МІНІСТЕРСТВО ОСВІТИ ТА НАУКИ УКРАЇНИ  
Національний технічний університет   
«Харківський політехнічний інститут»  
Кафедра ГМКГ

**Лабораторна работа №1**

З дисципліни «Інтелектуальний аналіз даних»

Виконав:

Студент групи ІКМ-220 г.

Ульянов Кирило Юрійович

Перевірив:

Доц. Дашкевич А.О.

Харків 2023

**Мета роботи:** вивчення класифікації методом k-найближчих сусiдiв

**Завдання на роботу:** завантаження набору даних «Iриси Фiшера»,

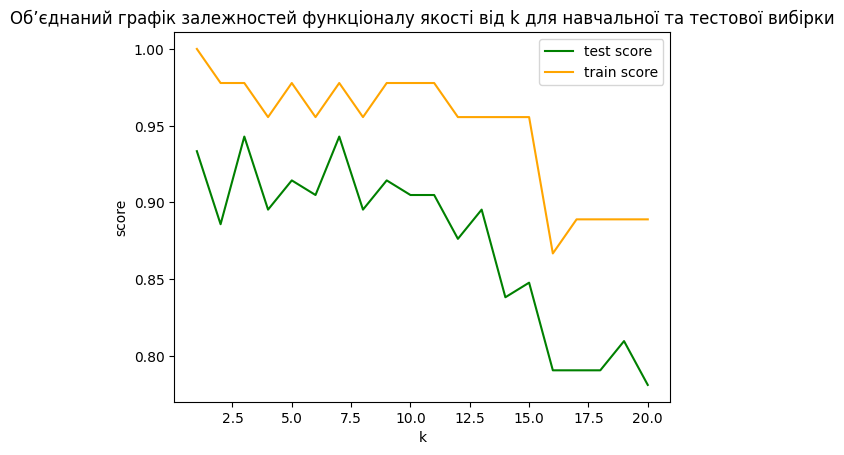
формування навчальної та тестової вибiрок, визначення функцiоналу якостi

класифiкатору за рiзних значень k, вiзуалiзацiя результатiв.

Таблиця результатів:

Встановив параметр розподілення датасету train\_size = 0.3

|  |  |  |
| --- | --- | --- |
| k | Точність train score | Точність test score |
| 1 | 1.0 | 0.9333333333333333 |
| 2 | 0.9777777777777777 | 0.8857142857142857 |
| 3 | 0.9777777777777777 | 0.9428571428571428 |
| 4 | 0.9555555555555556 | 0.8952380952380953 |
| 5 | 0.9777777777777777 | 0.9142857142857143 |
| 6 | 0.9555555555555556 | 0.9047619047619048 |
| 7 | 0.9777777777777777 | 0.9428571428571428 |
| 8 | 0.9555555555555556 | 0.8952380952380953 |
| 9 | 0.9777777777777777 | 0.9142857142857143 |
| 10 | 0.9777777777777777 | 0.9047619047619048 |
| 11 | 0.9777777777777777 | 0.9047619047619048 |
| 12 | 0.9555555555555556 | 0.8761904761904762 |
| 13 | 0.9555555555555556 | 0.8952380952380953 |
| 14 | 0.9555555555555556 | 0.8380952380952381 |
| 15 | 0.9555555555555556 | 0.8476190476190476 |
| 16 | 0.8666666666666667 | 0.7904761904761904 |
| 17 | 0.8888888888888888 | 0.7904761904761904 |
| 18 | 0.8888888888888888 | 0.7904761904761904 |
| 19 | 0.8888888888888888 | 0.8095238095238095 |
| 20 | 0.8888888888888888 | 0.780952380952381 |



**Код програми:**

**import matplotlib.pyplot as plt**

**import seaborn as sns**

**from matplotlib.colors import ListedColormap**

**from sklearn.model\_selection import train\_test\_split**

**from sklearn import datasets, neighbors**

***# Набiр даних «Iриси Фiшера»***

**iris = datasets.load\_iris()**

***# Розробка датасету***

**X = iris.data *# зберігає вимірні признаки ірісів***

**Y = iris.target *# зберігає класи ірісів***

**x\_train, x\_test, y\_train, y\_test = train\_test\_split(X, Y, train\_size=0.3,random\_state=0) *# тут за допомогою функції train\_test\_split розбиваємо датасет на тренінг та тест вибірки випадково у розмірі 30% - вивчення, 70% - тестування.***

***# Навчання класифiкатора***

**k = 1**

**k\_values = []**

**test\_score = [] *# зберігання оцінок тестової виборки***

**train\_score = [] *# зберігання оцінок навчальної виборки***

**while k <= 20:**

**k\_values.append(k)**

**knn = neighbors.KNeighborsClassifier(k)**

**knn.fit(x\_train, y\_train)**

**k += 1**

***# функцiонал якостi на навчальнiй та тестовiй вибiрцi для всiх значень k***

**test\_score\_1 = knn.score(x\_test, y\_test)**

**train\_score\_1 = knn.score(x\_train, y\_train)**

**test\_score.append(test\_score\_1)**

**train\_score.append(train\_score\_1)**

**plt.plot(k\_values, test\_score, color="green", label="test score")**

**plt.plot(k\_values, train\_score, color="orange", label="train score")**

**plt.xlabel('k')**

**plt.title('Об’єднаний графiк залежностей функцiоналу якостi вiд k для навчальної та тестової вибiрки')**

**plt.ylabel('score')**

**plt.legend()**

**plt.show()**

**print(test\_score) *# функціонал якості тестової вибірки***

**print(train\_score) *# функціонал якості навчальної вибірки***