

МИНИСТЕРСТВО ОБРАЗОВАНИЯ И НАУКИ РОССИЙСКОЙ ФЕДЕРАЦИИ

ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ АВТОНОМНОЕ ОБРАЗОВАТЕЛЬНОЕ

УЧРЕЖДЕНИЕ ВЫСШЕГО ОБРАЗОВАНИЯ

«Санкт-Петербургский национальный исследовательский университет

информационных технологий, механики и оптики»

Факультет информационных технологий и программирования

Кафедра информационных систем

Интеллектуальные системы и технологии

Лабораторная работа 1

Выполнила студентка группы № М3403

Слобода Полина

САНКТ-ПЕТЕРБУРГ

2020

Задача регрессии – необходимо выдать не дискретную метку класса, а одно из значений аналогового множества значений.

От задачи классификации переходим к задаче регрессии.

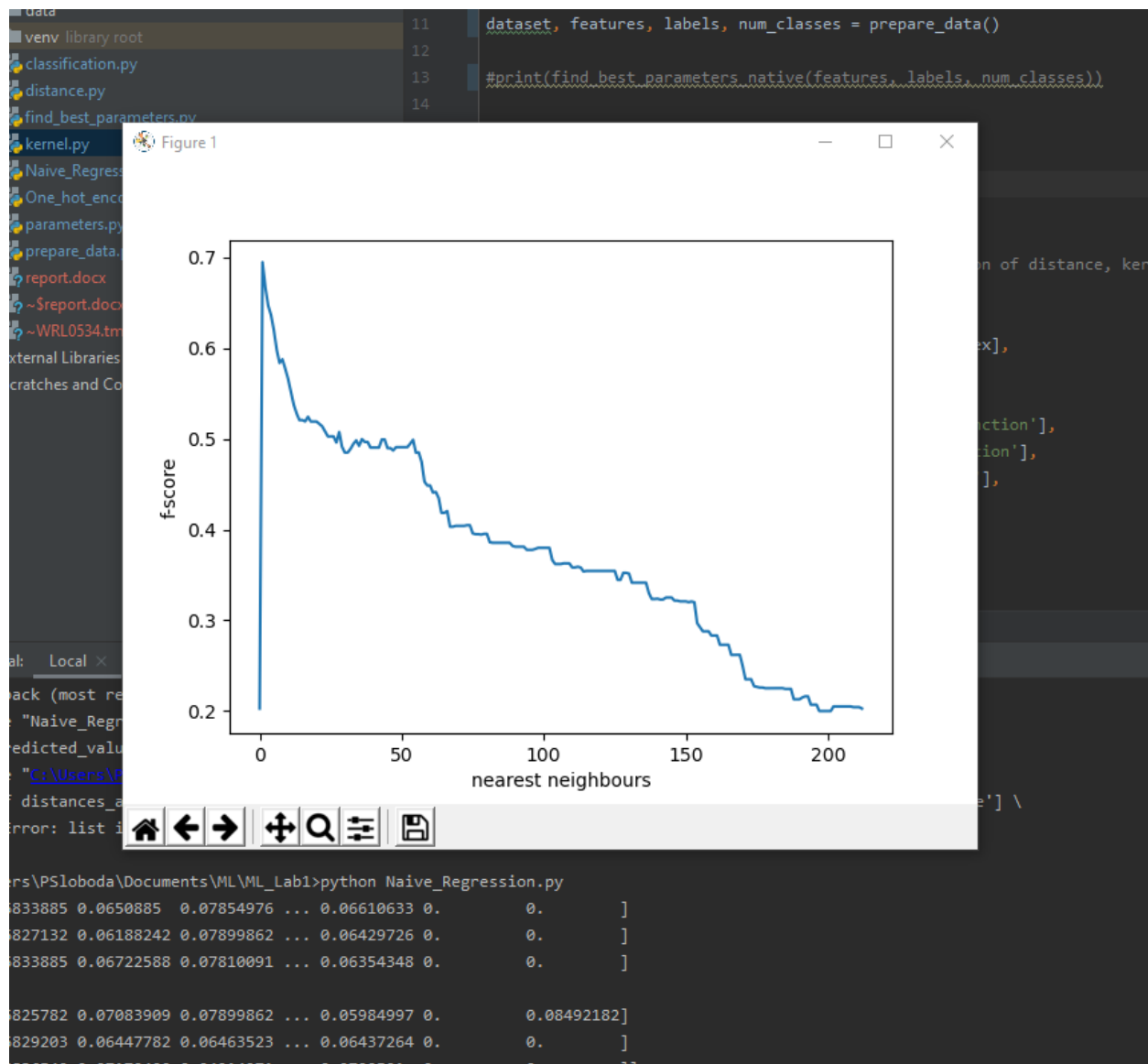
Каждое значение класса представляется одним числом. Во время предсказания полученный ответ округляется до ближайшего целого числа.

Скриншот части вывода данных долгой обработки данных для поиска лучших параметров:

На скриншоте можно заметить, что для нескольких значений параметров f -мера одинакова и максимальна. При этом среди всех лидеров функция расстояния манхеттеновская, а метод окна – взятие радиуса по ближайшим соседям (параметр 1, что логично). Функция ядра может быть различной, максимальные значения получились при вариантах 'uniform', 'triangular', 'epanechnikov', 'quartic', 'triweight', 'tricube', 'cosine'. Ядра 'gaussian', 'logistic', 'sigmoid' показали результат хуже.

Зафиксируем лучшими параметрами комбинацию **'manhattan'**, **'triweight'**, **'variable'**.

После получения лучшей комбинации для данного датасета производится построение графика зависимости f-меры от числа ближайших соседей (так как в данном случае используется метод 'variable'):

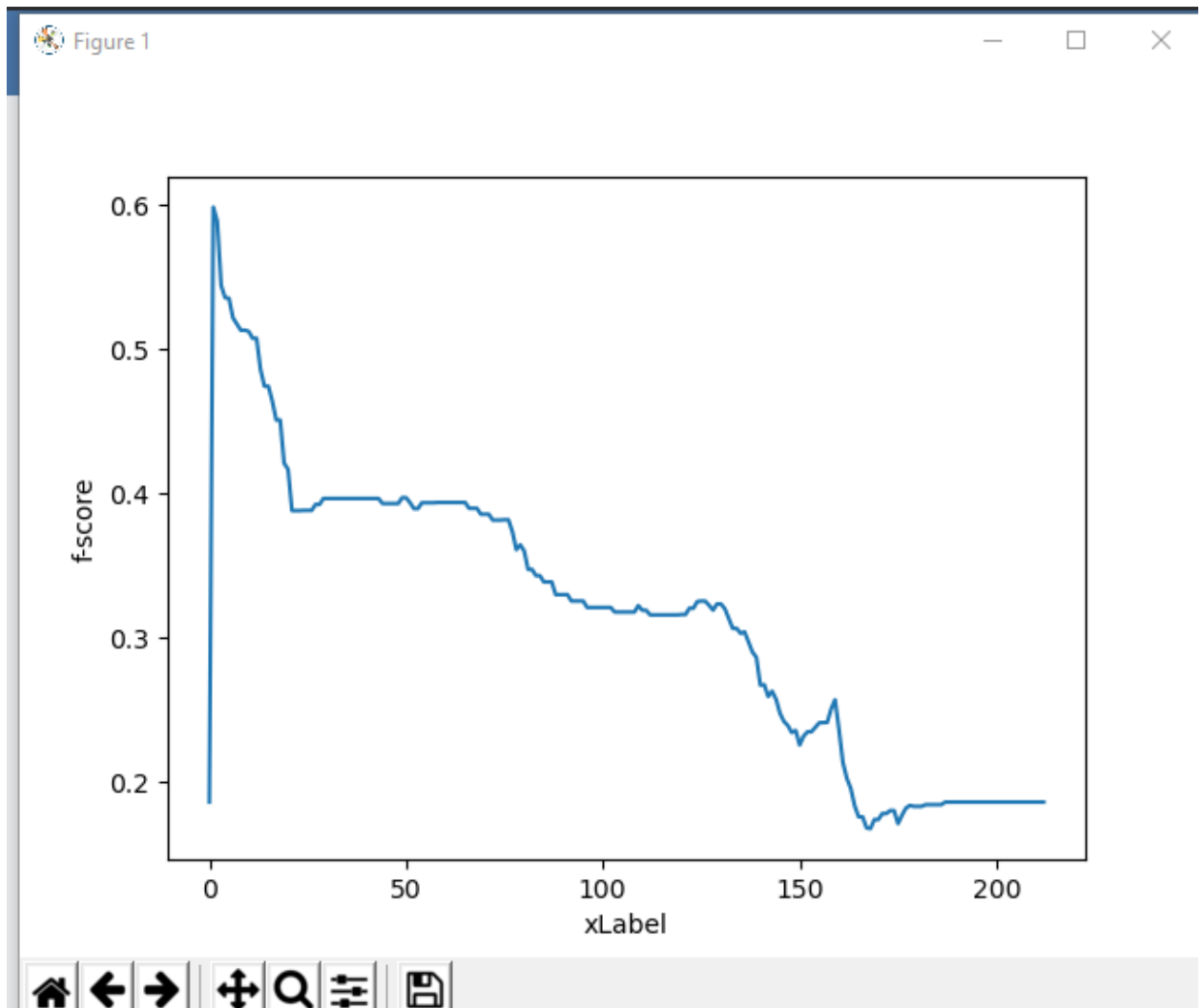


На графике видно, что при нуле соседей f-мера нулевая, при одном – максимальна, далее убывает и стремится к нулю, что вполне соответствует здравому смыслу. Параметр окна 1, как и в поиске лучших параметров, так и в проверке на графике, выдает максимальную f-меру.

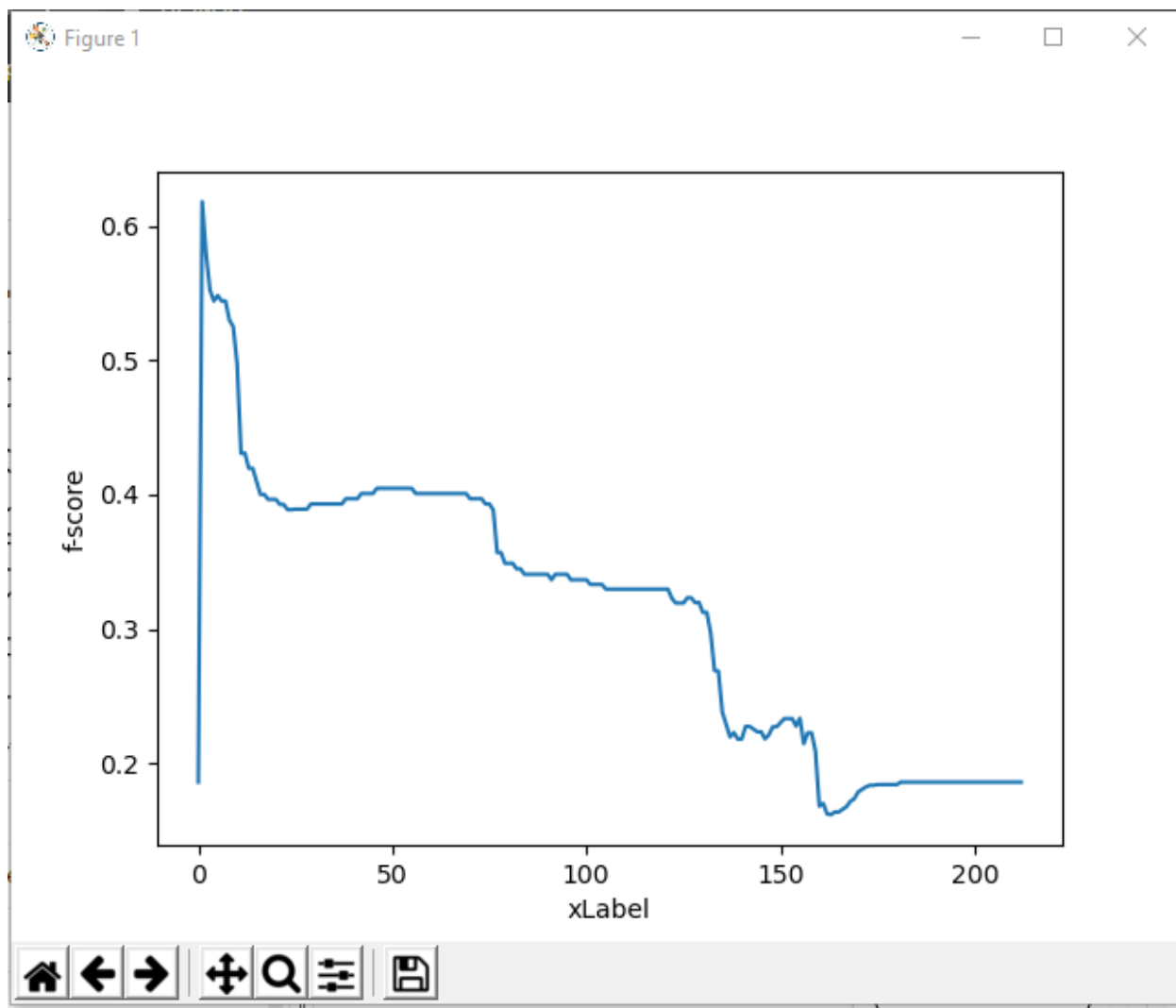
Ради интереса построила еще график с использованием другой функции ядра ('manhattan', 'uniform', 'variable'):

В данном случае одинаковых мер получилось гораздо больше. Функции расстояния в лучших комбинациях встречаются все три, функции ядра тоже все (все 10 вариантов). Неизменным остался метод подсчета окна по количеству соседей с параметром 1.

Выберем, к примеру, одну из наилучших комбинаций **'manhattan', 'sigmoid', 'variable'**



И тоже сравнить график с получаемым от другой функции расстояний (**'euclidean', 'sigmoid', 'variable'**):



Разница в использовании различных функций расстояния небольшая.