import numpy as np

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

from sklearn.preprocessing import StandardScaler

from sklearn.neighbors import KNeighborsClassifier

from sklearn.model\_selection import cross\_val\_score

from sklearn.model\_selection import LeaveOneOut

import matplotlib.pyplot as plt

# from sklearn import datasets

from sklearn.svm import SVC

import pandas as pd

url = "C:/Users/fox/Desktop/ii/data/seeds\_dataset.data"

names = [

    "площадь A",

    "периметр P",

    "компактность C ",

    "длина",

    "ширина",

    "асимметрия",

    "Длина канавки ядра",

    "Сорт"

] #название атрибутов

dataset = pd.read\_csv(url, names=names) #чтение файла с данными

dataset.head()

X = dataset.iloc[:, :-1] #выбор вс

Y = dataset.iloc[:,7]

#print(dataset.shape)

metrics = ['euclidean', 'manhattan', 'chebyshev', 'minkowski']

X\_train, X\_test, Y\_train, Y\_test = train\_test\_split( X, Y, test\_size=0.40, random\_state = 1)

scaler  = StandardScaler()

scaler.fit(X\_train)

X\_train = scaler.transform(X\_train)

X\_test = scaler.transform(X\_test)

#выбор метрик

for j in range(4):

    print( '\n' +(metrics[j-1]))

    metricIndex = j-1

#выыод расчета с соответствующими метриками

    for i in range(10):

        classifier = KNeighborsClassifier(n\_neighbors = i+1, metric = metrics[metricIndex])

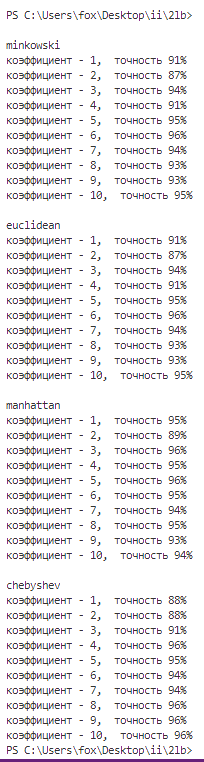
        classifier.fit(X\_train,Y\_train)

        Y\_pred = classifier.predict(X\_test)

        score = cross\_val\_score(classifier, X\_test, Y\_pred)

print ('коэффициент - {}, '.format(i+1), end = ' ')

        print ('точность {}%'.format(int(round(np.mean(score), 2) \* 100)))



|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | Среднее |
| minkowski | 91% | 87% | 94% | 91% | 95% | 96% | 94% | 93% | 93% | 95% | 93 |
| manhattan | 95% | 89% | 96% | 95% | 96% | 95% | 94% | 95% | 93% | 94% | 94 |
| euclidean | 91% | 87% | 94% | 91% | 95% | 96% | 94% | 93% | 93% | 95% | 93 |
| chebyshev | 88% | 88% | 91% | 96% | 95% | 94% | 94% | 96% | 96% | 96% | 93 |

Вывод:

Исходя из полученных данных можно сделать вывод, что метрики manhattan показывают лучший результат.