

Регрессия

Беляков Дмитрий

December 29, 2021

1 Описание задачи

Задача: исследовать и оценить алгоритмы линейной, полиномиальной и экспоненциальной регрессии на линейном, полиномиальном и экспоненциальном наборах данных

2 Описание решения

Используем метод `numpy.polynomial` для полиномиальной регрессии. Для экспоненциальной регрессии сделаем следующее:

- Нужно найти экспоненту описывающую кривую ab^x
- Возьмем $\ln(ab^x) = \ln(a) + \ln(b) * x$
- При помощи линейной регрессии логарифма Y найдем $\ln(a)$ и $\ln(b)$
- Восстановим функцию по найденным данным

3 Описание данных

Рассмотрим 5 функций:

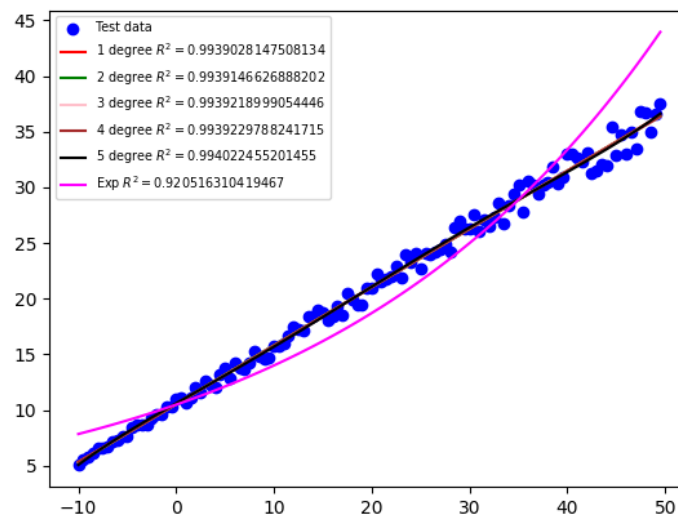
- $x^5 * 1/10^6 - x^4 * 1/10^4 + 1000$
- $x^4 + 120 * x^3 + 540 * x^2 - 1080x - 10^6$
- $x^3 - 120x^2 + 1000$
- $x^2 + 10$
- $x/2 + 10$
- $(\sqrt{e})^x / 1000$ Добавим помехи (умножим на случайное число от 1 до $1+N$)

Для каждой из них рассмотрим полиномиальные регрессии 1-5 степени, а также экспоненциальную регрессию

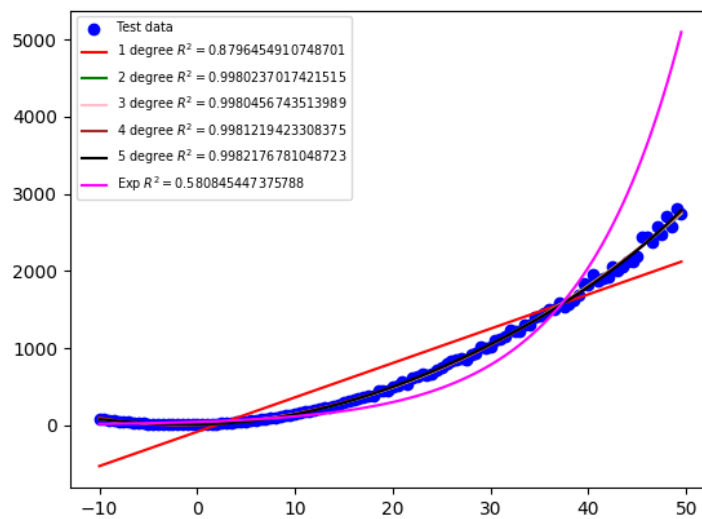
3.1 Оценка результатов

Для оценки результатов будем использовать коэффициент детерминации:
 $R^2 = 1 - \frac{D[y|x]}{D[y]} = 1 - \frac{\sigma^2}{\sigma_y^2}$, Чем ближе коэффициент к 1, тем эффективнее регрессия

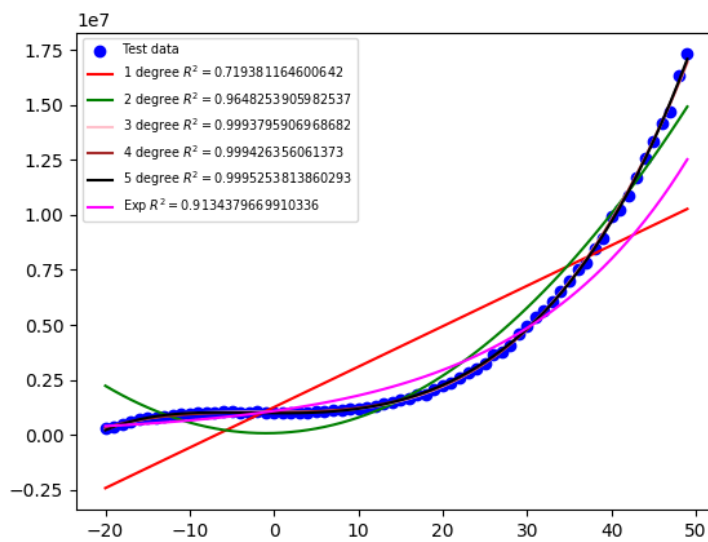
4 Тесты



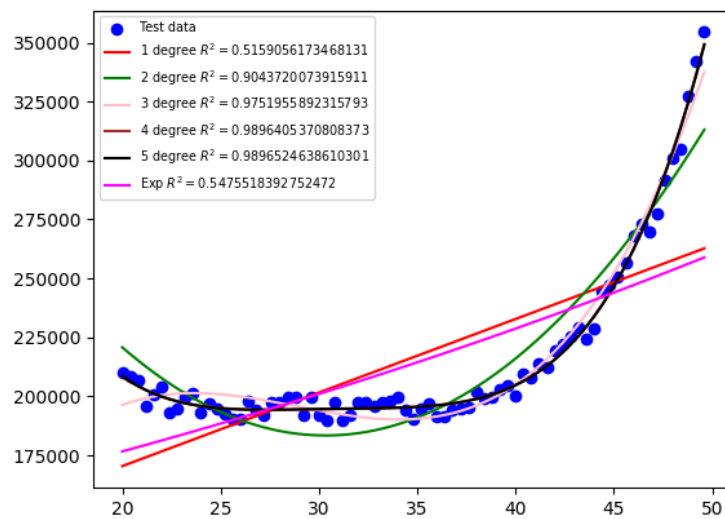
Линейная функция



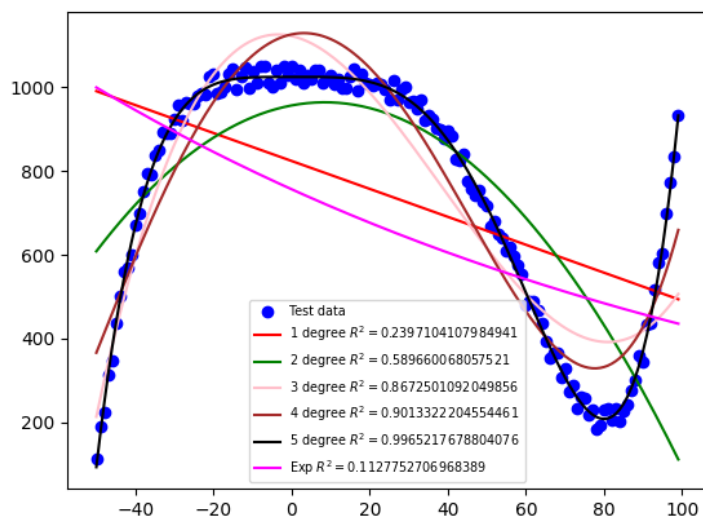
Многочлен 2 степени



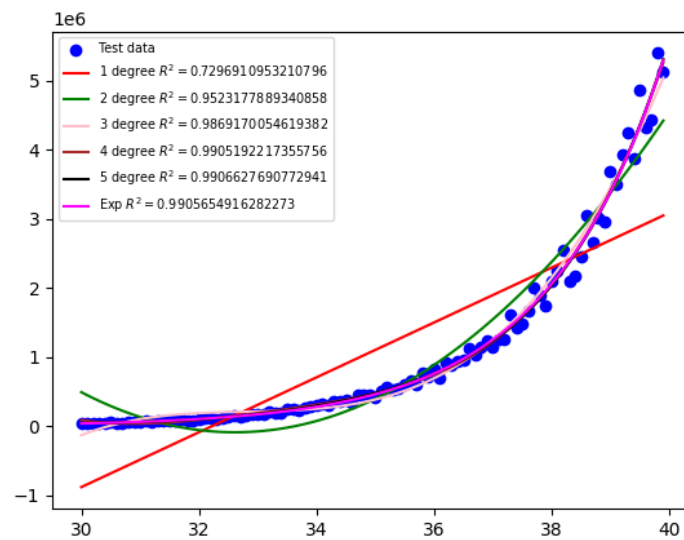
Многочлен 3 степени



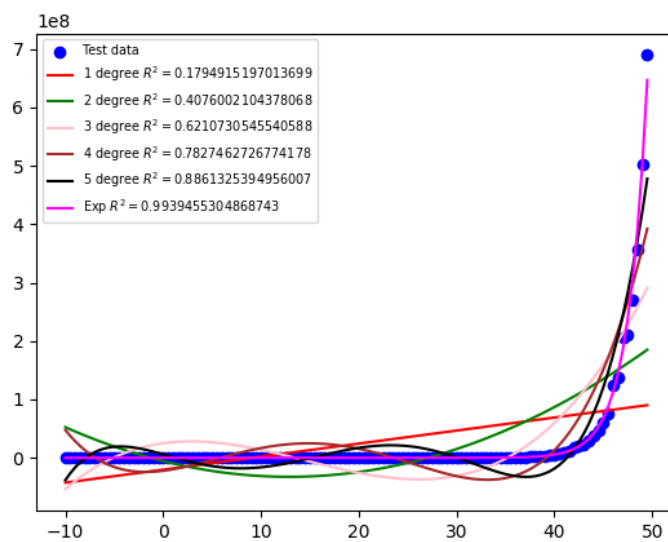
Многочлен 4 степени



Многочлен 5 степени



Малый отрезок экспоненты



Большой отрезок экспоненты

5 Вывод

Как видно из полученных данных, чем выше степень полиномиальной регрессии, тем большую точность она дает. Экспоненциальная регрессия может достаточно точно описывать гладкие отрезки многочленов, но не лучше чем полиномиальная регрессия. Так же и наоборот - экспоненциальная регрессия описывает экспоненту точнее других функций