МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ РОССИЙСКОЙ ФЕДЕРАЦИИ

федеральное государственное бюджетное образовательное учреждение

высшего образования

«УЛЬЯНОВСКИЙ ГОСУДАРСТВЕННЫЙ ТЕХНИЧЕСКИЙ УНИВЕРСИТЕТ»

Кафедра «Измерительно-вычислительные комплексы»

«Методы искусственного интеллекта»

Отчёт по лабораторной работе №4

Выполнил:

студент группы ИСТбд-42

Енилов Дмитрий

Ульяновск

2022

Лабораторная работа № 4. Метод к – ближайших соседей (k-NN)

Пошаговое описание

1. Создать симулированный набор данных и записать его на диск в виде csv файла с параметрами.



Рис. 1 Код для создания файла данных

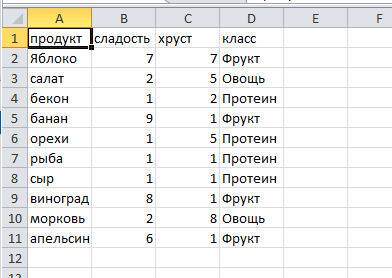


Рис. 2 Файл данных в программе excel

2. Запрограммировать метрический классификатор по методу k-NN. Для проверки решить ту же задачу методом k-NN библиотеки sklearn.



Рис. 3 Код метрического классификатора по методу k-NN

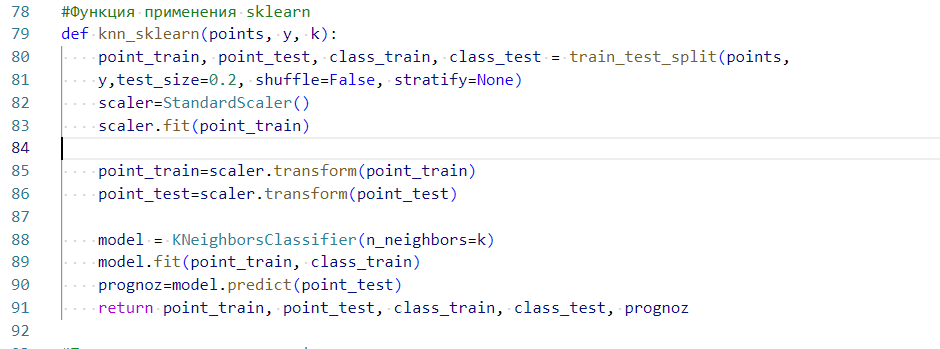


Рис. 4 Код метрического классификатора по методу k-NN с библиотекой sklearn

3. Прочитать сгенерированный набор данных. Настроить классификатор. Провести эксперимент по классификации с контролем для подготовленных примеров.

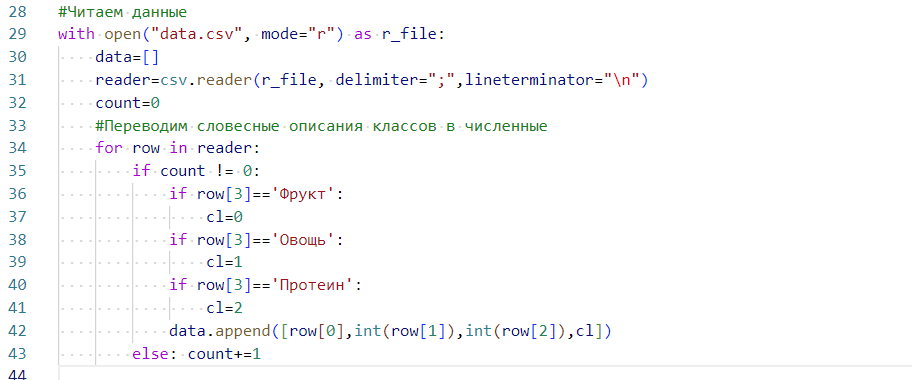


Рис. 5 Код чтения данных

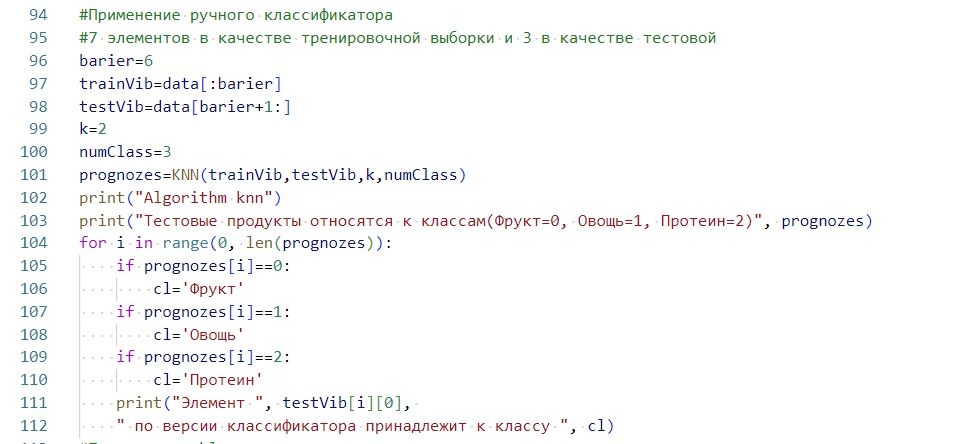


Рис. 6 Код прогнозирования классов данных с помощью ручного классификатора

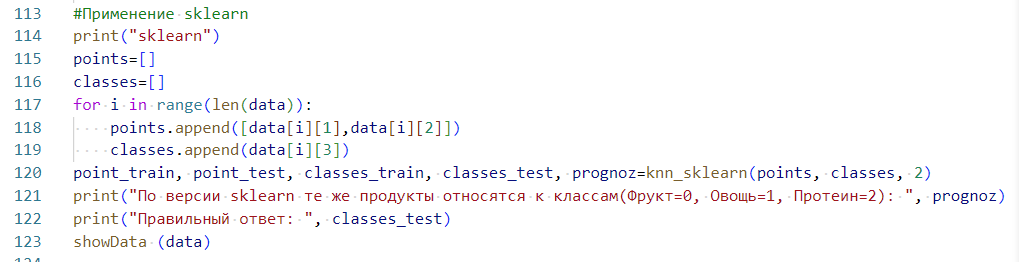


Рис. 7 Код прогнозирования классов данных с помощью классификатора библиотеки sklearn

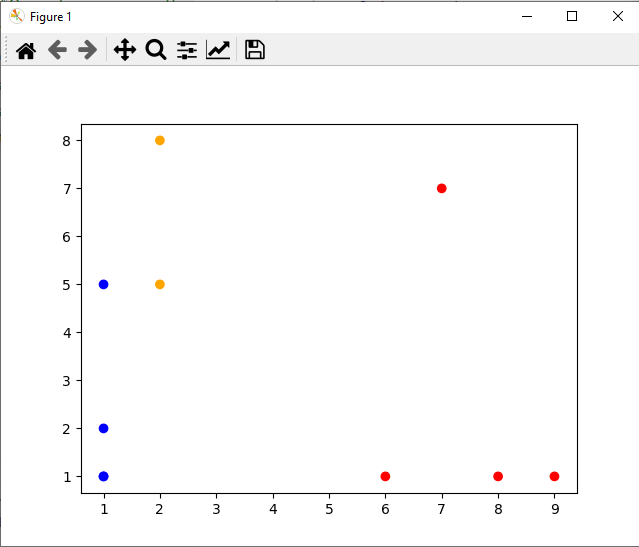


Рис. 8 Отображение продуктов и их классов

Ось x - сладость, ось y – хруст, красным цветом обозначены фрукты, рыжим овощи, а синим – протеины. Результаты показаний классификаторов при тренировочной выборке в 7, тестовой в 3 и количестве ближайших соседей k = 2:

**Вывод ручного классификатора**

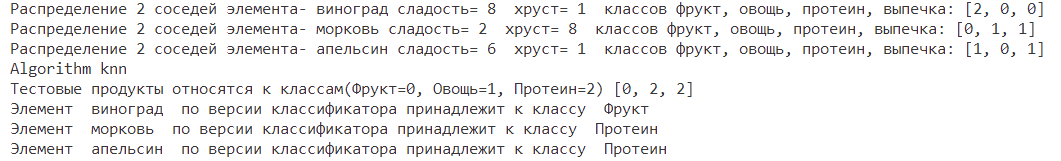


Рис. 9 Вывод в консоль ручн. клас. при k=2

Так как выборка данных достаточно мала, при чем учитывались только ближайшие 2 соседа, точность ручного алгоритма мала. Это объясняется еще и тем, что у соседей не предусмотрены весовые коэффициенты. Они только отбираются по весовым коэффициентам, но после отбора - весов нет. В итоге он смог угадать лишь одного из двух. Но если учитывать лишь самого ближайшего соседа (k=1):

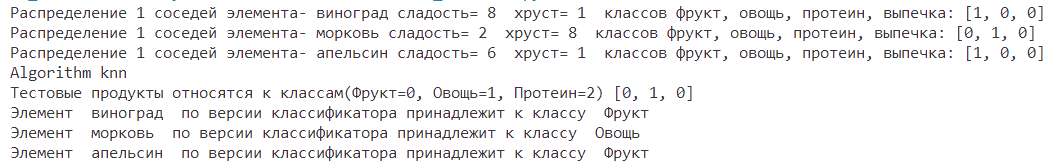


Рис. 10 Вывод в консоль ручн. клас. при k=1

Точность выше, он смог угадать всех, так как здесь учитывался лишь самый ближайший сосед.

**Вывод классификатора sklearn**



Рис. 11 Вывод в консоль клас. sklearn при k=2

Те же продукты из тестовой выборки при k=2 ближайших соседей были предсказаны очень точно, думаю это из-за учитывания близости соседей на всех этапах прогноза.



Рис. 12 Вывод в консоль клас. sklearn при k=1

При количестве учитываемых соседей в 1 результат тот же, стопроцентная точность.

5. Ввести в набор данных и примеры продукты еще одного класса (возможно изменив набор параметров) и повторить эксперимент.

Я добавил класс – выпечку, а также дополнил данные до 19 значений:



Рис. 13 Код создания файла дополненных данных



Рис. 14 Файл данных в программе excel

Код чтения данных не приведен, так как он практически не изменился.

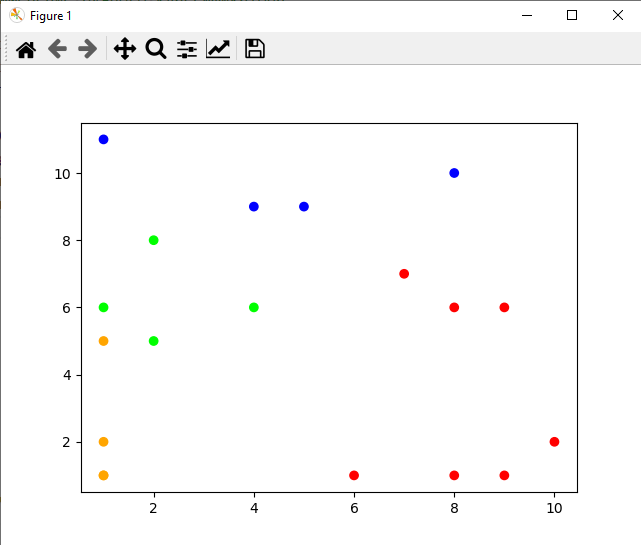


Рис. 15 Отображение продуктов и их классов

Ось x - сладость, ось y – хруст, красным цветом обозначены фрукты, рыжим - протеины, синим – выпечка, а зеленым - овощи. Результаты показаний классификаторов при тренировочной выборке в 15, тестовой в 4 и количестве ближайших соседей k = 3:

**Вывод ручного классификатора**

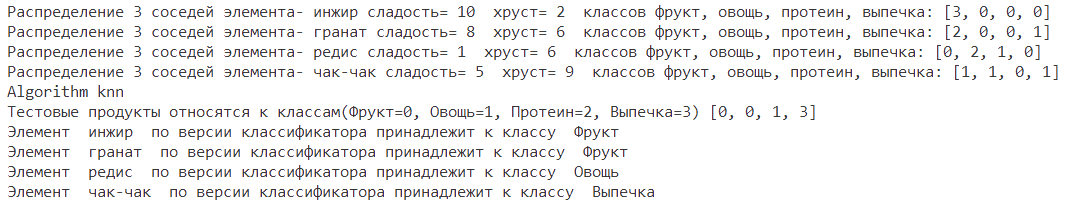


Рис. 16 Вывод в консоль ручн. клас. при k=3

Точность классификатора достаточно высока из-за большего количества данных.

**Вывод классификатора sklearn**



Рис. 17 Вывод в консоль клас. sklearn при k=3

Точность не 100%, думаю это из-за недостаточного количества k соседей.



Рис. 18 Вывод в консоль клас. sklearn при k=4

При увеличении k ближайших соседей точность возрасла.