Міністерство освіти і науки України

Харківський національний університет радіоелектроніки

Кафедра Штучного інтелекту

Звіт з лабораторної роботи №4

з курсу Машинне навчання

на тему “ Реализация и исследование алгоритма KNN”

Виконав:

студент гр. ІТШІ-17-1

Гура А.О.

Прийняла:

Кулішова Н.Є.

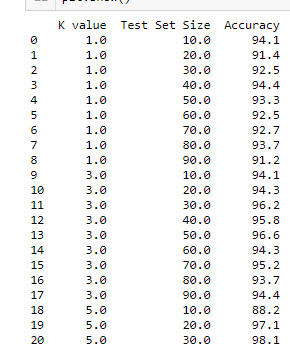
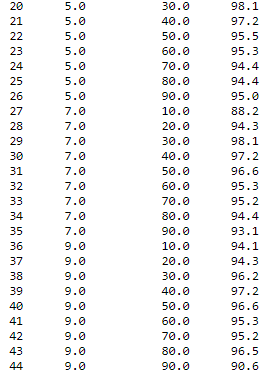
Харків 2019

**1.Цель работы:** Реализовать различные варианты алгоритма Ближайших Соседей и провести исследования зависимости точности алгоритма от различных параметров.

**2.Общее описание проделанной работы:** В ходе выполнения этой лабораторной работы я реализовал алгоритм KNN в нескольких вариантах(в классическом и взвешенном виде). По факту отличается только лишь функция для выбора ответа(голосования) и для предсказания. Данные алгоритмы были применены к таким выборкам: Iris, Wine, Breast Cancer. Для каждой из выборок я запускал алгоритм в 4 разных вариациях(обычной, с нормализацией данных, с использованием взвешенного алгоритма(взвешенного относительно расстояния), и комбинации взвешенного с нормализацией). Таким образом я получил 12 графиков. В каждом из графиков были проверены все комбинации алгоритма с различным значением К(1.3.5.7.9). и различным размером тестовой выборки(10, 20, …, 90). Итого в каждой из 12 вариаций алгоритм был запущен по 44 раза с различными параметрами.

**3.Числовые значения точности классификации:**

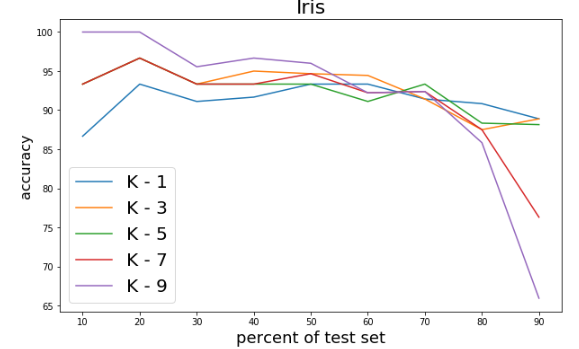
Как я и сказал выше, у меня имеется в наличии 12 таблиц 3х44. Учитывая такое их количество я продемонстрирую в отчете лишь несколько(одну), а остальные покажу на ноутбуке.

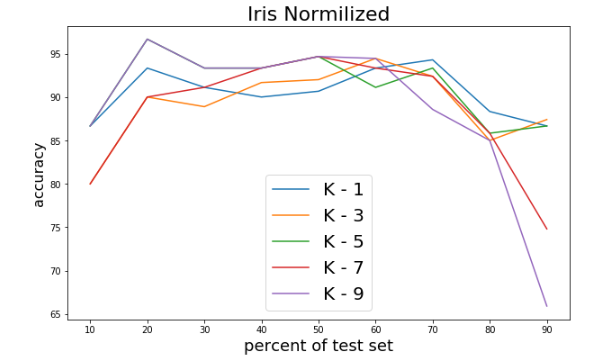
**4. Графические зависимости:**

Как я сказал выше, у меня всего построенно 12 графиков, поэтому в отчет я вставлю лишь парочку(4):

- Ирис(простой)

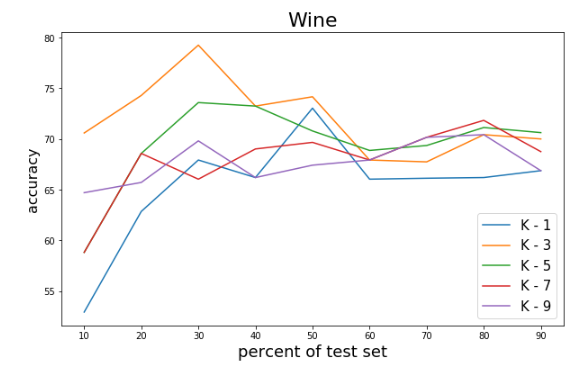


- Ирис(с нормализацией)

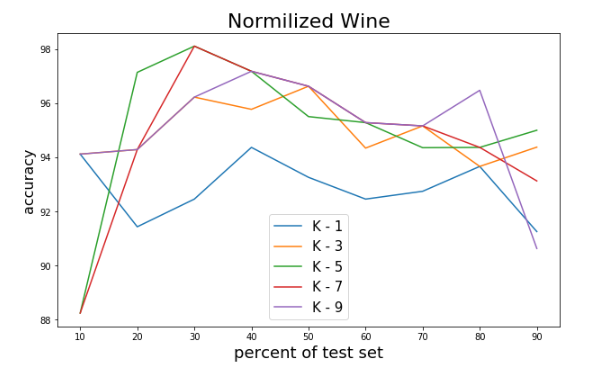


Как вы можете заметить точность при нормализации упала. По моему опыту в это работе такое произошло лишь для Ириса. Я считаю что это неудивительно, ведь тут данные по дефолту лежат в нормальном интервале(2 – 6 кажется примерно).

- Вино(простое)



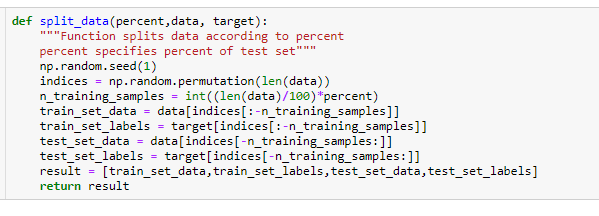
- Вино(с нормализацией)



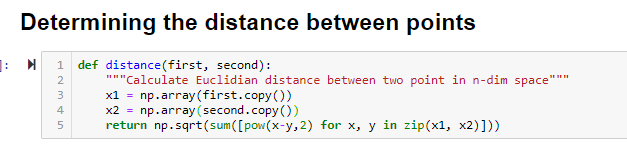
Как можно заметить тут нормализация дала нам очень хороший выигрыш в точности, я бы сказал крайне хороший.

**5. Тексты скриптов:**

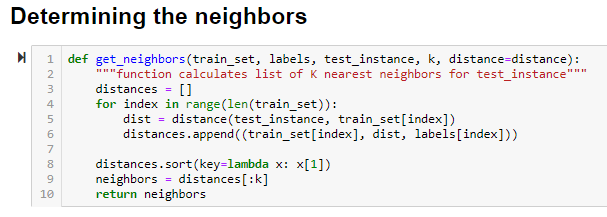
- Функция для разбиения выборки и перемешки

****

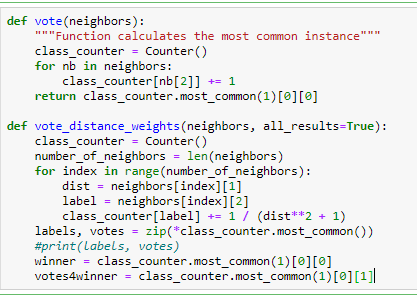
- Для расчета расстояния



- Для нахождения соседей



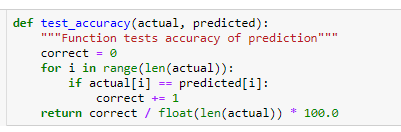
-Для голосования за значение целевой функции(обычное и взвешенное)



- Для классификации тестовой выборки



- Для оценки точности



**6. Выводы**

В ходе выполнения данной лабораторной работы я ознакомился с алгоритмом KNN и его разновидностями(классический, взвешенный по атрибутам и по расстоянию), а также реализовал и использовал на практике 1й и 3й из них.

Кроме того, я увидел влияние применения нормализации для данных. Также я увидел как изменяется точность классификации в зависимости от числа К соседей, и размера выборки для тестирования.