**1.3 Непараметрическое оценивание с помощью ядер**

**1.1 Одномерная аппроксимация плотности с помощью ядер**

Исследуем метод оценивания **kde** с прямоугольными ядрами. Для этого зададим достаточный объем обучающей выборки из 1000 примеров.

На рисунке 1 представлен график примеров обучающей выборки.

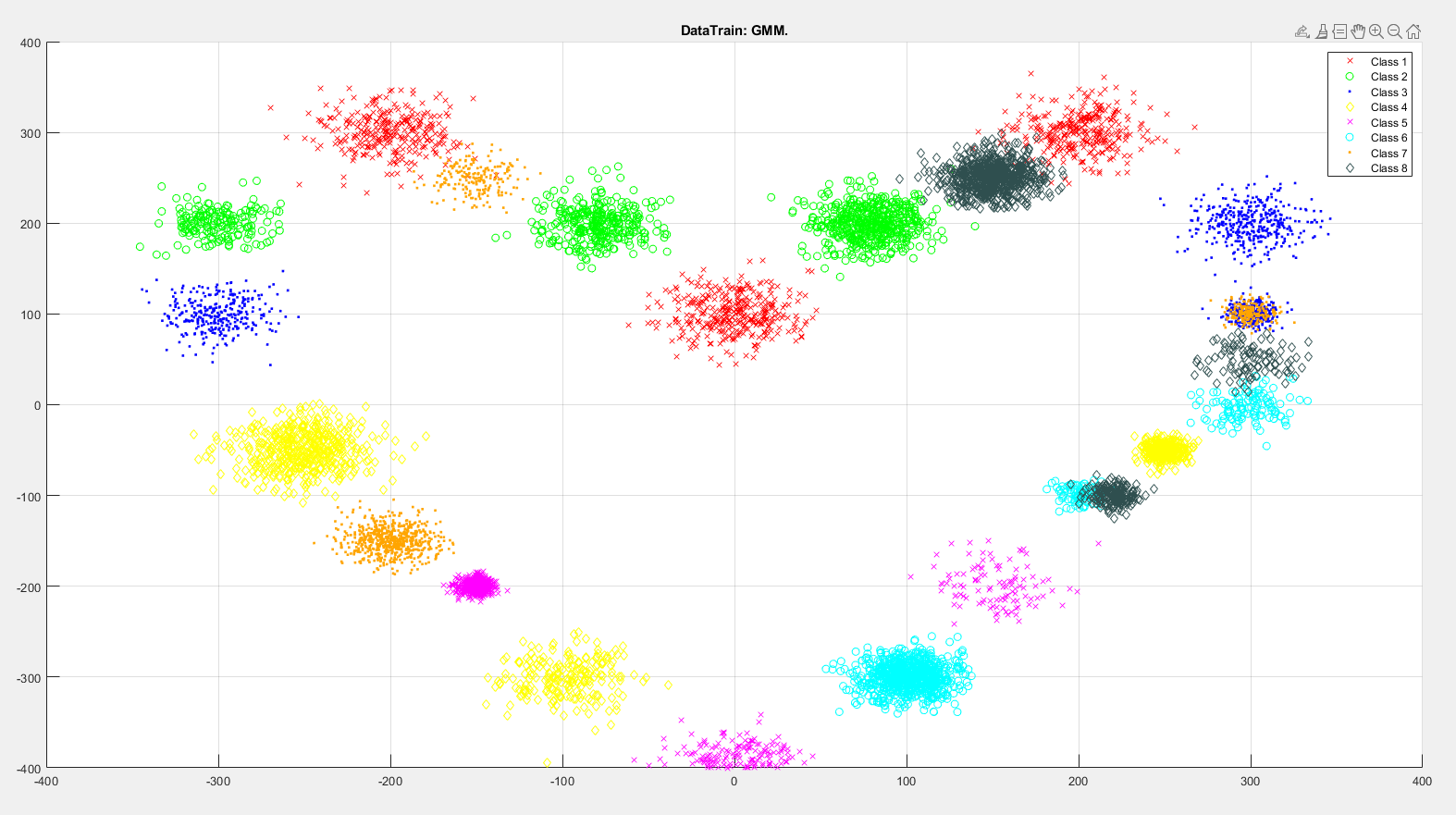


Рисунок 1 – График примеров обучающей выборки

**1.1.2 Исследуем влияние ширины окна h на качество оценивания**

1. Наиболее корректное оценивание (h1 = 1, h2 = 1)

Среднее суммарное отклонение между плотностями **64.6432**

Одномерные плотности для обеих координат (В качестве примера приведен первый случай):

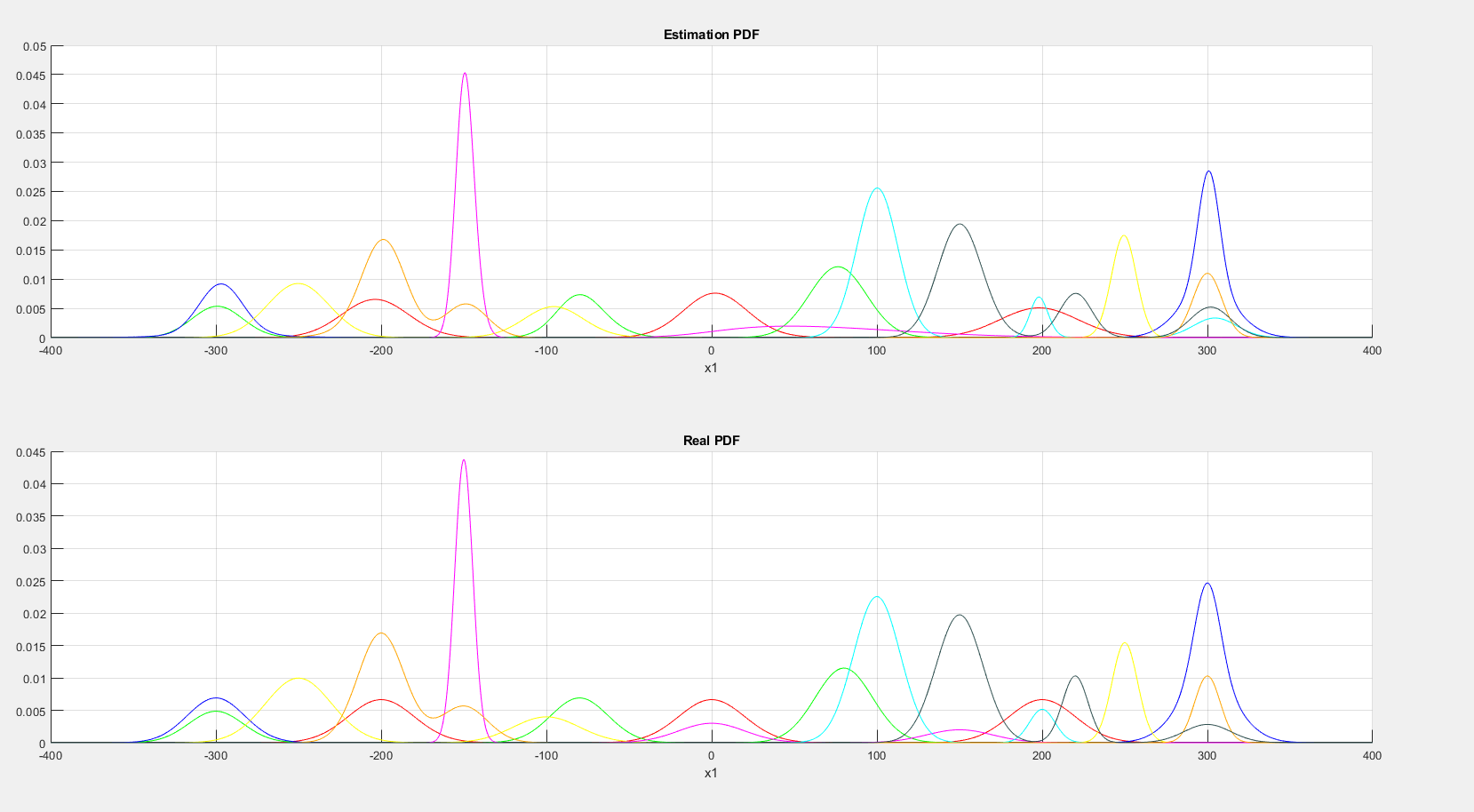


Рисунок 2 – Одномерная плотность для X1

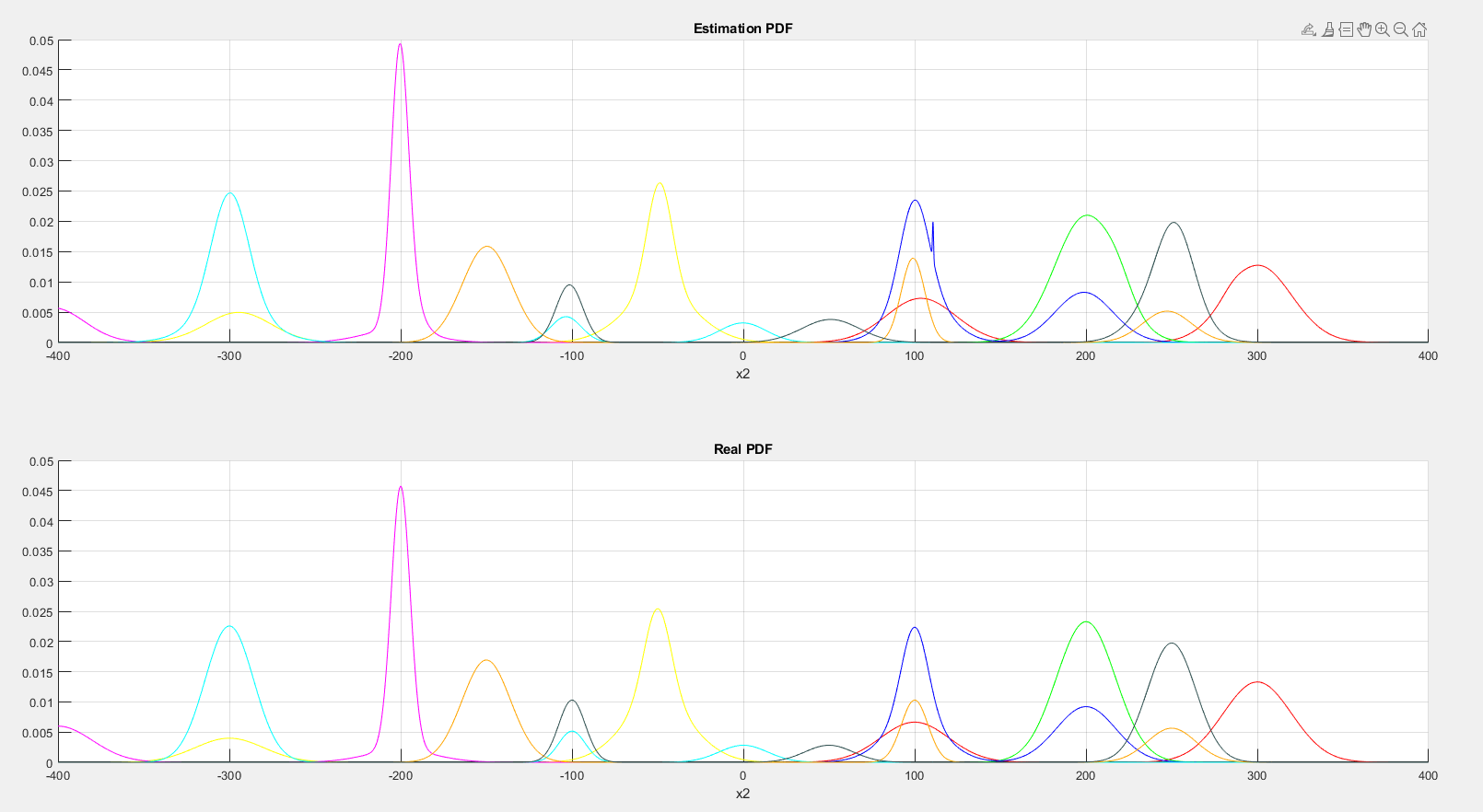


Рисунок 3 – Одномерная плотность для X2

1. Оценивание в виде отдельных бугорков (h1 = 0.1, h2 = 0.1)

Среднее суммарное отклонение между плотностями **168.9471**

1. Оценивание чересчур гладкой функцией (h1 = 3, h2 = 3)

Среднее суммарное отклонение между плотностями **115.8653**

При использовании ширины окна h = 1, мы достигаем наивысшей точности приближения исходного распределения. Однако, если выбрать меньшее значение h, плотность становится слишком неровной, а при больших значениях h - слишком гладкой. Это означает, что для наилучшей классификации важно подбирать оптимальное значение ширины окна, учитывая исходное распределение данных.

Построим график зависимости качества аппроксимации от параметров **h** для признаков **x1**, **x2**, рассмотрев диапазоны h от 0.01 до 10.

|  |  |
| --- | --- |
| **h** | Ошибка аппроксимации |
| 0.1 | 168.9471 |
| 1 | 64.6432 |
| 2 | 87.4068 |
| 3 | 115.8653 |
| 4 | 134.9643 |
| 5 | 141.4574 |
| 6 | 146.5771 |
| 7 | 147.8412 |
| 8 | 147.6574 |
| 9 | 148.7825 |
| 10 | 148.0654 |

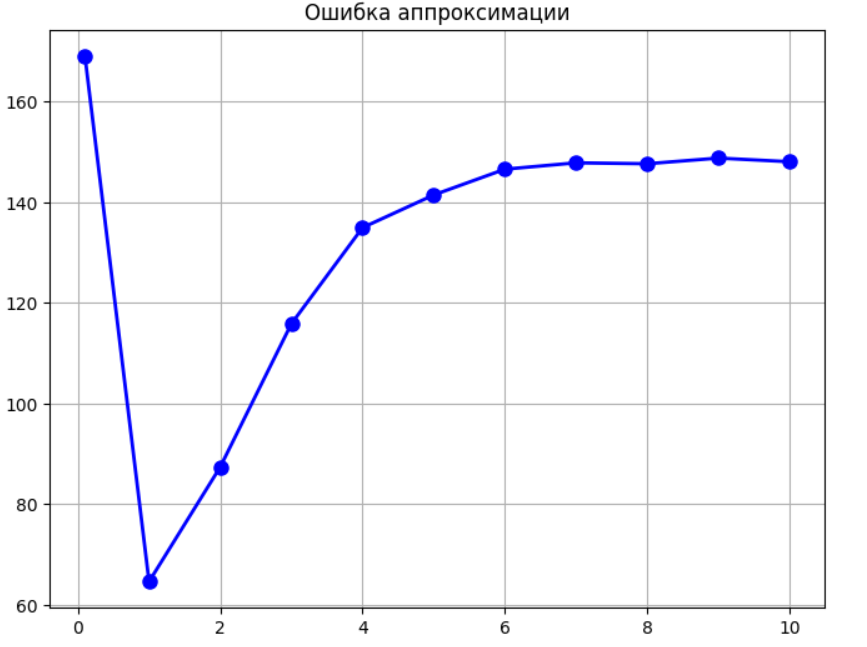


Рисунок 4 – Ошибка аппроксимации

Значения **h** исходя из процедур параметрического выбора:

|  |  |  |  |
| --- | --- | --- | --- |
|  | h1 | h2 | Ошибка аппроксимации |
| hopt1 | 1.1543 | 0.6543 | 69.0741 |
| hopt2 | 0.96432 | 0.47651 | 65.6108 |
| hMLCV | 1 | 1 | 64.6432 |

Использование специальных алгоритмов для подбора параметров ширины окна показало хорошие результаты аппроксимации - **hMLCV** подобрало наилучшее значение, которое было получено вручную (h1 = 1, h2 = 1).

Изменяя объем обучающей выборки в логарифмическом диапазоне от 100 до 10 тысяч примеров на каждый класс, попробуем для каждого признака подобрать наилучшие значение h1, h2.

Значения h в зависимости от объема обучающей выборки:

|  |  |  |  |
| --- | --- | --- | --- |
| **Объем обучающей выборки** | **h1** | **h2** | **Ошибка аппроксимации** |
| 10 | 2.382 | 1.309 | 111.858 |
| 100 | 1.523 | 0.797 | 76.875 |
| 200 | 1.354 | 0.707 | 68.422 |
| 300 | 1.332 | 0.657 | 99.526 |
| 400 | 1.168 | 0.600 | 60.855 |
| 500 | 1.127 | 0.586 | 58.670 |
| 800 | 1.117 | 0.546 | 54.476 |
| 1600 | 0.884 | 0.462 | 48.675 |
| 2000 | 0.858 | 0.441 | 46.807 |
| 3200 | 0.780 | 0.403 | 43.359 |
| 6000 | 0.670 | 0.351 | 40.590 |

Оптимальное число примеров 3200, т.к расчет ошибок производится довольно быстро и размер ошибки аппроксимации удовлетворителен

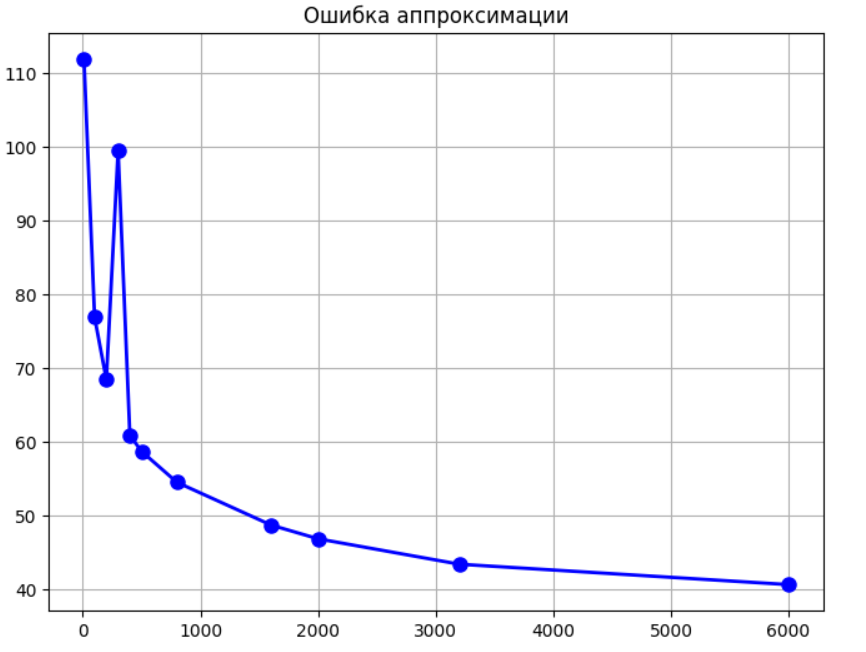


Рисунок 5 – Ошибка аппроксимации

Исследуем методы оценивания с помощью других ядер (гауссово, треугольное и др.). Построим типовые графики плотности, определим наиболее удачные типы ядер, значения **h**, алгоритмы нахождения **h**, объем выборки, достаточный для оценивания плотности с заданной точностью

|  |  |
| --- | --- |
| Тип ядра | Ошибка аппроксимации |
| Прямоугольное (равномерное) | **45.0975** |
| Гауссово | **46.9654** |
| Экспоненциальное (Лапласа) | **101.9642** |
| Коши | **112.3421** |
| Треугольное | **74.1864** |
| Восстанавливающий фильтр | **124.7531** |

Наиболее удачными типами ядер оказались прямоугольное и Гауссово.

Оптимальный размер обучающей выборки составляет 3200 примеров. При увеличении количества примеров ошибка аппроксимации уменьшается, но это уменьшение незначительно. Этот выбор размера обучающей выборки основан на производительности компьютера и уровне ресурсов, доступных для выполнения вычислений.

**2. Двумерная аппроксимация плотности с помощью ядер**

Рассмотрим метод многомерного оценивания в виде суммы произведений ядер.