**МИНИСТЕРСТВО ОБРАЗОВАНИЯ И НАУКИ РФ**

**Московский государственный технический университет**

**им. Н.Э. Баумана**

**(МГТУ им. Н.Э. Баумана)**

**Кафедра «Информационная безопасность» (ИУ8)**

Лабораторная работа № 1.2

По дисциплине: «Машинное обучение»

Тема: «Множественная линейная регрессия на чистых данных»

Выполнил: Веденеев А.А.,

Студент группы ИУ8-92

Проверила: Коннова Н.С.,

Преподаватель каф. ИУ8

г. Москва 2024 г.

# Практическая часть лабораторной работы 1.2

**Цель работы:** познакомиться с основными приемами работы с множественными регрессионными моделями с использованием библиотеки sklearn.

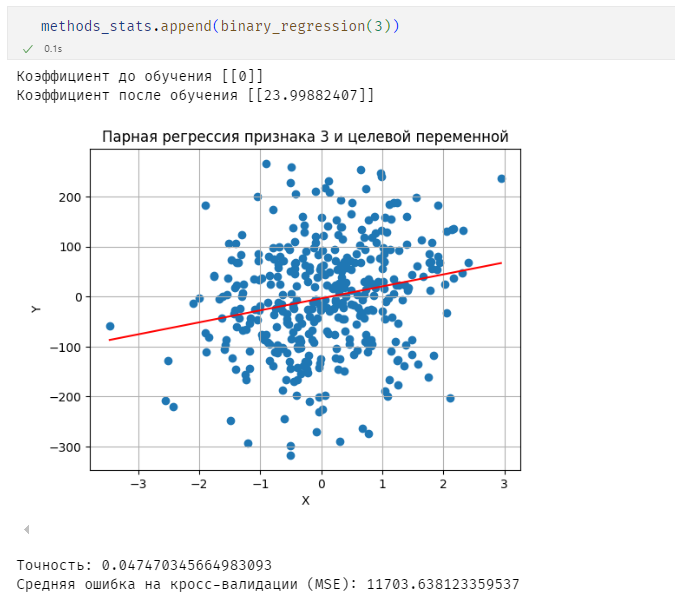
**Ход работы:**

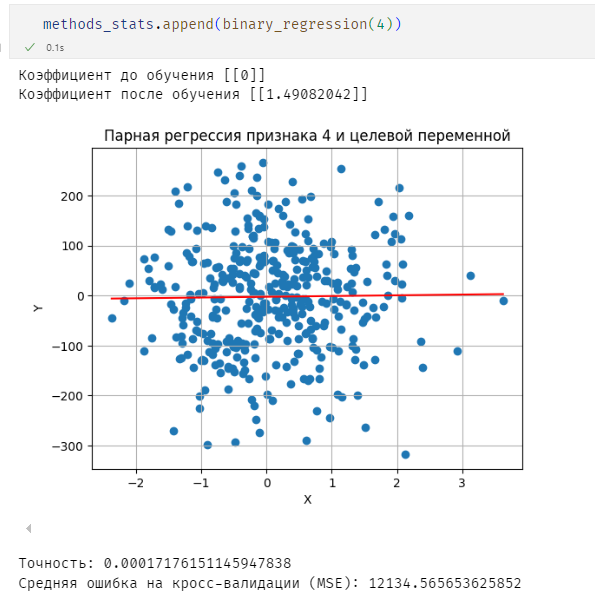




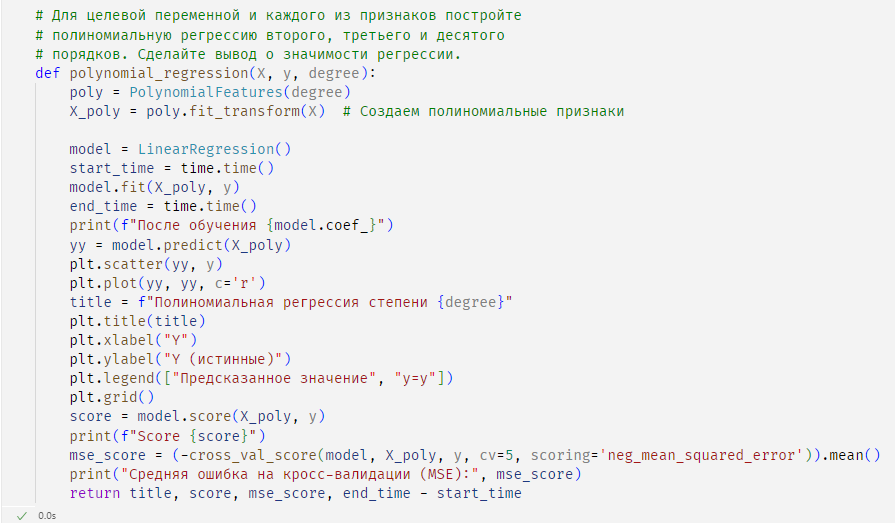
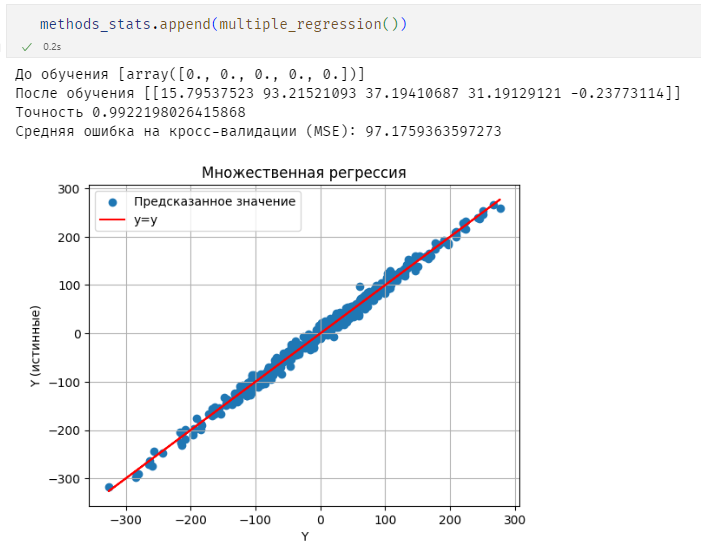
A graph with blue dots and red line

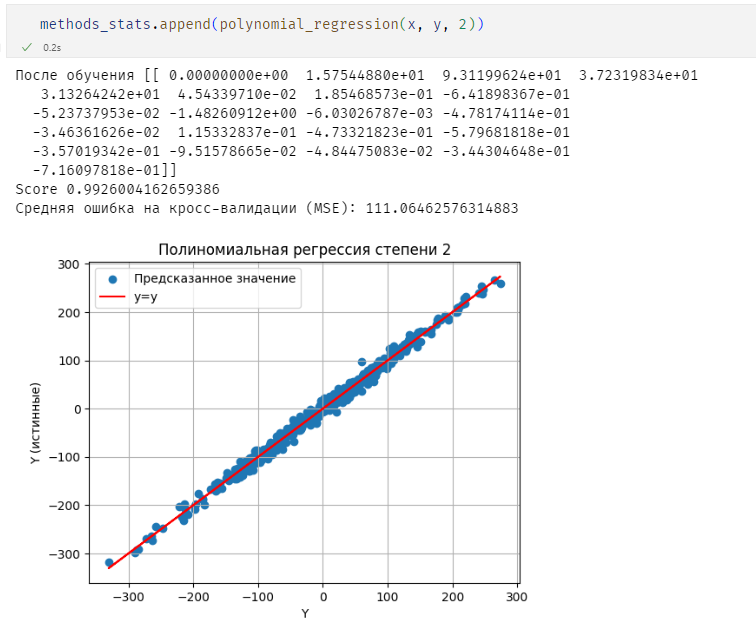
Description automatically generatedA screen shot of a graph

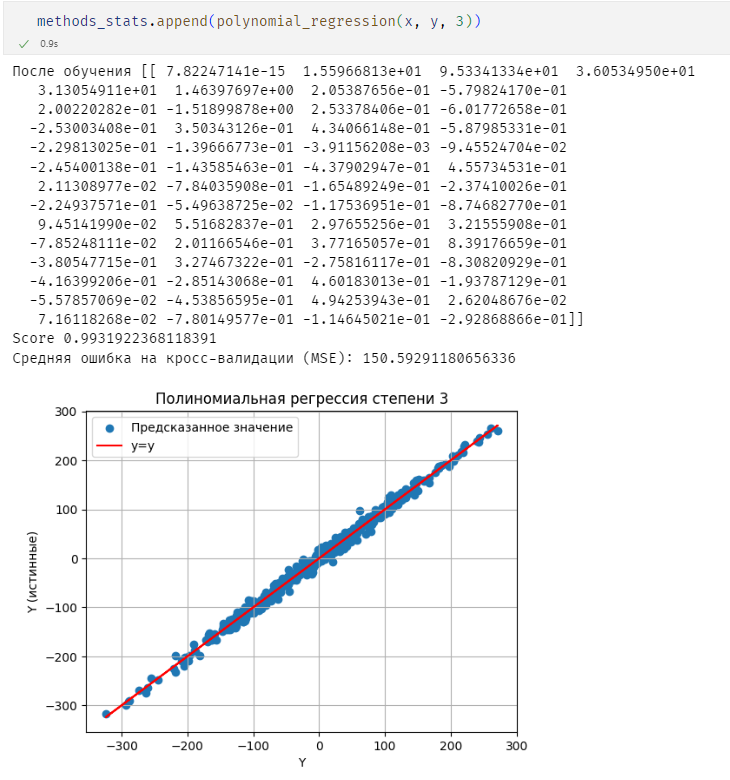
Description automatically generated

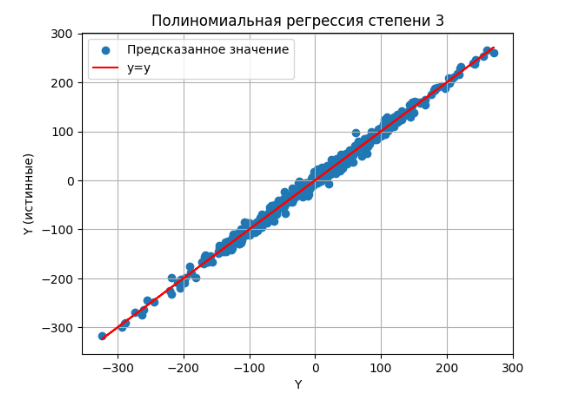
A screen shot of a graph

Description automatically generated









A graph with a line

Description automatically generated

A screenshot of a computer program

Description automatically generated

## Выводы

В рамках работы была выполнена задача построения предиктивной модели для прогнозирования целевой переменной на основе нескольких факторов. Были загружены и изучены исходные данные, проведен их количественный анализ, включая описание количества точек данных, признаков и базовой статистики.

Для каждого признака была построена парная регрессионная модель, что позволило визуально и количественно оценить связь между целевой переменной и отдельными факторами. На следующем этапе была построена модель множественной линейной регрессии, которая учитывала все признаки одновременно.

Анализ модели показал удовлетворительное качество регрессии и подтвердил значимость ряда факторов для прогнозирования целевой переменной. Работа продемонстрировала важность предварительного анализа данных и этапа визуализации для успешного построения и оценки модели.

Контрольные вопросы

1. Как должны выглядеть данные для модели машинного обучения?

Данные для задачи множественной регрессии должны быть в табличной форме, где каждая строка представляет один наблюдаемый объект (образец), а каждый столбец — это переменная (признак или целевая переменная).

Требования к данным:

* Признаки (фичи): несколько числовых или категориальных переменных, которые используются для предсказания целевой переменной. Например, возраст, доход, уровень образования и т.д.
* Целевая переменная: числовая переменная, которую модель должна предсказать.
* Данные должны быть очищены от пропусков или их следует заполнить.
* Признаки должны быть нормализованы или стандартизированы, если их масштабы сильно различаются.
* Признаки должны быть релевантны целевой переменной и коррелировать с ней в разумных пределах.

1. В чем физический смысл коэфициентов линейной регрессии?

Коэффициенты линейной регрессии показывают, как изменение каждого признака (фактора) влияет на целевую переменную.

Физический смысл:

1. Коэффициент при переменной: Показывает, на сколько изменится целевая переменная при изменении данной переменной на единицу, если остальные переменные остаются неизменными.

2. Знак коэффициента:

• Положительный — переменная и целевая переменная связаны прямо (рост переменной приводит к росту целевой переменной).

• Отрицательный — переменная и целевая переменная связаны обратно (рост переменной приводит к снижению целевой переменной).

1. Чем множественная регрессия отличается от парной?

Множественная регрессия использует несколько признаков для предсказания целевой переменной.

Парная регрессия использует только один признак для предсказания целевой переменной.

1. Какую функцию ошибки использует модель линейной регрессии?

Модель линейной регрессии использует среднеквадратичную ошибку (MSE) в качестве функции ошибки. MSE минимизирует квадраты отклонений, что позволяет лучше учитывать большие ошибки и повышает точность модели.

1. Какие метрики эффективности можно применять для оценки регрессионных моделей?

* MSE — среднеквадратичная ошибка.
* RMSE — квадратный корень из MSE.
* MAE — средняя абсолютная ошибка.
* R² — доля вариации, объяснённая моделью.
* MAPE — средняя абсолютная процентная ошибка.
* Huber loss — устойчивость к выбросам.

1. Какие еще регрессоры есть в библиотеке sklearn?

* LinearRegression — линейная регрессия.
* Ridge, Lasso, ElasticNet — линейные модели с регуляризацией.
* DecisionTreeRegressor — регрессия на основе деревьев решений.
* RandomForestRegressor — ансамбль деревьев.
* SVR — регрессия с опорными векторами.
* KNeighborsRegressor — метод ближайших соседей.
* GradientBoostingRegressor — бустинг деревьев.
* MLPRegressor — многослойный перцептрон.

1. Почему для визуализации модели нельзя использовать предсказанные значения по датасету?

Использовать предсказанные значения для визуализации модели неправильно, так как модель могла переобучиться на этом датасете. Это приводит к завышенной оценке её качества и не отражает способности обобщать данные на новых примерах.