

## **Применение регрессионного анализа и нейронных сетей для построения модели рынка подержанных телефонов iPhone**

*Мителиков Антон Владимирович*

*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 подержанных телефонов марки iPhone. Экспериментальные данные были взяты сайтов [www.ebay.com](http://www.ebay.com), [www.avito.ru](http://www.avito.ru), [www.farpost.ru](http://www.farpost.ru), [www.macstocks.ru](http://www.macstocks.ru). В результате исследования получены модели, позволяющие спрогнозировать стоимость покупки либо продажи подержанного телефона в зависимости от определенных параметров, которые считаются важными.

**Ключевые слова:** регрессионный анализ, нейронная сеть, рынок подержанных автомобилей

## **Using regression analysis and neural networks to build model of the market for second-hand phones iPhone**

*Mitelikov Anton Vladimirovich*

*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**

The article deals with methods of regression analysis in the environment Gretl and methods of neural network of Neural Network Wizard. The research is based on data from the market of used the iPhone phone model. Experimental data were taken from the sites [www.ebay.com](http://www.ebay.com), [www.avito.ru](http://www.avito.ru), [www.farpost.ru](http://www.farpost.ru), [www.macstocks.ru](http://www.macstocks.ru). The study resulted in models allowing to predict the cost of

buying or selling a second-hand phone depending on certain parameters, which are considered important.

**Keywords:** regression analysis, neural network, the market for used cars

Мобильные телефоны давно стали не только необходимым средством связи, но и средством хранения, передачи и преобразования информации. В наше время трудно представить жизнь без этого девайса. Но какой телефон выбрать? В данный момент на рынке прочно закрепились мобильные телефоны с двумя основными системами – это система android и iOS. Безусловным лидером, по мнению большинства экспертов, являются телефоны, работающие на платформе iOS, называются такие телефоны iPhone. Эти телефоны более стабильны и удобны в работе. Телефоны марки iPhone не только самые надёжные но и одни из самых дорогих, поэтому лучшим вариантом купить подержанный iPhone. Какую сумму правильно будет заплатить за этот бывший в употреблении телефон? Чтобы правильно дать ответ, необходимо учесть определенные параметры: год выпуска, объем встроенной памяти и состояние в котором находится телефон. На основе данных по рынку предложений можно построить различные модели для прогнозирования цены. Для этого хорошо подходят регрессионный анализ и искусственные нейронные сети.

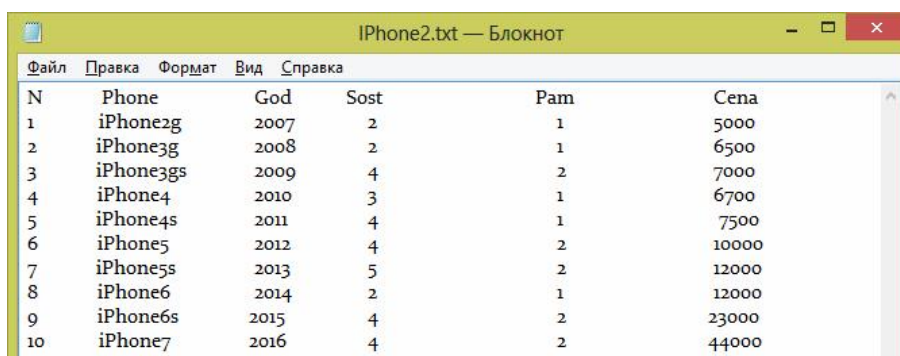
Ряд ученых изучает проблемы и методы построения различных регрессионных моделей и нейронных сетей. Использование априорной информации о месторасположении объектов лица для оценки его биометрических признаков нейронной сетью показал А.Д.Варламов [1]. Рекуррентную нейронную сеть с двумя сигнальными системами представил В.Ю.Осипов [2]. А.А.Арзамасцев и В.П.Рыков разработали модель искусственной нейронной сети с реализацией модульного принципа обучения [3]. В.П.Мешалкин и др. исследовали искусственные нейронные сети, используя их для моделирования свойств создаваемых композиционных наноматериалов [4]. Моделирование расхождения судов в зоне чрезмерного сближения нейронными сетями провели В.А.Седов и Н.А.Седова [5]. И.Г.Генералов и др. применили информационные технологии при статистической оценке конкурентной среды на региональных продуктовых рынках [6]. Р.И.Баженов и др. занимались проблемами прогнозирования с помощью нейронных сетей и регрессионных моделей [7-10]. Д. Ситтон и др. исследовали модель быстрой классификации почв для постройки модульных зданий с использованием нейронных сетей [11]. А.Каскарди и Ф.Миселли создали модель искусственных нейронных сетей для прогнозирования поведения армированных волокон полимера (FRP) бетона при осевой нагрузке [12].

Рассмотрим разработку регрессионной модели в среде Gretl и нейронную сеть в Neural Network Wizard для нахождения зависимости стоимости телефона от данных параметров.

По данным взятым с сайтов [www.ebay.com](http://www.ebay.com), [www.avito.ru](http://www.avito.ru), [www.farpost.ru](http://www.farpost.ru), [www.macstocks.ru](http://www.macstocks.ru) за март 2017 г. был осуществлен сбор

экспериментальных наблюдений. Рассматривались объявления о продаже мобильных телефонов марки iPhone. Были выбраны следующие критерии: год, состояние, объём памяти, цена.

После анализа объявлений полученных с данных сайтов создаём текстовый документ с данными обозначениями: N-номер телефона, Phone-модель телефона, God-год выпуска, Sost-состояние телефона (численный показатель варьируется от 2 до 5, где 2-это изрядно потрёпанное состояние, возможны даже трещины на экране, а 5-это отличное, практически идеальное состояние (показатель с цифрой 1 мы использовать в наших данных не будем, как правило телефоны в таком состоянии годны лишь на запчасти)), следующий показатель Pam - это объем встроенной памяти телефона (как правило телефоны выпускают с различными объемами памяти, современные это 32 Гб, 68 Гб и 128 Гб, а на старых версиях эти показатели 4 Гб, 8 Гб и 16 Гб), Cena – цена (была взята средняя цена так как на разных сайтах цена варьируется в большую или меньшую) (рис. 1).

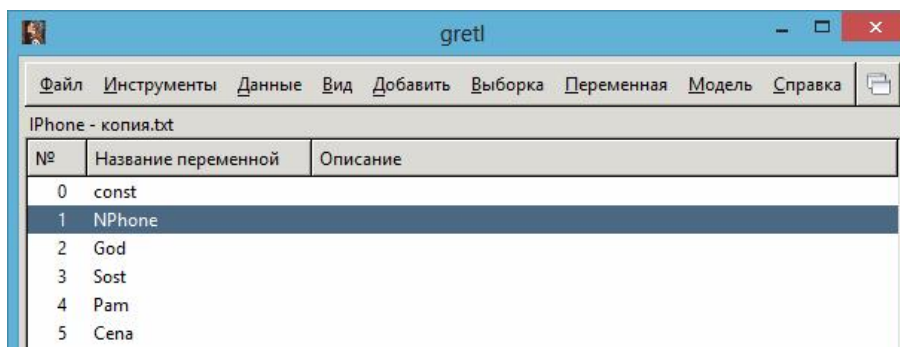


N	Phone	God	Sost	Pam	Cena
1	iPhone2g	2007	2	1	5000
2	iPhone3g	2008	2	1	6500
3	iPhone3gs	2009	4	2	7000
4	iPhone4	2010	3	1	6700
5	iPhone4s	2011	4	1	7500
6	iPhone5	2012	4	2	10000
7	iPhone5s	2013	5	2	12000
8	iPhone6	2014	2	1	12000
9	iPhone6s	2015	4	2	23000
10	iPhone7	2016	4	2	44000

Рисунок 1- Данные созданные в текстовом документе

Следующий шаг – импортировать данные текстового документа в Gretl.

В окне программы Gretl появляются переменные, которые необходимы, чтобы построить регрессионную модель (рис. 2).



№	Название переменной	Описание
0	const	
1	NPhone	
2	God	
3	Sost	
4	Pam	
5	Cena	

Рисунок 2 - Рабочее окно программы Gretl

Необходимо просмотреть получившуюся таблицу (рис. 3-4).

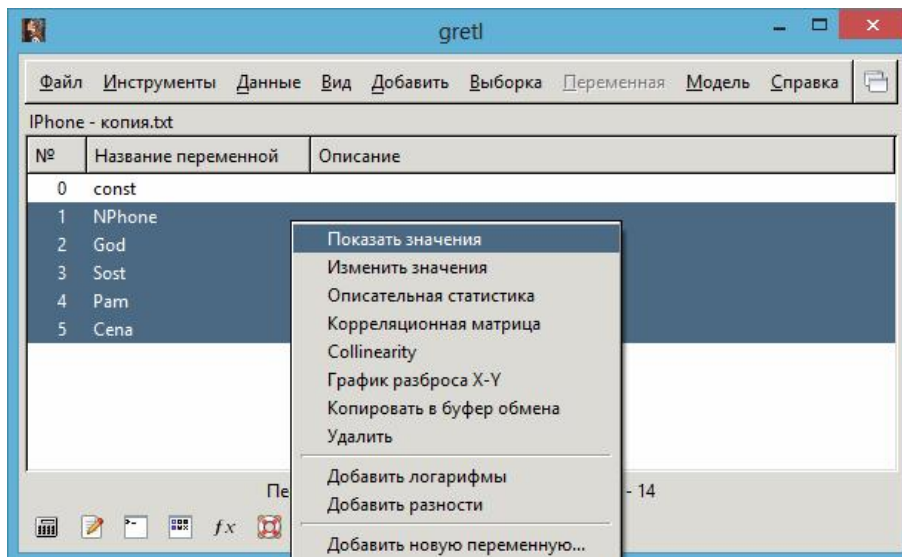


Рисунок 3 - Контекстное меню выделенных переменных



Рисунок 4 - Просмотр таблицы данных

Для решения нашей задачи найдем регрессионную модель, используя метод наименьших квадратов (рис. 5, 6).

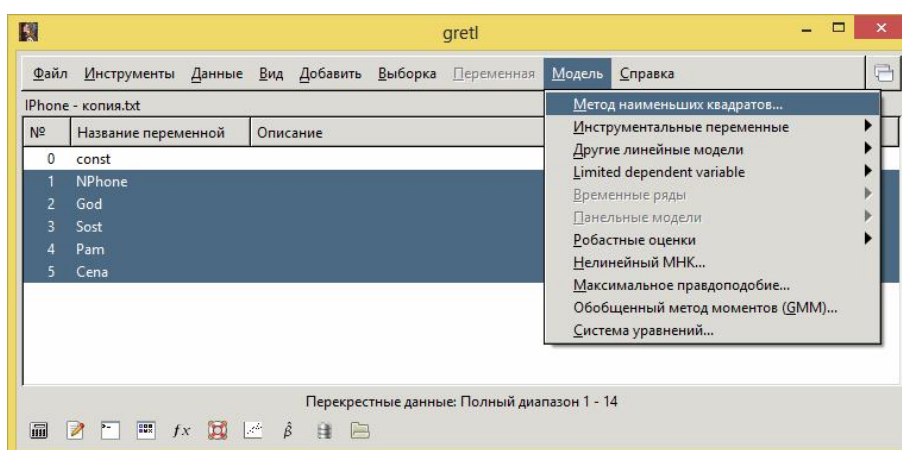


Рисунок 5 - Меню Модель

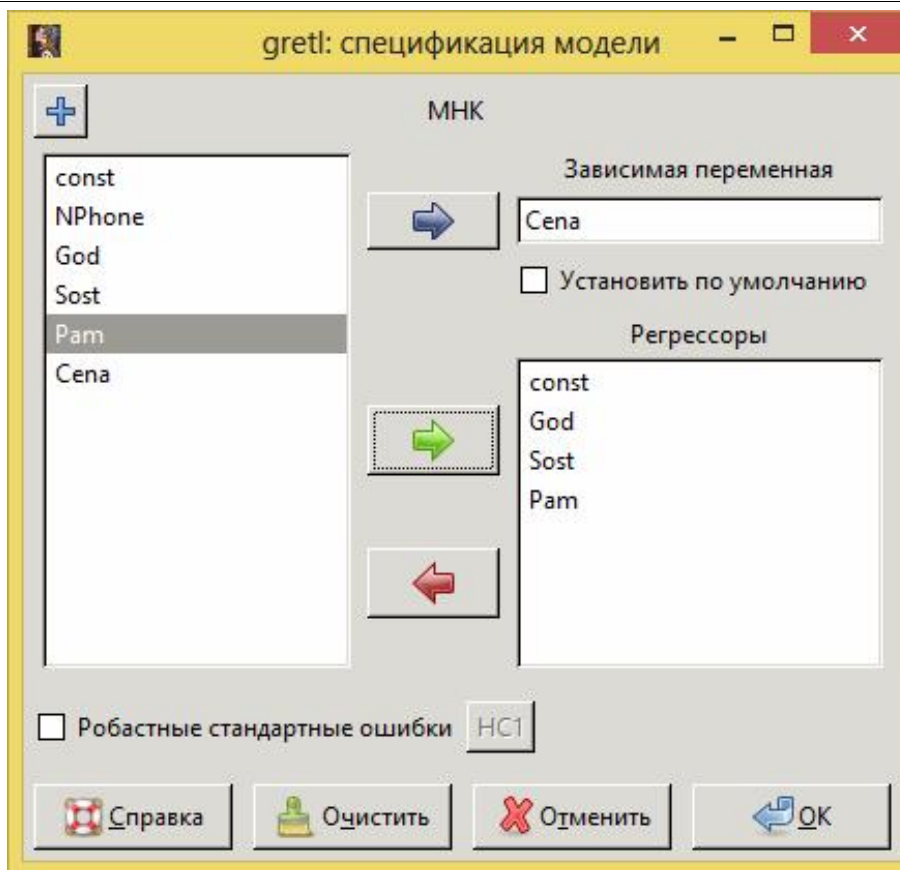


Рисунок 6 - Окно спецификации модели

Получившаяся модель и ее описательные статистики показаны на рис.

7.

	Коэффициент	Ст. ошибка	t-статистика	P-значение
const	-6,94107e+06	2,08385e+06	-3,331	0,0076 ***
God	3458,18	1038,26	3,331	0,0076 ***
Sost	-4470,33	3976,71	-1,124	0,2872
Pam	9069,51	6894,83	1,315	0,2177

Среднее зав. перемен	16642,86	Ст. откл. зав. перемен	13759,74
Сумма кв. остатков	7,99e+08	Ст. ошибка модели	8939,521
R-квадрат	0,675313	Испр. R-квадрат	0,577907
F(3, 10)	6,932963	P-значение (F)	0,008339
Лог. правдоподобие	-144,8852	Крит. Акаике	297,7703
Крит. Шварца	300,3265	Крит. Хеннана-Куинна	297,5337

Исключая константу, наибольшее p-значение получено для переменной 3 (Sost)

Рисунок 7 - Регрессионная модель

По значению исправленного R-квадрата можно судить о доле вариации результативного признака с учетом воздействия изучаемых факторов. В данной модели 57,7907% вариации переменной Cena зависит от влияния включенных факторов. Если коэффициент выше 80%, то модель считается достаточно хорошей.

Получившаяся модель имеет коэффициент всего 57,7907%, поэтому считается весьма не точной.

Добавим переменные  $\ln(\text{God})$  и  $\ln(\text{Cena})$ (рис. 8).

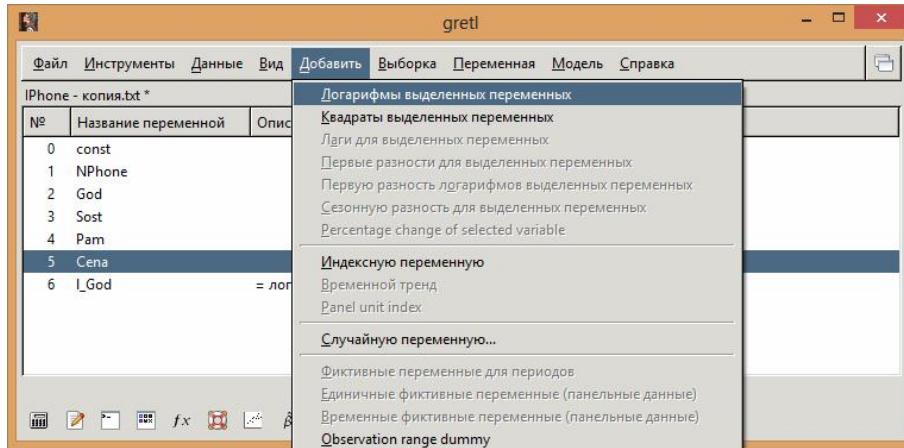


Рисунок 8 - Добавление логарифмов от переменных  
Найдем зависимость  $\ln(\text{God})$  от  $\ln(\text{Cena})$ , NPhone, Sost, Pam (рис.9, 10).

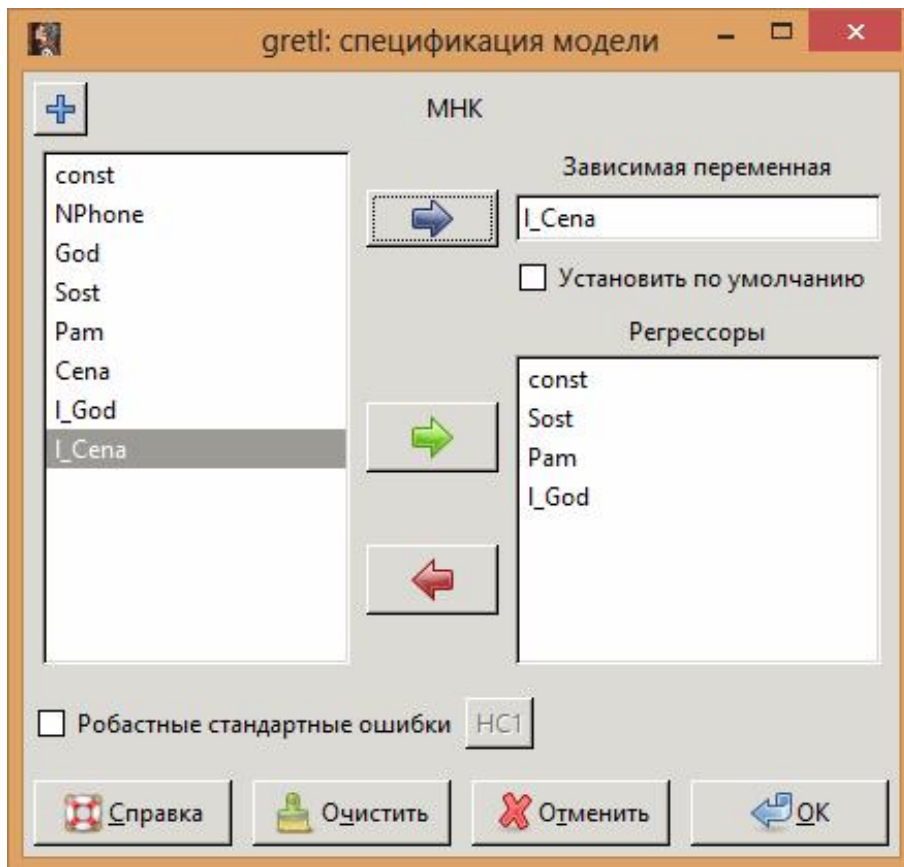


Рисунок 9

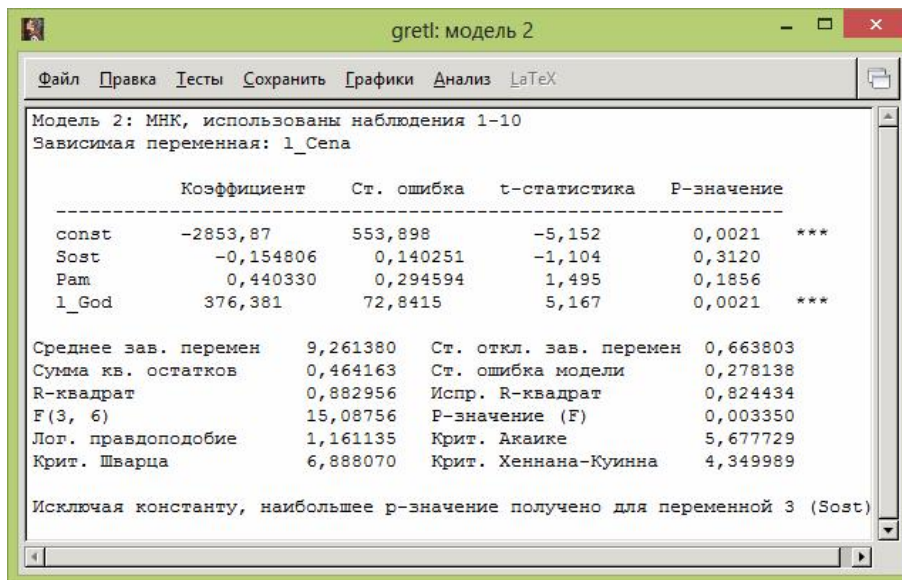


Рисунок 10 - Новая регрессионная модель

В этой модели исправленный R-квадрат=82,4434%. Значит, модель более точная, чем предыдущая.

Перейдем к решению уравнения:

$$\ln(\text{Cena}) = a \cdot \text{Sost} + b \cdot \text{Pam} + c \cdot \ln(\text{God}) + \text{const}$$

Введем в Excel данные const, sost, pam, l\_God, l\_Cena. В ячейку l\_God пишем формулу =LN(год модели телефона), будем прогнозировать модель 2009 года поэтому пишем формулу =LN(2009) (Рис.11). В таблицу Excel вводим необходимые параметры нашего телефона.

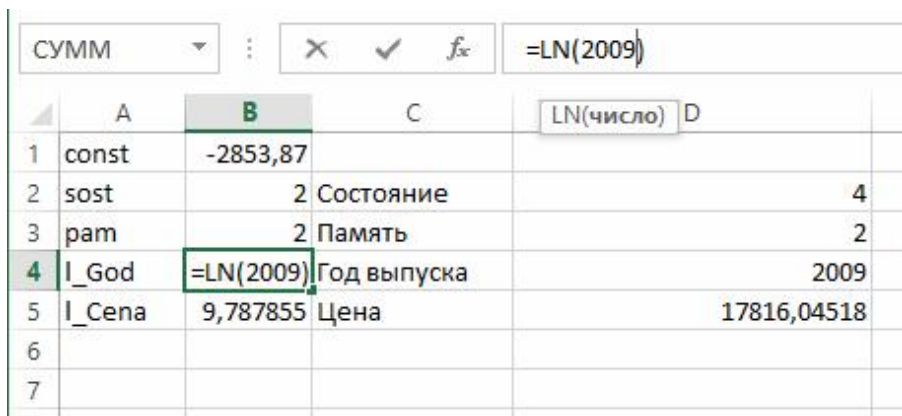


Рисунок 11-заполнение ячейки l\_God

В ячейку l\_Cena вводим формулу (Рис. 12).

СУММ		X		✓		fx		=-0,154806*B2+0,44033*B3+37	
	A	B	C	D					
1	const	-2853,87							
2	sost	2	Состояние						4
3	ram	2	Память						2
4	l_God	7,605392	Год выпуска						2009
5	l_Cena	*B4+B1	Цена						10160,18227
6									

Рисунок 12- заполнение ячейки l\_Cena

Далее, создадим нейронную сеть в программе Neural Network Wizard. Для этого будем использовать созданный ранее текстовый файл с данными о телефонах (рис. 1).

Откроем его в программе (рис. 13).



Рисунок 13 - Программа Neural Network Wizard

Далее, нужно указать, что поле Cena — целевое. То есть, нейронная сеть будет пытаться определить каким образом значения полей God, Sost, Ram влияют на поле Cena (рис.14).



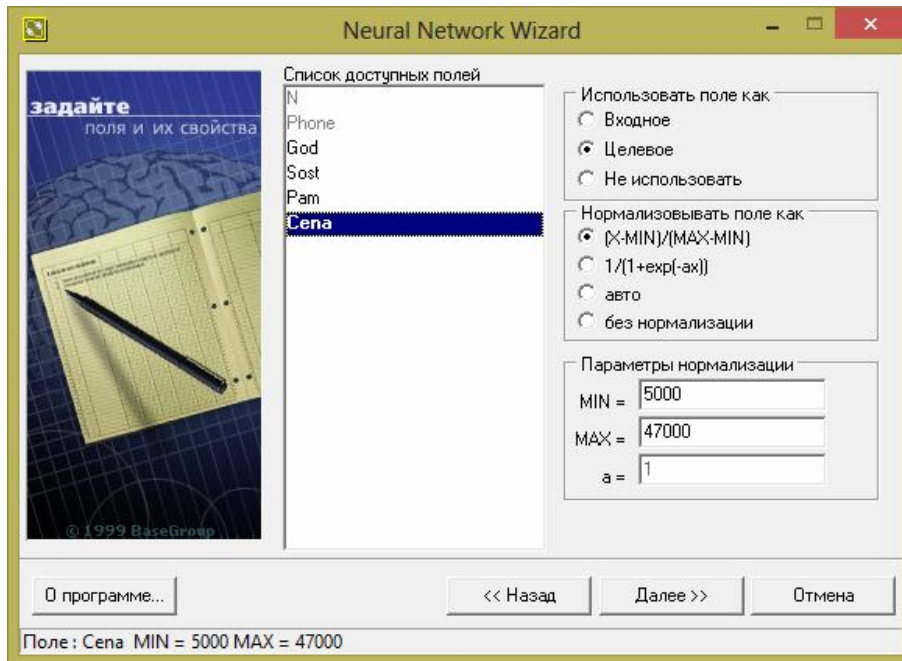


Рисунок 14 – Окно программы со списком полей

Определим конфигурацию нейронной сети. Зададим количество скрытых слоев – 1. Количество элементов в первом слое – 3. Параметр сигмоиды зададим 1,66 (рис. 15).

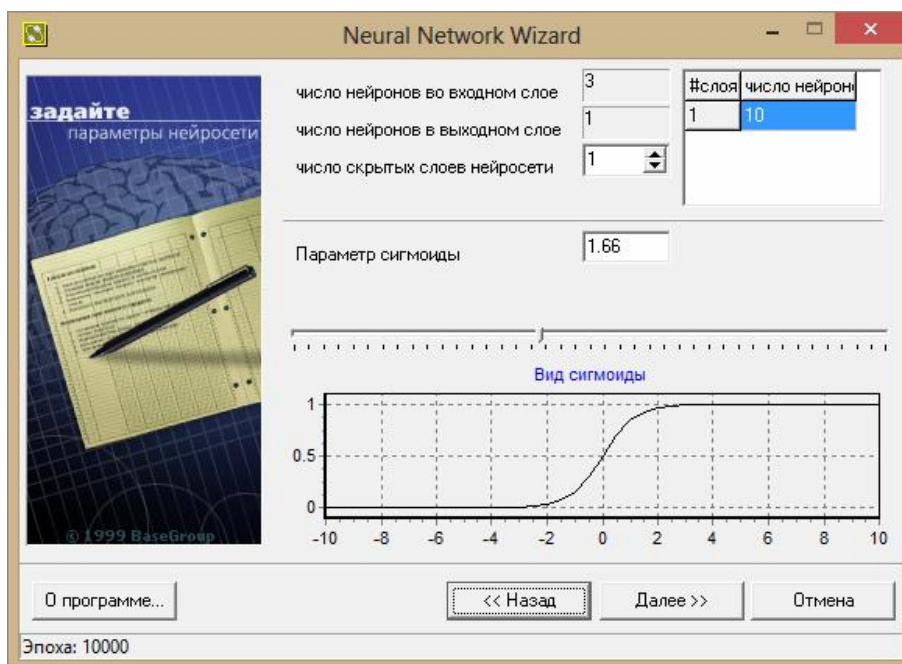


Рисунок 15 - Окно конфигурации нейронной сети

Определим параметры обучения — остановить обучение по прошествии 10000 эпох (рис. 16).

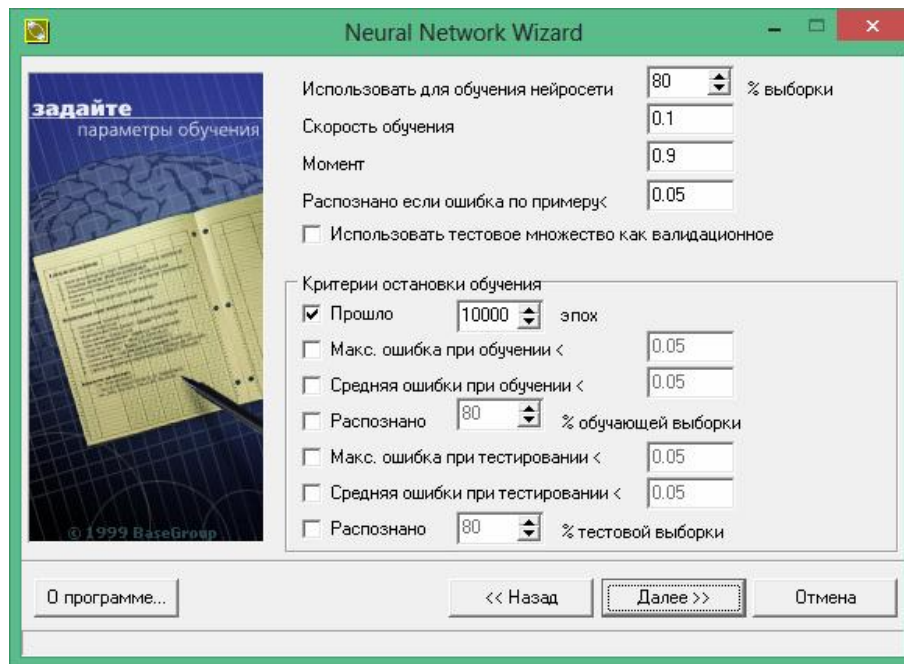


Рисунок 16 - Окно параметров обучения

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

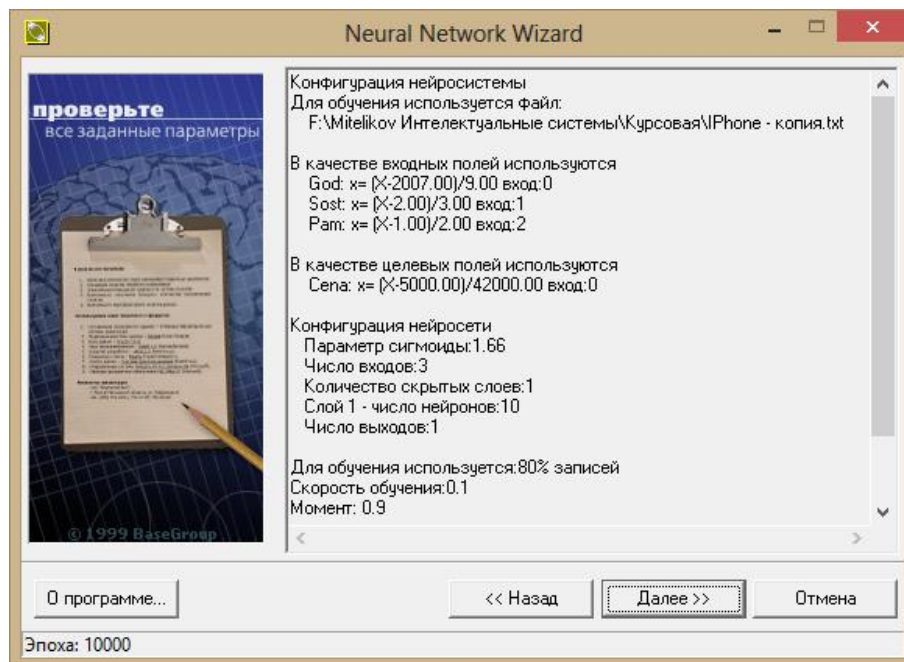


Рисунок 17 - Проверка конфигурации сети и параметров обучения

Запустим систему на обучение (рис. 18).

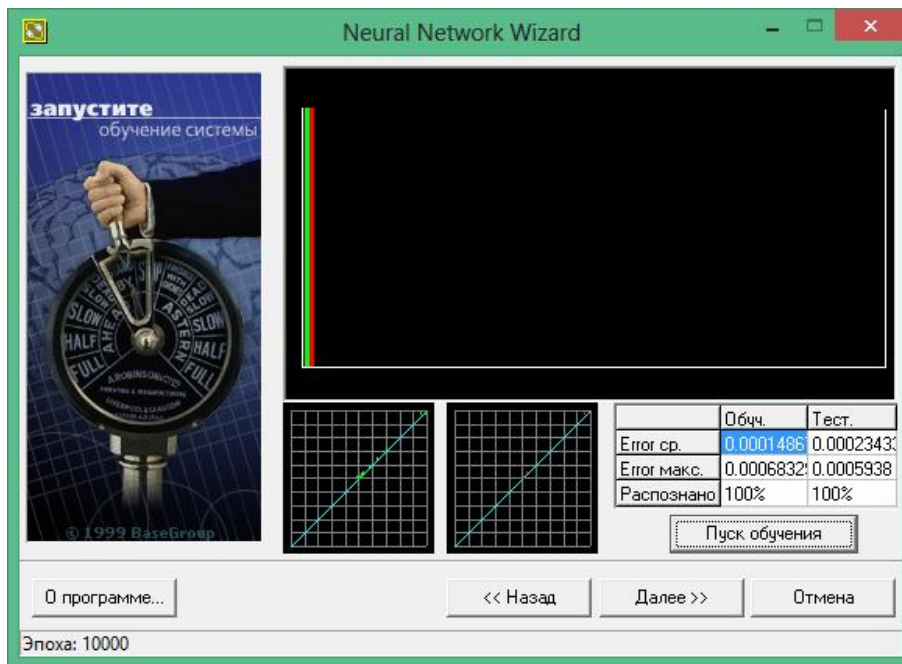


Рисунок 18 - Окно запуска обучения

По окончании обучения протестируем полученную модель. Обозначения входных параметров: год – God, состояние – Sost, память – Pam. Введем начальные параметры и рассчитаем стоимость телефона – Cena (рис. 19).

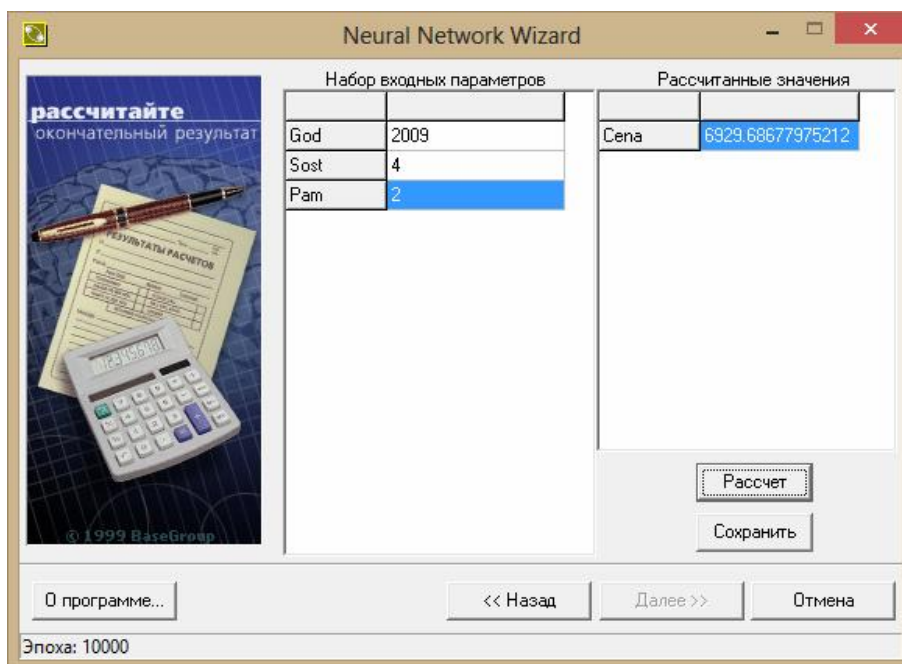


Рисунок 19 - Окно тестирования нейронной сети

Теперь сравним методы регрессионного анализа и нейронных сетей.

В таблицу внесем стоимости телефонов по одинаковым параметрам в столбцы «Регрессионный анализ» и «Нейронная сеть». Покажем разницу стоимости телефонов.

Таблица будет состоять из ячеек:

1. Внесённые параметры телефона- это данные которые мы будем вносить для получения стоимости телефона
2. Фактические данные- конкретные данные которые имеются в нашем текстовом документе о прогнозируемой модели телефона.
3. Логические данные- сколько должна стоить прогнозируемая модель с точки зрения человека.
4. Регрессионный анализ- данные полученные путём анализа в системе Gretl.
5. Нейронная сеть- данные полученные путём анализа в системе Neural Network Wizard.
6. Абсолютная разница- числовая величина рознящаяся от логических данных.
7. Относительная разница- числовая величина рознящаяся от логических данных в процентах.

В первом случае возьмем явные данные из нашего текстового документа и проверим какой из методов имеет большую погрешность.

Как видно из анализа нейронная сеть имеет меньшую погрешность по сравнению с регрессионным анализом, всего 1,1% против 3,5.

Далее будем использовать данные, не имеющиеся в нашем документе.

Таблица 1 – Сравнение данных экспериментов

Внесённые параметры телефона (God, Sost, Pam)	Фактические данные (God, Sost, Pam = Cena)	Логические данные (God, Sost, Pam = Cena)	Регрессионный анализ	Нейронная сеть	Абсолютная разница (регрессионный анализ)	Абсолютная разница (нейронная сеть)	Относительная разница, % (регрессионный анализ)	Относительная разница, % (нейронная сеть)
God 2009 Sost 4 Pam 2	God 2009 Sost 4 Pam 2 =7000	God 2009 Sost 4 Pam 2 =7000	7256	7071	256	71	3,5	1,0
God 2012 Sost 2 Pam 2	God 2012 Sost 4 Pam 2 =10000	God 2012 Sost 2 Pam 2 =7800	17816	7438	10540	362	59	4,8
God 2014 Sost 2 Pam 2	God 2014 Sost 2 Pam 1 =12000	God 2014 Sost 4 Pam 2 =16500	19000	17126	4000	626	21	3,65

По данным табл.1 видно, что регрессионный метод не подходит для анализа стоимости телефонов. Полученную регрессионную модель не рекомендуется использовать для покупки и продажи мобильных телефонов

iPhone, в то время как нейронную сеть можно использовать для данной задачи.

### Библиографический список

1. 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.
2. Осипов В.Ю. Рекуррентная нейронная сеть с двумя сигнальными системами // Информационно-управляющие системы. 2013. № 4 (65). С. 8-15.
3. 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.
4. 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.
5. 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.
6. 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.
7. Мельниченко С.Л., Баженов Р.И. Построение в среде gretl регрессионной модели стоимости подержанных электрогитар в регионах Дальневосточного федерального округа и Восточной Сибири // Постулат. 2016. № 5. С. 15.
8. Трубочанинов В.Н., Баженов Р.И. Построение регрессионной модели в среде Gretl для оценки стоимости внутренних жёстких дисков в Дальневосточном федеральном округе // Постулат. 2016. № 5. С. 5.
9. Войтешко О.А., Баженов Р.И. Применение регрессионного анализа и нейронных сетей для построения моделей рынка квартир в г.Биробиджан // Постулат. 2016. № 11 (13). С. 35.
10. Козич В.Г., Бондаренко В.В., Баженов Р.И. Применение регрессионного анализа и нейронных сетей для построения моделей рынка подержанных автомобилей Toyota Prius. // Постулат. 2015. № 1 (1). С. 4.
11. Ситтон Д., Яейнали Я., Бретт А. Исследование модели быстрой классификации почв для постройки модульных зданий с использованием нейронных сетей // Construction and Building Materials. 2017. №138. С. 214–221.

12. Каскарди А., Миселли Ф. Создание модели искусственных нейронных сетей для прогнозирования поведения армированных волокон полимера (FRP) бетона при осевой нагрузке // Engineering Structures. 2017. №140. С. 199-208.