

Решение задачи прогнозирования с помощью линейной регрессии

Лекция 12 (1 час)

Емельянова М.Г.

Понятие линейной регрессии

Линейная регрессия представляет собой модель зависимости между входными и выходными переменными с линейной функцией связи.

Множественной называют линейную регрессию, в модели которой число независимых переменных две или более.

Уравнение множественной линейной регрессии имеет вид:

$$Y = b_0 + b_1x_1 + b_2x_2 + \dots + b_nx_n$$

Поиск $b_0, b_1, b_2, \dots, b_n!$

<https://wiki.loginom.ru/articles/linear-regression.html>

<https://wiki.loginom.ru/articles/multiple-linear-regression.html>

Примеры применения

Анализ эластичности спроса по цене, характеризующей реакцию потребительского спроса на изменение цены товара. Обработчик позволяет построить модель продаж, где в качестве входной переменной будет использоваться цена, а в качестве выходной – объем продаж.

Прогнозирование объёма продаж. Регрессия строится на основе временного ряда продаж за репрезентативный период. Данная модель является базисом для формирования оптимального плана закупок и товарно-финансовых планов.

Прогнозирование стоимости ценных бумаг. Регрессионная модель строится на основе таких показателей, как чистая прибыль компании, доход, рентабельность выручки компании, балансовой стоимости и прочее.

Прогнозирование загруженности веб-сервиса. Резкие всплески интереса могут повысить нагрузку на серверы и отрицательно сказаться на качестве работы веб-сервиса. Эта задача особенно актуальная, если вычислительные ресурсы арендуются в облаке, где имеется возможность гибко управлять мощностями доступных серверов.

Пример решения задачи прогнозирования с помощью линейной регрессии

Набор данных Trade.txt.

При таком наборе данных, когда значения представлены в одном столбце, необходима трансформация к скользящему окну.

Запуск обработчика «Линейная регрессия» осуществляется после применения обработчика «Скользящее окно» (см. лабораторную работу 5, лекцию 11).

Пример решения задачи прогнозирования с помощью линейной регрессии

На первом шаге задаём назначение исходных столбцов. Предположим, что на прогноз влияет информация за 3 прошлых месяца, тогда укажем входными столбцами поля: «Количество – 3», «Количество – 2», и «Количество – 1».

В качестве выходного поля укажем столбец «Количество». Остальные поля сделаем информационными (рисунок 2).

The screenshot shows a software window titled "Мастер обработки - Линейная регрессия (2 из 8)". The main heading is "Настройка назначений столбцов" (Column Assignment Configuration) with the instruction "Задайте назначения исходных столбцов данных" (Specify the assignments of the original data columns). A list of columns is on the left, with "Количество" (Quantity) selected. On the right, the configuration for the selected column is shown: "Имя столбца" (Column Name) is COL2, "Тип данных" (Data Type) is Вещественный (Real), "Назначение" (Assignment) is Выходное (Output), and "Вид данных" (Data View) is Непрерывный (Continuous). A "Статистика" (Statistics) section displays: Minimum: 1075589,03358277; Maximum: 2380312,1698561; Average: 1864843,52986291; Standard Deviation: 354553,479266148. At the bottom, there are buttons for "< Назад" (Back), "Далее >" (Next), and "Отмена" (Cancel).

Column Name	Type	Assignment	View
Количество	Вещественный	Выходное	Непрерывный

Statistic	Value
Минимум	1075589,03358277
Максимум	2380312,1698561
Среднее	1864843,52986291
Стандартное откл.	354553,479266148

Пример решения задачи прогнозирования с помощью линейной регрессии

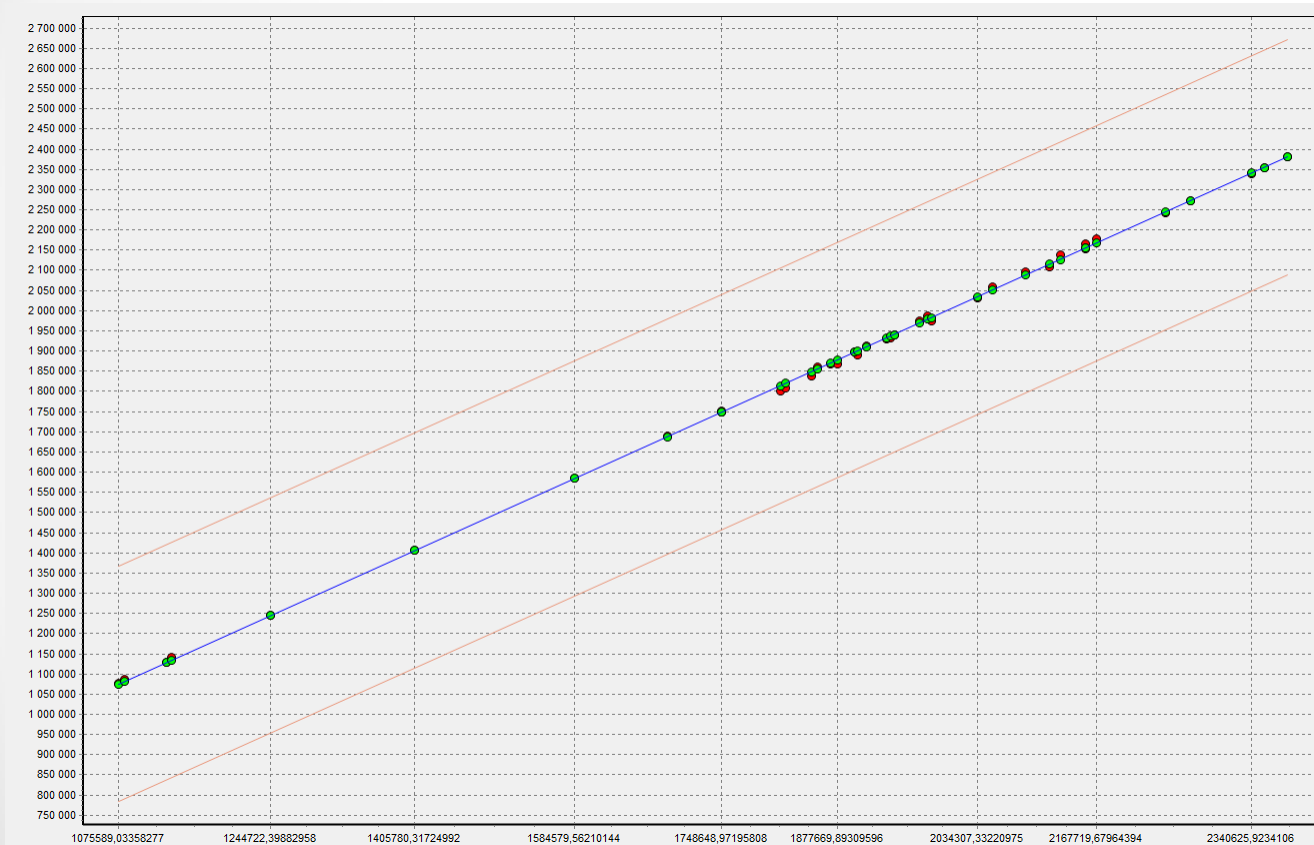
Далее происходит настройка обучающего и тестового множеств, ограничение диапазона входных значений (не изменяем).

Далее – запуск процесса обучения.

После выполнения процесса – отображение визуализаторов «Коэффициенты регрессии», «Отчёт по регрессии», «Диаграмма рассеяния», «Что-если».

Пример решения задачи прогнозирования с помощью линейной регрессии

Визуализатор «Диаграмма рассеяния».



Пример решения задачи прогнозирования с помощью линейной регрессии

Фрагмент визуализатора «Отчёт по регрессии».

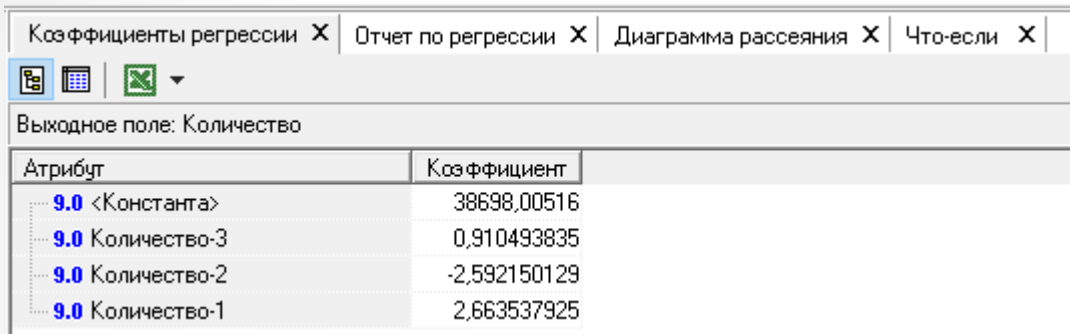
Регрессия "1"					
Множ. коэффициент корреляции, R	Коэффициент детерминации, R ²	Скоррект. коэффициент детерминации	Стандартное отклонение	Размер выборки	Метод отбора переменных
0,9998	0,9997	0,9996	6782,9458	36	Полное включение

Если коэффициент детерминации близок к 1, это указывает на то, что модель работает очень хорошо (имеет высокую значимость), а если к 0, то это означает низкую значимость модели.

<https://wiki.loginom.ru/articles/coefficient-of-determination.html>

Пример решения задачи прогнозирования с помощью линейной регрессии

Визуализатор «Коэффициенты регрессии».



Атрибут	Коэффициент
9.0 <Константа>	38698,00516
9.0 Количество-3	0,910493835
9.0 Количество-2	-2,592150129
9.0 Количество-1	2,663537925

Полученное уравнение регрессии имеет вид:

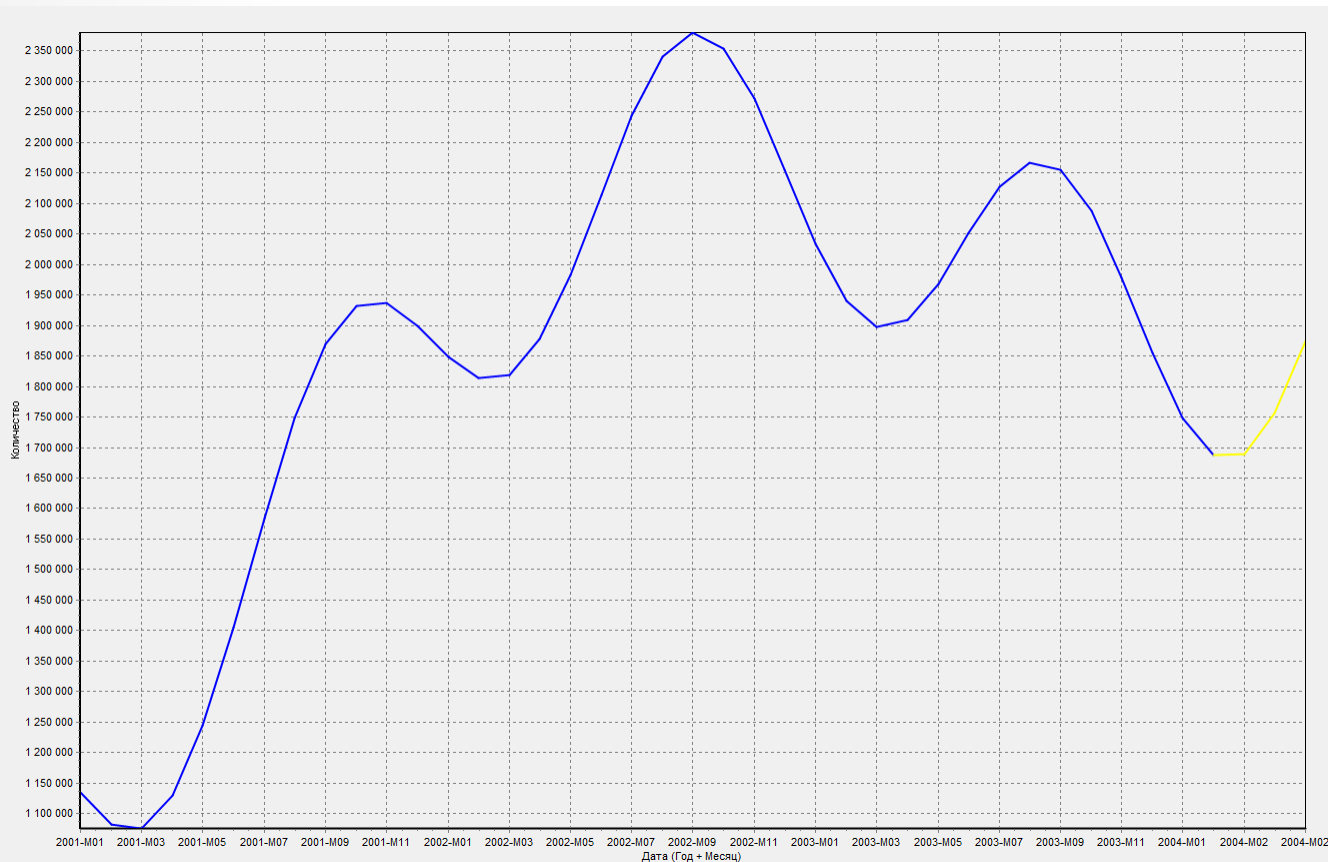
$$Y=38698,00516+0,910493835x_1 -2,592150129x_2+2,663537925x_3.$$

Пример решения задачи прогнозирования с помощью линейной регрессии

Обработчик «Прогнозирование».

Например, горизонт прогноза 3.

Визуализатор «Диаграмма прогноза».



Вопросы для проверки

1. Что собой представляет линейная регрессия?
2. Какой вид имеет множественное уравнение регрессии?
3. Приведите примеры использования линейной регрессии.
4. Что показывает коэффициент детерминации?
5. Для чего используются визуализаторы «Коэффициенты регрессии», «Отчёт по регрессии», «Диаграмма рассеяния»?
6. Для чего используются диаграмма прогноза?