ЛАБОРАТОРНАЯ РАБОТА 10. ПОЛИНОМИАЛЬНАЯ РЕГРЕССИЯ

**Исследование набора данных**

**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;
* освоение модификаций kNN-метода.

**Индивидуальное задание**

**Описание набора данных**

Указанный набор данных предназначен для построения модели классификации. Данные о 150 экземплярах ириса (рис. 1), по 50 экземпляров из трёх видов – Ирис щетинистый (Iris setosa), Ирис виргинский (Iris virginica) и Ирис разноцветный (Iris versicolor). Для каждого экземпляра измерялись четыре характеристики (в сантиметрах):

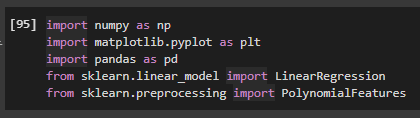
1. длина наружной доли околоцветника (sepal length);
2. ширина наружной доли околоцветника (sepal width);
3. длина внутренней доли околоцветника (petal length);

4) ширина внутренней доли околоцветника (petal width).

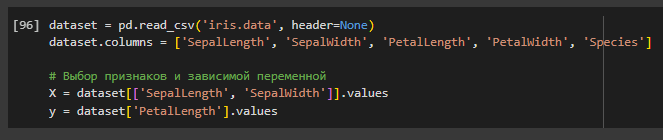
**Индивидуальное задание**

Постройте модель полиномиальной регрессии с использованием. Проанализируйте кривые аппроксимации при различных степенях полинома.

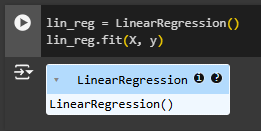
1 Подключение библиотек



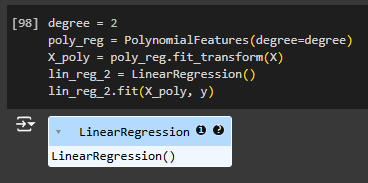
2 Загрузка данных



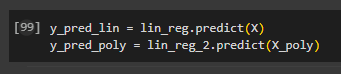
3 Обучение линейной модели



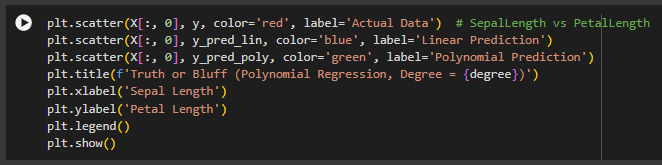
4 Обучение полиномиальной модели



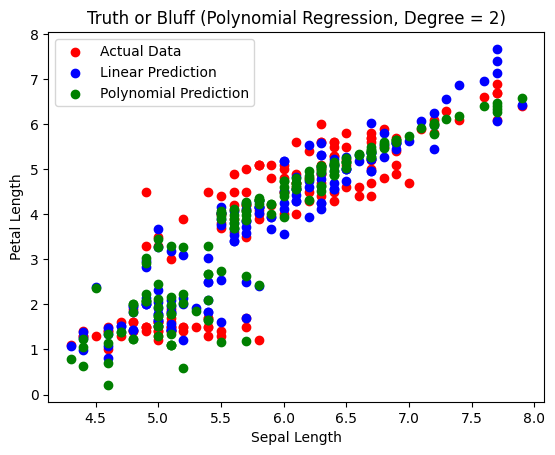
5 Предсказание



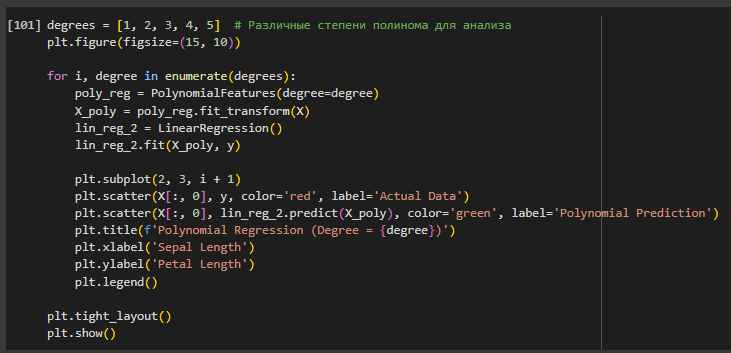
6 Визуализация результатов



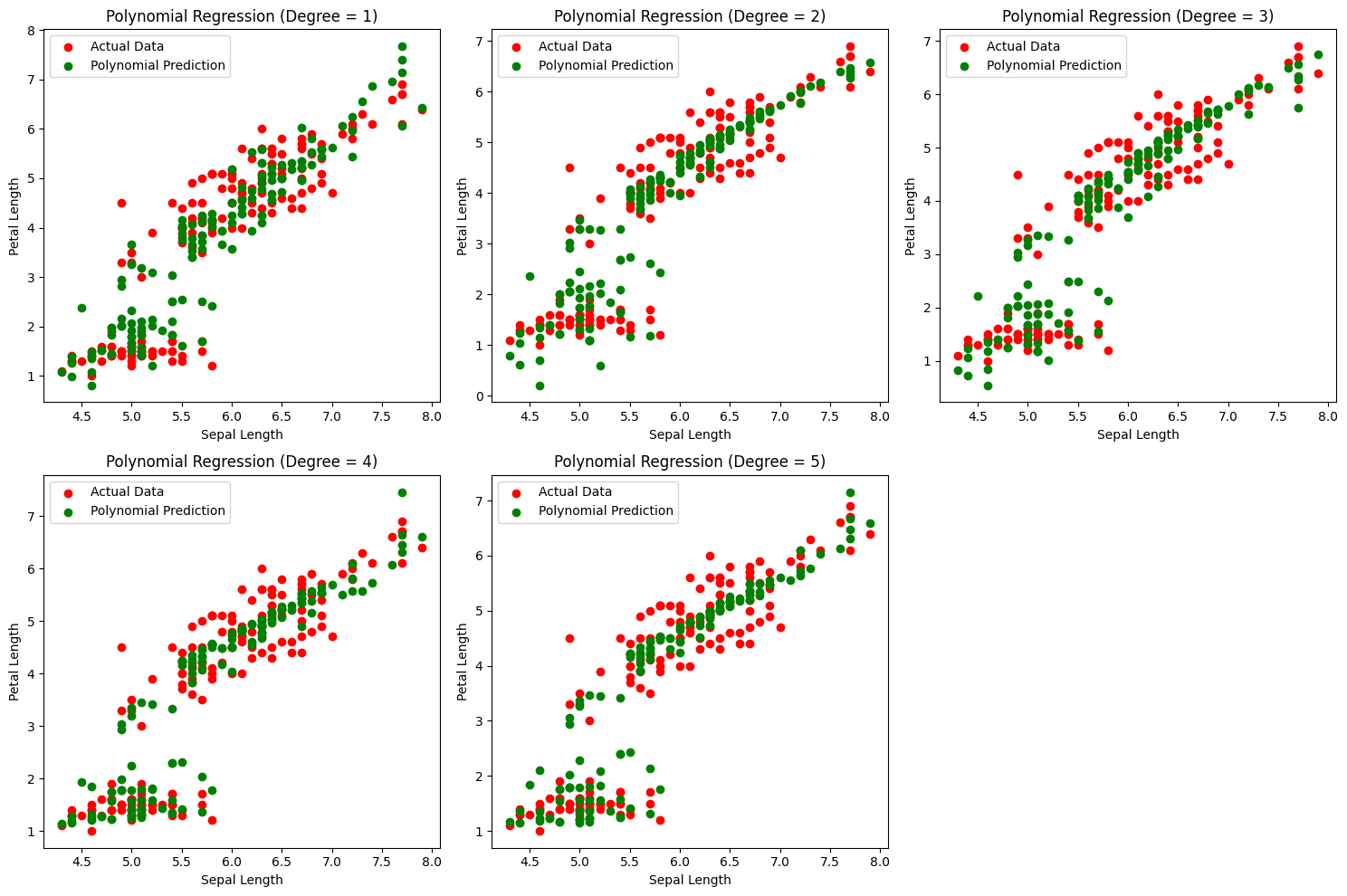
Результат:



7 Анализ кривых аппроксимации при различных степенях полинома



Результат:



**Контрольные вопросы**

1. Фиктивный признак (или константный признак) добавляется для учета свободного члена в уравнении регрессии. Он позволяет модели учитывать смещение, обеспечивая возможность корректного предсказания, даже если все остальные признаки равны нулю.
2. Фиктивная переменная — это бинарная переменная, используемая для представления категориальных данных в числовом формате. При перекодировке категориального признака с (k) категориями создаются (k-1) фиктивных переменных, чтобы избежать проблемы мультиколлинеарности, когда одна переменная может быть выражена через другие.
3. Модель полиномиальной регрессии создается с использованием класса PolynomialFeatures из библиотеки sklearn.preprocessing, который позволяет генерировать полиномиальные и взаимодействующие признаки.
4. Принцип преобразования признаков при построении полиномиальной регрессии заключается в создании новых признаков, которые являются степенями исходных признаков (например, (x^2), (x^3) и т.д.) и их взаимодействиями. Это позволяет модели захватывать нелинейные зависимости.
5. Да, технологии масштабирования признаков, такие как стандартизация или нормализация, могут применяться при реализации полиномиальной регрессии. Это помогает улучшить сходимость алгоритмов оптимизации и может повысить качество модели.