

Московский авиационный институт  
(национальный исследовательский университет)

Факультет информационных технологий и прикладной  
математики

Кафедра вычислительной математики и программирования

Лабораторная работа №1 по курсу «Искусственный интеллект»

Студент: В. И. Лобов  
Преподаватель: С. Х. Ахмед  
Группа: М8О-306Б  
Дата:  
Оценка:  
Подпись:

Москва, 2019

# Лабораторная работа №1

**Задача:** Ваша задача познакомиться с платформой Azure Machine Learning, реализовав полный цикл разработки решения задачи машинного обучения, используя три различных алгоритма, реализованные на этой платформе.

**Требования:**

1. Уникальность решения
2. Обоснованность выбора той или иной операции
3. В отчете должны быть указаны алгоритмы, которые применялись, результаты применения этих алгоритмов, а также скрины некоторых этапов обработки данных

**Выбранные алгоритмы:**

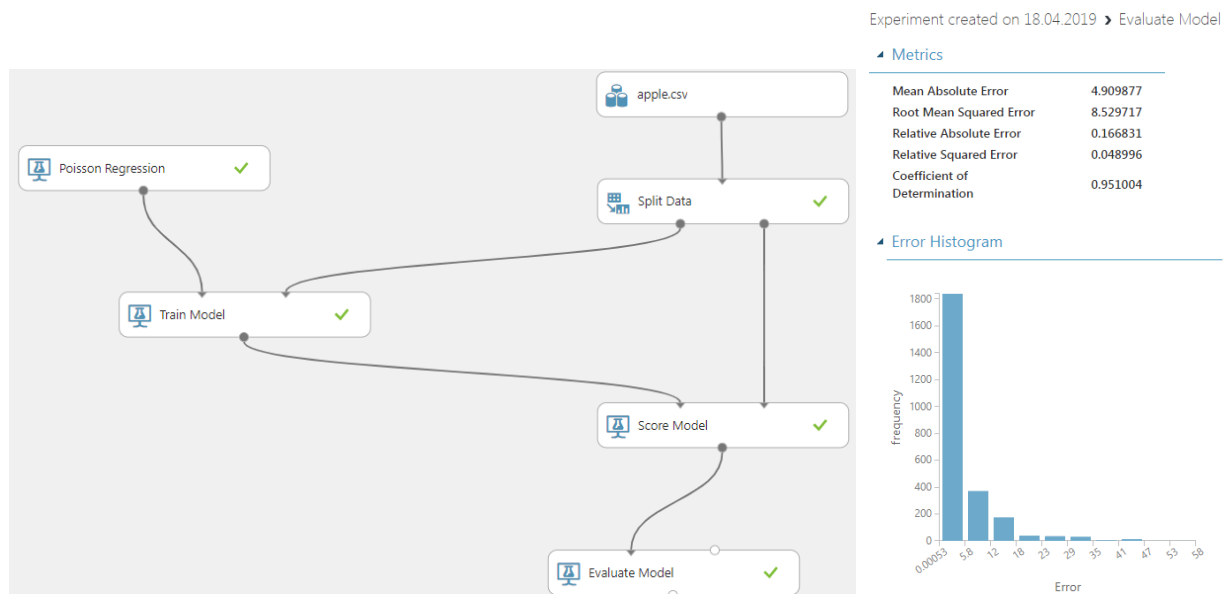
1. Poisson Regression
2. Linear Regression
3. Neural Network Regression

# 1 Описание

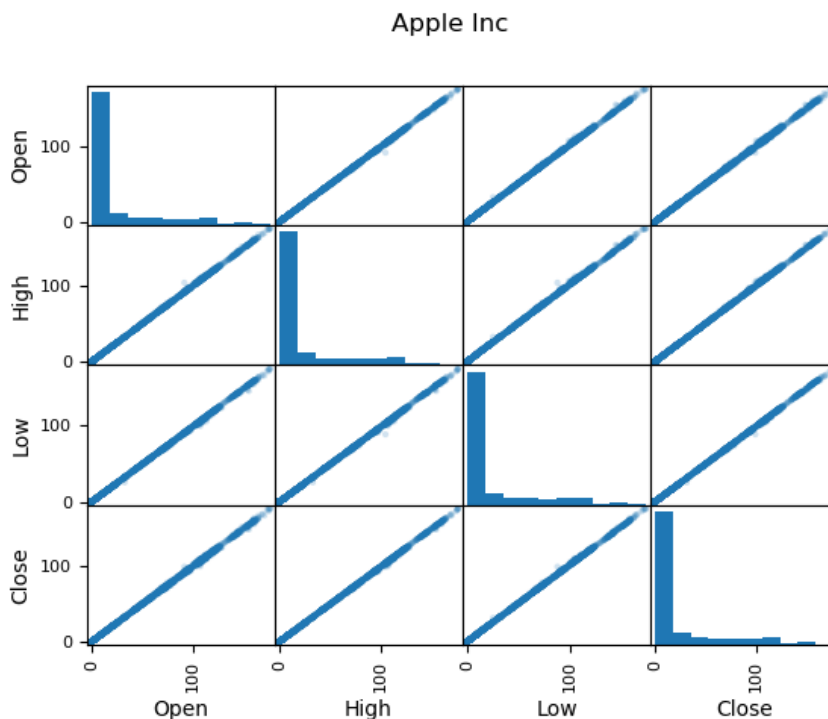
Имея на руках датасет с котировками акций американских компаний, попробуем смоделировать цену акций Apple на момент открытия биржи. Рассмотрим 3 различных модели регрессии и проанализируем, какие из них лучше всего справляются с задачей. Разделим данные котировок в отношении 7:3, при этом обучим модель на 70% данных.

## Poisson Regression.

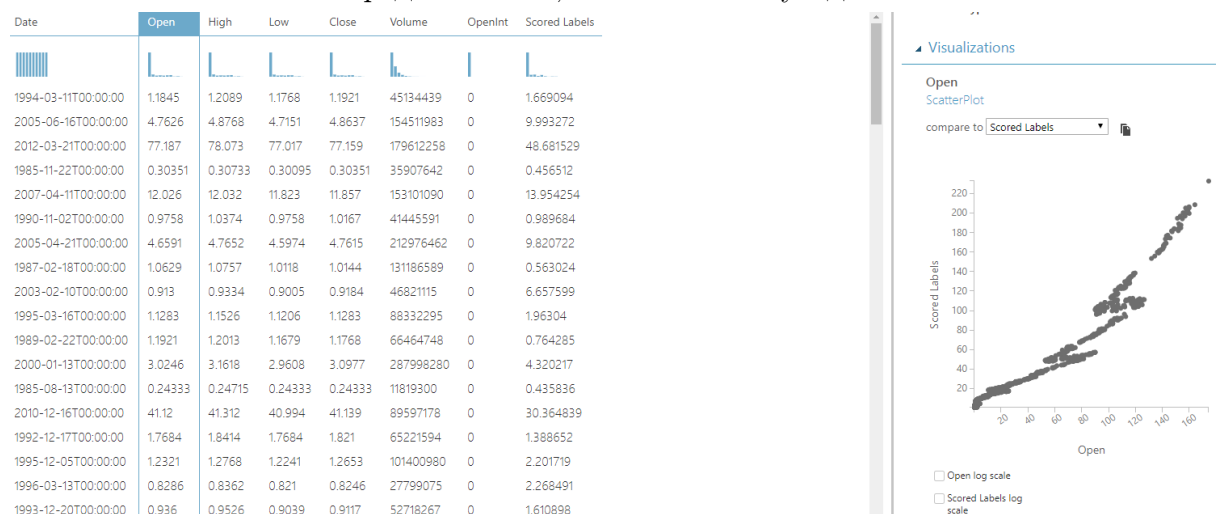
Пуассоновская регрессия - обобщённая форма модели линейной регрессии. Предполагается, что оцениваемая величина  $Y$  распределена по Пуассоновскому закону ( $Y \sim P(\lambda)$ ) и что логарифм от математического ожидания  $Y$  может быть представлен в виде линейной комбинации некоторых параметров  $\theta$ , поэтому она также называется линеарифметической:  $\log(E(Y|x)) = \log(exposure) + \theta x$ .



Распределение ошибки похоже на Пуассоновское, и сама величина ошибки достаточно велика. Взглянув на распределение признака Open всего датасета, видим:



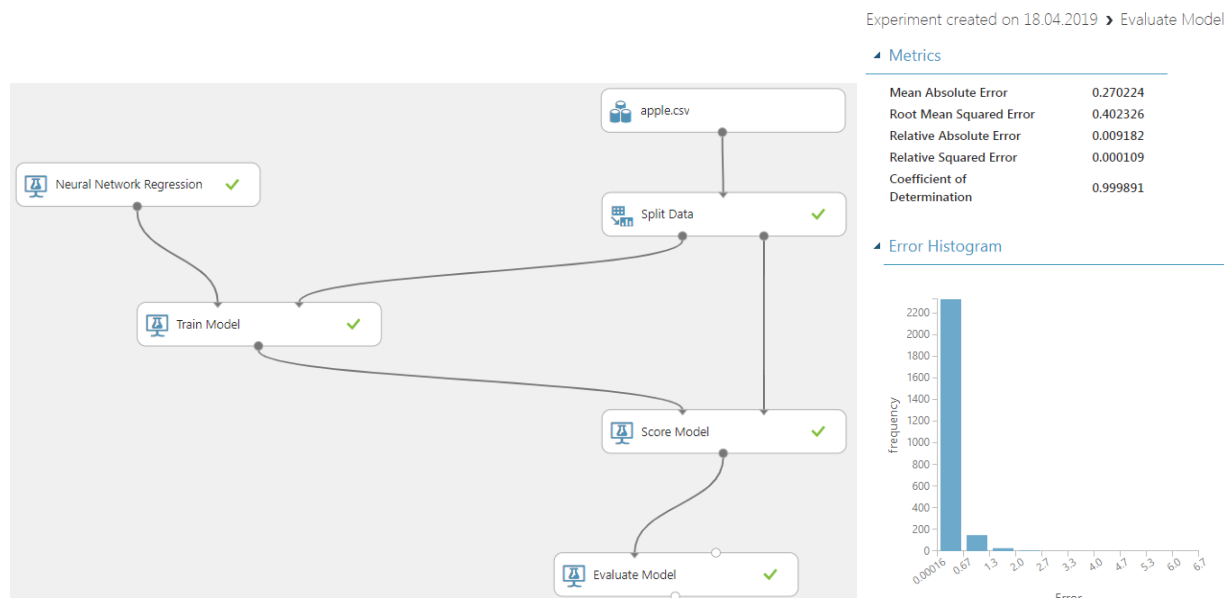
Распределение не похоже на Пуассоновское, следовательно, в данном случае Пуассоновская не позволяет добиться высокой точности. Построив зависимость предсказываемой величины от предсказанной, можно в этом убедиться:



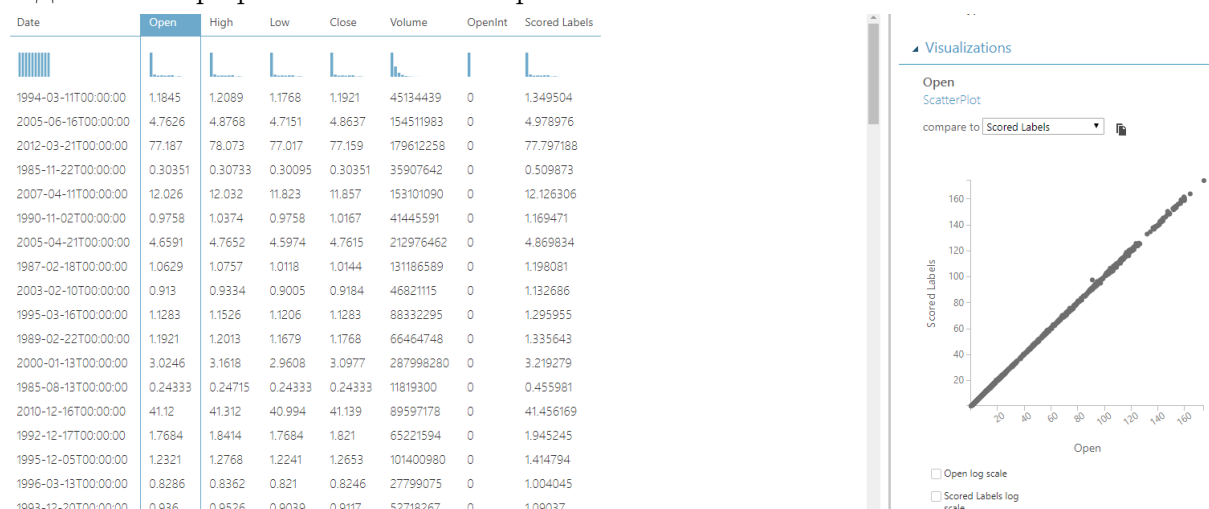
## Neural Network Regression.

Регрессия на основе нейронных сетей - адаптивное обучение нейронных сетей, решающих сложные задачи глубокого обучения или распознавания изображений, к задачам регрессии. Любой класс статистической модели может быть представлен в виде нейронной сети, если она использует адаптивные веса и может аппроксимировать нелинейные

функции на входе. Эта модель может помочь тогда, когда традиционные алгоритмы регрессии не справляются с задачей.



В отличие от Пуассоновской регрессии, ошибка значительно меньше, что подтверждается на графике зависимости признаков:



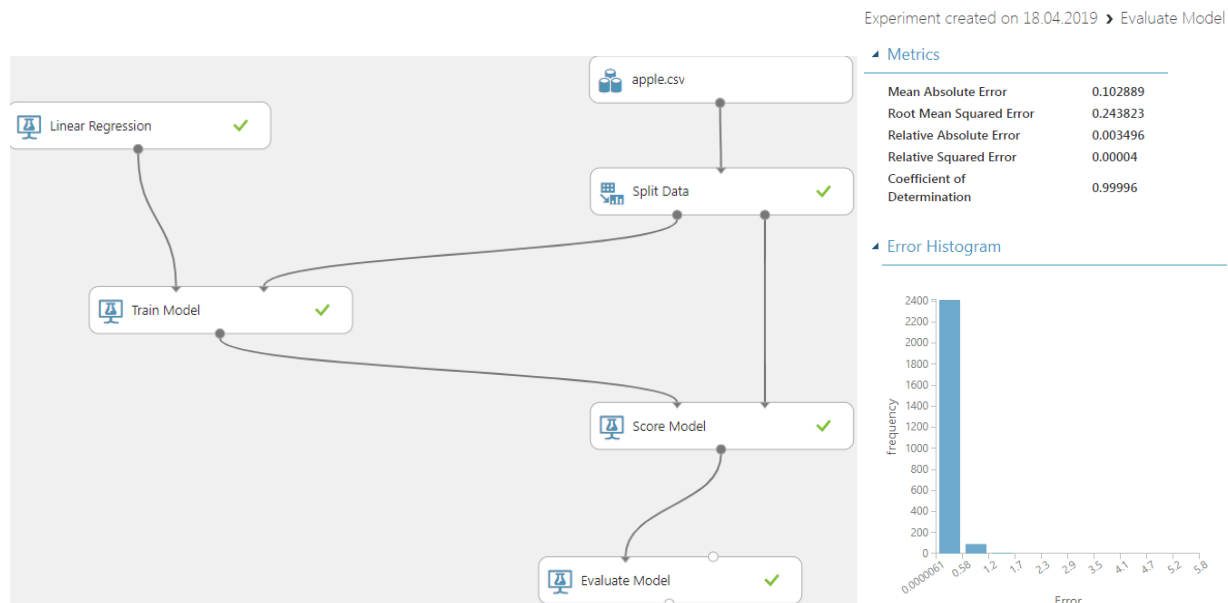
Тем не менее, среднее отклонение составляет 0.27\$, а это около 17,26 рублей по текущему курсу. Погрешность достаточно высока, но хотя бы не Пуассон.

## Linear Regression.

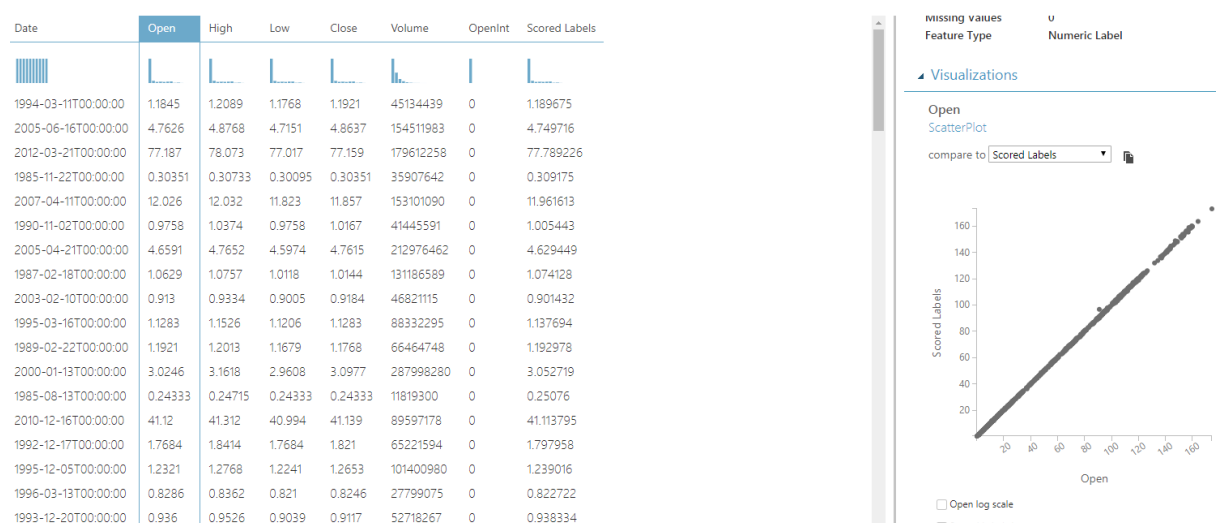
Линейная регрессия - используемая в статистике регрессионная модель зависимости одной (объясняемой, зависимой) переменной  $y$  от другой или нескольких других переменных (факторов, регрессоров, независимых переменных)  $x$  с линейной функцией зависимости. Модель линейной регрессии является часто используемой и наиболее

изученной в эконометрике (наука, изучающая количественные и качественные экономические взаимосвязи с помощью математических и статистических методов и моделей). С эконометрической точки зрения более важное значение имеет линейность по параметрам, чем линейность по факторам модели.

В случае акций Apple можно сделать вывод, что данная модель подходит лучше всего, так как необходимые нам признаки линейны относительно друг друга. Построим данную модель:



В сравнении с нейронной сетью, ошибка стала более чем в 2 раза меньше: 0.1\$. Благодаря этому график зависимости предсказываемой величины от предсказанной будет похож на линейную функцию с углом наклона  $\frac{\pi}{4}$ :



Сравнив работу 3-х различных моделей регрессии, можно сделать вывод, что наиболее подходящей для решения задач является именно линейная модель. К сожалению, это утверждение верно не для всех котировок акций. В большинстве случаев зависимость признаков друг от друга линейная, но это не есть правило. Таким образом, задача предсказания котировок акций достаточно сложна и характеризуется не только данными о предыдущих ценах, но и другими зависимостями.