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:

Лабораторная работа №1

Задача: Ваша задача познакомиться с платформой Azure Machine Learning, реализовав полный цикл разработки решения задачи машинного обучения, использовав 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:

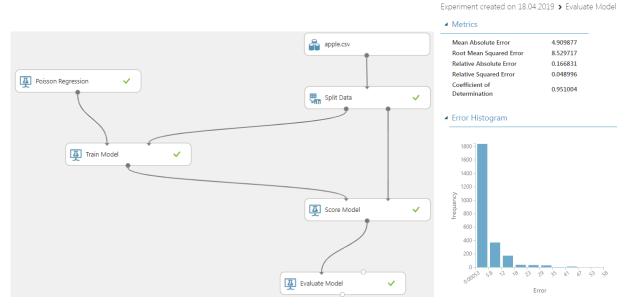
- 1. Poisson Regression
- 2. Linear Regression
- 3. Neural Network Regression

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.

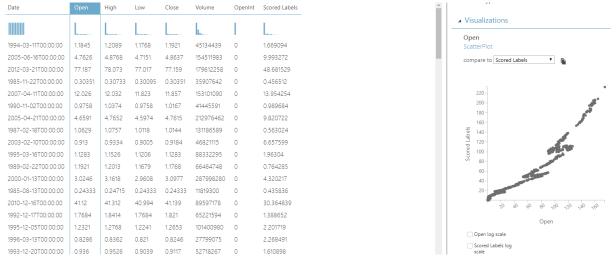
Пуассоновская регрессия - обобщённая форма модели линейной регрессии. Предполагается, что оцениваемая величина Y распределена по Пуассоновскому закону $(Y|P(\lambda))$ и что логарифм от математического ожидания Y может быть представлен в виде линейной комбинации некоторых параметров θ , поэтому она также 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 Ореп всего датасета, видим:

Apple Inc

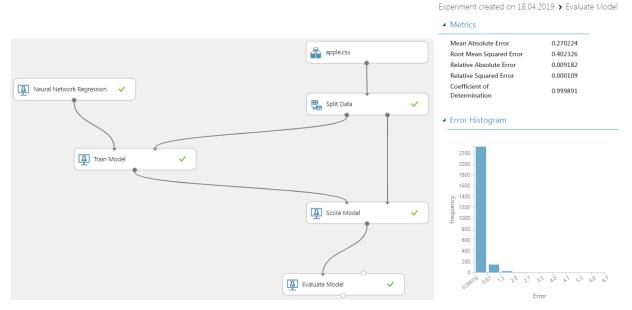
Распределение не похоже на Пуассоновское, следовательно, в данном случае Пуассоновская не позволяет добиться высокой точности. Построив зависимость предсказываемой величины от предсказанной, можно в этом убедиться:



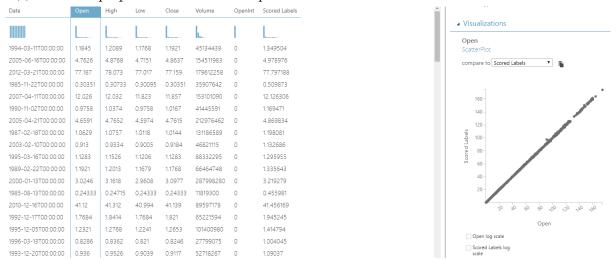
Neural Network Regression.

Регрессия на основе нейронных сетей - адаптирование 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 апроксимировать 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:



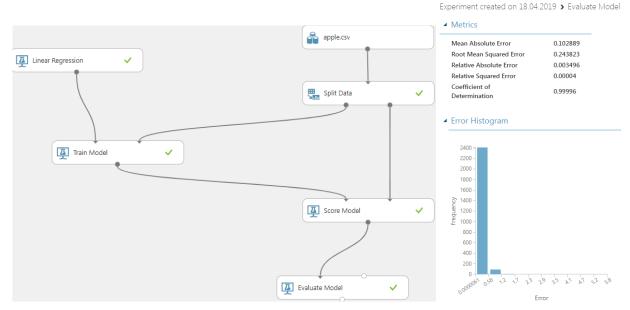
Тем не менее, среднее отклонение составляет 0.27\$, а это около 17,26 рублей по текущему курсу. Погрешность достаточно высока, но хотя бы не Пуассон.

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

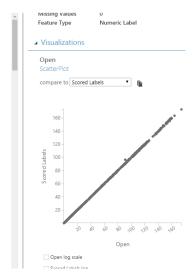
изученной в эконометрике (наука, изучающая количественные и качественные экономические взаимосвязи с помощью математических и статистических методов и моделей). С эконометрической точки зрения более важное значение имеет линейность по параметрам, чем линейность по факторам модели.

В случае акций Apple можно сделать вывод, что данная модель подходит лучше всего, так как необходимые нам признаки линейны отонсительно 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 $\frac{\pi}{4}$:

| Date | Open | High | Low | Close | Volume | OpenInt | Scored Labels |
|---------------------|---------|---------|---------|---------|-----------|---------|---------------|
| | <u></u> | <u></u> | | | l. | | |
| 1994-03-11T00:00:00 | 1.1845 | 1.2089 | 1.1768 | 1.1921 | 45134439 | 0 | 1.189675 |
| 2005-06-16T00:00:00 | 4.7626 | 4.8768 | 4.7151 | 4.8637 | 154511983 | 0 | 4.749716 |
| 2012-03-21T00:00:00 | 77.187 | 78.073 | 77.017 | 77.159 | 179612258 | 0 | 77.789226 |
| 1985-11-22T00:00:00 | 0.30351 | 0.30733 | 0.30095 | 0.30351 | 35907642 | 0 | 0.309175 |
| 2007-04-11T00:00:00 | 12.026 | 12.032 | 11.823 | 11.857 | 153101090 | 0 | 11.961613 |
| 1990-11-02T00:00:00 | 0.9758 | 1.0374 | 0.9758 | 1.0167 | 41445591 | 0 | 1.005443 |
| 2005-04-21T00:00:00 | 4.6591 | 4.7652 | 4.5974 | 4.7615 | 212976462 | 0 | 4.629449 |
| 1987-02-18T00:00:00 | 1.0629 | 1.0757 | 1.0118 | 1.0144 | 131186589 | 0 | 1.074128 |
| 2003-02-10T00:00:00 | 0.913 | 0.9334 | 0.9005 | 0.9184 | 46821115 | 0 | 0.901432 |
| 1995-03-16T00:00:00 | 1.1283 | 1.1526 | 1.1206 | 1.1283 | 88332295 | 0 | 1.137694 |
| 1989-02-22T00:00:00 | 1.1921 | 1.2013 | 1.1679 | 1.1768 | 66464748 | 0 | 1.192978 |
| 2000-01-13T00:00:00 | 3.0246 | 3.1618 | 2.9608 | 3.0977 | 287998280 | 0 | 3.052719 |
| 1985-08-13T00:00:00 | 0.24333 | 0.24715 | 0.24333 | 0.24333 | 11819300 | 0 | 0.25076 |
| 2010-12-16T00:00:00 | 41.12 | 41.312 | 40.994 | 41.139 | 89597178 | 0 | 41.113795 |
| 1992-12-17T00:00:00 | 1.7684 | 1.8414 | 1.7684 | 1.821 | 65221594 | 0 | 1.797958 |
| 1995-12-05T00:00:00 | 1.2321 | 1.2768 | 1.2241 | 1.2653 | 101400980 | 0 | 1.239016 |
| 1996-03-13T00:00:00 | 0.8286 | 0.8362 | 0.821 | 0.8246 | 27799075 | 0 | 0.822722 |
| 1993-12-20T00:00:00 | 0.936 | 0.9526 | 0.9039 | 0.9117 | 52718267 | 0 | 0.938334 |



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.