Федеральное государственное автономное

образовательное учреждение

высшего образования

«СИБИРСКИЙ ФЕДЕРАЛЬНЫЙ УНИВЕРСИТЕТ»

Институт Космических и информационных технологий

институт

Кафедра «Информатика»

кафедра

**ОТЧЕТ ПО САМОСТОЯТЕЛЬНОЙ РАБОТЕ**

Исследование непараметрической оценки прямой регрессии

тема

Преподаватель \_\_\_\_\_\_\_\_\_\_\_ А.С. Михалев

подпись, дата инициалы, фамилия

Студент КИ15–16Б, 031510065 \_\_\_\_\_\_\_\_\_\_\_ Т.В. Радионов

номер группы, зачетной книжки подпись, дата инициалы, фамилия

Красноярск 2018

# Цель исследования

Цель данного задания провести исследование непараметрической оценки прямой регрессии путем реализации компьютерной программы.

# Теория

Регрессией называют первый начальный условный момент [1]:

|  |  |
| --- | --- |
|  | (1) |

Оценка регрессии строится на основе серии измерений выхода и входа объекта [2]

|  |  |
| --- | --- |
|  | (2) |
|  |  |

Ядро нормировано на 1 на системе экспериментальных точек.

Нормированность ядер приводит к условию , которое говорит о существовании полосы, за пределы которой не выходит непараметрическая оценка регрессии.

Усечённостьнормированных ядер(в силу усечённости ядра)позволяет при построении оценки регрессии в каждой фиксированнойточке *x* учитывать только несколько близлежащих значений и не "перелопачивать" всю выборку.

Основное влияние на оценку регрессии оказывает положительная константа *c*, но зависимость от *c* при возрастании *n* ослабевает. Форма ядра усеченная параболическая. Константа *c*, определяющая коэффициент размытости, вычисляется по выборке путём минимизации эмпирических показателей (характеризующих наилучшее сглаживание экспериментальных данных).

Считаем, что выборке измерения входа находятся на равных расстояниях друг от друга, а объем выборки фиксирован. Перейдем от размерного параметра *c* (его размерность, обратная размерности *x*) к безразмерному :

|  |  |
| --- | --- |
|  | (3) |

Оценка регрессии приобретает вид:

|  |  |
| --- | --- |
|  | (4) |
|  |  |

При  0 оценка регрессии *n* (*x*) не зависит от *x*. Такой вариант, хотя и редко, но возможен. Выбранный вход объекта не оказывает влияния на выход объекта.

При 1 оценка регрессии *n* (*x*) точно проходит через экспериментальные точки, т. е. оценка не осуществляет сглаживания экспериментальных дынных. Такой вариант тоже возможен, если сигнальная часть выхода объекта не зашумлена помехой.

При наличии помех в выходе объекта оценка должна сглаживать экспериментальные дынные. Этот наиболее распространённый вариант соответствует параметру , находящемуся внутри интервала [0; 1]. Для его вычисления необходимо строить критерии оптимальности [3, 4].

# Описание задачи

Для построения оценки регрессии будут использоваться входные данные X на задаваемом вручную интервале и шагом. Параметры для функций также могут задаваться вручную. Коэффициент beta тоже необходимо задавать вручную (рисунок 1). Также будет описана функция Y для линейной, нелинейной и синусоидальной модели, а также функция Y с помехой. Помеха будет рассчитываться по закону распределения центральной предельной теоремы с генерацией случайных чисел включительно.

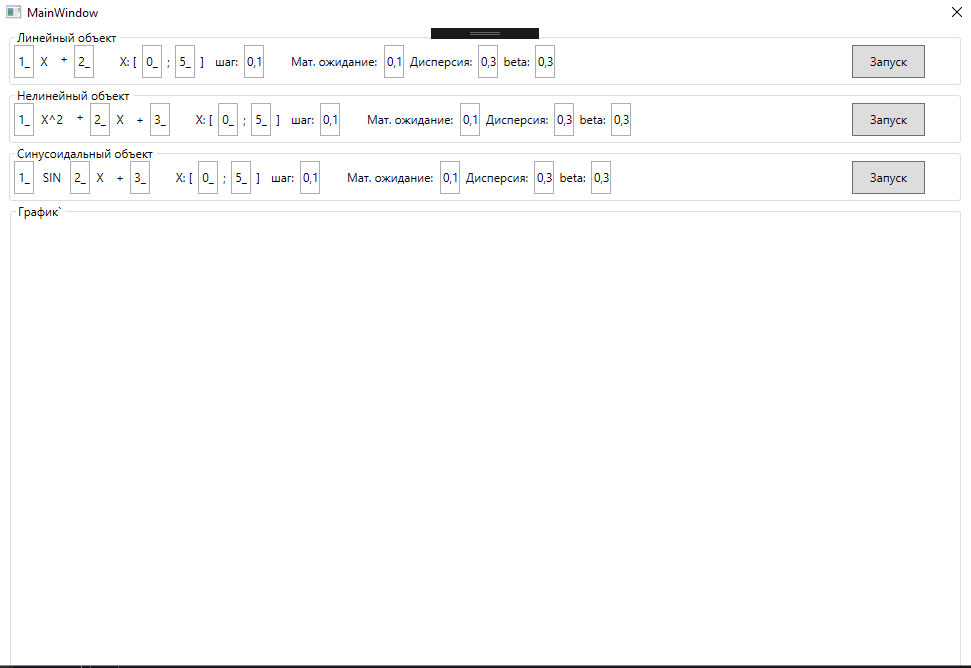


Рисунок 1 – Интерфейс программы

Далее будет описано четыре вида ядра: прямоугольное, треугольное, параболическое и кубическое. Для каждого из ядер будет составлена кривая оценки регрессии, где визуально необходимо будет установить наиболее оптимальные ядра.

# Ход работы

В ходе исследования будет изменяться только коэффициент beta для установления изменений оценки регрессии. Параметры M, D, X\_start, X\_end, X\_step везде будут одинаковыми.

Приступим к исследованию линейной функции .

В первом опыте был взят beta = 0.3. Как видно на рисунке 2, треугольное (далее – K1), параболическое (далее – K2) и кубическое (далее – K3) ведут себя практически одинаково и различия незначительны. График прямоугольного (далее – K0) отличается от вышеуказанных и он менее точен.

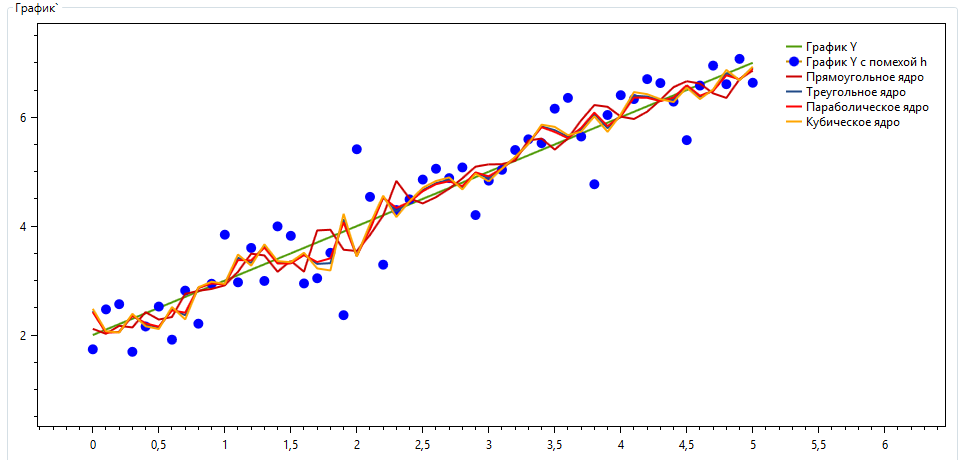


Рисунок 2 – Опыт 1

Проведем опыт 2 и изменим коэффициент beta = 0.1. Как видно и из рисунка 2 – все ядра показали различный, но при этом схожий результат.

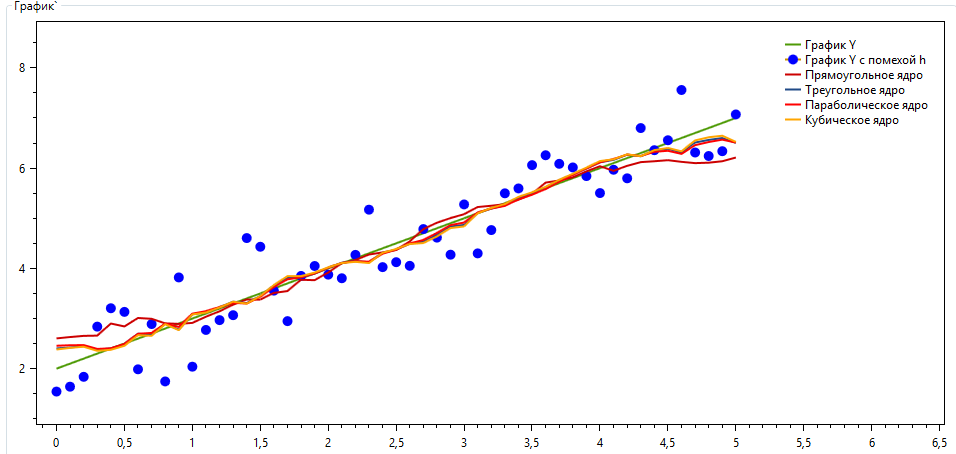


Рисунок 3 – Опыт 2

Проведем опыт 3 и увеличим beta = 0.5. Ядра K1, K2, K3 полностью совпали на рисунке 4, K0 отличается, однако его показания не такие неточные, как было при beta = 0.3.

Примечательно, что при увеличении beta графики становятся ломанного вида, а при уменьшении более гладкие и близкие к исходной функции.

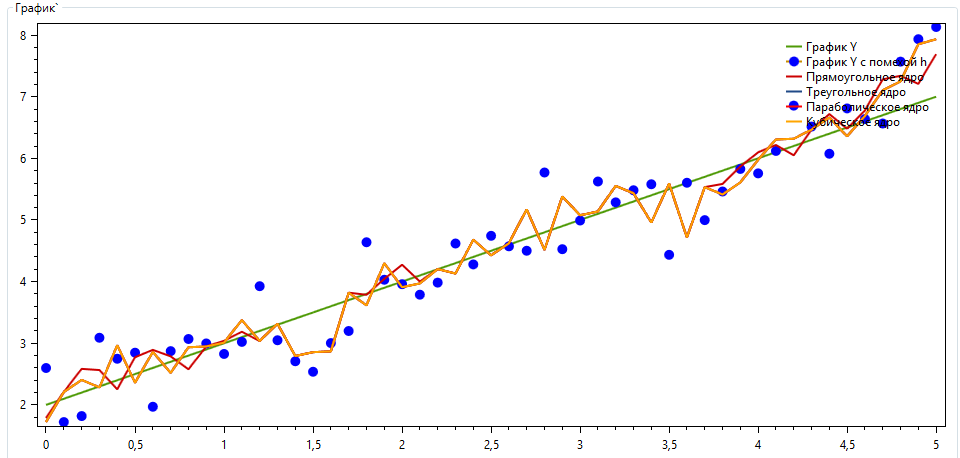


Рисунок 4 – Опыт 3

Проведем теперь исследования для нелинейной функции .

В опыте 4 возьмем beta = 0.3. Как видно на рисунке 5, разброс точек невелик, а все ядра показали почти одинаковый результат.

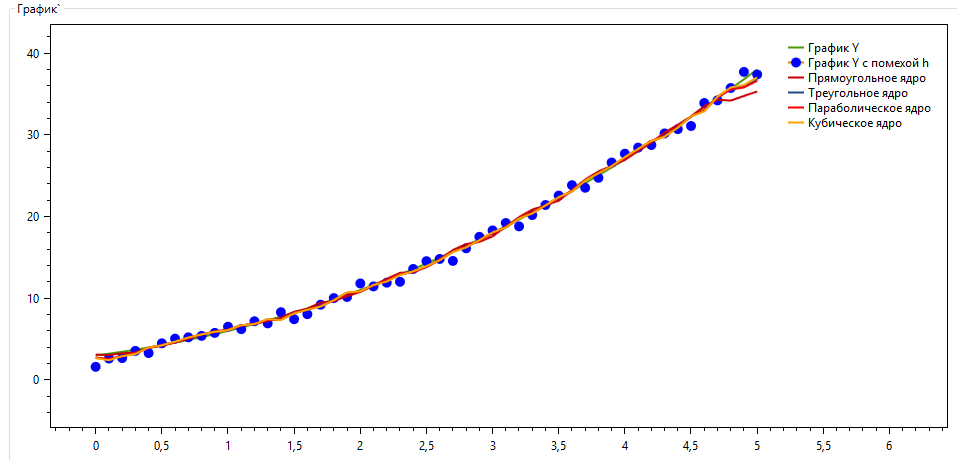


Рисунок 5 – Опыт 4

В опыте 5 уменьшим beta = 0.1. Как видно на рисунке 6, ядро K0 показало худшие результаты, остальные ядра ведут себя схоже, однако результат хуже, чем при beta = 0.3.

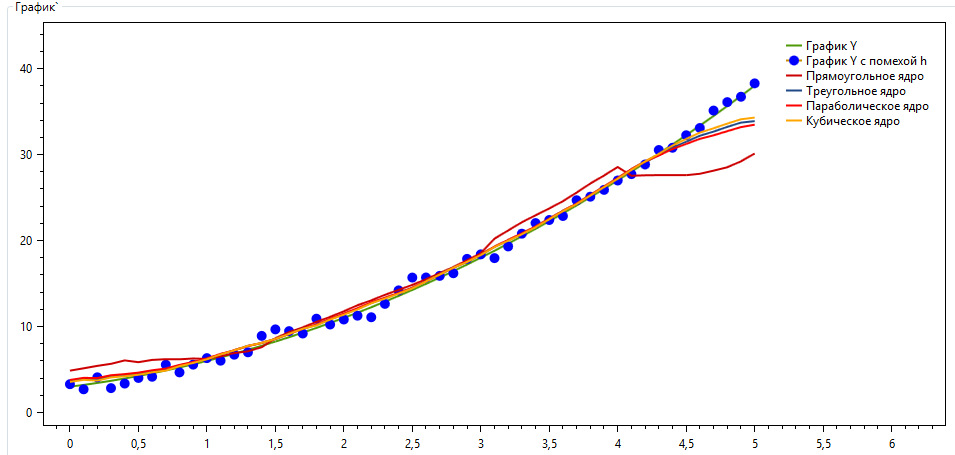


Рисунок 6 – Опыт 5

В опыте 6 увеличим beta = 0.5. На рисунке 7 видно, что K0 при X = 4 имеет отклонение, однако в остальных случаях ядра ведут себя одинаково схоже.

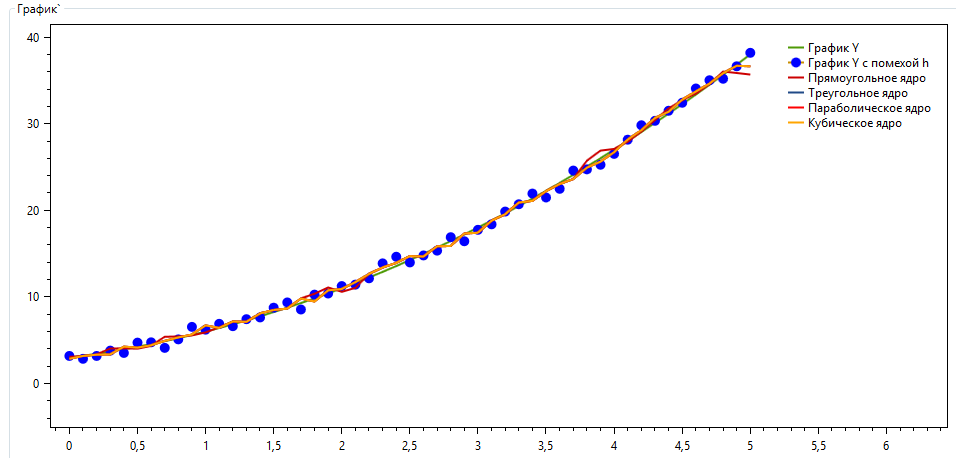


Рисунок 7 – Опыт 6

Проведем теперь исследования для синусоидальной функции .

В опыте 7 возьмем beta = 0.3. В целом оценка для всех ядер не такая точная, как для предыдущей функции, однако вполне приемлема. K0 вновь отличен от остальных ядер (рисунок 8).

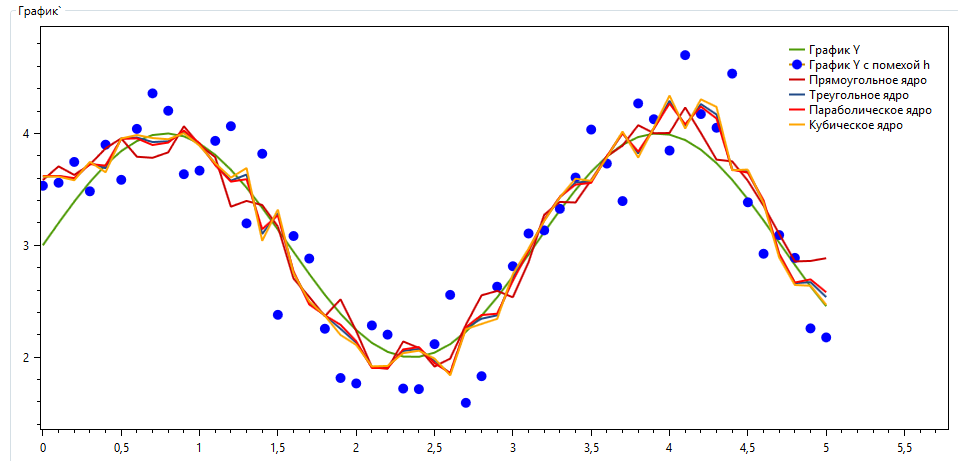


Рисунок 8 – Опыт 7

Проведем опыт 8 с beta = 0.1. На рисунке 9 видно, что ядро K0 значительно отличается от остальных и оценка регрессии неточна настолько, что сложно было бы сказать, что это график синуса по ней.

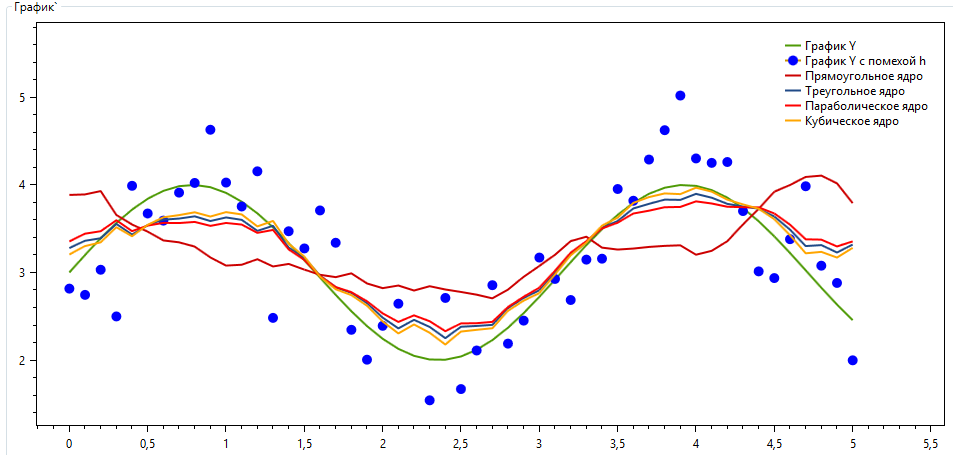


Рисунок 9 – Опыт 8

Проведем последний опыт 9 с beta = 0.5. Как видно на рисунке 10, графики оценок приобретают ломанный характер, K0 отличается при некоторых значениях X, однако в основном графики наложились друг на друга.

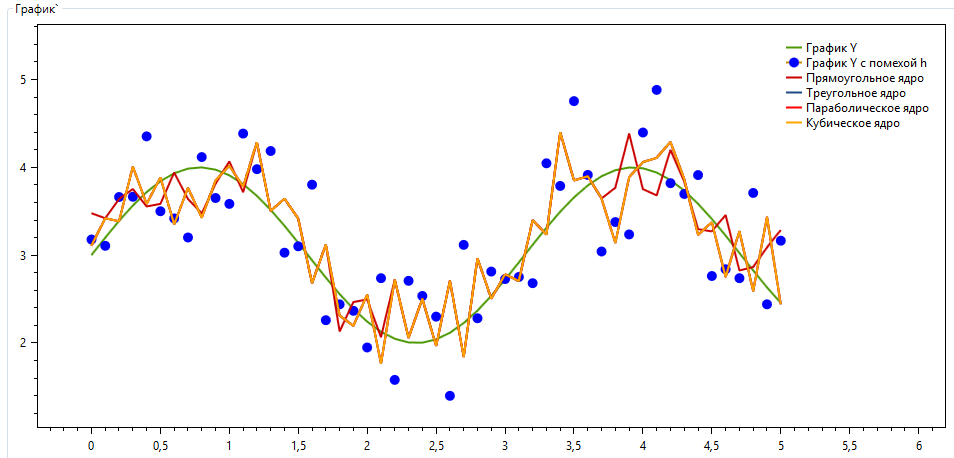


Рисунок 10 – Опыт 9

# Выводы

В ходе исследования были сделаны следующие выводы: ядерные функции треугольного, параболического и кубического вида ведут себя практически одинаково во всех опытах и невозможно сказать, какая из них лучшая. Однако в большинстве опытов прямоугольное ядро показывало результаты хуже остальных ядер. Также примечателен тот факт, что при увеличении коэффициента beta характер графиков становился ломанным и близким к точкам, что показывает правдивость утверждения о том, что при beta = 1 ломаная будет проходить через точки и не сглаживать полученные данные. При уменьшении beta график становится более гладким и схожим с графиком исходной функции без помехи.

В целом исследования проведены и зависимость оценки от beta установлена, а также определены оптимальные и не оптимальные ядра.

# Список использованных источников

x

|  |  |
| --- | --- |
| 1. | Тарасенко Ф.П. Непараметрическая статистика. |
| 2. | Хардле В. Прикладная непараметрическая регрессия. Файл. |
| 3. | Рубан А.И. Методы анализа данных. Красноярск: Центр обучающих систем ИнТК СФУ, 2012. Электронный ресурс. |
| 4. | Рубан А.И., Кузнецов А.В. Методы анализа данных. Учебное пособие по циклу расчетно-графических работ по курсу "Методы анализа данных". Красноярск: 2007. |

x