Построение регрессионной модели.

Владимирова Элина

23 декабря 2020 г.

1 Резюме

Была решена задача вычисления коэффициентов в задаваемой модели регрессии для генерируемых внутри программы точек на плоскости.

Входные данные - требуемый тип регрессии, в случае выбора полиномиальной модели предлагается также задать степень искомого многочлена. Результат работы программы - вектор коэффициентов и график уравнения полученного регрессора.

2 Постановка задачи

Для заданных точек на плоскости ($\leqslant$ 1000 пар координат) необходимо построить регрессионную модель 3 типов: линейную, полиномиальную и экспоненциальную простого вида, т.е. вычислить коэффициенты в соответствующих уравнениях так, чтобы конечный график регрессионной кривой представлял собой оптимальное усреднение имеющихся точечных значений.

3 Допущения

- Количество точек считается заданным изначально.
- При запуске программы предлагается осуществить выбор типа регрессии из трех возможных.
- При выборе полиномиального регрессора предлагается задать степень нужного многочлена.
- Координаты точек генерируются внутри программы.
- Для случая экспоненциальной регрессии предполагается уравнение вида $y(x) = e^{ax+b}$.
- Также для вычисления кривой при экспоненциальной регрессии используются только точки с положительными координатами.

4 Описание решения

Вычисление коэффициентов для регрессионных моделей осуществляется с использованием матриц с координатами точек согласно формуле $\omega = (A^T * A)^{-1} * A^T * y$, где

ω - вектор коэффициентов регрессионной кривой, упорядоченных по возрастанию соответсвующей им степени переменной

A - матрица, содержащая x-координаты данных точек в соответствующих типу регрессии степенях

A^T - транспонированная матрица A

y - столбец, содержащий y-координаты данных точек

Для экспоненциального случая используется модификация указанной формулы: $\omega = (A^T * A)^{-1} * A^T * log(y)$.

Все графики строятся с использованием библиотеки *matplotlib.pyplot*.

5 Результаты

Для наборов точек, соответсвующих изложенным выше ограничениям, были построены уравнения линейной, полиномиальной и экспоненциальной моделей регрессии. Правильность полученных графиков проверена путем настройки генератора точек на уравнения, способные быть с большой точностью заданными тестируемой моделью регрессии, а также визуально.

Рис. 1: Линейная модель для 10 точек с минимальным и несколько большим отклонением от общей траектории, а также для 100 точек с увеличивающимся случайным разбросом.

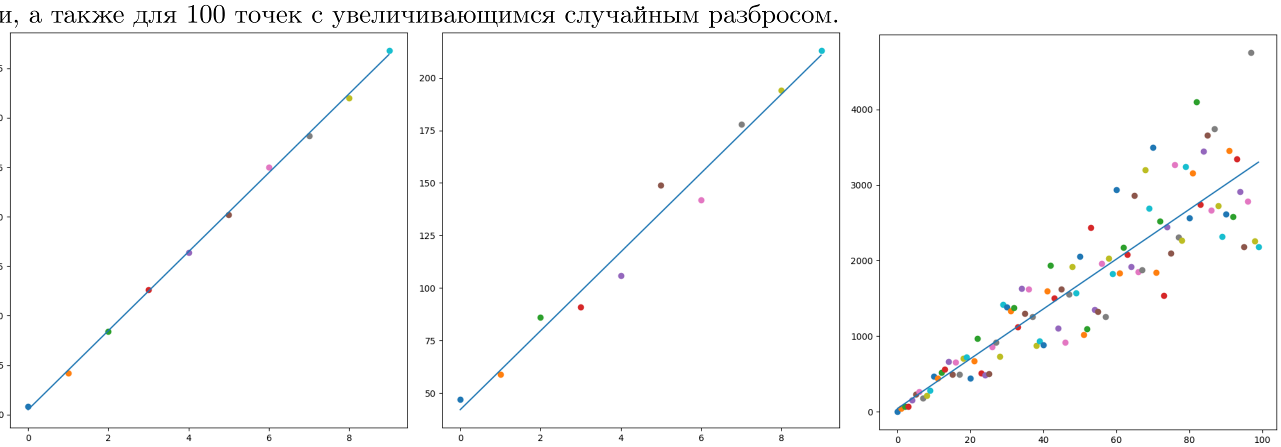
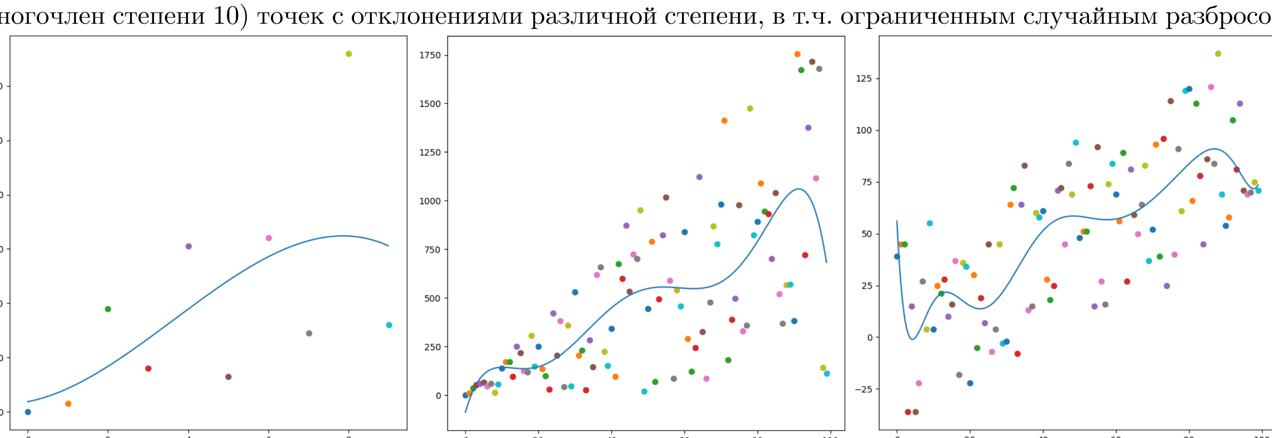


Рис. 2: Полиномиальная модель для 10 (многочлен степени 3), 100 (многочлены степеней 6 и 10) и 1000 (многочлен степени 10) точек с отклонениями различной степени, в т.ч. ограниченным случайным разбросом.



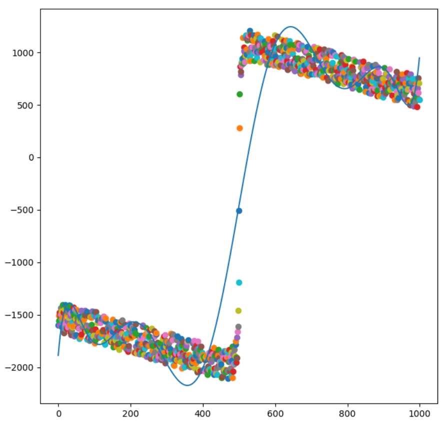


Рис. 3: Экспоненциальная модель для 10, 100 и 1000 точек с минимальным отклонением от общей траектории и с увеличивающимся случайным разбросом.