

## ОБРАБОТКА В СИСТЕМЕ SPSS СТАТИСТИЧЕСКИХ ДАННЫХ МАГАЗИНА КОМПЬЮТЕРНОЙ ТЕХНИКИ

Р. Д. Ульянов – студент, Институт информационных технологий и радиоэлектроники, кафедра информационных технологий и радиоэлектроники, группа ПРИ-118, E-mail: [mstr.calling@gmail.com](mailto:mstr.calling@gmail.com)

С.Ю. КИРИЛЛОВА – научный руководитель, к.т.н, доц., проф. кафедры ИСПИ, Институт информационных технологий и радиоэлектроники, Email: [svkir@mail.ru](mailto:svkir@mail.ru)

**Аннотация:** В статье описан анализ в системе SPSS статистических данных о товарах ИС «Магазин компьютерной техники» с применением регрессионного анализа и нейронной сети «многослойный персептрон».

**Ключевые слова:** SPSS, регрессия, анализ, нейронная сеть.

### Введение

В настоящее время необходимо уметь прогнозировать данные для получения результатов. При помощи прогноза можно получить вероятность появления определенного события. Предсказывание событий в мире имеет не последнее место по важности: предсказывание поведения курса валют, погоды, осадков, даже поведения акций компании на рынке позволяет рассчитать вероятность возникновения неблагоприятного исхода и предпринять меры по его предотвращению. Поскольку расчет прогноза при помощи ручки и листа бумаги занимает много времени и уступает в скорости исполнения и точности результата компьютерным вычислениям, люди создали специальные программы, позволяющие анализировать данные при помощи уже записанных методов анализа. Одна из таких программ – программа SPSS Statistic – позволяет производить обработку данных при помощи описательной статистики.

### Анализ данных в системе SPSS при помощи регрессионного анализа и нейронной сети

В данной статье рассматривается пример применения анализа данных при помощи методов программы SPSS в рамках предметной области «Магазин компьютерной техники». База данных, разработанная в рамках данной предметной области, состоит из 7 сущностей, из которых три являются ссылочными (должности, способ доставки, фирмы поставщики) и четыре

ссылающиеся (состояние заказа, заказы, товары, сотрудники компании). Программа SPSS Statistics предоставляет богатый набор функций для поведения анализа: создание отчетов, методы описательной статистики, корреляционный, кластерный и регрессионный анализы, прогнозирование данных, расчет надежности модели, создание настраиваемых таблиц, расчет выживаемости и анализ данных при помощи нейронной сети типа многослойный персептрон. Ниже приведена схема данных сущностей предметной области:

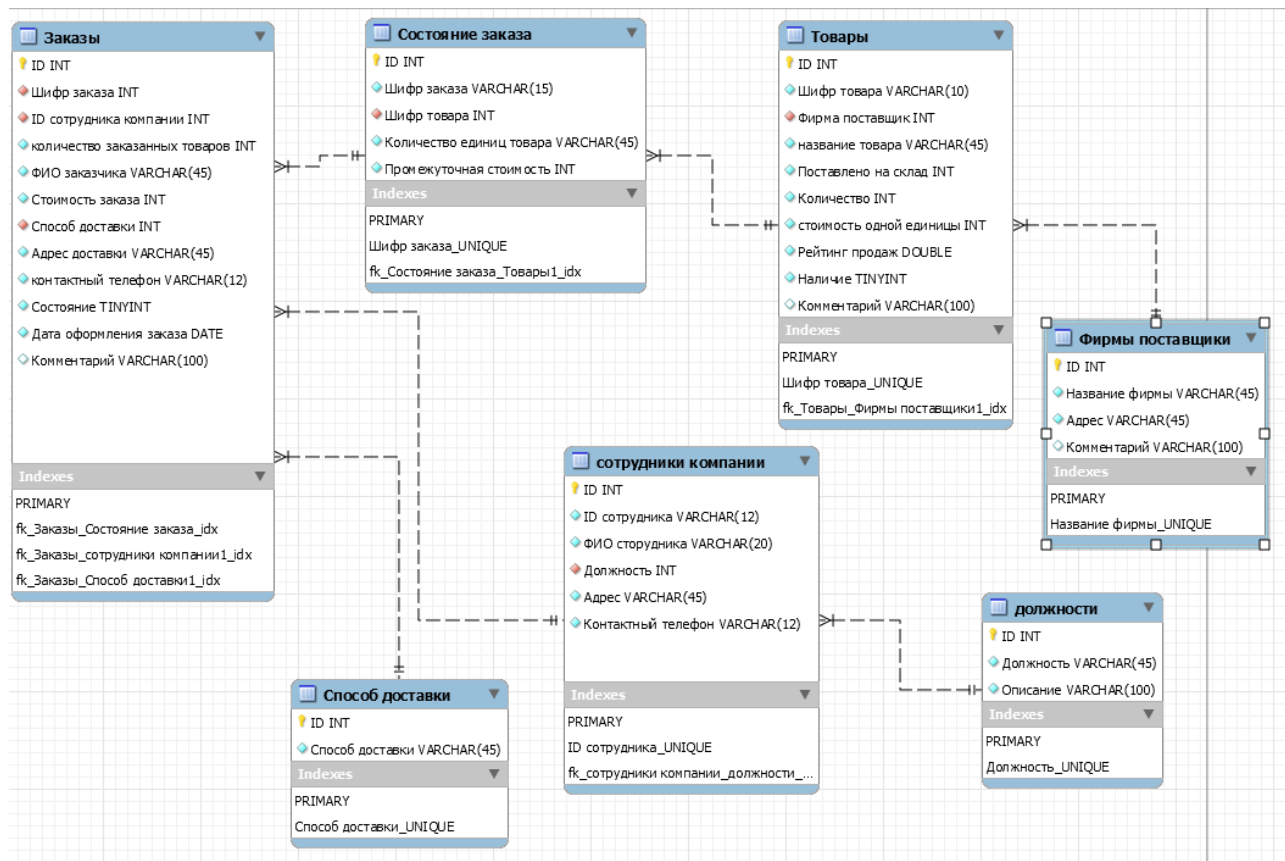


Рисунок 1. Сущности в предметной области

Поскольку демонстрация всех возможностей анализа данных программы SPSS Statistics превышает объем статьи, был проведен анализ данных сущности товары при помощи регрессионного анализа и нейронных сетей «многослойный персептрон». Целью анализа является прогнозирование времени окончания товара. Регрессионный анализ позволяет составить модель данных на основе зависимой переменной (переменной отклика) и независимой переменной (объясняющей переменной). Система SPSS позволяет выбрать множество вариантов регрессионного анализа, таких как линейная регрессия, анализ методом наименьших квадратов, полиномиальная регрессия, порядковая регрессия и др. Нейронные сети позволяют анализировать данные на основе набора, который служил для ее обучения. Тип нейронной сети «многослойный персептрон», используемый в программе SPSS, является одной из наиболее популярных типов сетей прямого распространения. Суть работы сети состоит в

следующем: при анализе данных входной сигнал распространяется в них постепенно, от слоя к слою, общий результат формируется как результат разности между ответами, полученными от каждого слоя в процессе анализа. Чем больше слоев в сети, тем выше точность.

В системе SPSS сущность товары представлена в следующем виде:

	cipher	supplercompany	productname	countall	balance	price	rathing	exists	explanation
1	AA001	Asus	Телефон Asus T3 42GB	200	200	4500,00	7,50	1	товар поль...
2	AA002	Xiaomi	Телефон Xiaomi redmi Notepad 5	4000	4000	6500,00	8,00	1	товар поль...
3	AA003	Honor	Телефон Honor T3	250	250	4000,00	8,90	1	товар поль...
4	AA001	Asus	Телефон Asus T3 64GB	7800	7650	42000,00	9,25	1	товар поль...
5	AA005	Samsung	Веб камера Samsung View 360	3500	3250	1250,00	7,65	1	товар поль...
6	AA006	Honor	Телефон Honor T4 Pro	4500	3825	7500,00	8,82	1	товар поль...
7	AA007	Asus	Телефон Asus T3 32GB Ultra	7500	1250	4800,00	7,80	1	товар поль...
8	AA008	Xiaomi	Телефон Xiaomi redmi Notepad 4	3250	0	71500,00	6,76	0	товар поль...
9	AA009	Honor	Телефон Honor T3 256GB	3500	3250	7500,00	8,71	1	товар поль...
10	AA010	Honor	Клавиатура Honor Game+ Механика	4500	2500	4700,00	7,25	1	товар поль...
11	AA011	Samsung	Клавиатура Samsung for office	3250	3000	7520,00	8,70	1	товар поль...
12	AA012	LG	Клавиатура LG Game механика	7500	7000	4250,00	8,60	1	товар поль...
13	AA013	Asus	Мышь Asus Game 32000 DPI	7000	6520	4800,00	7,25	1	товар поль...
14	AA014	Samsung	Мышь Samsung for office	3600	1250	256,00	7,87	1	товар поль...
15	AA015	Honor	Мышь Asus Game 16000 DPI	7800	7250	3250,00	7,25	1	товар поль...
16	AA016	Asus	ПК Asus Pro 16 GB	400	0	5250,00	6,80	0	товар поль...
17	AA017	Xiaomi	Телефон Xiaomi Redmi Notepad 6	7900	7820	7500,00	7,89	1	товар поль...
18	AA018	HP	ПК HP Ultra Plus 128GB SSD	3800	3620	7500,00	8,90	1	товар поль...
19	AA019	Asus	ПК Asus T3 128GB	6000	5200	7500,00	7,83	1	товар поль...
20	AA020	Xiaomi	Телефон Xiaomi redmi nitepad 5	4000	4000	6500,00	8,00	1	товар поль...
21	AA021	Xiaomi	Телефон Samsung Galaxy A5 64 GB	4000	3960	13500,00	8,70	1	товар поль...
22	AA022	Samsung	Телефон Samsung Galaxy A5 32 GB	3500	3480	8100,00	7,12	1	товар поль...
23	AA023	Xiaomi	Телефон Xiaomi redmi notepad 7 Plus 64GB	7200	7000	6500,00	8,00	1	товар поль...
24	AA024	Honor	Телефон Honor T2 32GB	8500	3500	6500,00	7,89	1	товар поль...
25	AA025	Asus	Телефон Asus T3 16GB	6500	6000	1250,00	7,50	1	товар поль...
26	AA026	LG	Планшетный ПК LG AXRSDF 128GB	4000	4000	7250,00	8,90	1	товар поль...
27	AA027	Honor	ПК Honor Turbo 250GB	4000	3500	8500,00	9,80	1	товар поль...
28	AA028	Asus	Телефон Asus T3 42GB	200	200	4500,00	7,50	1	товар поль...
29	AA029	Xiaomi	Телефон Xiaomi redmi nitepad 5	4000	4000	6500,00	8,00	1	товар поль...
30	AA030	Asus	Телефон Asus Z5467KL 32GB	9000	8250	4000,00	8,50	1	товар поль...
31	AA031	Asus	Телефон Asus T3 42GB	4000	3850	4500,00	7,50	1	товар поль...
32	AA032	Xiaomi	Телефон Xiaomi redmi nitepad 5	4000	4000	6500,00	8,00	1	товар поль...
33	AA033	HP	ПК HP ProBook s55 Series	2800	2550	3900,00	8,50	1	товар поль...
34	AA034	Asus	Телефон Asus T3 42GB	200	200	4500,00	7,50	1	товар поль...
35	AA035	LG	Компьютер LG Vega	300	210	1500,00	7,25	1	товар поль...
36	AA036	Honor	Телефон Honor T1	250	250	4000,00	8,30	1	товар поль...
37	AA037	HP	Ноутбук HP PROBook H350 Serie	4520	4000	4500,00	7,50	1	товар поль...
38	AA038	Xiaomi	Телефон Xiaomi redmi notepad 3	8000	8000	6500,00	8,00	1	товар поль...

Рисунок 2. Данные в системе SPSS.

Сущность товары включает в себя атрибуты: шифр товара (cipher), фирма поставщик (supplercompany), название товара (productname), общее число завезенных товаров на склад (totalcount), текущее количество товара (balance), цену одной единицы товара (price), рейтинг товара (rating), доступность товара для заказа (exists), комментариев к товару (explanation). Целью регрессионного анализа стоит в получении зависимости между зависимой переменной balance и переменных отклика totalcount и price. Анализ будет считаться успешным, если у полученной модели значения R квадрат более 80%. Регрессионный анализ производится методом наименьших квадратов. Суть метода наименьших квадратов [2] состоит в нахождении коэффициентов линейной зависимости, с учетом которых функция двух переменных  $a$  и  $b$   $F(a, b) = \sum_{i=1}^n (y_i - (ax_i + b))^2 \rightarrow \min$ , т.е. принимает наименьше значение. Для проведения

анализа методом наименьших квадратов в программе SPSS следует выбрать анализ>регрессия>двухэтапный МНК (рисунок 3). В появившемся окне выбираем зависимую переменную (balance) и переменные отклика (totalcount и price). В параметрах анализа зададим вывод анализа, остатков и ковариацию (рисунок 4).

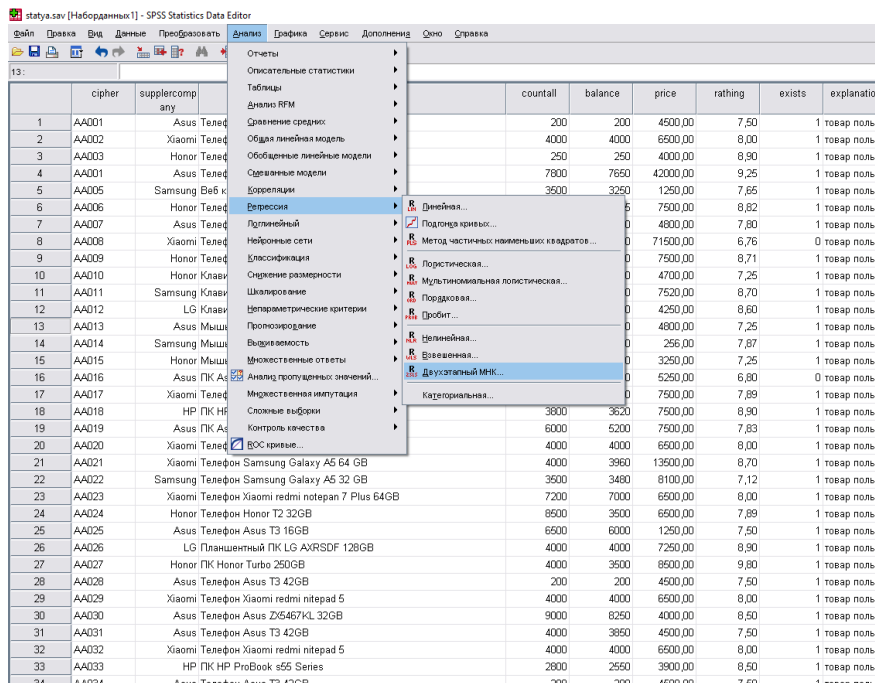
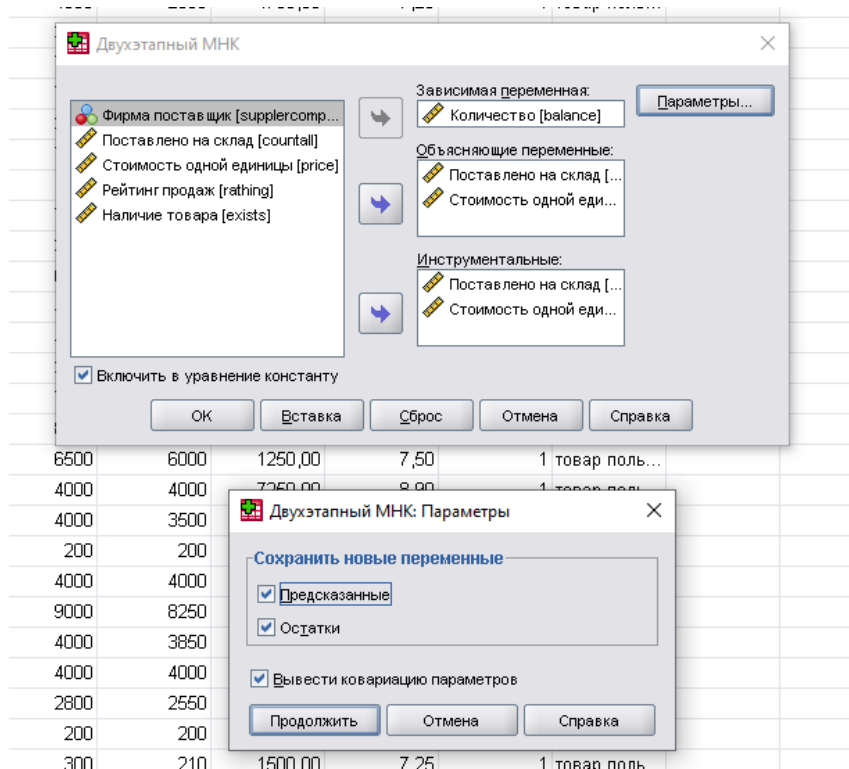


Рисунок 3. Регрессионный анализ методом наименьших квадратов, установка зависимости между зависимой переменными balance и переменными отклика price и totalcount.



#### Рисунок 4. Настройка анализа методом наименьших квадратов

В результате анализа можно увидеть, что зависимость между переменными умеренная, поскольку значение коэффициента детерминации ( $R^2$ ) составляет приблизительно 71%, коэффициент множественной корреляции составляет 87%. Это означает, что переменная balance на 71% зависит от факторов price и totalcount, а в остальных 29% - от других факторов. Ниже представлен результат проведенного анализа

[Наборданных1] C:\Users\Роман\Documents\statya.sav

Описание модели		
		Тип переменной
Уравнение 1	balance	зависимая
	countall	предиктор и инструментальная
	price	предиктор и инструментальная

MOD\_2

Сводка модели		
Уравнение 1	Множественный R	,873
	R-квадрат	,762
	Скорректированный R-квадрат	,749
	Стд. ошибка оценки	1270,512

Дисперсионный анализ						
		Сумма квадратов	Ст. св.	Средний квадрат	F	Знач.
Уравнение 1	Регрессия	191129408,132	2	95564704,066	59,202	,000
	Остаток	59725463,468	37	1614201,715		
	Итого	250854871,600	39			

Коэффициенты						
		Нестандартизованные коэффициенты		Бета	t	Знач.
		B	Стд. ошибка			
Уравнение 1	(Константа)	206,611	398,879		,518	,608
	countall	,837	,077	,875	10,871	,000
	price	-,022	,017	-,105	-1,302	,201

Коэффициенты корреляции				
Уравнение 1	Корреляции		countall	price
				countall
		price	-,077	1,000
	Ковариации	countall	,006	-9,995E-5
		price	-9,995E-5	,000

Рисунок 5. Результат анализа зависимости переменной balance от влияния на нее факторов price и totalcount.

Также над данной выборкой можно провести анализ при помощи нейронной сети. Как писалось ранее, преимущество применения нейронных сетей при анализе данных является то, что она способна сравнивать данные на основе выборки, которая служила для обучения. Затруднением для применения нейронных сетей является то, что при изменении выборки нейронную сеть необходимо переобучать, что занимает достаточно длительное время. Ниже будет показан пример обучения нейронной сети «многослойный персептрон» на примере обучающей выборки из 40 элементов, созданной ранее.

Для начала необходимо запустить анализ нейронной сети, перейдя по вкладке анализ>нейронные сети>многослойный персептрон. Откроется окно настройки нейронной сети.

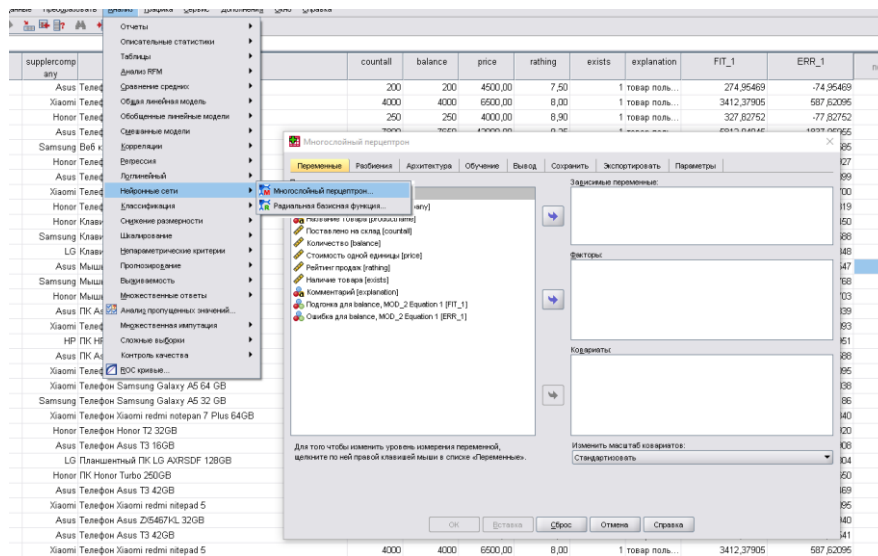


Рисунок 6. Окно настройки нейронной сети.

Поместим в зависимые переменные переменную balance, в ковариаты – переменные отклика totalcount и price (рисунок 7). Далее необходимо настроить сам персептрон для анализа: задать число слоев сети (больше слоев обеспечивает большую точность вычислений), функцию активации (задает входное значение нейрона в зависимости от результата взвешенной суммы входных переменных и предельного значения), а также время обучения сети (время, которое сеть будет анализировать данные). Сигмоидная функция задается уравнением  $f(x) = \sigma(x) = \frac{1}{1+e^{-x}}$ , имеет область значений [0;1] и не аппроксимируется. На рисунках 6-8 показаны демонстрация настроенных параметров нейронной сети

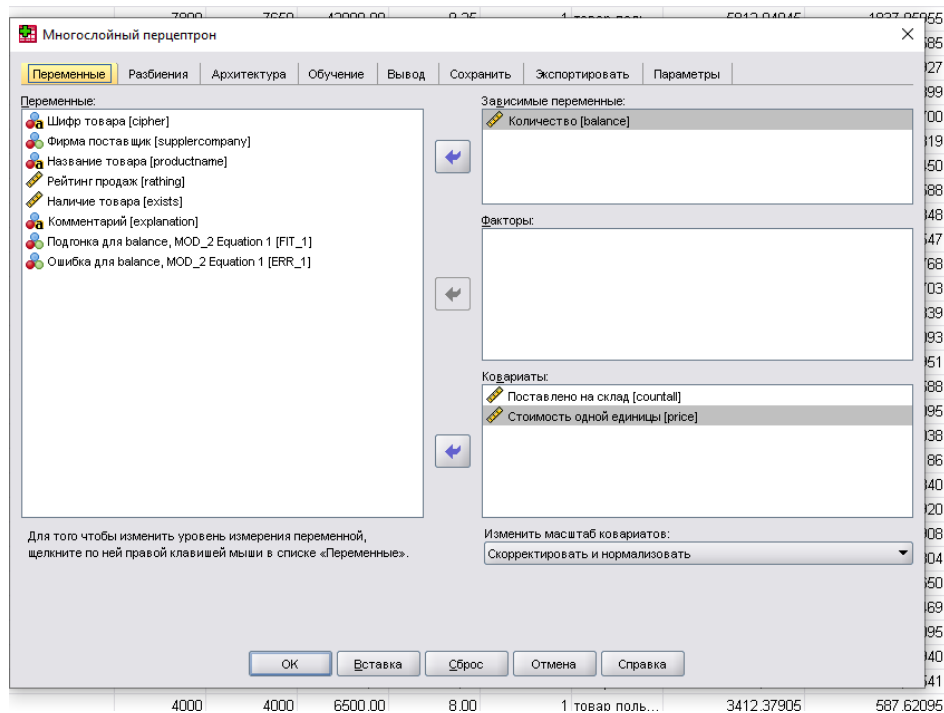


Рисунок 7. Настройка нейронной сети: установка зависимых переменных и ковариаты.

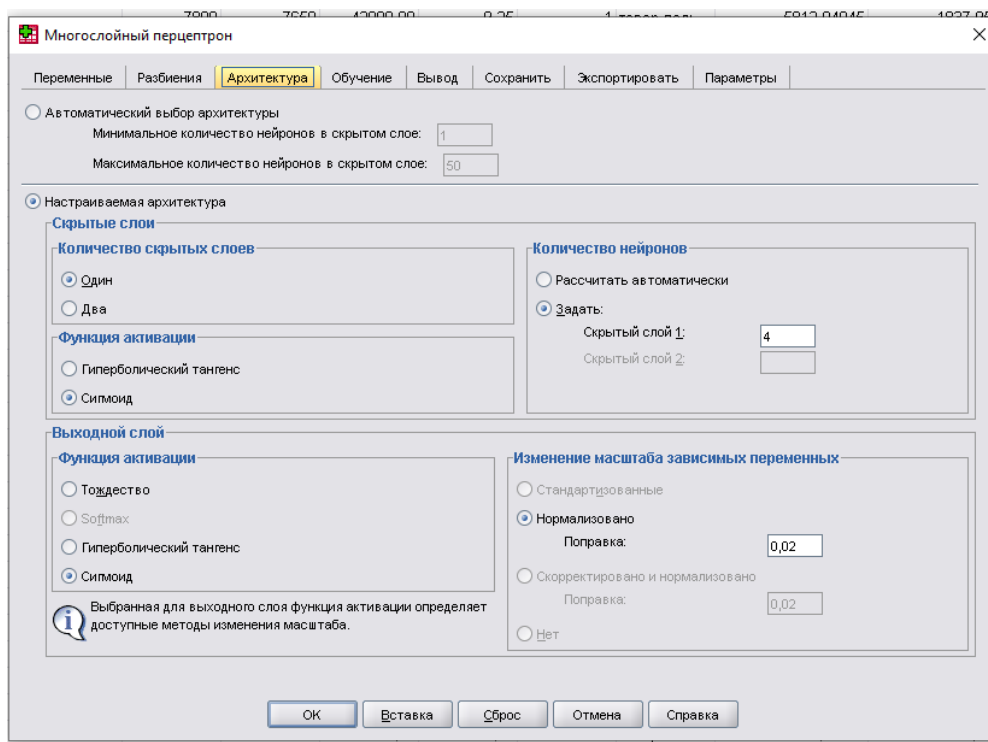


Рисунок 8. Настройка нейронной сети: установка количества слоев, числа нейронов, тип функции активации и нормализованной поправки.

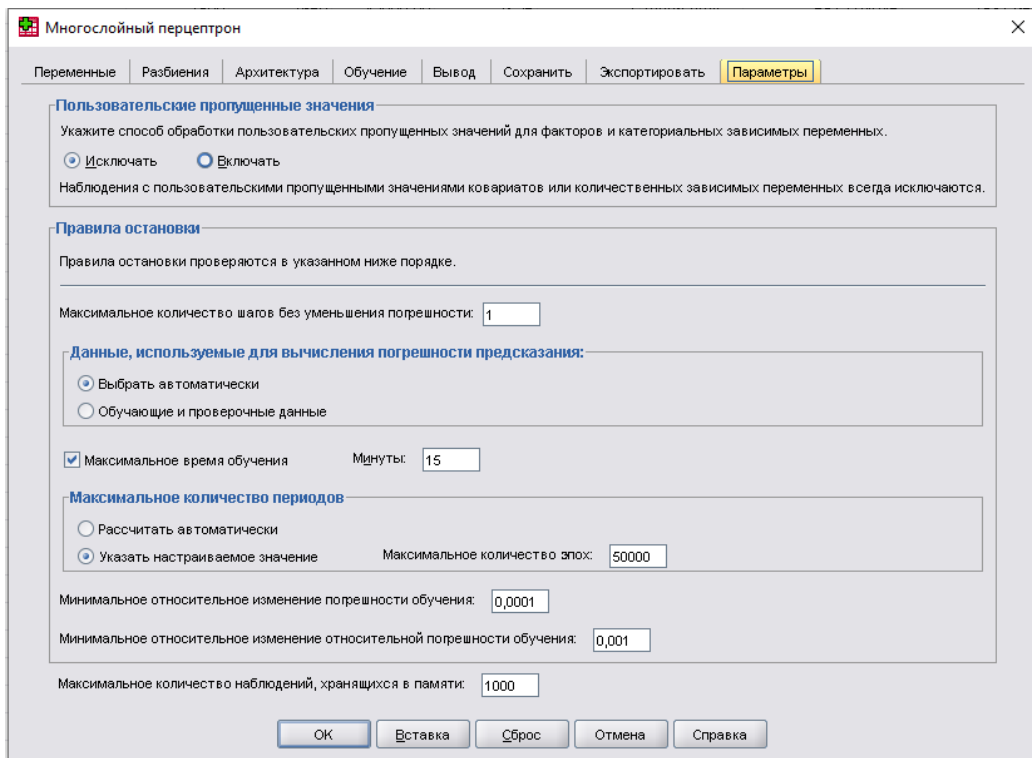


Рисунок 9. Настройка нейронной сети: настройка дополнительных параметров: максимальное время обучения, максимального количества проходов

Настроив параметры таким образом мы получаем отчет о настройке нейронной сети – время настройки, граф, отчет об ошибках. В отчете необходимо проанализировать значения переменных «относительная ошибка»[1,2] и «ошибка суммы квадратов»[1,2]. Первая из них вычисляется по формуле  $\delta_x = \frac{\Delta x}{x_d} * 100\%$ , где  $\Delta x$  – абсолютная погрешность,  $x_d$  – действительное значение. Последняя вычисляется по формуле  $skr = \sqrt{\frac{v^2}{n-1}}$ , где  $n$  – число измерений,  $v^2$  – сумма квадратов вероятностей ошибок. Ошибки суммы квадратов составили 0,019, а относительная ошибка 0,049. Эти значения достаточно малы, что свидетельствует о том, что о нейронная сеть хорошо обучена. Ниже на рисунке представлен отчет с результатами обучения нейронной сети и ее граф.



**Сводка для модели**

Обучающая	Ошибка суммы квадратов	,418
	Относительная ошибка	,267
	Использованное правило остановки	Количество последовательных шагов без уменьшения ошибки: 1 <sup>a</sup>
	Время обучения	0:00:00.003
Проверочная	Ошибка суммы квадратов	,019
	Относительная ошибка	,049

Зависимая переменная: Количество

а. При вычислении ошибок используется проверочная выборка.

**Оценки параметра**

Предиктор		Предсказанное				Выходной слой balance
		Скрытый слой 1				
		H(1:1)	H(1:2)	H(1:3)	H(1:4)	
Входной слой	(Смещение)	,823	-1,214	-,136	-,285	
	countall	4,117	-5,913	-1,549	-1,817	
	price	-,041	1,159	,032	-,172	
Скрытый слой 1	(Смещение)					6,153
	H(1:1)					-3,569
	H(1:2)					-3,418
	H(1:3)					-3,831
	H(1:4)					-3,162

Рисунок 10. Результат обучения нейронной сети: отчет об ошибке и смещение каждого обучающего нейрона.

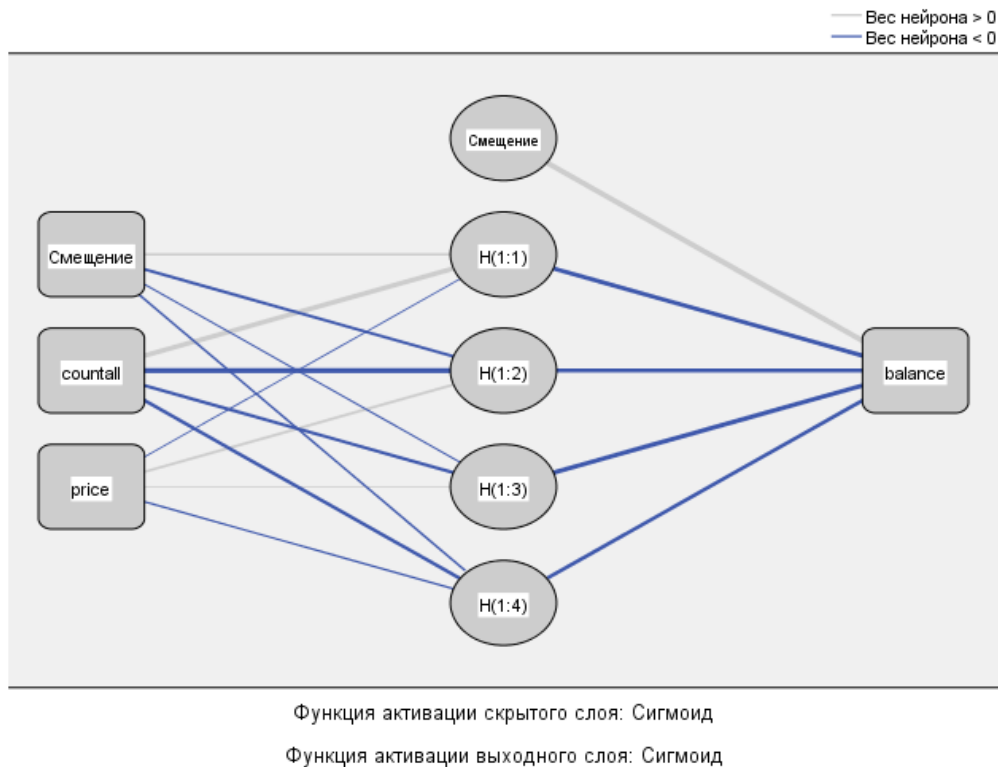


Рисунок 11. Схема работы функции активации сигмоид.

**Вывод**

В данной статье было рассмотрено применение регрессионного анализа и нейронной сети к данным о товарах в предметной области «Магазин компьютерной техники».

#### БИБЛИОГРАФИЧЕСКИЙ СПИСОК

1. Козич В.Г., Бондаренко В.В., Баженов Р.И. Исследование с помощью интеллектуальных методов программы SPSS рынка подержанных автомобилей Toyota Prius // Современные научные исследования и инновации. 2016. № 6 [Электронный ресурс]. URL: <http://web.snauka.ru/issues/2016/06/69138> (дата обращения: 03.02.2020)
2. Елисеева И.И., Юзбашев М.М. Общая теория статистики-2005. УДК 311(075ю8)