Министерство науки и высшего образования Российской Федерации

Федеральное государственное автономное образовательное учреждение высшего образования

«СЕВЕРО-КАВКАЗСКИЙ ФЕДЕРАЛЬНЫЙ УНИВЕРСИТЕТ»

|  |  |
| --- | --- |
| Институт | математики и компьютерных наук |
| Кафедра | компьютерной безопасности |

ОТЧЕТ

ЛАБОРАТОРНАЯ РАБОТА 8.

ПОЛИНОМИАЛЬНАЯ РЕГРЕССИЯ

Выполнил: Окунев Николай Александрович,

студент 2 курса

группы КМБ-с-о-23-1

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

«\_\_\_\_» \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_2025 г.

Проверено с оценкой:

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

«\_\_\_\_» \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_2025 г.

Ставрополь, 2025

1. **Цели и задачи**

Цель лабораторной работы: научиться применять разработанный пайплайн для тиражирования кода с целью решения задачи полиномиальной регрессии. Основные задачи: – получение навыков рефакторинга кода в проектах машинного обучения; – изучение поведения модели полиномиальной регрессии при изменении степени полинома; – исследование свойств набора данных в рамках задачи полиномиальной регрессии.

1. **Теоретическое обоснование**

Линейная и параболическая модели являются частными случаями более сложной модели – полиномиальной. Построить модель регрессии – это значит найти параметры той функции, которая будет в ней фигурировать. Для линейной регрессии – два параметра: коэффициент и свободный член. Полиномиальная регрессия может применяться в математической статистике при моделировании трендовых составляющих временных рядов. Временной ряд – это, по сути, ряд чисел, которые зависят от времени. Например, средние значения температуры воздуха по дням за прошедший год, или доход предприятия по месяцам. Порядок моделируемого полинома оценивается специальными методами, например, критерием серий. Цель построения модели полиномиальной регрессии в области временных рядов всё та же – прогнозирование.

1. **Методика и порядок выполнения работы**

Перед выполнением индивидуального задания рекомендуется выполнить все пункты учебной задачи.

* 1. Учебная задача Задание. На основе разработанного пайплайна для линейной одномерной регрессии разработать полиномиальную модель регрессии. Решение. Для разработки модели необходимо реализовать следующий код:

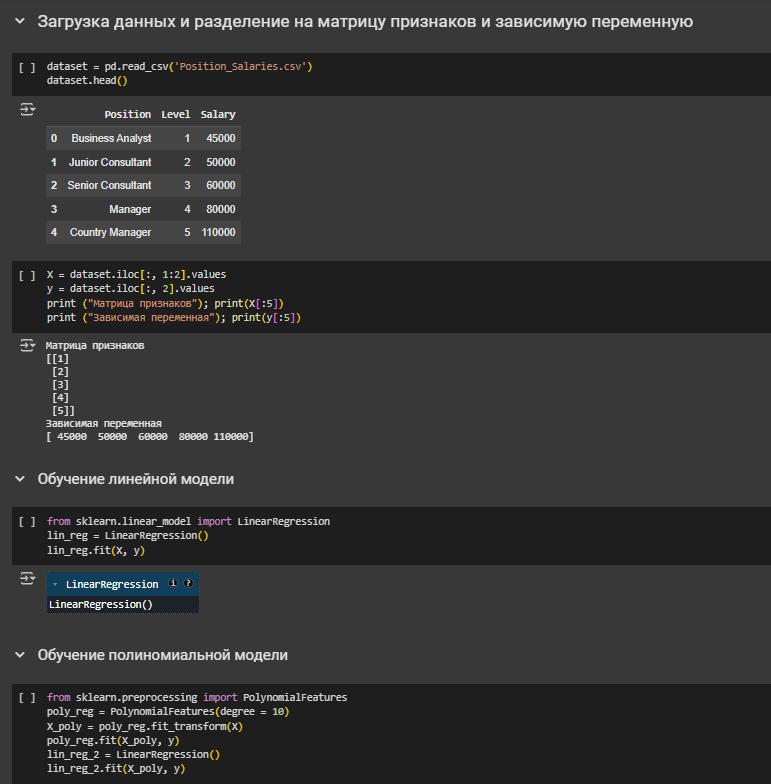


Рисунок 8.1 – Код Python для построения модели полиномиальной регрессии

Рассмотрим как изменяется модель при изменении степени апроксимирующего полинома (рисунок 8.2). n = 15 n = 2 n = 3 n = 6

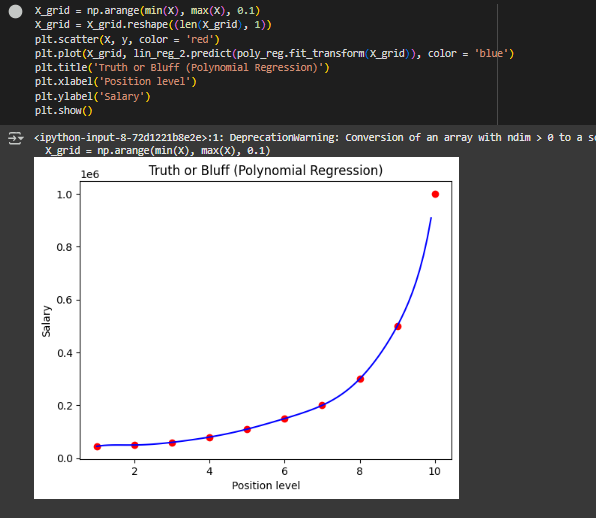
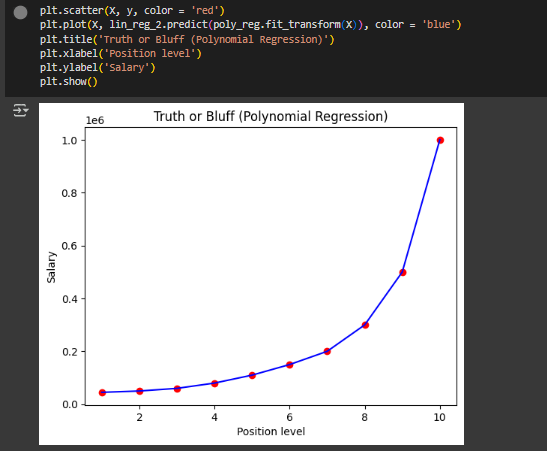
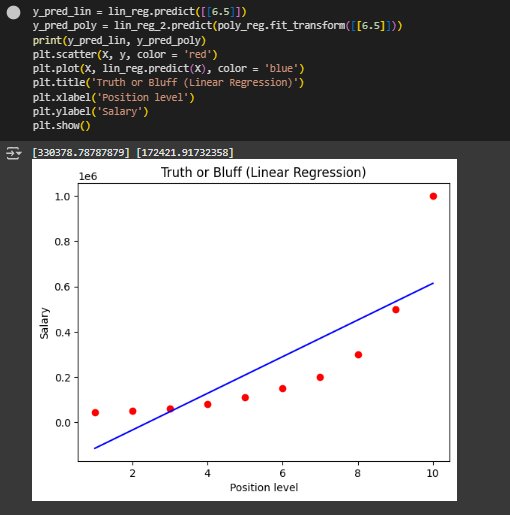


Рисунок 8.2 – Анализ модели полиномиальной регрессии

Очевидно, что при реализации полиномиальной регрессии нет необходимости в немотивированном увеличении степени аппроксимации.

**5. Контрольные вопросы**

1)Почему при реализации многомерной линейной регрессии необходимо добавить фиктивный признак с единственным значением 1.0?

Фиктивный признак (также называемый "intercept" или "bias term") добавляется для того, чтобы модель могла иметь ненулевое смещение. Без фиктивного признака модель будет всегда проходить через начало координат (0, 0, ..., 0), что может быть неоптимально. Фиктивный признак позволяет модели "сдвигать" линию регрессии, чтобы лучше соответствовать данным.

2)Что такое фиктивная переменная? Поясните причину удаления одной фиктивной переменной, возникающей при перекодировке категориального признака.

Фиктивная переменная (dummy variable) — это бинарная переменная (0 или 1), которая используется для представления категориального признака в модели регрессии. При перекодировке категориального признака с k категориями создается k фиктивных переменных. Однако, для избежания мультиколлинеарности (линейной зависимости между признаками) необходимо удалить одну из фиктивных переменных. Причина удаления: Если мы оставим все k фиктивных переменных, то сумма этих переменных будет всегда равна 1. Это означает, что одна из переменных может быть выражена через остальные, что приводит к мультиколлинеарности. Мультиколлинеарность может привести к нестабильности оценок коэффициентов модели и затруднить их интерпретацию. Удаление одной фиктивной переменной делает оставшиеся переменные линейно независимыми и позволяет получить корректные оценки коэффициентов.

3)С использованием какого класса создается модель полиномиальной регрессии?

Для создания модели полиномиальной регрессии обычно используются два класса из библиотеки scikit-learn:

PolynomialFeatures: Этот класс используется для преобразования исходных признаков в полиномиальные. Он создает новые признаки, представляющие собой степени исходных признаков и их произведения.

LinearRegression: Этот класс используется для построения линейной модели на основе преобразованных признаков.

4)Поясните принцип преобразования признаков при построении полиномиальной регрессии.

При построении полиномиальной регрессии исходные признаки преобразуются в полиномиальные признаки.

Степени признаков: Для каждого признака создаются новые признаки, представляющие собой его степени (например, x, x^2, x^3 и т.д.).

Произведения признаков: Создаются новые признаки, представляющие собой произведения исходных признаков (например, x1 \* x2, x1 \* x3 и т.д.).

Например, если у нас есть два признака x1 и x2, и мы хотим построить полиномиальную регрессию степени 2, то мы получим следующие признаки: 1, x1, x2, x1^2, x2^2, x1 \* x2

5)Возможно ли применение технологий масштабирования признаков при реализации полиномиальной регрессии?

Да, применение технологий масштабирования признаков при реализации полиномиальной регрессии не только возможно, но и часто необходимо. Причина: Полиномиальные признаки могут иметь очень разный масштаб. Например, если исходный признак x имеет значения от 0 до 1, то x^2 будет иметь значения от 0 до 1, x^3 — от 0 до 1, и так далее. Однако, если исходный признак x имеет значения от 100 до 200, то x^2 будет иметь значения от 10000 до 40000, x^3 — от 1000000 до 8000000, и так далее. Такой разброс масштабов может привести к проблемам при обучении модели. Масштабирование признаков позволяет привести все признаки к одному масштабу, что улучшает сходимость алгоритма оптимизации и повышает устойчивость модели

**Вывод:** научились применять разработанный пайплайн для тиражирования кода с целью решения задачи полиномиальной регрессии.