Государственное бюджетное общеобразовательное учреждение города Москвы "Школа № 1532"

Сравнение непараметрических методов моделирования с параметрическими:

Исследование эффективности непараметрических методов (например регрессия, основанная на оценке Надарая-Ватсона) с параметрическими методами (например аппроксимация с подгонкой по функции) для различных выборок.

10 класс, ГБОУ Школа №1532,

Гришин Илья Андреевич

Руководитель: учитель …, ГБОУ Школа №1532,

…

Москва, 2024

**Оглавление**

[Введение 0](#_Toc33143531)

1. [Актуальность работы 0](#_Toc33143534)

2. [Цели и задачи проекта 0](#_Toc33143532)

3. [Методика выполнения работы 0](#_Toc33143533)

3.1. [Подготовка эксперементальных данных 0](#_Toc33143533)

3.2. [Парамметрические методы 0](#_Toc33143533)

3.3. [Непараметрические методы 0](#_Toc33143533)

3.4. [Сравнение и итоги 0](#_Toc33143533)

4. [Результаты 0](#_Toc33143534)

5. [Выводы 0](#_Toc33143535)

6. [Список используемой литературы 0](#_Toc33143536)

7. [Приложения 0](#_Toc33143533)

**Введение**

В работе рассмотрены методы статистического моделирования, на различных выборках с известным параметром и нет. Для решения данной задачи применяется регрессионная модель, основанная на непараметрической оценке Надарая-Ватсона, а также аппроксимация с подгонкой по функции.

Современные методы моделирования данных предлагают широкий спектр инструментов для анализа и прогнозирования различных явлений. Среди них особое внимание привлекают параметрические и непараметрические методы, каждый из которых имеет свои особенности и применимость в различных сценариях.

В данном проекте проводится сравнительный анализ эффективности непараметрических методов, таких как регрессия на основе оценки Надарая-Ватсона, с параметрическими методами, включая аппроксимацию с подгонкой по функции, на различных выборках данных. Путем сравнительного анализа мы стремимся обеспечить более глубокое понимание того, какие методы моделирования следует предпочитать в различных контекстах и при различных условиях данных.

**Актуальность работы**

Современная область анализа данных и статистики стремительно развивается, и в силу этого актуальность исследований по сравнению методов моделирования остается на высоком уровне. В контексте данного проекта, где основное внимание уделяется сопоставлению непараметрических и параметрических методов, существует несколько ключевых моментов, которые делают данное исследование весьма актуальным:

1. Гибкость и адаптивность методов: Сложные структуры данных могут поддаваться более успешному моделированию с использованием непараметрических методов, в то время как параметрические методы могут быть предпочтительны в случаях, когда структура данных более предсказуема. Исследование эффективности каждого из подходов становится важным шагом для выбора оптимального метода в зависимости от особенностей данных.

2. Практическая применимость: С популяризацией методов машинного обучения и статистического моделирования, важно понять, какие методы наиболее подходят для конкретных задач. Исследование различных методов на разнообразных выборках данных предоставляет ценную информацию для практикующих специалистов в области анализа данных.

3. Разнообразие областей применения: Методы моделирования широко используются в различных областях, таких как экономика, медицина, биология и социальные науки. Сравнение методов на разнообразных выборках позволяет обобщить результаты и делает исследование более универсальным в контексте различных дисциплин.

4. Оптимизация ресурсов: Эффективное использование ресурсов, таких как вычислительная мощность, время и данные, является критическим вопросом в современных исследованиях. Понимание, какие методы более эффективны для конкретных сценариев, может значительно сэкономить ресурсы и повысить эффективность аналитических процессов.

С учетом этих факторов, данное исследование о сравнении непараметрических и параметрических методов моделирования представляет собой актуальный вклад в развивающуюся область анализа данных и статистики.

**Цель и задачи проекта**

Провести сравнительный анализ между непараметрическими и параметрическими методами моделирования с целью выявления их преимуществ, недостатков и областей применения. Исследование направлено на определение эффективности каждого метода в различных контекстах и создание основы для рекомендаций по выбору подходящего метода в зависимости от конкретных задач и данных.

Были поставлены следующие задачи работы:

1. Создание экспериментальных данных.
2. Реализовать параметрический метод моделирования.
3. Реализовать непараметрические методы моделирования.
4. Рассмотреть каждый метод для определённой выборки.
5. Подвести итоги проделанной работы.

# 

# Методика выполнения исследования

***Первый этап – Подготовка экспериментальных данных***

*Создадим различные имитации выборок как для 3D, так и для 2D моделирования и запишем их в отдельный файл.*

Мы начнем с генерации некоторых случайных точек 2D-данных с помощью библиотеки [***NumPy***](https://www.geeksforgeeks.org/numpy-tutorial/). Каждый пример будем записывать в отдельный .txt файл с соответствующим названием.

Реализуем программу:

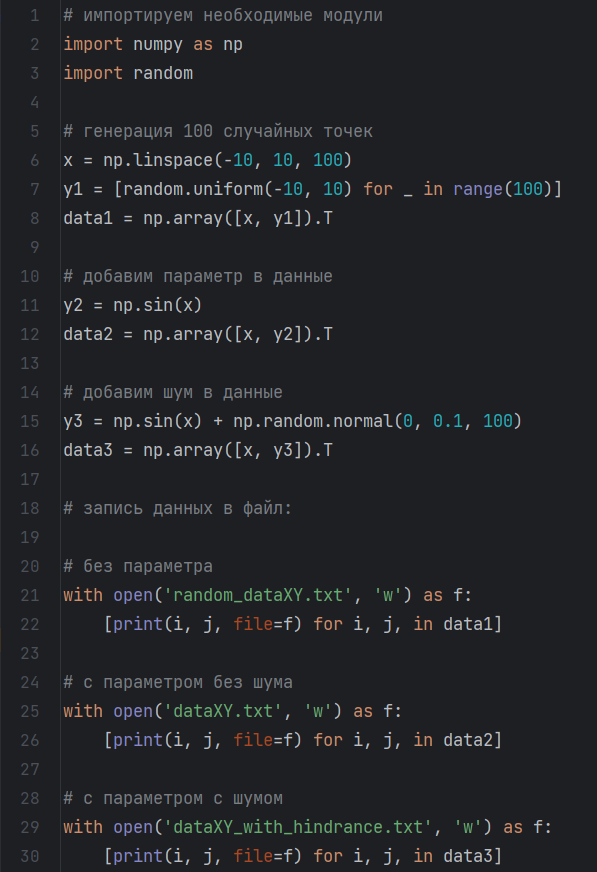


Рисунок 1 – Скрипт для генерации данных в 2D пространстве

Теперь аналогично сгенерируем точки в 3D пространстве

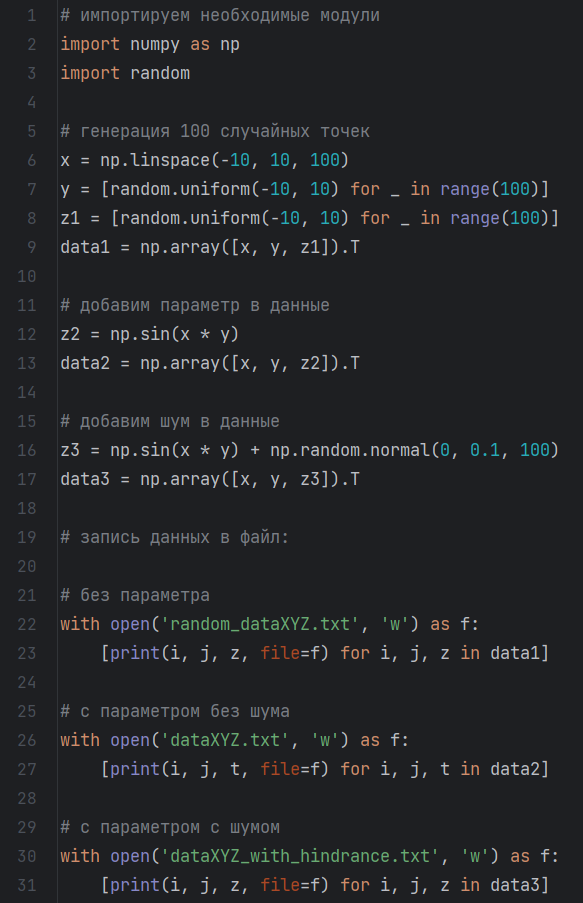


Рисунок 2 – Скрипт для генерации данных в 3D пространстве

Что у нас получилось:

Мы создали шесть файлов с примерами различных выборок, состоящих из 100 точек, сгенерированных по-разному.

Далее приведены все варианты с описанием

‘random\_dataXY.txt’ - два независящих друг от друга массива 2D

‘dataXY.txt’ - два массива с параметром без шума

‘dataXY\_with\_hindrance.txt’ - два массива с параметром с шумом

‘random\_dataXYZ.txt’ - три независящих друг от друга массива 3D

‘dataXYZ.txt’ - три массива с параметром без шума

‘dataXYZ\_with\_hindrance.txt’ - три массива с параметром с шумом

Теперь давайте сделаем программу для удобного получения данных из файла.

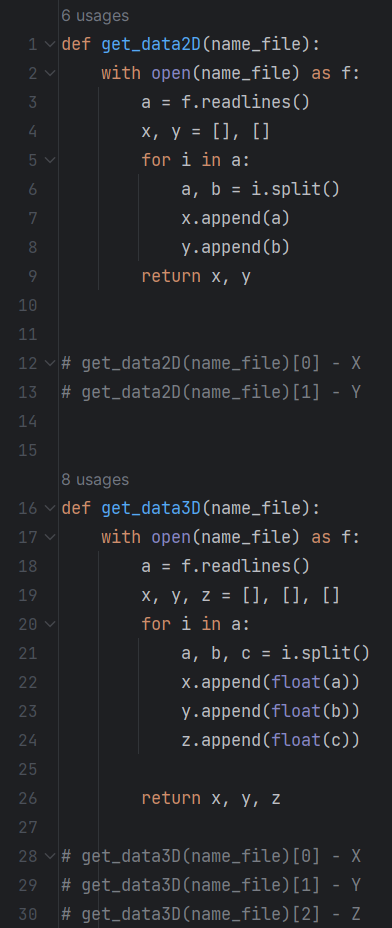


Рисунок 3 – Программа для получения данных

Здесь реализовано две функции для получения 2D и 3D данных. На вход принимается название файла и возвращается массивы с данными для каждой оси, которые можно получить, вызвав функцию и указав индекс необходимого массива, данные для X находятся под индексом *0,* для Y по индексом *1.*

***Второй этап – Параметрические методы***

**Параметрическое моделирование** — [моделирование](https://ru.wikipedia.org/wiki/%D0%9C%D0%BE%D0%B4%D0%B5%D0%BB%D0%B8%D1%80%D0%BE%D0%B2%D0%B0%D0%BD%D0%B8%D0%B5) ([проектирование](https://ru.wikipedia.org/wiki/%D0%9F%D1%80%D0%BE%D0%B5%D0%BA%D1%82%D0%B8%D1%80%D0%BE%D0%B2%D0%B0%D0%BD%D0%B8%D0%B5)) с использованием [параметров](https://ru.wikipedia.org/wiki/%D0%9F%D0%B0%D1%80%D0%B0%D0%BC%D0%B5%D1%82%D1%80) элементов [модели](https://ru.wikipedia.org/wiki/%D0%9C%D0%BE%D0%B4%D0%B5%D0%BB%D1%8C) и [соотношений](https://ru.wikipedia.org/wiki/%D0%A1%D0%BE%D0%BE%D1%82%D0%BD%D0%BE%D1%88%D0%B5%D0%BD%D0%B8%D0%B5) между этими параметрами. [Параметризация](https://ru.wikipedia.org/w/index.php?title=%D0%9F%D0%B0%D1%80%D0%B0%D0%BC%D0%B5%D1%82%D1%80%D0%B8%D0%B7%D0%B0%D1%86%D0%B8%D1%8F&action=edit&redlink=1) позволяет за короткое время «проиграть» (с помощью изменения параметров или геометрических соотношений) различные конструктивные [схемы](https://ru.wikipedia.org/wiki/%D0%A1%D1%85%D0%B5%D0%BC%D0%B0_(%D0%BC%D0%B0%D1%82%D0%B5%D0%BC%D0%B0%D1%82%D0%B8%D0%BA%D0%B0)) и избежать принципиальных ошибок.

[*https://ru.wikipedia.org/wiki/Параметрическое\_моделирование*](https://ru.wikipedia.org/wiki/Параметрическое_моделирование)

Нам необходимо создать функции, которые будет удобно применить к каждому из примеров на языке *Python*

Реализуем аппроксимацию для 2D пространства*:*

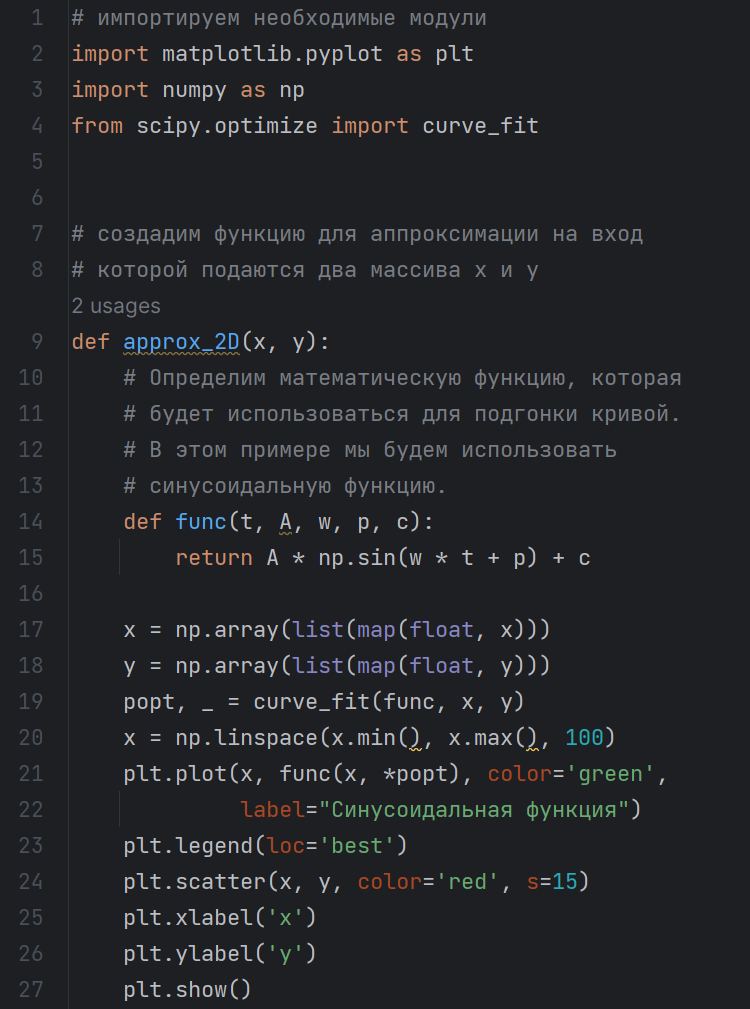


Рисунок 4 – Функция для аппроксимации в 2D пространстве.

Мы реализовали функцию для аппроксимации в 2D ‘*approx\_2D’*.

На вход принимается два массива x и y. Выводится график с данными точками и аппроксимируемой кривой. Кривая подгоняется с помощью библиотеки *scipy* функции *curve\_fit*. Функция для подгонки, заданная нами синусоидальная.

Теперь приступим к реализуем аппроксимацию для 3D пространства*:*

Создадим, аналогичную прошлой, функцию с некоторыми нюансами

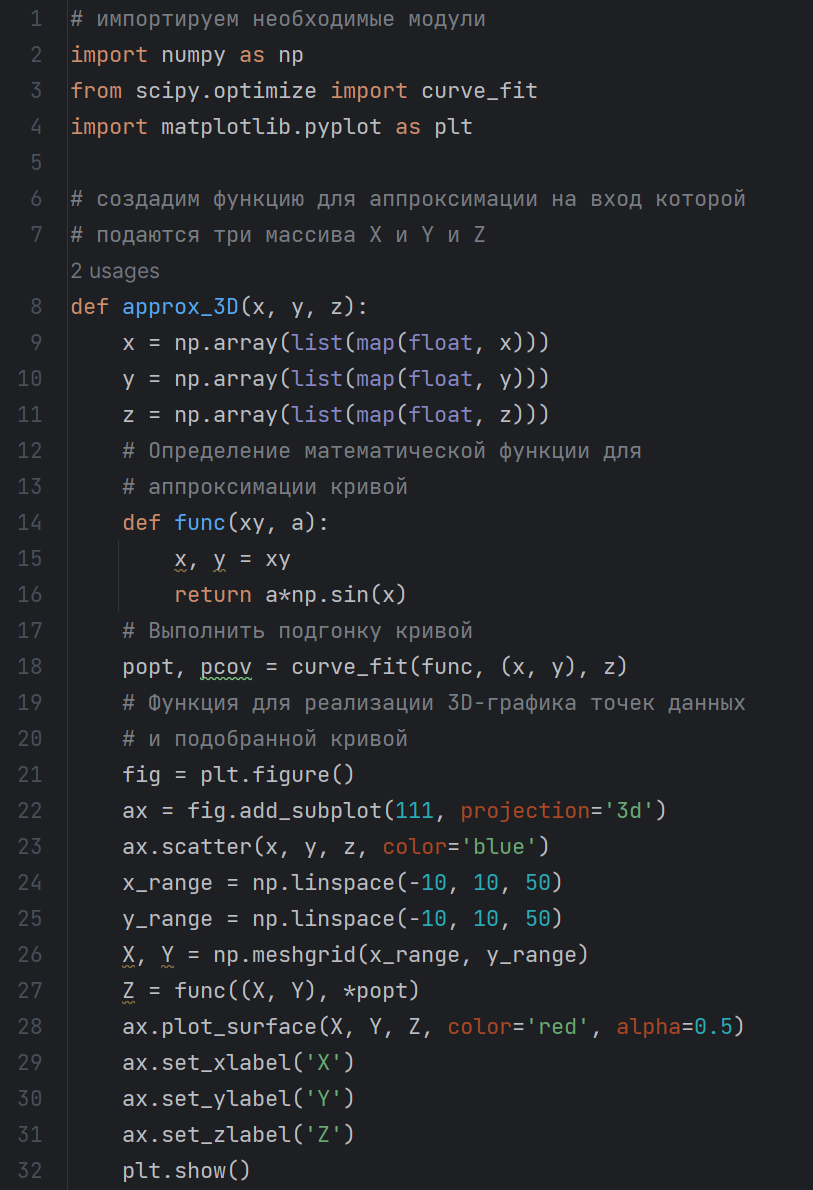


Рисунок 5 – Функция для аппроксимации в 3D пространстве.

На вход принимается уже три массива x, y и z. Выводится также график с данными точками и аппроксимирующим рельефом. Подгоняется с помощью библиотеки *scipy* функции *curve\_fit*. Функция для подгонки, заданная нами также синусоидальная.

***Третий этап – Непараметрические методы***

*Углубимся в непараметрические методы*

Непараметрические методы – это количественные методы статистической обработки данных, применение которых не требует знания закона распределения изучаемых признаков в совокупности и вычисления их основных параметров. - *ПРОВЕРКА СТАТИСТИЧЕСКИХ ГИПОТЕЗ Смирнова З.М., Крейнина М.В.*

В данном случае детально рассмотрим непараметрическую оценку регрессии Надарая-Ватсона.

Формула для непараметрической оценки регрессии Надарая-Ватсона

Реализуем непараметрический метод моделирования для 2D пространства:



Рисунок 6 – Функция для непараметрической регрессии в 2D

Реализуем непараметрические методы моделирования для 3D пространства:

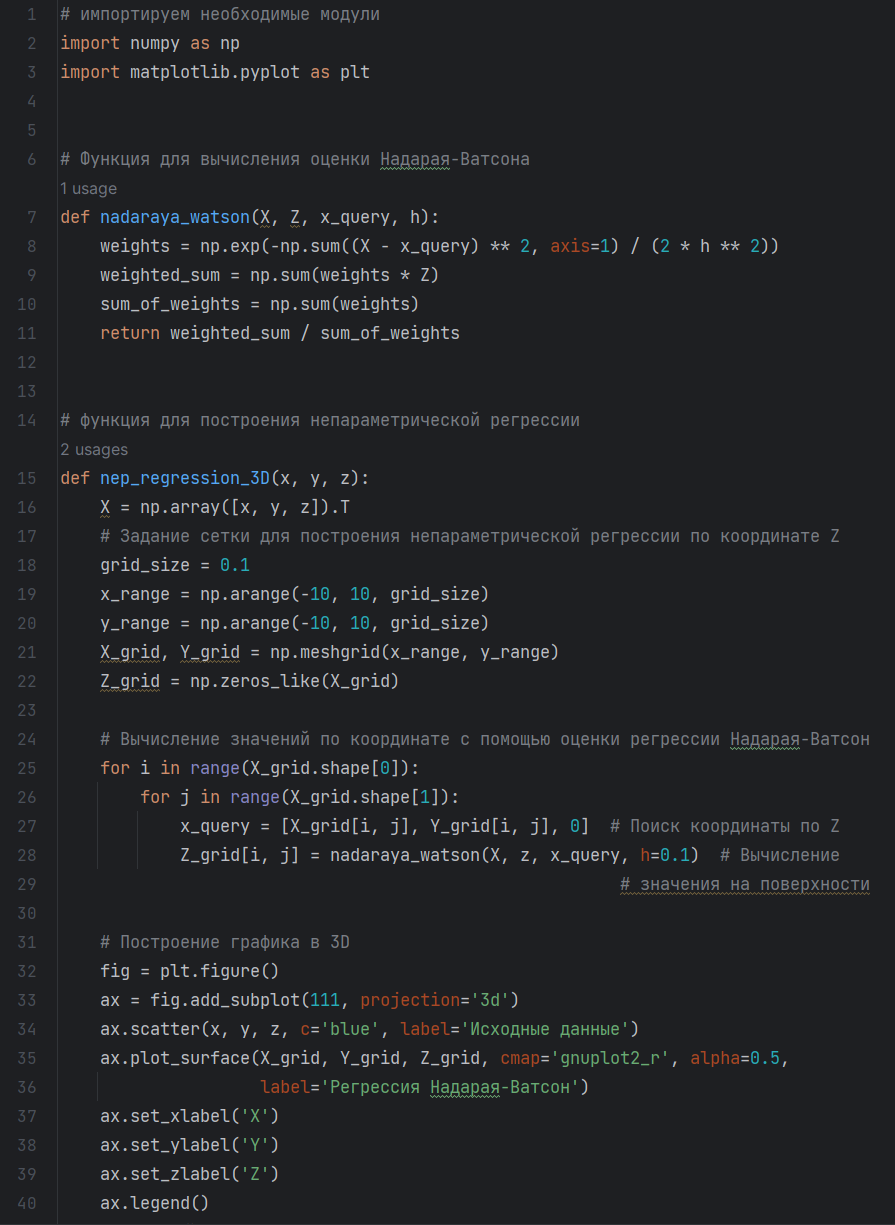


Рисунок 7 – Функция для непараметрического моделирования в 3D

Описание реализации:

Мы создаём функцию по формуле Надарая-Ватсона, далее проходимся по каждой точке и с помощью функции вычисляем значения, по которым далее строим график.

***Четвёртый этап – Сравнение методов***

*Применим методы к каждому из примеров и выявим их особенности и область применения*

Начнём с 2D методов. Напишем программы для запуска каждого метода.

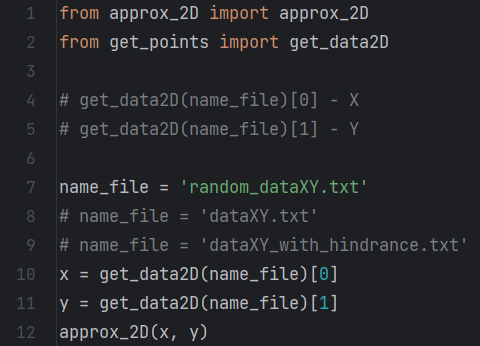


Рисунок 8 – Программа для запуска аппроксимации функции в 2D

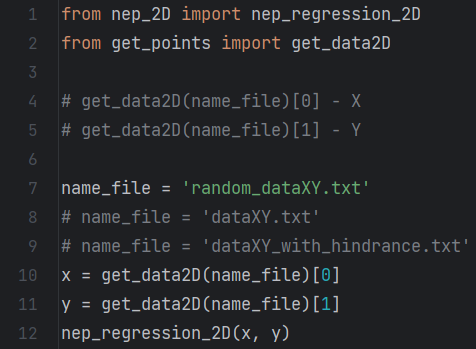
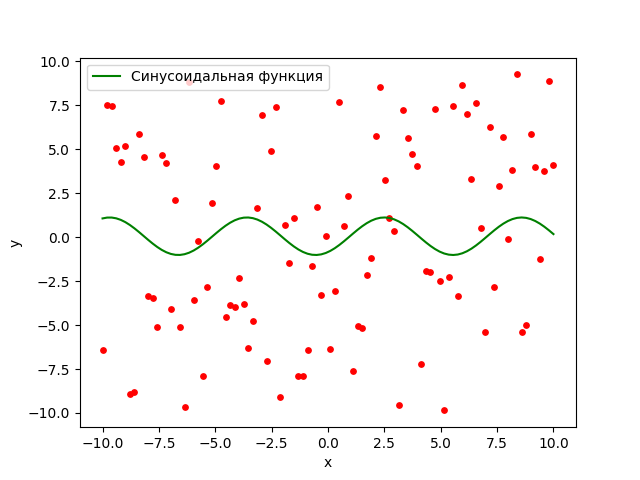
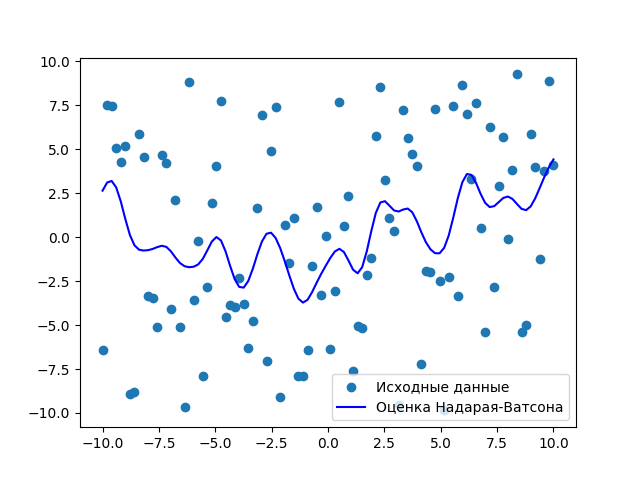


Рисунок 9 – Программа для запуска непараметрической регрессии функции в 2D

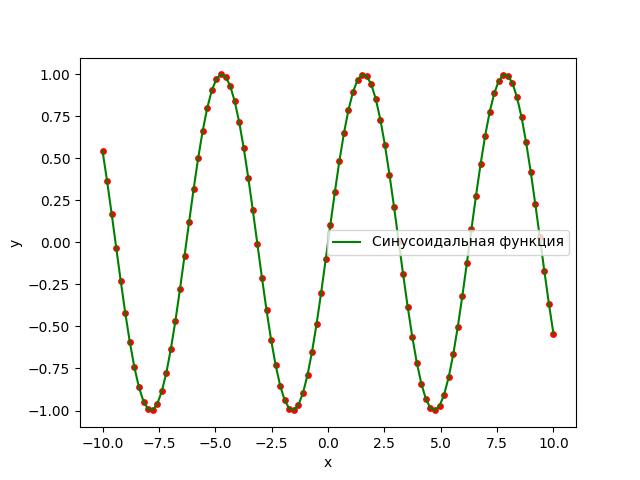
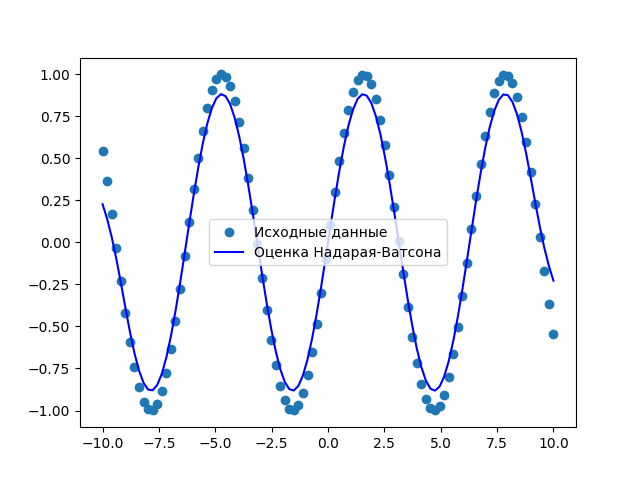
Рассмотрим результаты:

Для файла 'random\_dataXY.txt':

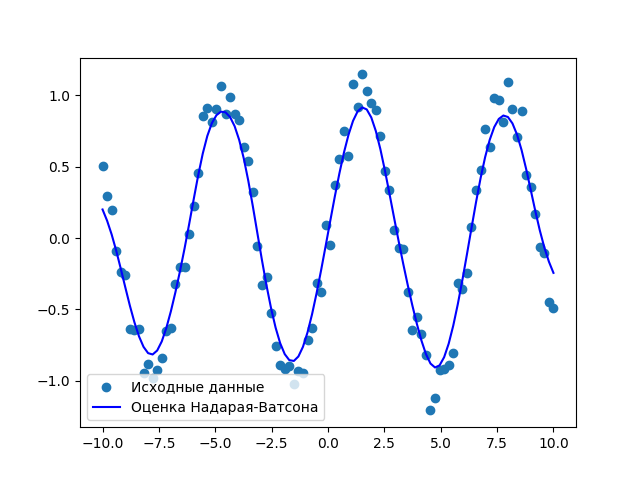
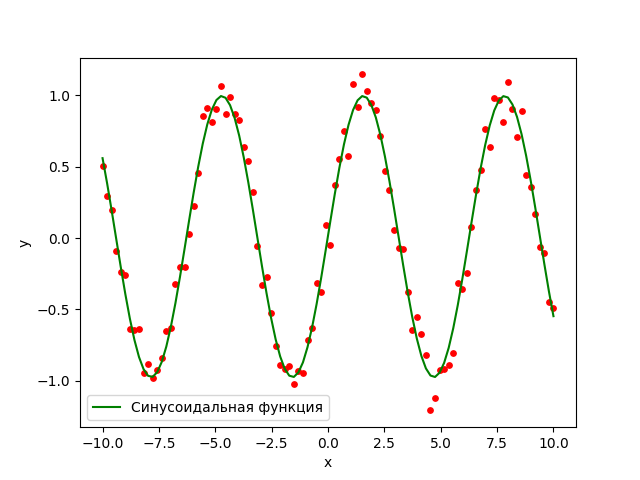
………

Для файла 'dataXY.txt':

………

Для файла 'dataXY\_with\_hindrance.txt':



………

Теперь рассмотрим 3D методы:



Рисунок 10 – Программа для запуска аппроксимации функции в 3D

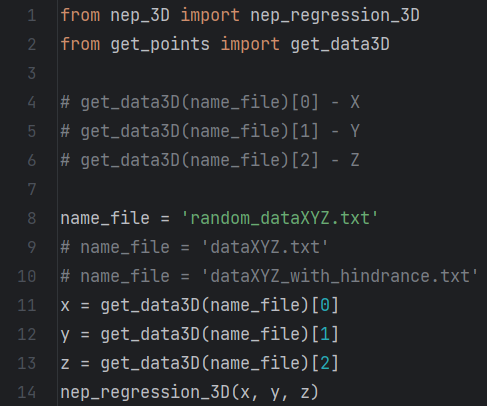
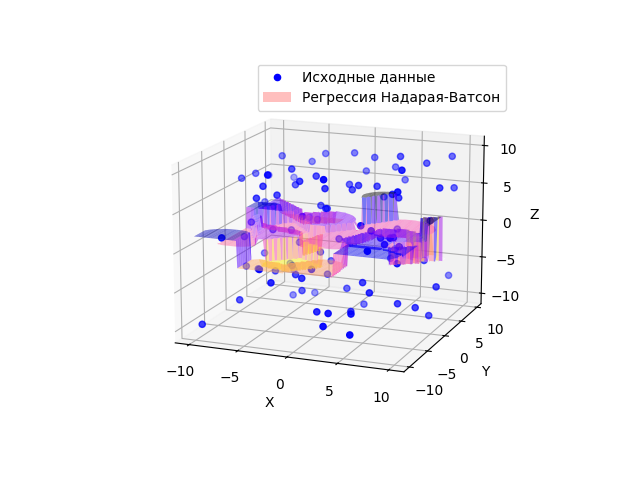
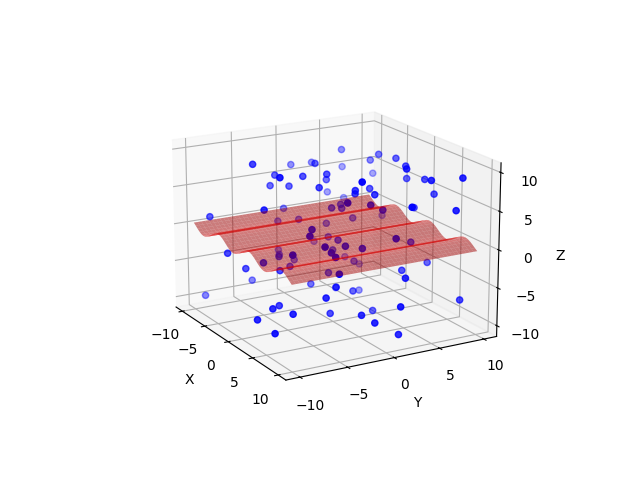


Рисунок 11 – Программа для запуска непараметрической регрессии функции в 3D

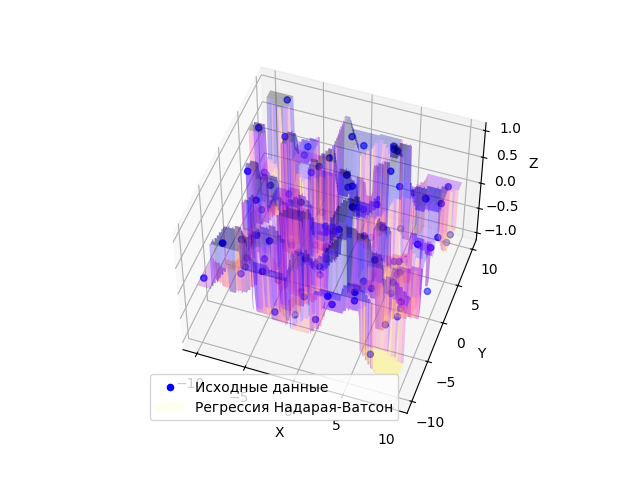
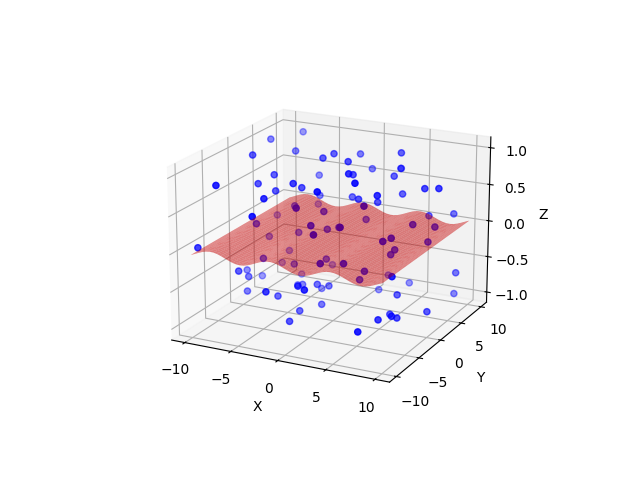
Рассмотрим результаты:

Для файла 'random\_dataXYZ.txt':



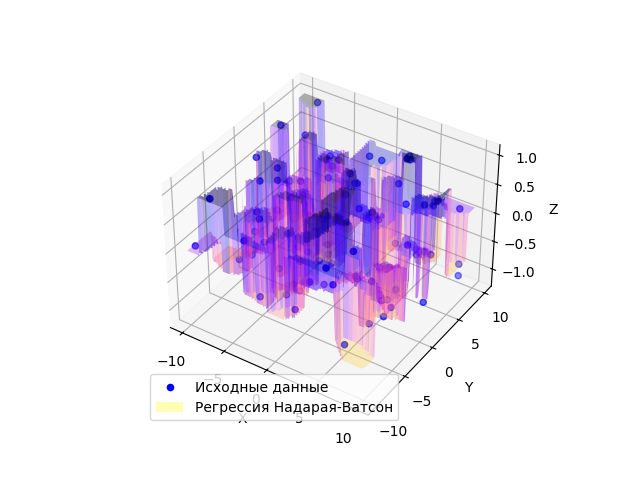
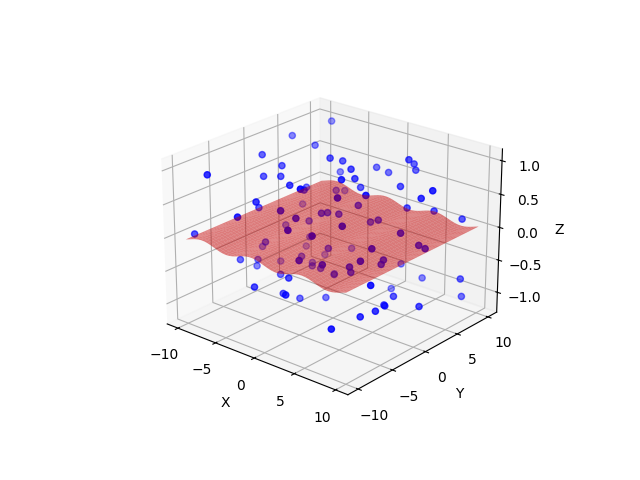
………

Для файла 'dataXYZ.txt':



………

Для файла 'dataXYZ\_with\_hindrance.txt':



………

# Результаты и обсуждение

Отличия параметрического и непараметрического моделирования:

1. Предположения о распределении данных:

• Параметрическое моделирование: Основано на предположении о конкретной функциональной форме или распределении данных, например, нормальном или экспоненциальном.

• Непараметрическое моделирование: не требует априорных предположений о распределении данных, что делает его более гибким и универсальным.

2. Число параметров модели:

• Параметрическое моделирование: имеет фиксированное число параметров, которые нужно оценить, основываясь на данных.

• Непараметрическое моделирование: Число параметров модели зависит от размера выборки, что позволяет модели гибко адаптироваться к разнообразным формам данных.

3. Устойчивость к выбросам и аномалиям:

• Параметрическое моделирование: может быть чувствительным к выбросам в данных, особенно если выбранная функциональная форма недостаточно гибка.

• Непараметрическое моделирование: более устойчиво к выбросам, так как не предполагает конкретной формы данных и может лучше адаптироваться к аномальным наблюдениям.

4. Интерпретируемость:

• Параметрическое моделирование: часто более легко интерпретируемо, так как параметры модели имеют конкретные смысловые интерпретации.

• Непараметрическое моделирование: может быть менее интерпретируемым из-за отсутствия явных параметров, хотя некоторые методы, такие как ядерная регрессия, могут предоставлять некоторую интерпретируемость.

5. Сложность модели:

• Параметрическое моделирование: часто более простое в понимании и реализации, так как требует определения конкретной функциональной формы.

• Непараметрическое моделирование: может быть более сложным и требовать более высокого уровня алгоритмического понимания для его применения.

Выбор между параметрическим и непараметрическим моделированием зависит от конкретного контекста задачи, характера данных и требований к модели.

# Выводы

В данном проекте было проведено исследование эффективности непараметрических методов моделирования, таких как регрессия, основанная на оценке Надарая-Ватсона, с параметрическими методами, например аппроксимация с подгонкой по функции.

Цель исследования заключалась в сравнении этих методов на различных выборках. Для этого были выбраны несколько наборов данных с разной структурой и характером. Затем были применены непараметрические и параметрические методы к каждой выборке, и произведено сравнение результатов.

Результаты исследования показали, что эффективность непараметрических методов может значительно различаться в зависимости от выборки. В некоторых случаях непараметрические методы показали более точные и надежные результаты, особенно если выборка имела сложную структуру или сильные выбросы. Однако в других случаях параметрические методы показали более стабильные и устойчивые результаты.

Таким образом, выбор между непараметрическими и параметрическими методами моделирования должен основываться на характеристиках конкретной выборки и целях исследования. Непараметрические методы могут быть предпочтительными в случаях, когда данные имеют сложную структуру или несимметричное распределение, в то время как параметрические методы могут быть более подходящими для простых и симметричных выборок.

Однако необходимо отметить, что эффективность методов может зависеть не только от выборки, но и от других факторов, таких как объем выборки, точность измерений и выбор функции подгонки. Поэтому для получения более точных результатов рекомендуется провести дополнительные исследования и сравнения на большем объеме данных.

# Список используемой литературы

1. Бронштейн, И. Н. Справочник по математике для инженеров и учащихся втузов [Текст] / И. Н. Бронштейн, К.А. Семендяев. – М.: Наука. Главная редакция физико-математической литературы, 1981. – 720с.
2. Бесстремянная, Г. Е. Применение ядерных и параметрических регрессий для оценки влияния страховых медицинских организаций на качество региональных систем здравоохранения [Текст] / Г. Е. Бесстремянная, 2015. - 18 c.
3. Математический энциклопедический словарь [Текст] / Гл. ред. Ю. В. Прохоров. - М.: Советская энциклопедия, 1988. - 847 с.
4. Хиценко, В. Е. Непараметрическая статистика в задачах защиты информации. Конспект лекций [Текст] / В. Е. Хиценко, 2012. ­­- 196 c.