

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	Ошибка! Закладка не определена.
1. Актуальность работы	3
2. Цели и задачи проекта	5
3. Методика выполнения работы	5
3.1. Подготовка экспериментальных данных	0
3.2. Параметрические методы.....	0
3.3. Непараметрические методы	0
3.4. Сравнение и итоги.....	0
4. Результаты.....	Ошибка! Закладка не определена.
5. Выводы	6
6. Список используемой литературы.....	13
7. Приложения	Ошибка! Закладка не определена.

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*.

```
1  # импортируем необходимые модули
2  import numpy as np
3
4  # генерация 100 случайных точек
5  x = np.random.random(100)
6  y = np.random.random(100)
7  data = np.array([x, y]).T
8  # запись данных в файл
9  with open('random_dataXY.txt', 'w') as f:
10     for i, j in data:
11         print(i, j, file=f)
```

Рисунок 1 – Скрипт для генерации случайных данных.

Теперь добавим параметр в данные.

```
1  # импортируем необходимые модули
2  import numpy as np
3
4  # генерация 100 точек
5  x = np.random.random(100)
6  y = np.sin(x)
7  # добавим шум в данные
8  y2 = np.sin(x) + np.random.normal(0, 0.1, size=100)
9  data = np.array([x, y]).T
10  data2 = np.array([x, y2]).T
11
12  # запись данных в файл
13  with open('dataXY.txt', 'w') as f:
14     for i, j in data:
15         print(i, j, file=f)
16
17  with open('dataXY_with_hindrance.txt', 'w') as f:
18     for i, j in data2:
19         print(i, j, file=f)
```

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.

```
1  # импортируем необходимые модули
2  import numpy as np
3
4  # генерация 100 случайных чисел
5  x = np.random.random(100)
6  y = np.random.random(100)
7  z = np.random.random(100)
8  data = np.array([x, y, z]).T
9  # запись данных в файл
10 with open('random_dataXYZ.txt', 'w') as f:
11     for i, j, t in data:
12         print(i, j, t, file=f)
```

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

```
1  # импортируем необходимые модули
2  import numpy as np
3
4  # генерация 100 точек
5  x = np.random.random(100)
6  y = np.random.random(100)
7  z = np.sin(x * y)
8  data = np.array([x, y, z]).T
9  # добавим шум в данные
10 z2 = np.sin(x * y) + np.random.normal(0, 0.1, size=100)
11 data2 = np.array([x, y, z2]).T
12 # запись данных в файл
13 with open('dataXYZ.txt', 'w') as f:
14     for i, j, r in data:
15         print(i, j, r, file=f)
16
17 with open('dataXYZ_with_hindrance.txt', 'w') as f:
18     for i, j, r in data2:
19         print(i, j, r, file=f)
```

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 (проектирование) с использованием параметров элементов модели и соотношений между этими параметрами. Параметризация позволяет за короткое время «проиграть» (с помощью изменения параметров или геометрических соотношений) различные конструктивные схемы и избежать принципиальных ошибок.

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:

```
1 # импортируем необходимые модули
2 import matplotlib.pyplot as plt
3 import numpy as np
4 from scipy.optimize import curve_fit
5
6
7 # создадим функцию для аппроксимации на вход которой подаются два массива X и Y
8 def approx(x, y, function):
9     popt, _ = curve_fit(function, x, y)
10    x = np.linspace(x.min(), x.max(), 100)
11    plt.plot(x, func(x, *popt), color='green', label="Полиномиальная функция степени 3")
12    plt.legend(loc='best')
13    plt.scatter(x, y, color='red', s=15)
14    plt.xlabel('x')
15    plt.ylabel('y')
16    plt.show()
```

Рисунок 3 – Функция для аппроксимации в 2D пространстве.

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

```
1  # импортируем необходимые модули
2  import numpy as np
3  from scipy.optimize import curve_fit
4  import matplotlib.pyplot as plt
5
6
7  1 usage
8  def threeD_approx(x, y, z, function):
9      # Выполнить подгонку кривой
10     popt, pcov = curve_fit(function, (x, y), z)
11
12     # Функция для реализации 3D-графика точек данных и подобранной кривой
13     fig = plt.figure()
14     ax = fig.add_subplot(111, projection='3d')
15     ax.scatter(x, y, z, color='blue')
16     x_range = np.linspace(0, 1, 50)
17     y_range = np.linspace(0, 1, 50)
18     X, Y = np.meshgrid(x_range, y_range)
19     Z = func((X, Y), *popt)
20     ax.plot_surface(X, Y, Z, color='red', alpha=0.5)
21     ax.set_xlabel('X')
22     ax.set_ylabel('Y')
23     ax.set_zlabel('Z')
24     plt.show()
```

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. Бесстремьянная, Г. Е. Применение ядерных и параметрических регрессий для оценки влияния страховых медицинских организаций на качество региональных систем здравоохранения [Текст] / Г. Е. Бесстремьянная, 2015. - 18 с.
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 с.