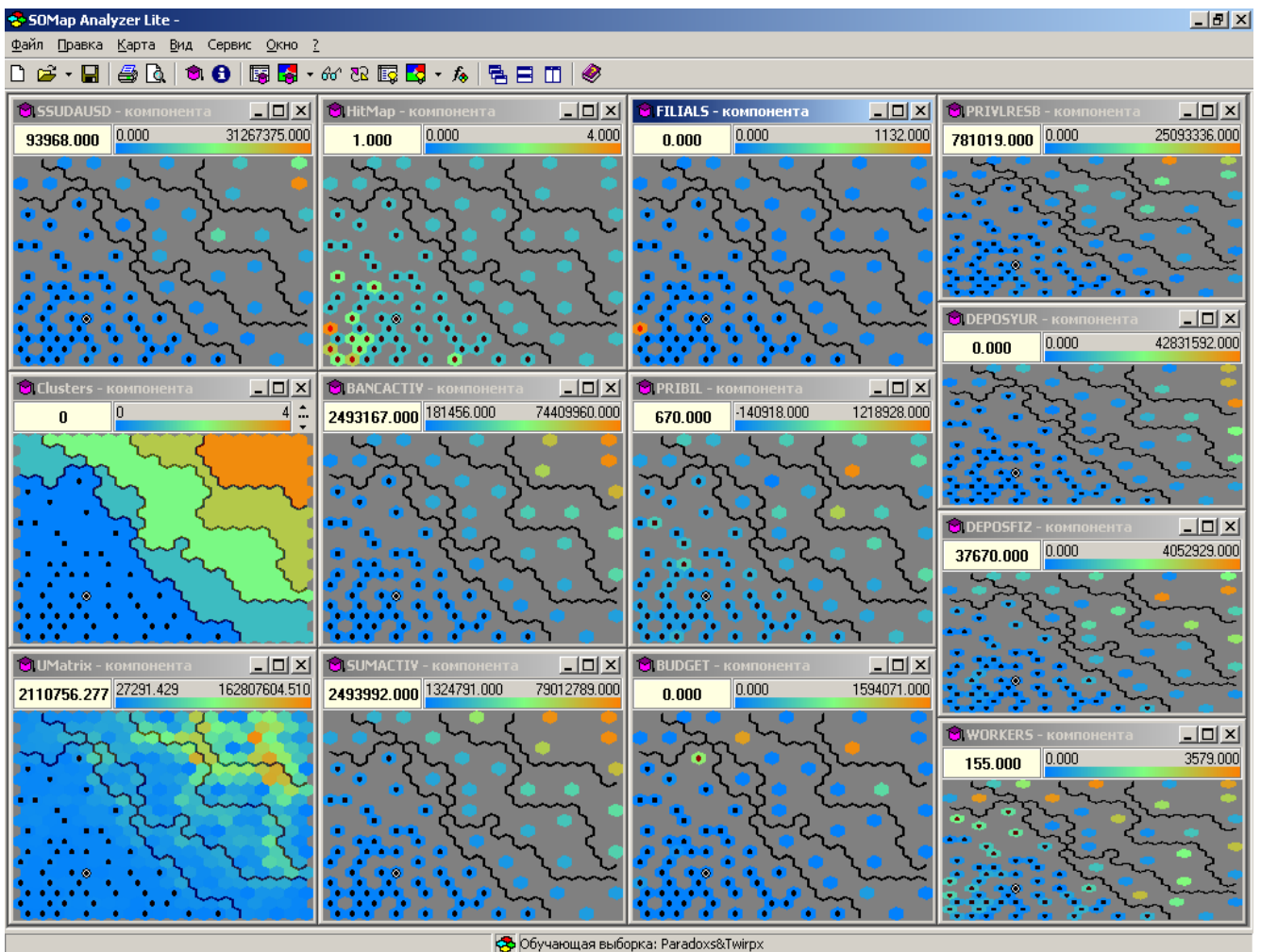
Clusters - Кластера:



QError – Ошибка квантирования



Кластеризация по обучающей выборке:



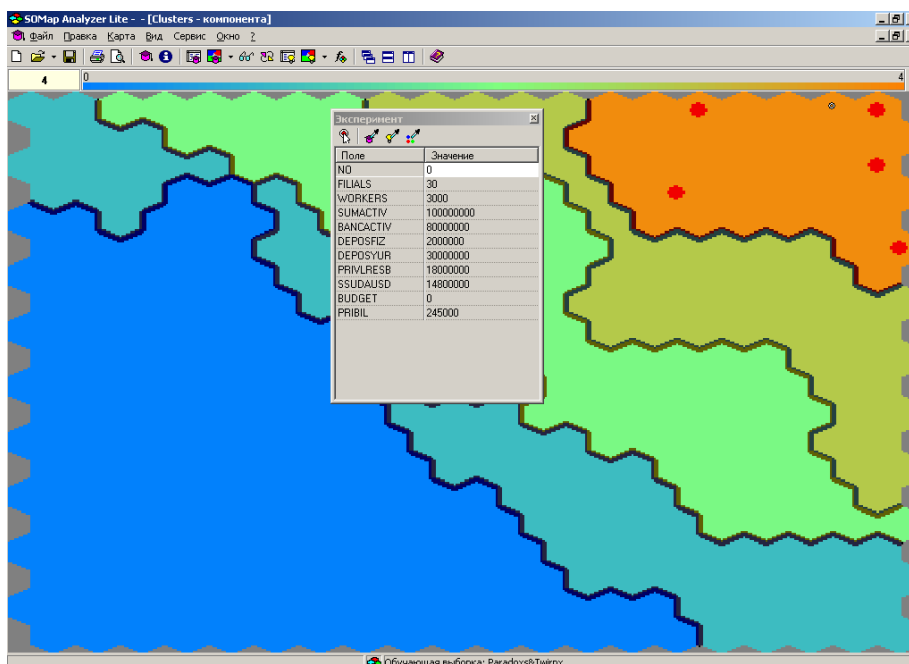
	DEPOSFIZ	WORKERS	PRIBIL	BUTGET	BANCAKTIV	DEPOSUR	SUMACTIV	FILIALS	SSUDAUSD	PRIVLRESB
1	Н	С	Н	Н	Н	Н	Н	Н	Н	Н
2	Н	С	Н	Н	Н	Н	Н	Н	Н	Н
3	С	С	С	С	С	Н	С	С	Н	Н
4	С	С	С	Н	С	С	С	С	С	С
5	В	В	В	С	В	В	В	С	В	В

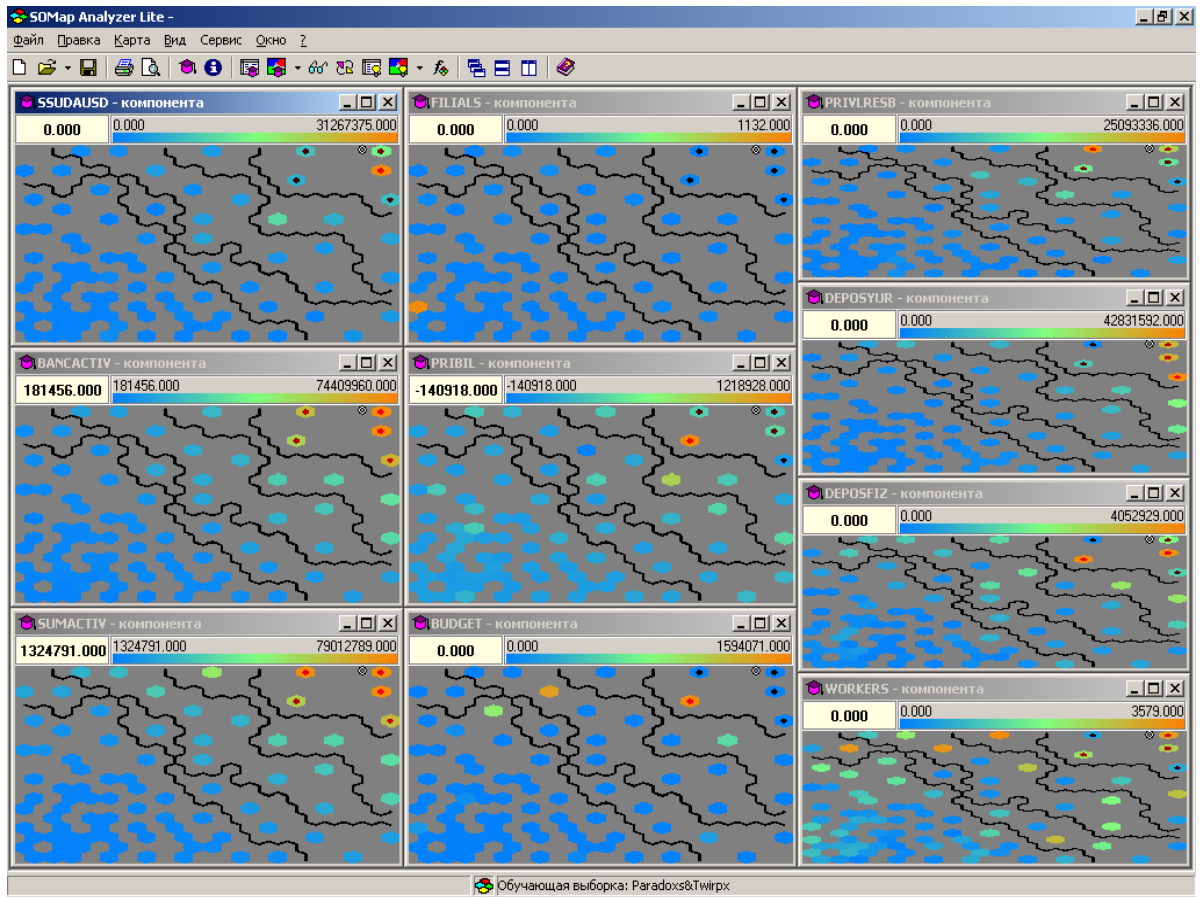
В – высокий показатель

С – средний показатель

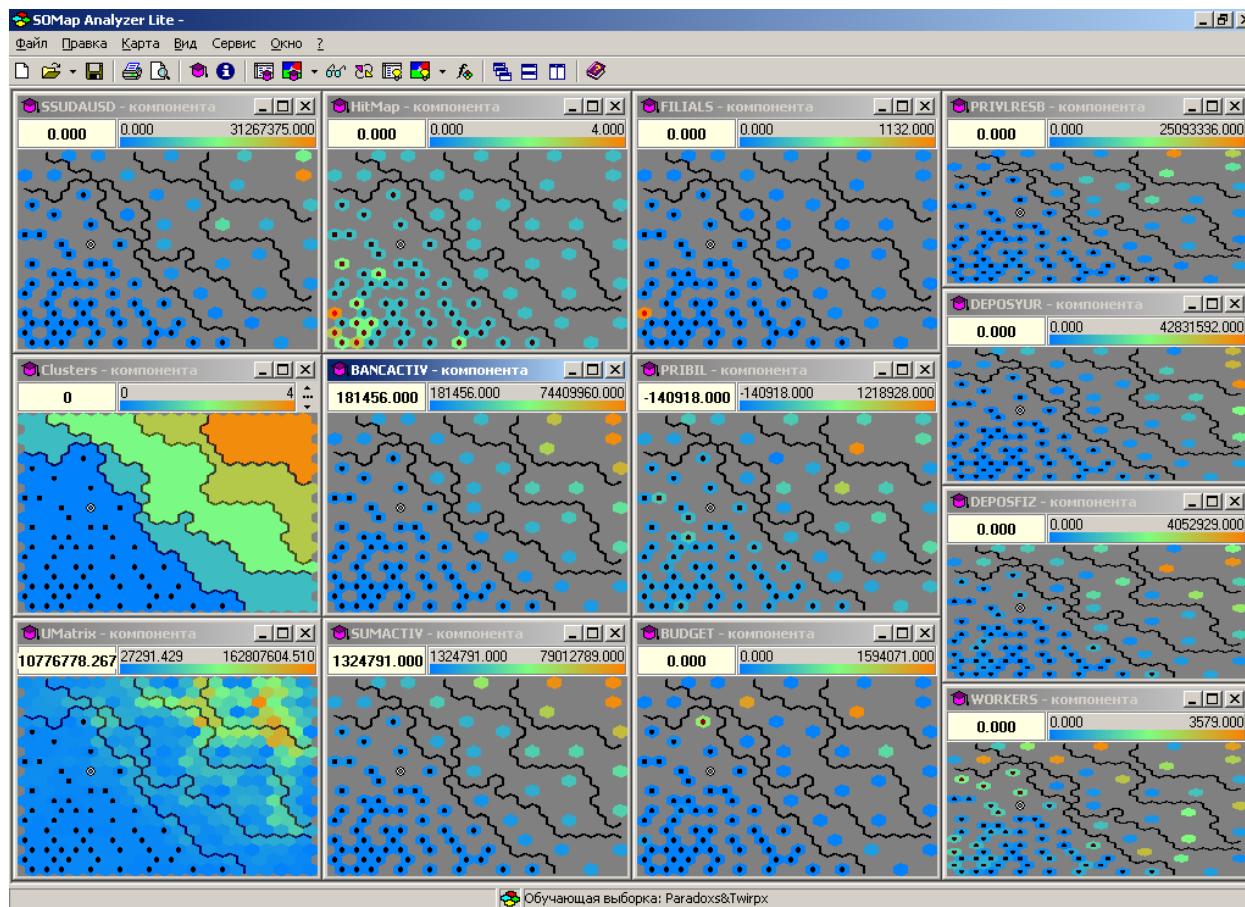
Н – низкий показатель

Эксперимент №1 по рабочей выборке:





Эксперимент №2 по рабочей выборке:



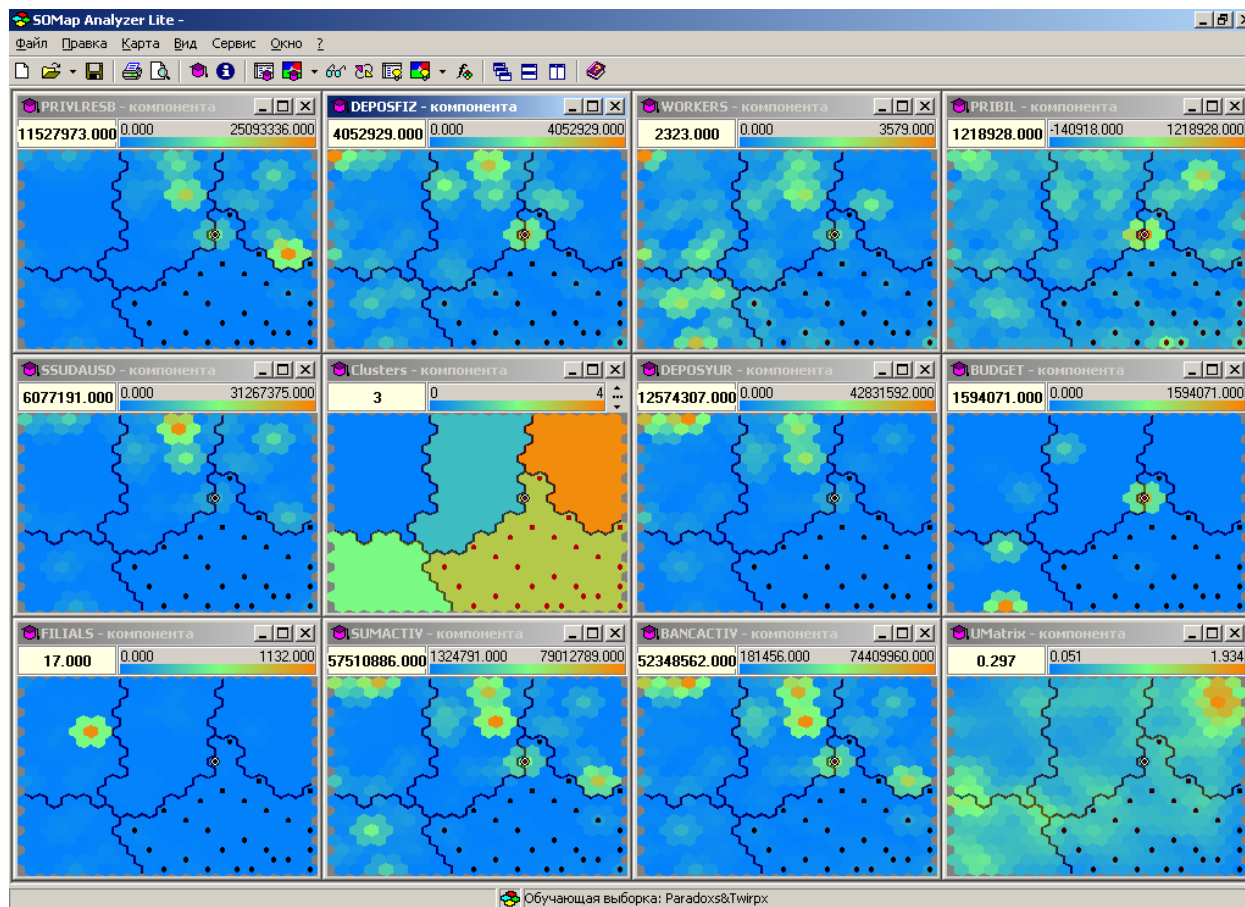
Обучение сети по новым параметрам:

Информация о карте

Поля	Параметры
<input type="checkbox"/> NO	Количество эпох. Грубая подстройка: 500
<input type="checkbox"/> REGNO	Тонкая подстройка: 50
<input checked="" type="checkbox"/> FILIALS	Скорость обучения. Грубая подстройка: 0.2
<input checked="" type="checkbox"/> WORKERS	Тонкая подстройка: 0.05
<input checked="" type="checkbox"/> SUMACTIV	Радиус обучения. В начале обучения: 10
<input type="checkbox"/> SOBSACTIV	В конце обучения: 1
<input checked="" type="checkbox"/> BANCACTIV	Модификация скорости обучения: Линейно-убывающая
<input checked="" type="checkbox"/> DEPOSFIZ	Инициализация весов векторов: Примерами из выборки
<input checked="" type="checkbox"/> DEPOSYUR	<input checked="" type="checkbox"/> Нормализовать входной вектор
<input checked="" type="checkbox"/> PRIVLRESB	Размер карты. По горизонтали: 20
<input type="checkbox"/> SSUDARUB	По вертикали: 20
<input checked="" type="checkbox"/> SSUDAUSD	Отображать на карте: Максимальное в ячейке
<input type="checkbox"/> SREDSTVVB	Форма ячеек: Шестиугольники
<input type="checkbox"/> MEGFILMONE	Цветовая палитра по умолчанию: Цветная палитра
<input checked="" type="checkbox"/> BUDGET	Количество кластеров: 5
<input checked="" type="checkbox"/> PRIBIL	Размер ячеек: 6
<input type="checkbox"/> CARDS	<input checked="" type="checkbox"/> Показывать границы кластеров
<input type="checkbox"/> GOSOBL	<input type="checkbox"/> Сглаживать цвета карты
<input type="checkbox"/> SREDSTV_DU	

Ок

Результат кластеризации:



	DEPOSFIZ	WORKERS	PRIBIL	BUTGET	BANCAKTIV	DEPOSUR	SUMACTIV	FILIALS	SSUDAUSD	PRIVLRESB
1	В	С	С	Н	В	В	В	С	Н	Н
2	С	В	С	С	Н	Н	С	Н	Н	Н
3	В	С	С	Н	В	С	В	Н	С	С
4	Н	Н	В	С	Н	Н	Н	Н	Н	Н
5	С	Н	В	Н	С	Н	С	Н	Н	В

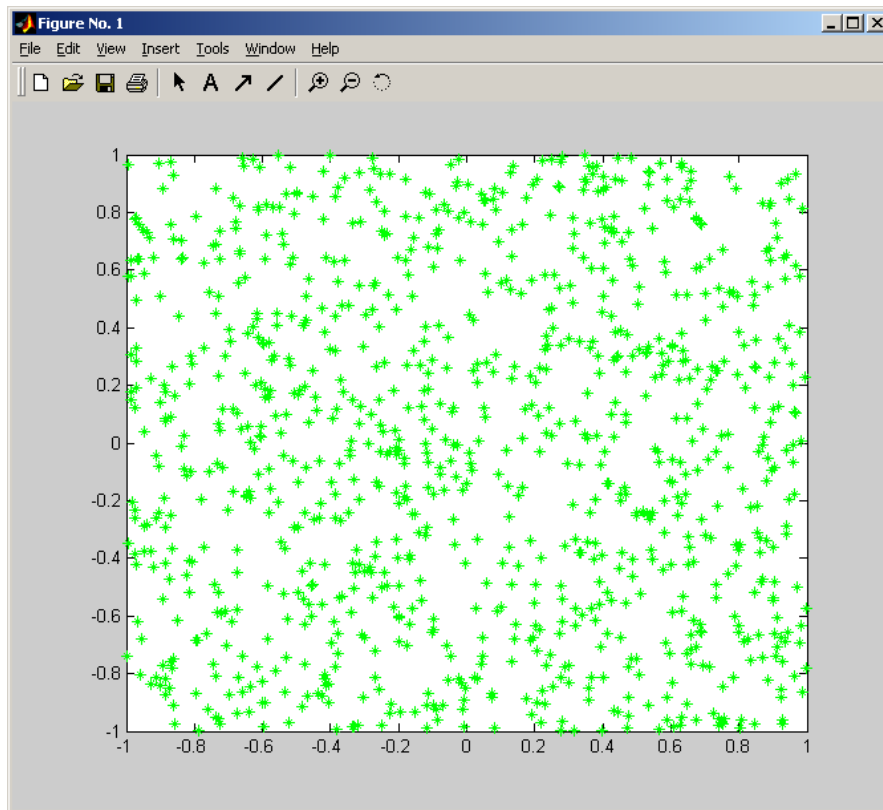
В – высокий показатель

С – средний показатель

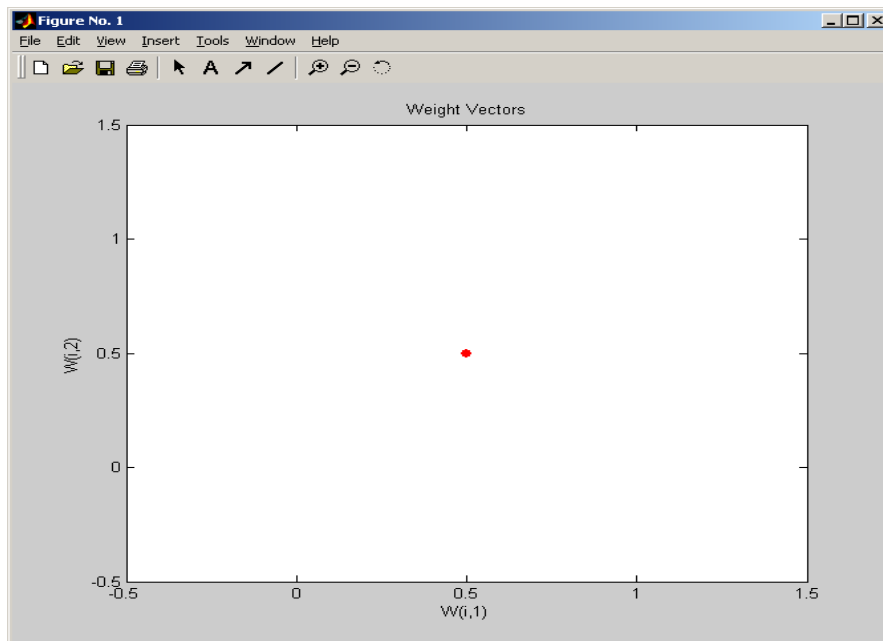
Н – низкий показатель

Сети Кохонена и алгоритм обучения без учителя с использованием MatLab:

```
>> p=rand(2,1000);
>> plot(p(1,:),p(2,:), '*g');
```

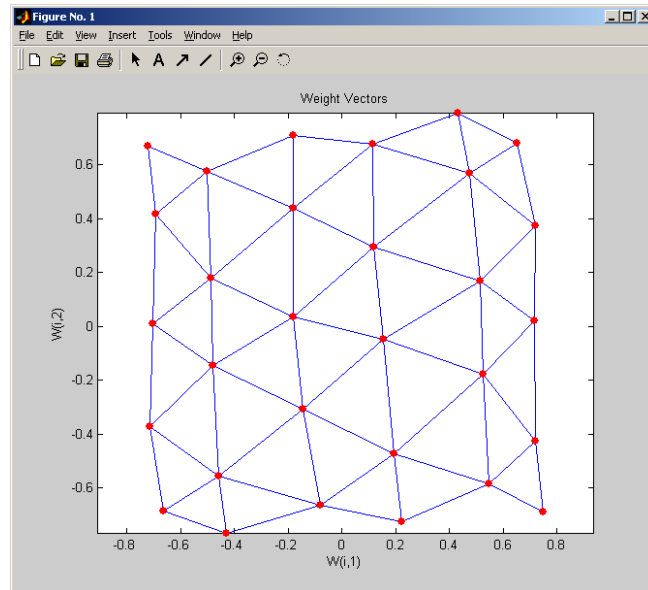


```
>> net=newsom([0 1; 0 1],[5 6]);
>> plotsom(net.iw{1,1},net.layers{1}.distances);
```



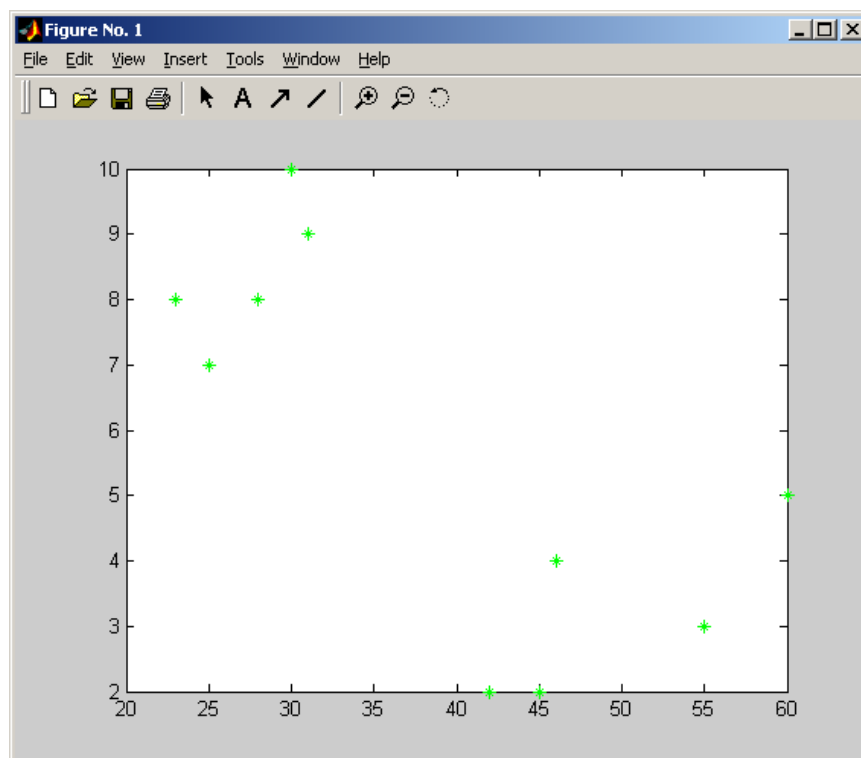
```
>> net.trainParam.epochs = 100;
>> net = train(net, p);
TRAINR, Epoch 0/100
TRAINR, Epoch 25/100
TRAINR, Epoch 50/100
TRAINR, Epoch 75/100
TRAINR, Epoch 100/100
TRAINR, Maximum epoch reached.
```

```
>> plotsom(net.iw{1,1},net.layers{1}.distances);
```

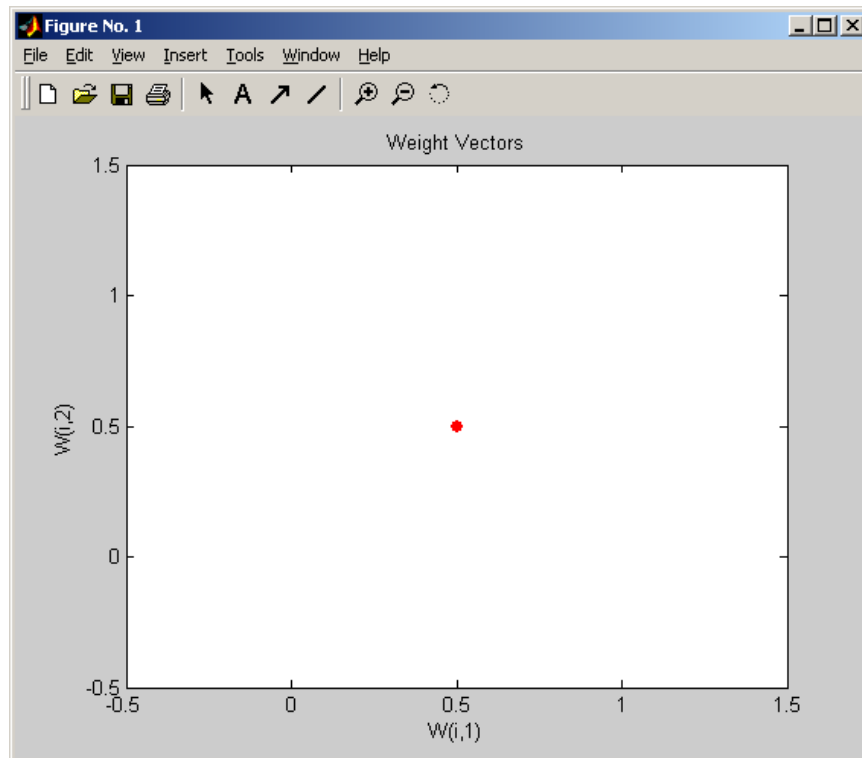


Исходная обучающая выборка:

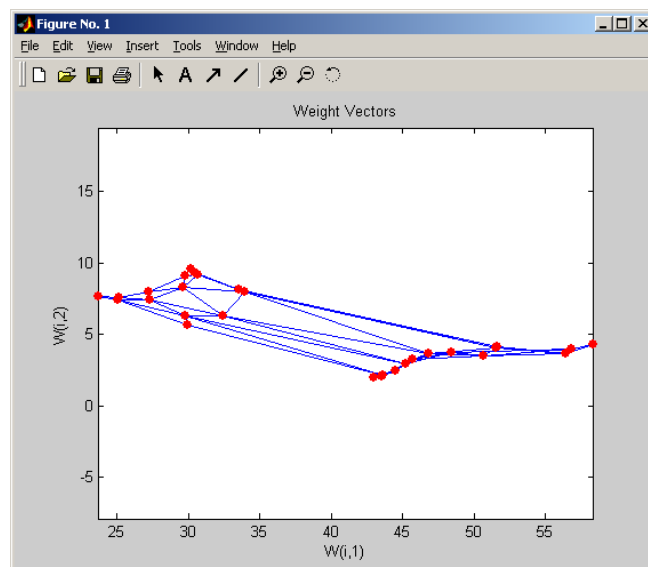
```
>> p = [28 31 25 23 30 60 55 45 46 42; 8 9 7 8 10 5 3 2 4 2];  
>> plot(p(1,:),p(2,:), '*g');
```



```
>> net=newsom([0 1; 0 1],[5 6]);  
>> plotsom(net.iw{1,1},net.layers{1}.distances);
```



```
>> net.trainParam.epochs = 100;  
>> net = train(net, p);  
TRAINR, Epoch 0/100  
TRAINR, Epoch 25/100  
TRAINR, Epoch 50/100  
TRAINR, Epoch 75/100  
TRAINR, Epoch 100/100  
TRAINR, Maximum epoch reached.  
>> plotsom(net.iw{1,1},net.layers{1}.distances);
```



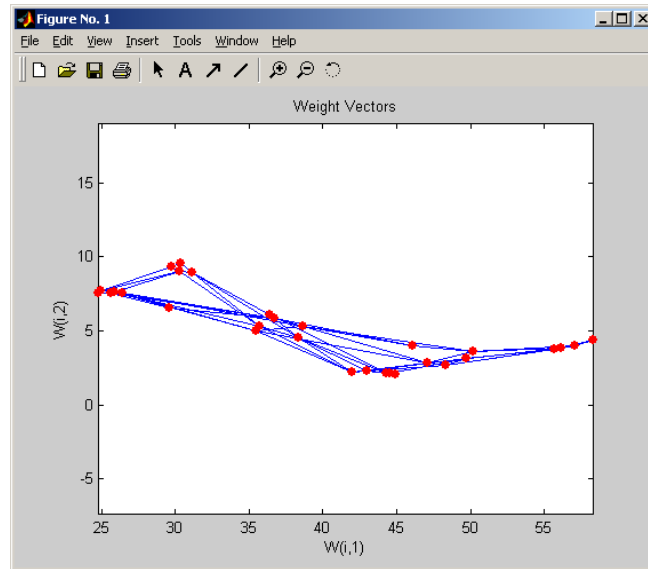
```

>> test_all = [39 40 42; 5 6 2];
>> a = sim(net, test_all);

>> net = train(net, p);
TRAINR, Epoch 0/100
TRAINR, Epoch 25/100
TRAINR, Epoch 50/100
TRAINR, Epoch 75/100
TRAINR, Epoch 100/100
TRAINR, Maximum epoch reached.

>> plotsom(net.iw{1,1},net.layers{1}.distances);

```



Контрольные вопросы

1. Что представляет собой нейронная сеть?
2. Какие функции активации вы знаете?
3. Какие методы обучения нейронных сетей существуют?
4. Как функционируют самоорганизующиеся сети Кохонена?

Список литературы:

1. Балдин, К. В. Информационные системы в экономике : учебник / К. В. Балдин, В. Б. Уткин. - 9-е изд., стер. - Москва : Дашков и К°, 2021. - 395 с. : ил., табл. - URL: <https://biblioclub.ru/index.php?page=book&id=684194> (дата обращения 24.02.2022) . - Режим доступа : по подписке. - Текст : электронный.

2. Лисяк, В.В. Разработка информационных систем : учебное пособие / В.В. Лисяк. – Ростов-на-Дону ; Таганрог : Южный федеральный университет, 2019. – 97 с. –URL: <https://biblioclub.ru/index.php?page=book&id=577875> (дата обращения: 03.02.2021). – Режим доступа : по подписке. – Текст : электронный.

3. Грекул, В. И. Проектирование информационных систем : учебное пособие / В. И. Грекул, Г. Н. Денищенко, Н. Л. Коровкина. - Москва : Интернет-Университет Информационных Технологий, 2005. - 304 с. - (Основы информационных технологий). - URL: <http://biblioclub.ru/index.php?page=book&id=233071> (дата обращения 14.11.2022) . - Режим доступа : по подписке. - Текст : электронный.

4. Антонов, В. Ф. Методы и средства проектирования информационных систем : учебное пособие / В. Ф. Антонов, А. А. Москвитин. – Ставрополь : Северо-Кавказский Федеральный университет (СКФУ), 2016. – 342 с. – URL: <https://biblioclub.ru/index.php?page=book&id=458663> (дата обращения: 15.11.2022). – Режим доступа : по подписке. – Текст : электронный.

5. Золотов, С. Ю. Проектирование информационных систем : учебное пособие / С. Ю. Золотов. - Томск : Эль Контент, 2013. - 88 с. - URL: https://biblioclub.ru/index.php?page=book_red&id=208706 (дата обращения 13.09.2022) . - Режим доступа : по подписке. - Текст : электронный.