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 3](#_Toc33143531)

1. [Актуальность работы 3](#_Toc33143534)

2. [Цели и задачи проекта 5](#_Toc33143532)

3. [Методика выполнения работы 5](#_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 4](#_Toc33143534)

5. [Выводы 14](#_Toc33143535)

6. [Список используемой литературы 15](#_Toc33143536)

7. [Приложения 3](#_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/).

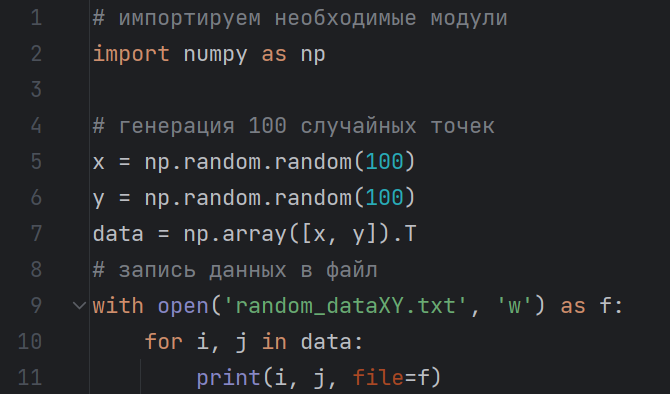


Рисунок 1 – Скрипт для генерации случайных данных.

Теперь добавим параметр в данные.

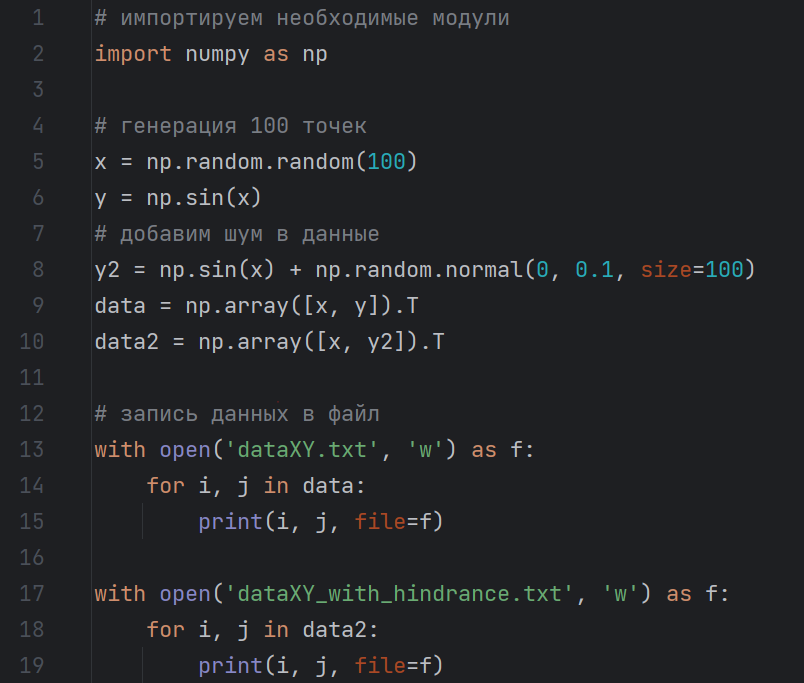


Рисунок 2 – Скрипт для генерации данных с параметром, а также с добавлением шума и без

Мы сгенерировали случайные данные, состоящие из 100 случайных точек в 2D-пространстве, где точки X и Y никак не зависят друг от друга, или же в ином случае имеют некий параметр в виде зависимости Y от X, где Y определяется как функция числа X с некоторым добавлением шума и без.

Теперь создадим данные по тому же принципу только в 3D пространстве.

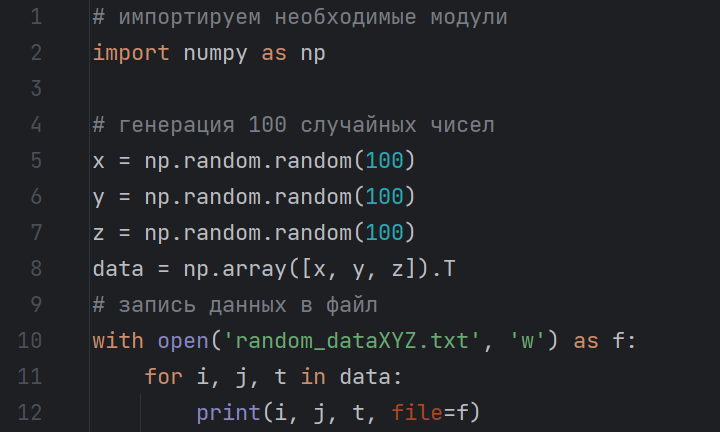


Рисунок 3 - Скрипт для генерации не зависимых друг от друга данных в 3D пространстве.

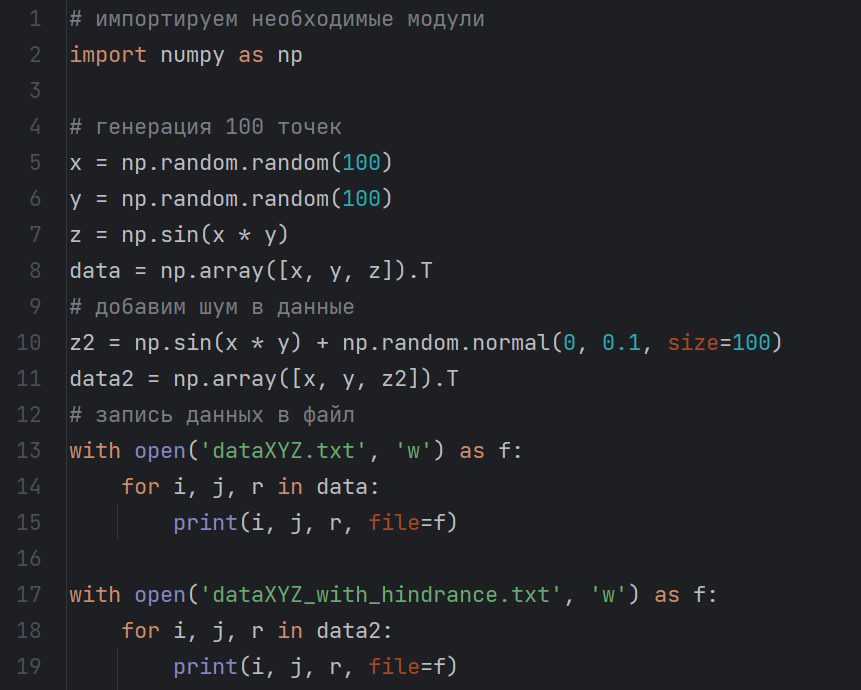


Рисунок 4 - Скрипт для генерации данных с параметром, а также с добавлением шума и без, в 3D пространстве.

random\_dataXY.txt = два независящих друг от друга массива 2D

dataXY.txt = два массива с параметром без помех

dataXY\_with\_hindrance.txt = два массива с параметром с помехой

random\_dataXYZ.txt = три независящих друг от друга массива 3D

dataXYZ.txt = три массива с параметром без помех

dataXYZ\_with\_hindrance.txt = три массива с параметром с помехой

***Второй этап – Параметрические методы***

*Рассмотрим подробно параметрические методы на каждом примере.*

**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

Реализуем аппроксимацию на языке *Python* для 2D пространства*:*

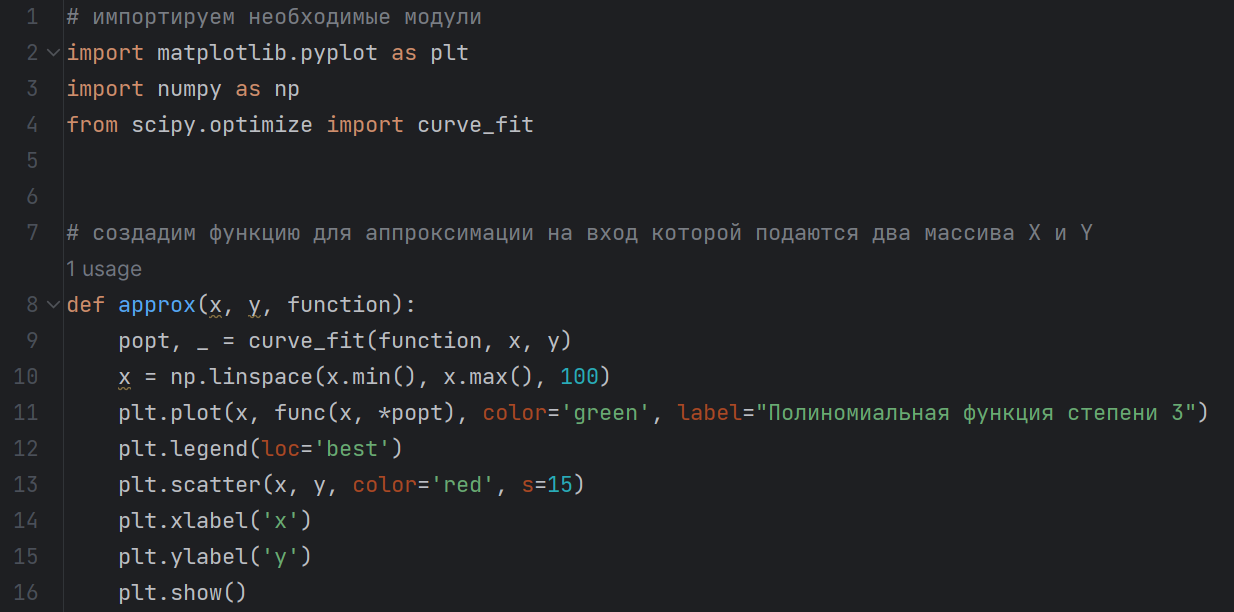
**

Рисунок 3 – Функция для аппроксимации в 2D пространстве.

Реализуем аппроксимацию на языке *Python* для 3D пространства*:*

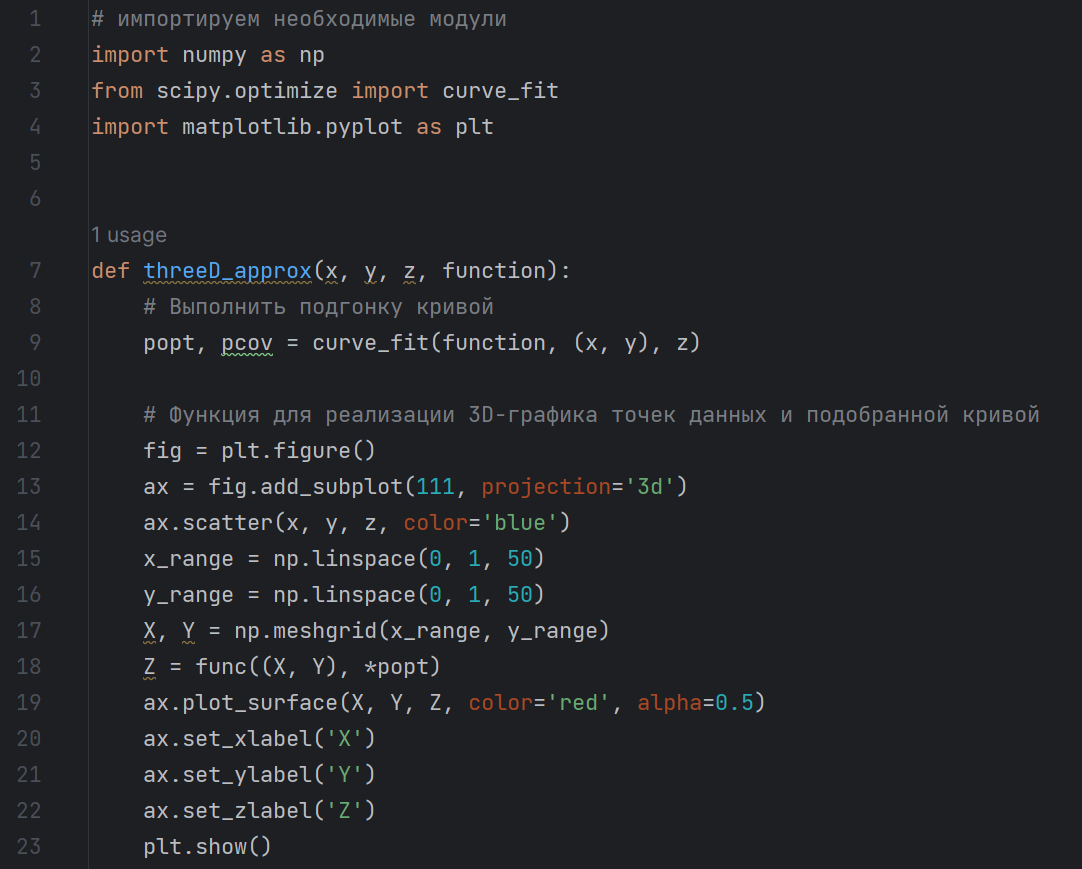
**

Рисунок 4 – Функция для аппроксимации в 3D пространстве.

***Третий этап – Непараметрические методы***

*Углубимся в параметрические 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.

Реализуем непараметрические методы моделирования для 2D пространства:

# 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

# 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.