

## **Применение регрессионного анализа и нейронных сетей для построения моделей рынка квартир в г.Биробиджан**

*Войтешко Олег Альбертович*

*Приамурский государственный университет имени Шолом-Алейхема  
Студент*

*Баженов Руслан Иванович*

*Приамурский государственный университет им.Шолом-Алейхема  
к.п.н., доцент, зав.кафедрой информационных систем, математики и  
методик обучения*

### **Аннотация**

В данной статье рассматриваются методы регрессионного анализа в среде Gretl и построения нейронной сети в Neural Network Wizard. Исследование производится на основе данных рынка 1,2,3 комнатных квартир в г. Биробиджан. Экспериментальные данные были взяты с сайта avito.ru. В результате исследования получены модели, позволяющие спрогнозировать стоимость покупки либо продажи квартиры в зависимости от определенных параметров, которые считаются наиболее значимыми.

**Ключевые слова:** регрессионный анализ, нейронная сеть, рынок квартир

## **The use of regression analysis and neural networks for the construction of apartments market models in Birobidzhan**

*Voiteshko Oleg Albertovich*

*Sholom-Aleichem Priamursky State University  
Student*

*Bazhenov Ruslan Ivanovich*

*Sholom-Aleichem Priamursky State University  
Candidate of pedagogical sciences, associate professor, Head of the Department  
of Information Systems, Mathematics and teaching methods*

### **Abstract**

This article discusses the methods of regression analysis in the environment of Gretl and building neural network Neural Network Wizard. The study is based on market data 1,2,3 bedroom apartments in the city of Birobidzhan. The experimental data were taken from avito.ru site. The study produced models to predict the cost of the purchase or sale of the apartment, depending on certain parameters, which are considered the most significant. Keywords: regression analysis, neural network, market apartments. This article discusses the methods of regression analysis Gretl environment and building a neural network in the Master

network Neural. The study is based on market data 1,2,3 bedroom apartments in the city of Birobidzhan. The experimental data were taken from avito.ru site. The study produced models to predict the cost of the purchase or sale of the apartment, depending on certain parameters that are considered important.

**Keywords:** regression analysis, neural network, market apartments

В настоящее время люди, начинают задумываться о покупке квартиры которое, удовлетворяло всем их требованиям. Для большинства граждан предпочтительнее приобрести квартиру, которая будет недалеко от работы, школы, детского садика. Какую сумму правильно будет заплатить за эту поддержанную или только построенную, отремонтированную квартиру? Чтобы правильно дать ответ, необходимо учесть определенные параметры: количество комнат, район, площадь, стоимость. На основе данных по рынку предложений можно построить различные модели для прогнозирования цены. Для этого хорошо подходят регрессионный анализ и искусственные нейронные сети.

Ряд ученых изучает проблемы и методы построения различных регрессионных моделей и нейронных сетей. С.Г.Радченко [1] показал необходимость использования эвристических решений в статистических задачах и привел разработанные эвристики в планировании эксперимента, регрессионном анализе. В.В.Манжула, Д.С.Федяшов [2] использовали интеллектуальные анализы данных, а именно нейронные сети Кохонена и нечеткие нейронные сети, выявлены достоинства и недостатки данных сетей. Д.В.Голов, Л.В.Красовская [3] рассматривали нейронные сети и распознавание рукописных цифр на основе искусственных нейронных сетей. В.П.Зимин, М.М.Филиппов [4] исследовали задачи аппроксимации семейства экспериментальных вольтамперных характеристик термоэмиссионного преобразователя слоистыми искусственными нейронными сетями. Ф.М.Льянова [5] рассматривала использование методов корреляционно-регрессионного анализа в анализе хозяйственной деятельности предприятий. А.Л.Мячин [6] использовал методы общей теории паттернов и регрессионного анализа для исследования показателей инновационной деятельности отдельных регионов Российской Федерации. М.П.Кривенко [7] рассмотрел задачи прогнозирования значений одной переменной по значениям другой методами регрессионного анализа. Д.К.Тюмиков [8] продемонстрировал взаимодействие предикторных переменных в регрессионном анализе нелинейных объектов. Использование априорной информации о месторасположении объектов лица для оценки его биометрических признаков нейронной сетью показал А.Д.Варламов [9]. Рекуррентную нейронную сеть с двумя сигнальными системами представил В.Ю.Осипов [10]. А.А.Арзамасцев и В.П.Рыков разработали модель искусственной нейронной сети с реализацией модульного принципа обучения [11]. В.П.Мешалкин и др. исследовали искусственные нейронные сети, используя их для моделирования свойств создаваемых композиционных наноматериалов [12]. Моделирование расхождения судов в зоне чрезмерного

сближения нейронными сетями провели В.А.Седов и Н.А.Седова [13]. И.Г.Генералов и др. применили информационные технологии при статистической оценке конкурентной среды на региональных продуктовых рынках [14]. Р.И.Баженов и др. занимались проблемами прогнозирования с помощью нейронных сетей и регрессионных моделей [15-20].

Рассмотрим разработку регрессионной модель в среде Gretl и нейронную сеть в Neural Network Wizard для определения зависимости стоимости квартир от заданных параметров.

По данным сайта [www.avito.ru](http://www.avito.ru) за октябрь, ноябрь 2016 г. Был произведен сбор экспериментальных наблюдений. Просматривались объявления о продаже квартир по г.Биробиджан. Были подобраны следующие критерии: количество комнат, район, площадь, стоимость,

Создадим таблицу данных в MS Excel. Используя данные 48 объявлений (рис. 1).

1	Количество комнат	Район, ул.	Площадь, кв. м	Стоимость, тыс. руб.
2	1	Шолом-Алейхема, 94	30	1450
3	1	Шолом-Алейхема, 26	30,9	1700
4	1	Шолом-Алейхема, 83а	36,4	1670
5	1	Шолом-Алейхема, 75а	34,6	1200
6	1	40 лет Победы 10	36,9	1800
7	1	40 лет Победы, 15	30	1600
8	1	40 лет Победы, 25	29,1	1390
9	1	40 лет Победы, 19	30	1150
10	1	Пионерская, 17	30	1500
11	1	Пионерская, 80а	30	1200
12	1	Пионерская, 54	32,5	1200
13	1	Пионерская, 77а	36	1700
14	1	Советская, 68	31	1400
15	1	Советская, 64	32	1400
16	1	Советская, 70	30,1	1150
17	1	Советская, 68	32	1700
18	2	Шолом-Алейхема, 48	43,5	1950
19	2	Шолом-Алейхема, 80	46	1700
20	2	Шолом-Алейхема, 77	44	1600
21	2	Шолом-Алейхема, 27	57	2250
22	2	40 лет Победы, 8	46	2400
23	2	40 Лет Победы, 10	57	3000
24	2	40 Лет Победы, 9	50	2200

Рисунок 1 - Часть данных в Excel

Введем обозначения переменных: количество комнат – К, район – R, площадь – P, стоимость – S (рис. 2)

1	К	R	P	S
2	1	Шолом-Алейхема, 94	30	1450
3	1	Шолом-Алейхема, 26	30,9	1700
4	1	Шолом-Алейхема, 83а	36,4	1670
5	1	Шолом-Алейхема, 75а	34,6	1200

Рисунок 2 – Наименования переменных

Следующий шаг – Закодировать столбец Район из текстового значения в числовое. Выведем четыре наиболее значимых района и обозначим их: Шолом-Алейхема 1-30 = 1, Шолом-Алейхема 31-150=2, 40 лет Победы 1-30 = 3, 40 лет Победы 31-150= 4, Пионерская 1-30= 5, Пионерская 31-150= 6, Советская 1-30= 7, Советская 31-150= 8. Далее нужно импортировать данные таблицы MS Excel в Gretl. В окне программы Gretl появляются переменные, которые необходимы, чтобы построить регрессионную модель (рис. 3,4,5).

1	Район, ул.	Кодирование	Район	Кодировка
2	Шолом-Алейхема, 94	2	ША 1-30	1
3	Шолом-Алейхема, 26	1	ША 31-150	2
4	Шолом-Алейхема, 83а	2	40 лет 1-30	3
5	Шолом-Алейхема, 75а	2	40 лет 31-150	4
6	40 лет Победы, 10	3	Пионерская 1 - 30	5
7	40 лет Победы, 15	3	Пионерская 31 - 150	6
8	40 лет Победы, 25	3	Советская 1-30	7
9	40 лет Победы, 19	3	Советская 31-150	8
10	Пионерская, 17	5		
11	Пионерская, 80а	6		
12	Пионерская, 54	6		
13	Пионерская, 77а	6		
14	Советская, 68	8		
15	Советская, 64	8		
16	Советская, 70	8		
17	Советская, 68	8		
18	Шолом-Алейхема, 48	2		
19	Шолом-Алейхема, 80	2		
20	Шолом-Алейхема, 77	2		
21	Шолом-Алейхема, 27	1		
22	40 лет Победы, 8	3		
23	40 лет Победы, 10	3		
24	40 лет Победы, 9	3		

Рисунок 3 – Кодирование переменных

	A	B	C	D
1	K	R	P	S
2	1	2	30	1450
3	1	1	30,9	1700
4	1	2	36,4	1670
5	1	2	34,6	1200
6	1	3	36,9	1800
7	1	3	30	1600
8	1	3	29,1	1390
9	1	3	30	1150
10	1	5	30	1500
11	1	6	30	1200
12	1	6	32,5	1200
13	1	6	36	1700
14	1	8	31	1400
15	1	8	32	1400

Рисунок 4 – Получившаяся таблица с данными

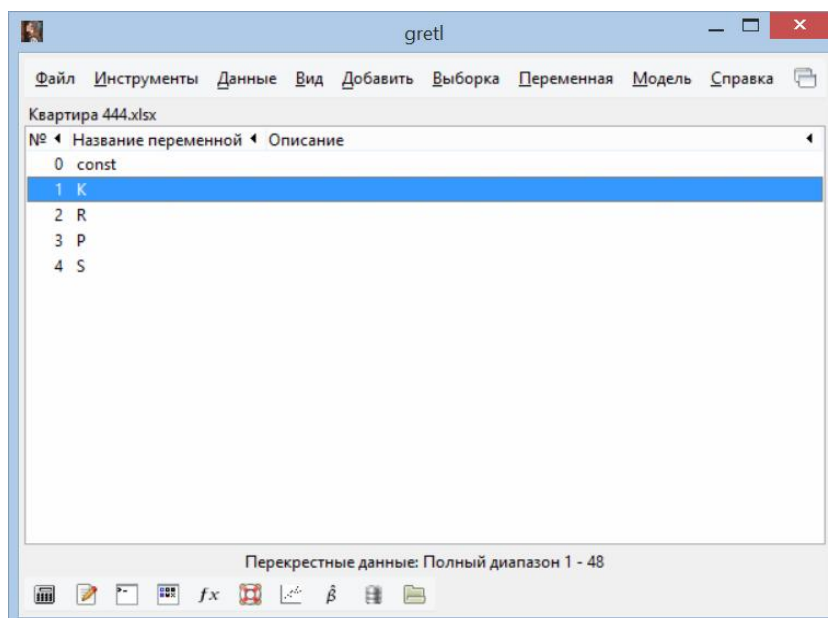


Рисунок 5 - Рабочее окно программы Gretl

Следует просмотреть получившуюся таблицу (рис. 6-7).

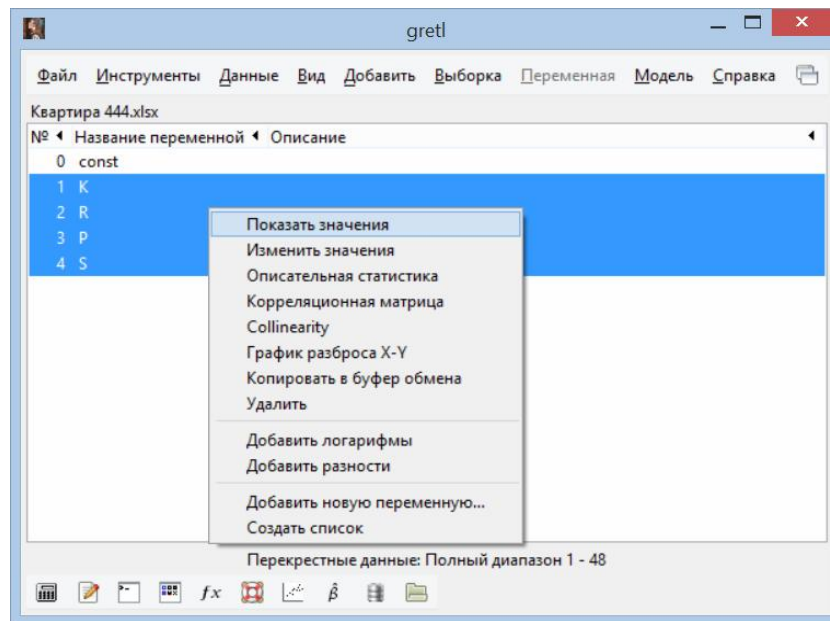


Рисунок 6 - Контекстное меню выделенных переменных

	K	R	P	S
1	1	2	30,0	1450
2	1	1	30,9	1700
3	1	2	36,4	1670
4	1	2	34,6	1200
5	1	3	36,9	1800
6	1	3	30,0	1600
7	1	3	29,1	1390
8	1	3	30,0	1150
9	1	5	30,0	1500
10	1	6	30,0	1200
11	1	6	32,5	1200
12	1	6	36,0	1700
13	1	8	31,0	1400
14	1	8	32,0	1400
15	1	8	30,1	1150
16	1	8	32,0	1700
17	2	2	43,5	1950
18	2	2	46,0	1700
19	2	2	44,0	1600
20	2	1	57,0	2250

Рисунок 7 - Просмотр таблицы данных

Перейдем к построению уравнения модели  $y = a \cdot x_1 + b \cdot x_2 + c \cdot x_3$ . Для решения нашей задачи найдем регрессионную модель, используя метод наименьших квадратов (рис. 8, 9).

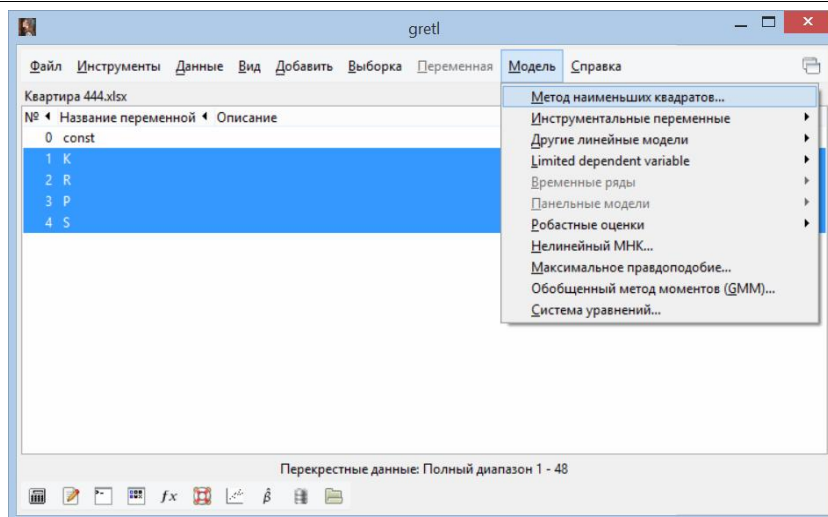


Рисунок 8 - Меню модель

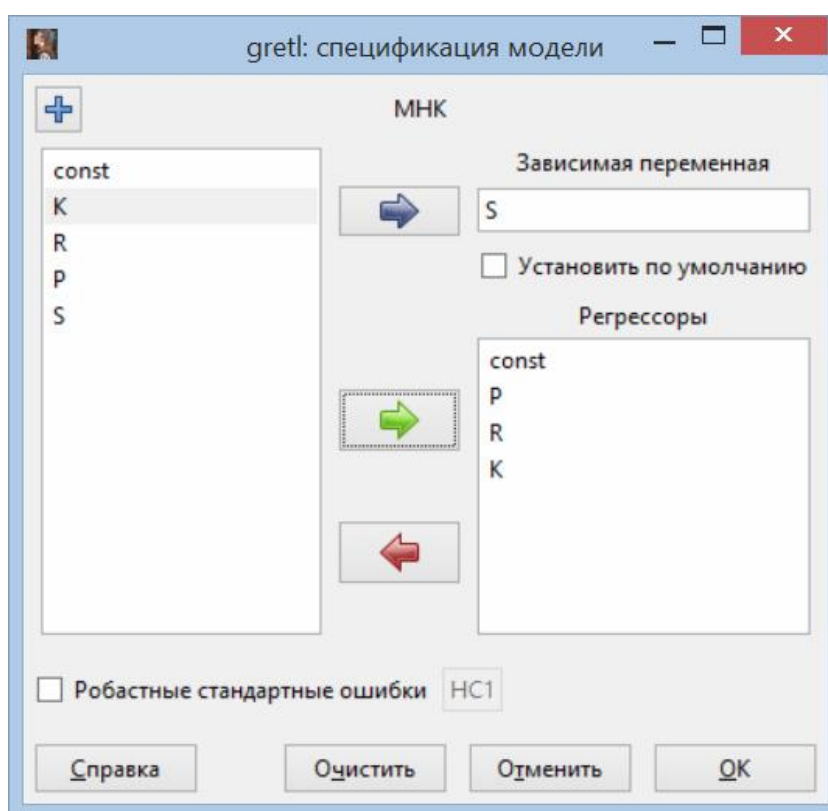


Рисунок 9 – Рабочее окно спецификации модели

Получившаяся модель и ее описательные статистики продемонстрированы на рис.10.

	Коэффициент	Ст. ошибка	t-статистика	P-значение
const	85,9837	213,636	0,4025	0,6893
P	55,7653	7,05430	7,905	5,52e-010 ***
R	-29,9372	22,7695	-1,315	0,1954
K	-262,354	135,826	-1,932	0,0599 *

Среднее зав. перемен	2112,708	Ст. откл. зав. перемен	805,6813
Сумма кв. остатков	6451297	Ст. ошибка модели	382,9104
R-квадрат	0,788543	Испр. R-квадрат	0,774125
F(3, 44)	54,69328	P-значение (F)	6,87e-15
Лог. правдоподобие	-351,5152	Крит. Акаике	711,0305
Крит. Шварца	718,5153	Крит. Хеннана-Куинна	713,8590

Исключая константу, наибольшее р-значение получено для переменной 2 (R)

Рисунок 10- Регрессионная модель

По значению исправленного R-квадрата можно судить о доле вариации результативного признака с учетом воздействия изучаемых факторов. В данной модели 77,4% вариации переменной S зависит от влияния включенных факторов. Если коэффициент выше 75%, то модель считается достаточно хорошей. Найдем зависимость S(цена) от P(Площадь), R(Район), K(Количество комнат).

Перейдем к решению уравнения

$$s = a \cdot p + b \cdot r + c \cdot k$$

Введем в Excel данные const, P, R, K. В ячейку, окрашенную в синий цвет, вводим формулу расчета (рис.11). В таблицу Excel вводим необходимые параметры нашей квартиры, которую мы хотим купить или продать, в ячейки, окрашенные в желтый цвет. После этого в ячейке окрашенную в синий цвет отобразится стоимость интересующей нас квартиры.

	A	B	C	D	E
1	const	85,98337			
2	P	55,7653	Площадь кв.м		35
3	R	-29,9372	Район, ул.		1
4	K	-262,354	Количество комнат		2
5					
6					
7	S	1483,12367			
8					

Рисунок 11 - Расчет стоимости квартиры по заданным параметрам

Если мы хотим узнать стоимость другой квартиры в г. Биробиджан, то требуется в ячейки, окрашенные в желтый цвет ввести другие параметры.

Далее, разработаем нейронную сеть в программе Neural Network Wizard. Для этого, предварительно, создадим текстовый файл с данными о квартирах (рис. 12).

К	R	P	S
1	2	30	1450
1	1	30.9	1700
1	2	36.4	1670
1	2	34.6	1200
1	3	36.9	1800
1	3	30	1600
1	3	29.1	1390
1	3	30	1150
1	5	30	1500
1	6	30	1200
1	6	32.5	1200
1	6	36	1700
1	8	31	1400
1	8	32	1400
1	8	30.1	1150
1	8	32	1700
2	2	43.5	1950
2	2	46	1700
2	2	44	1600
2	1	57	2250
2	3	46	2400
2	3	57	3000
2	3	50	2200
2	3	45	2100
2	6	56	3200
2	5	44	1750
2	6	44	1700
2	6	59.9	3100
2	8	40.9	1700

Рисунок 12 – Фрагмент данных в текстовом документе

Откроем его в программе

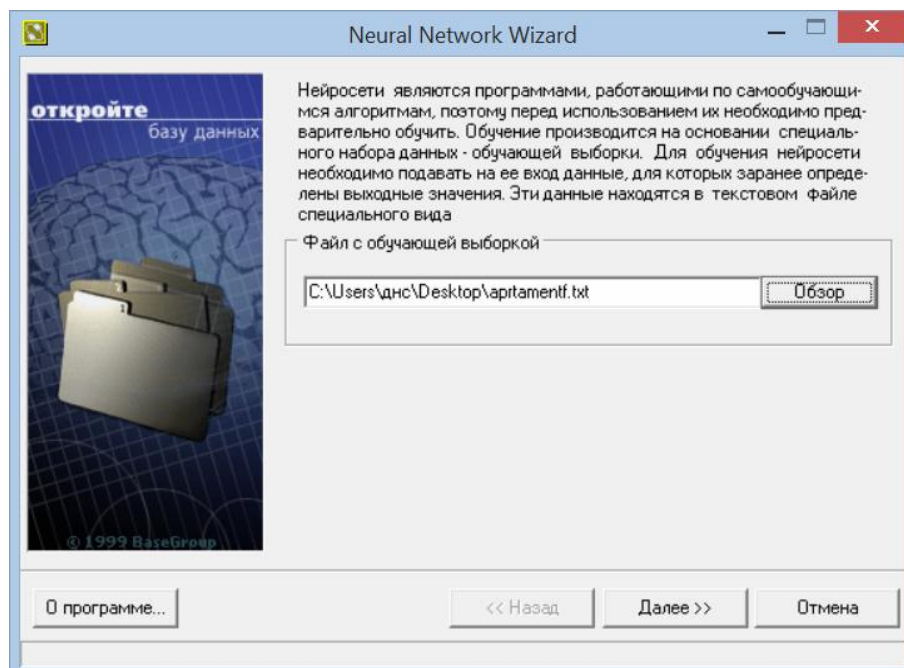


Рисунок 13 - Программа Neural Network Wizard



Далее, нужно указать, что поле  $s$  — целевое. То есть, нейронная сеть будет пытаться определить каким образом значения полей  $K$ ,  $R$ ,  $P$  влияют на поле  $S$  (рис. 14).

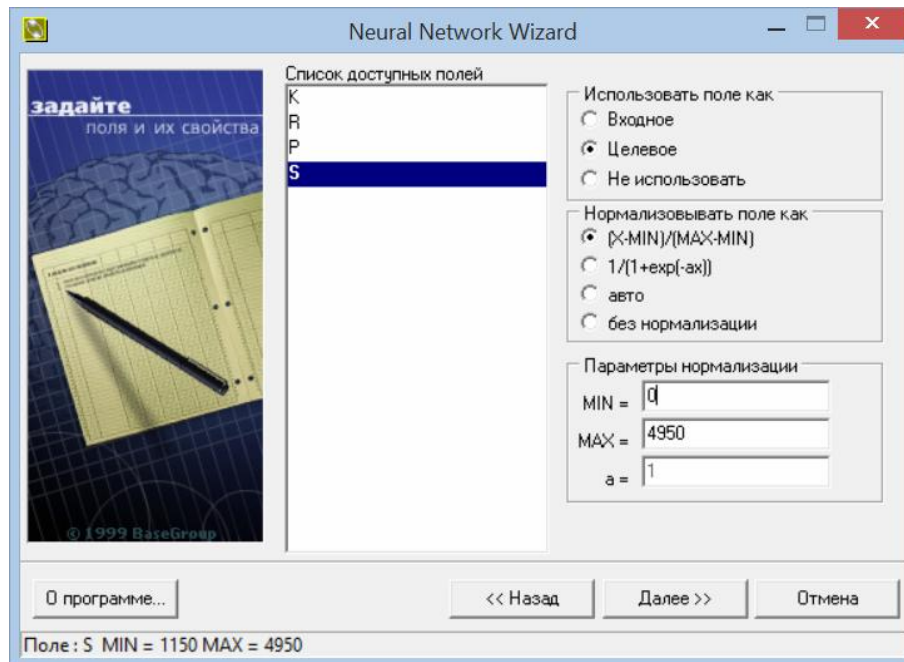


Рисунок 14 - Программа Neural Network Wizard

Определим конфигурацию нейронной сети. Зададим количество скрытых слоев — 1. Количество элементов в первом слое — 3. Параметр сигмоиды зададим 0,1 (рис. 15).

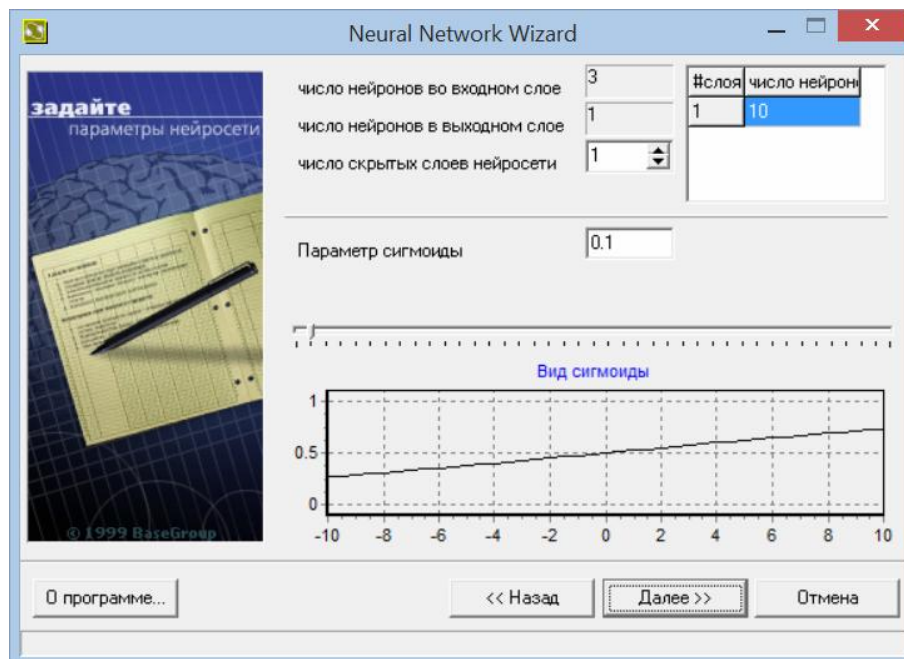


Рисунок 15 - Окно конфигурации нейронной сети

Определим параметры обучения — остановить обучение по прошествии 70000 эпох (рис. 16).

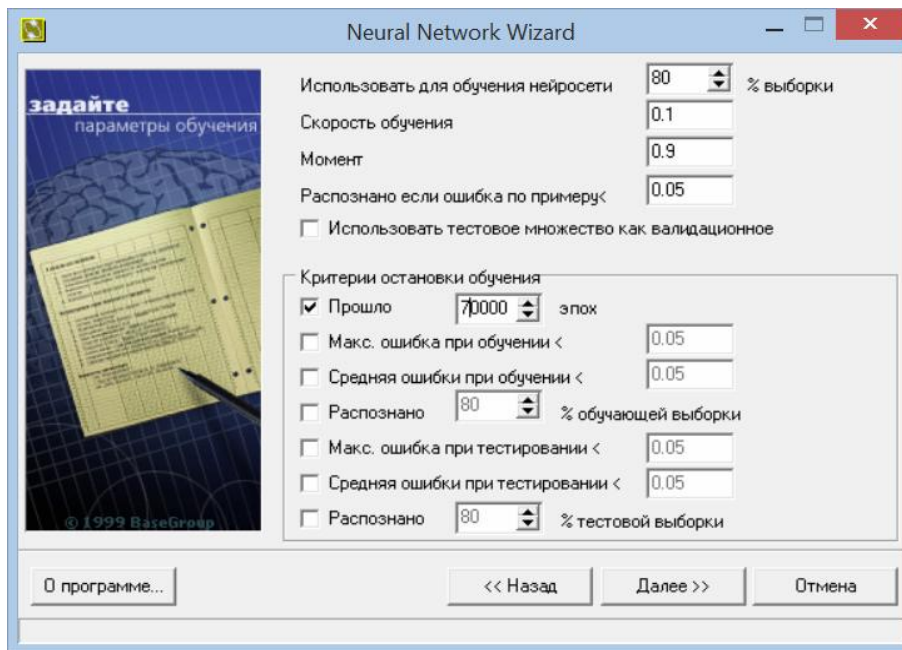


Рисунок 16 – Настройка параметров обучения

Проверим правильность конфигурации сети и параметров обучения (рис. 17).

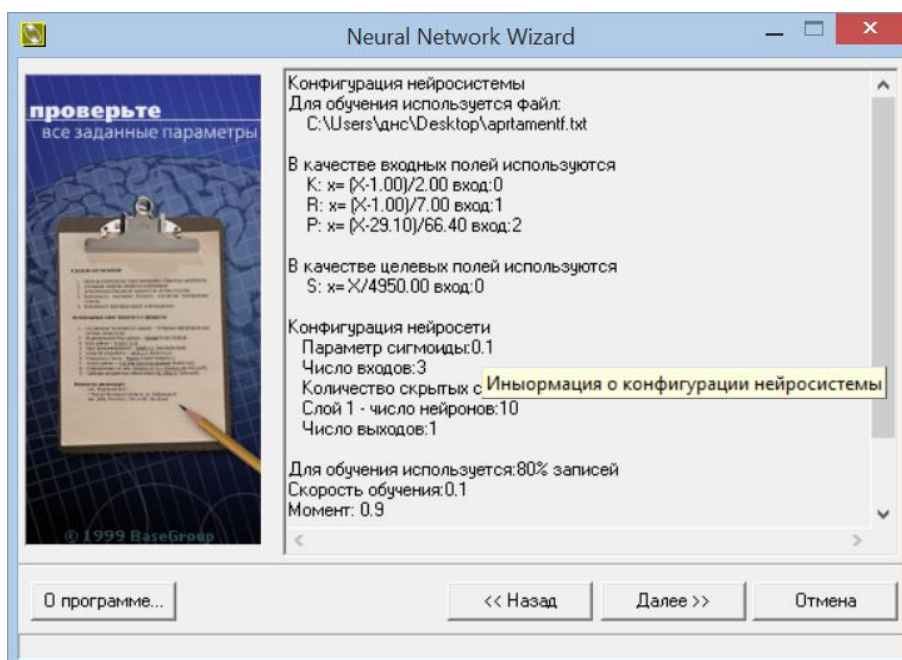


Рисунок 17 - Проверка конфигурации сети и параметров обучения

Запустим систему на обучение (рис. 18).

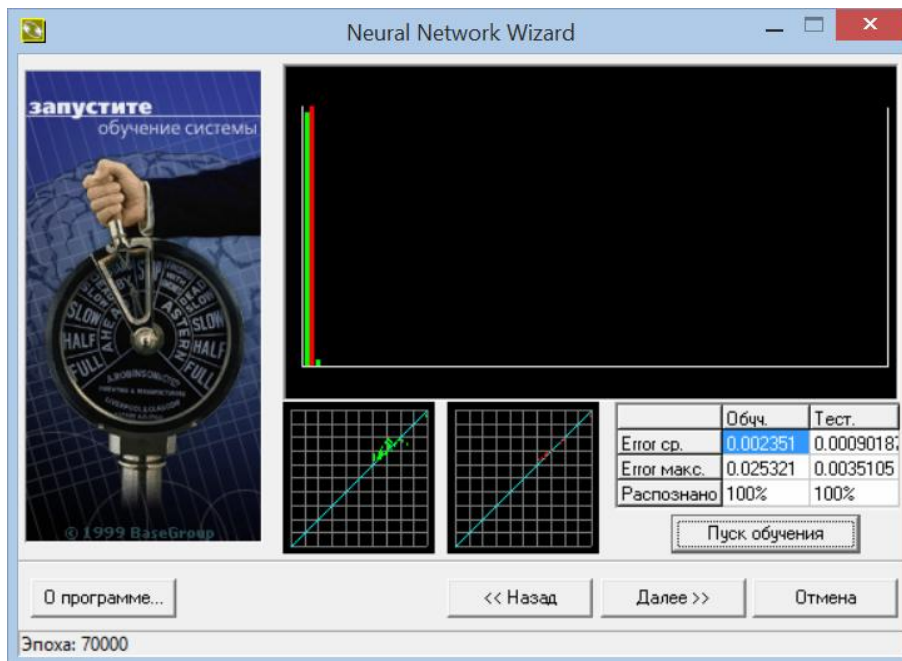


Рисунок 18 - Окно запуска обучения

По окончании обучения протестируем полученную модель. Обозначения входных параметров: количество комнат –  $K$ , район –  $R$ , площадь –  $P$ , стоимость –  $S$ . Введем начальные параметры и рассчитаем стоимость квартиры –  $S$  (рис. 19).

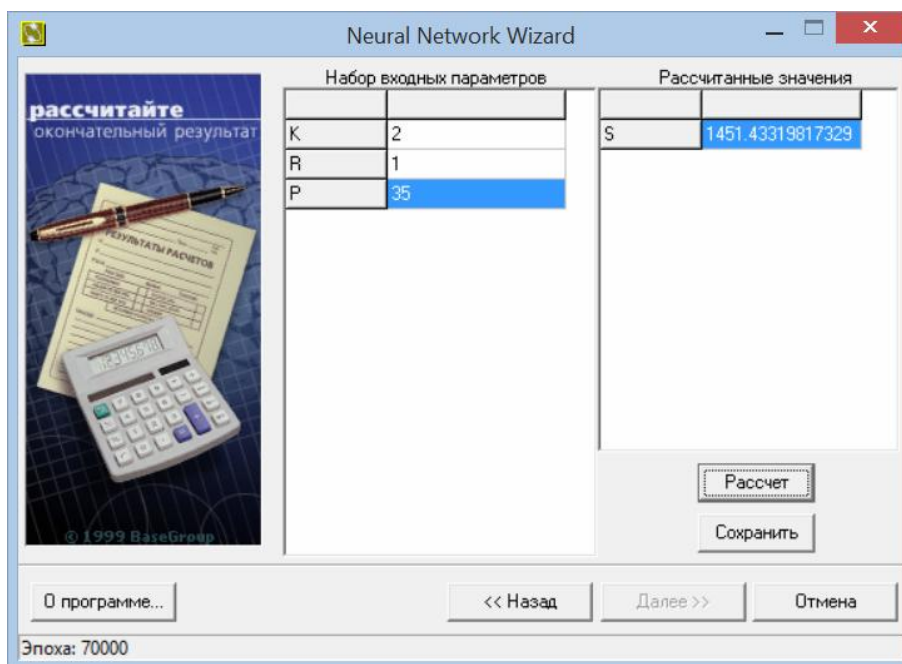


Рисунок 19 - Окно тестирования нейронной сети

Теперь сравним методы регрессионного анализа и нейронных сетей. В таблицу внесем стоимости квартир по одинаковым параметрам в столбцы «Регрессионный анализ» и «Нейронная сеть». Покажем разницу стоимости квартир (табл.1).

Таблица 1 - Данные экспериментов

<b>Регрессионный анализ</b>	<b>Нейронная сеть</b>	<b>Абсолютная разница</b>	<b>Относительная разница, %</b>
1483	1451	32	2
2040	1859	181	8
1900	2177	277	12
2694	3000	306	11

По данным таблица видно, что оба метода отлично подходят для установления стоимости квартиры.

Исследование рынка недвижимости может быть использовано организациями, занимающимися покупкой – продажей недвижимости, сдачей ее в аренду для определения лучшего соотношения цены и качества жилья, определения привлекательности недвижимости в зависимости от района расположения и наличия инфраструктуры (школ, больниц, транспортных остановок).

В настоящее время, когда налог на недвижимость рассчитывается от рыночной стоимости жилья, граждане имеют право обращаться в суд для пересчета налога, не соглашаясь с предложенной оценкой недвижимости. Исследование может позволить организациям, занимающимся независимой оценкой недвижимости, оценить стоимость квартир более справедливо при обращении граждан в данных случаях.

### **Библиографический список**

1. Радченко С.Г. Использование эвристики в планировании эксперимента и регрессионном анализе // Математические машины и системы . 2015. №3. с. 87-92.
2. Манжула В.Г., Федяшов Д.С. Нейронные сети Кохонена и нечеткие нейронные сети в интеллектуальном анализе данных // Фундаментальные исследования. 2011. №4. С. 108-114.
3. Голов Д.В., Красовская Л.В. Нейронные сети Кохонена и нечеткие нейронные сети в интеллектуальном анализе данных // Исследования технических наук . 2014. №4 (14) . С. 18-20.
4. Зимин В.П., Филиппов М.М. Исследование аппроксимации вольтамперных характеристик термоэмиссионного преобразователя искусственными нейронными сетями // Исследования технических наук . 2008. №5. С. 142-146.
5. Льянова Ф.М. Использование методов корреляционно-регрессионного анализа в анализе хозяйственной деятельности предприятий // Исследования технических наук . 2013. №3-4. С. 40-44.
6. Мячин А.Л. Совместное использования метода «pattern» и регрессионного анализа при анализе инновационной деятельности в регионах РФ // Научные труды московского гуманитарного университета . 2013. №9. С.

- 50-56.
7. Кривенко М.П. Сравнительный анализ процедур регрессионного анализа // Информатика и ее применения . 2014. №3. С. 70-78.
  8. Тюмиков Д.К. Эффекты взаимодействий предикторных переменных в регрессионном анализе нелинейных объектов // Вестник транспорта Поволжья . 2010. №2. С. 52-57.
  9. Варламов А.Д. Использование априорной информации о месторасположении объектов лица для оценки его биометрических признаков нейронной сетью // Нейрокомпьютеры: разработка, применение. 2015. № 3. С. 34-38.
  10. Осипов В.Ю. Рекуррентная нейронная сеть с двумя сигнальными системами // Информационно-управляющие системы. 2013. № 4 (65). С. 8-15.
  11. Арзамасцев А.А., Рыков В.П. Модель искусственной нейронной сети (ИНС) с реализацией модульного принципа обучения // Вестник Тамбовского университета. Серия: Естественные и технические науки. 2012. Т. 17. № 4. С. 1219-1224.
  12. Мешалкин В.П., Дли М.И., Стоянова О.В. Исследование искусственных нейронных сетей, используемых для моделирования свойств создаваемых композиционных наноматериалов // Известия высших учебных заведений. Серия: Химия и химическая технология. 2011. Т. 54. № 5. С. 124-127.
  13. Седов В.А., Седова Н.А. Моделирование расхождения судов в зоне чрезмерного сближения нейронными сетями // Научные проблемы транспорта Сибири и Дальнего Востока. – 2014. – №3. – С. 102-105.
  14. Генералов И.Г., Суслов С.А., Завиваев Н.С., Балдов Д.В. Применение информационных технологий при статистической оценке конкурентной среды на региональных продуктовых рынках // Управление экономическими системами: электронный научный журнал. 2015. № 9 (81). С. 12.
  15. Пивенко К.А., Баженов Р.И. Построение регрессионной модели в среде Gretl на примере рынка поддержанных автомобилей г. Биробиджана и г. Хабаровска // Экономика и менеджмент инновационных технологий. 2015. № 4-1 (43). С. 72-80.
  16. Лагунова А.А., Баженов Р.И. Разработка в среде gretl регрессионной модели рынка вторичного жилья г. Биробиджана // Nauka-Rastudent.ru. 2015. № 1 (13). С. 40.
  17. Широкова Н.А., Баженов Р.И. Применение корреляционного анализа для исследования данных спортивных показателей студентов в среде SPSS // Nauka-Rastudent.ru. 2015. № 6 (18). С. 25.
  18. Козич В.Г., Бондаренко В.В., Баженов Р.И. Применение регрессионного анализа и нейронных сетей для построения моделей рынка поддержанных автомобилей Toyota Prius // Постулат. 2015. № 1 (1). С. 4.
  19. Ступников А.В., Баженов Р.И. Прогнозирование цены легковых автомобилей с помощью нейронных сетей в среде Neural Network Wizard // Современная техника и технологии. 2015. № 7 (47). С. 3-10.

---

20. Николаев С.В., Пронина О.Ю., Баженов Р.И. Исследование методов интеллектуального анализа для формирования краткосрочного прогноза в программной среде Statistica // Экономика и менеджмент инновационных технологий. 2015. № 7 (46). С. 89-102.