ASSIGNMENT 01: K NEAREST NEIGHBORS

Name: Nguyen Trung Tam

Students ID: B1910697

Class: M02

1.1. Given a dataset as follows:

X1	X2	Class
0.376	0.488	0
0.312	0.544	0
0.298	0.624	0
0.394	0.6	0
0.506	0.512	0
0.488	0.334	1
0.478	0.398	1
0.606	0.366	1
0.428	0.294	1
0.542	0.252	1

Classifying the testset with 1NN, 3NN:

	X1	X2	Class
P1	0.55	0.364	?
P2	0.558	0.47	?
P3	0.456	0.45	?
P4	0.45	0.57	?

We will use Euclidean distance to classify the test set as below:

Euclidean distance

with $q = 2 => d$ is Euclidean distance

$$d(u, v) = \sqrt{(|u_1 - v_1|^2 + |u_2 - v_2|^2 + \cdots + |u_n - v_n|^2)}$$

First, we will need to calculate the distance between the test points and every points in the training set

1.1.1. Calculate the distance between training points and P1

X1	X2	Class	Distance
0.376	0.488	0	0.213663287
0.312	0.544	0	0.298402413
0.298	0.624	0	0.362082863
0.394	0.6	0	0.282899275
0.506	0.512	0	0.154402073
0.488	0.334	1	0.068876701
0.478	0.398	1	0.079624117
0.606	0.366	1	0.056035703
0.428	0.294	1	0.140655608
0.542	0.252	1	0.112285351

From the results above, we see that the nearest point is

0.606	0.366	1	0.056035703

So 1NN = 1 for P1

From the results above, we see that the 3 nearest point is

0.606	0.366	1	0.056035703
0.488	0.334	1	0.068876701
0.478	0.398	1	0.079624117

So 3NN = 1 for P1

1.1.2. Calculate the distance between training points and P1

X1	X2	Class	Distance
0.376	0.488	0	0.182887944
0.312	0.544	0	0.256889081
0.298	0.624	0	0.302185374
0.394	0.6	0	0.209274939
0.506	0.512	0	0.0668431
0.488	0.334	1	0.15295751
0.478	0.398	1	0.107628992
0.606	0.366	1	0.114542569
0.428	0.294	1	0.21880585
0.542	0.252	1	0.218586367

From the results above, we see that the nearest point is

0.506	0.512	0	0.0668431

So 1NN = 0 for P2

From the results above, we see that the 3 nearest point is

0.506	0.512	0	0.0668431
0.478	0.398	1	0.107628992
0.606	0.366	1	0.114542569

So 3NN = 1 for P2

1.1.3. Calculate the distance between training points and P3

X1	X2	Class	Distance
0.376	0.488	0	0.088566359
0.312	0.544	0	0.171965113
0.298	0.624	0	0.235031913
0.394	0.6	0	0.162308349
0.506	0.512	0	0.079649231
0.488	0.334	1	0.120332872
0.478	0.398	1	0.056462377
0.606	0.366	1	0.171918585
0.428	0.294	1	0.158492902
0.542	0.252	1	0.215870331

From the results above, we see that the nearest point is

0.478	0.398	1	0.056462377

So 1NN = 1 for P3

From the results above, we see that the 3 nearest point is

0.478	0.398	1	0.056462377
0.506	0.512	0	0.079649231
0.376	0.488	0	0.088566359

So 3NN = 0 for P3

1.1.4. Calculate the distance between training points and P4

X1	X2	Class	Distance
0.376	0.488	0	0.11045361
0.312	0.544	0	0.140427917
0.298	0.624	0	0.16130716
0.394	0.6	0	0.063529521
0.506	0.512	0	0.080622577
0.488	0.334	1	0.239039746
0.478	0.398	1	0.174264167
0.606	0.366	1	0.256811215
0.428	0.294	1	0.276875423
0.542	0.252	1	0.331040783

From the results above, we see that the nearest point is

0.394	0.6	0	0.063529521

So 1NN = 0 for P4

From the results above, we see that the 3 nearest point is

0.394	0.6	0	0.063529521
0.506	0.512	0	0.080622577
0.376	0.488	0	0.11045361

So 3NN = 0 for P4

To summary, we have the result as below

a. 1NN

	X1	X2	Class
P1	0.55	0.364	1
P2	0.558	0.47	0
P3	0.456	0.45	1
P4	0.45	0.57	0

b. 3NN

	X1	X2	Class
P1	0.55	0.364	1
P2	0.558	0.47	1
P3	0.456	0.45	0
P4	0.45	0.57	0

1.2. Implement kNN from scratch in Python

```
import numpy as np
import pandas as pd
from scipy.stats import mode
from sklearn import metrics


def euclidean_distance(point1, point2):
    # calculating Euclidean distance
    # using linalg.norm()
    dist = np.linalg.norm(point1 - point2)
    return dist


# Locate the most similar neighbors
def get_neighbors(train, test_row, num_neighbors):
    distances = list()
    for train_row in train:
        dist = euclidean_distance(test_row, train_row)
        distances.append(dist)
    distances = np.array(distances)
    neighbors = train[np.argpartition(distances,
num_neighbors)[:num_neighbors]]
    return neighbors


def predict(train, test, k):
    labels = []
    for item in test:
        neighbors = get_neighbors(train, item, k)
        pred_labels = neighbors[:, -1]
        # noinspection PyUnresolvedReferences
        labels.append(mode(pred_labels).mode[0])
    return labels


def report(train_f, test_f, k):
    train_set = pd.read_csv(train_f, sep='[,,\s]', header=None,
engine='python')
    test_set = pd.read_csv(test_f, sep='[,,\s]', header=None,
engine='python')

    x_train = train_set.values  # Get training data points (exclude class
value)
```

```
    num_row, num_col = train_set.shape
    test_num_row, test_num_col = test_set.shape

    x_test = test_set.values  # Get training data points (exclude class
value)
    y_test = test_set.iloc[:, test_set.shape[1] - 1].values  # Get training
class data points (the last column)

    pred = predict(x_train, x_test, k)

    print("-" * 24 + "CLASSIFY RESULT WITH K = %s" % k + "-" * 24)
    print("* TRAIN FILE: %s, WITH %d SAMPLES" % (train_f, num_row))
    print("* TEST FILE: %s, WITH %d SAMPLES" % (test_f, test_num_row))
    print("* ACCURACY SCORE: %d%%" % (metrics.accuracy_score(y_test, pred) *
100))
    print("* CONFUSION MATRIX:\n", metrics.confusion_matrix(y_test, pred))
    print("* CLASSIFICATION REPORT:\n", metrics.classification_report(y_test,
pred))
    print("-" * 24 + "END OF CLASSIFY RESULT WITH K = %s" % k + "-" * 24)


if __name__ == '__main__':
    report('data/faces/data.trn', 'data/faces/data.tst', 1)
```

1.2.1. Iris (K = 1)

```
------------------------CLASSIFY RESULT WITH K = 1------------------------
* TRAIN FILE: data/iris/iris.trn, WITH 100 SAMPLES
* TEST FILE: data/iris/iris.tst, WITH 50 SAMPLES
* ACCURACY SCORE: 100%
* CONFUSION MATRIX:
 [[17  0  0]
 [ 0 15  0]
 [ 0  0 18]]
* CLASSIFICATION REPORT:
              precision    recall  f1-score   support

           0       1.00      1.00      1.00        17
           1       1.00      1.00      1.00        15
           2       1.00      1.00      1.00        18

    accuracy                           1.00        50
   macro avg       1.00      1.00      1.00        50
weighted avg       1.00      1.00      1.00        50

------------------------END OF CLASSIFY RESULT WITH K = 1------------------------
```

1.2.2. Iris (K = 3)

```
------------------------CLASSIFY RESULT WITH K = 3------------------------
* TRAIN FILE: data/iris/iris.trn, WITH 100 SAMPLES
* TEST FILE: data/iris/iris.tst, WITH 50 SAMPLES
* ACCURACY SCORE: 100%
* CONFUSION MATRIX:
 [[17  0  0]
 [ 0 15  0]
 [ 0  0 18]]
* CLASSIFICATION REPORT:
              precision    recall  f1-score   support

           0       1.00      1.00      1.00        17
           1       1.00      1.00      1.00        15
           2       1.00      1.00      1.00        18

    accuracy                           1.00        50
   macro avg       1.00      1.00      1.00        50
weighted avg       1.00      1.00      1.00        50

------------------------END OF CLASSIFY RESULT WITH K = 3------------------------
```

1.2.3. Iris (K = 5)

```
------------------------CLASSIFY RESULT WITH K = 5------------------------
* TRAIN FILE: data/iris/iris.trn, WITH 100 SAMPLES
* TEST FILE: data/iris/iris.tst, WITH 50 SAMPLES
* ACCURACY SCORE: 100%
* CONFUSION MATRIX:
 [[17  0  0]
 [ 0 15  0]
 [ 0  0 18]]
* CLASSIFICATION REPORT:
              precision    recall  f1-score   support

           0       1.00      1.00      1.00        17
           1       1.00      1.00      1.00        15
           2       1.00      1.00      1.00        18

    accuracy                           1.00        50
   macro avg       1.00      1.00      1.00        50
weighted avg       1.00      1.00      1.00        50

------------------------END OF CLASSIFY RESULT WITH K = 5------------------------
```

1.2.4. Faces (K = 1)

```
------------------------CLASSIFY RESULT WITH K = 1------------------------
* TRAIN FILE: data/faces/data.trn, WITH 768 SAMPLES
* TEST FILE: data/faces/data.tst, WITH 192 SAMPLES
* ACCURACY SCORE: 100%
* CONFUSION MATRIX:
 [[17  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0]
 [ 0 10  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0]
 [ 0  0  7  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0]
 [ 0  0  0  4  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0]
 [ 0  0  0  0  6  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0]
 [ 0  0  0  0  0 10  0  0  0  0  0  0  0  0  0  0  0  0  0  0]
 [ 0  0  0  0  0  0  8  0  0  0  0  0  0  0  0  0  0  0  0  0]
 [ 0  0  0  0  0  0  0  4  0  0  0  0  0  0  0  0  0  0  0  0]
 [ 0  0  0  0  0  0  0  0  8  0  0  0  0  0  0  0  0  0  0  0]
 [ 0  0  0  0  0  0  0  0  0 19  0  0  0  0  0  0  0  0  0  0]
 [ 0  0  0  0  0  0  0  0  0  0  9  0  0  0  0  0  0  0  0  0]
 [ 0  0  0  0  0  0  0  0  0  0  0  8  0  0  0  0  0  0  0  0]
 [ 0  0  0  0  0  0  0  0  0  0  0  0 11  0  0  0  0  0  0  0]
 [ 0  0  0  0  0  0  0  0  0  0  0  0  0 12  0  0  0  0  0  0]
 [ 0  0  0  0  0  0  0  0  0  0  0  0  0  0  8  0  0  0  0  0]
 [ 0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  5  0  0  0  0]
 [ 0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  8  0  0  0]
 [ 0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0 21  0  0]
 [ 0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  7  0]
 [ 0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0 10]]
```

```
* CLASSIFICATION REPORT:
              precision    recall  f1-score   support

           0       1.00      1.00      1.00        17
           1       1.00      1.00      1.00        10
           2       1.00      1.00      1.00         7
           3       1.00      1.00      1.00         4
           4       1.00      1.00      1.00         6
           5       1.00      1.00      1.00        10
           6       1.00      1.00      1.00         8
           7       1.00      1.00      1.00         4
           8       1.00      1.00      1.00         8
           9       1.00      1.00      1.00        19
          10       1.00      1.00      1.00         9
          11       1.00      1.00      1.00         8
          12       1.00      1.00      1.00        11
          13       1.00      1.00      1.00        12
          14       1.00      1.00      1.00         8
          15       1.00      1.00      1.00         5
          16       1.00      1.00      1.00         8
          17       1.00      1.00      1.00        21
          18       1.00      1.00      1.00         7
          19       1.00      1.00      1.00        10

    accuracy                           1.00       192
   macro avg       1.00      1.00      1.00       192
weighted avg       1.00      1.00      1.00       192

-----------------------END OF CLASSIFY RESULT WITH K = 1-----------------------
```

1.2.5. Faces (K = 3)

```
------------------------CLASSIFY RESULT WITH K = 3------------------------
* TRAIN FILE: data/faces/data.trn, WITH 768 SAMPLES
* TEST FILE: data/faces/data.tst, WITH 192 SAMPLES
* ACCURACY SCORE: 100%
* CONFUSION MATRIX:
 [[17  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0]
 [ 0 10  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0]
 [ 0  0  7  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0]
 [ 0  0  0  4  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0]
 [ 0  0  0  0  6  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0]
 [ 0  0  0  0  0 10  0  0  0  0  0  0  0  0  0  0  0  0  0  0]
 [ 0  0  0  0  0  0  8  0  0  0  0  0  0  0  0  0  0  0  0  0]
 [ 0  0  0  0  0  0  0  4  0  0  0  0  0  0  0  0  0  0  0  0]
 [ 0  0  0  0  0  0  0  0  8  0  0  0  0  0  0  0  0  0  0  0]
 [ 0  0  0  0  0  0  0  0  0 19  0  0  0  0  0  0  0  0  0  0]
 [ 0  0  0  0  0  0  0  0  0  0  9  0  0  0  0  0  0  0  0  0]
 [ 0  0  0  0  0  0  0  0  0  0  0  8  0  0  0  0  0  0  0  0]
 [ 0  0  0  0  0  0  0  0  0  0  0  0 11  0  0  0  0  0  0  0]
 [ 0  0  0  0  0  0  0  0  0  0  0  0  0 12  0  0  0  0  0  0]
 [ 0  0  0  0  0  0  0  0  0  0  0  0  0  0  8  0  0  0  0  0]
 [ 0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  5  0  0  0  0]
 [ 0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  8  0  0  0]
 [ 0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0 21  0  0]
 [ 0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  7  0]
 [ 0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0 10]]
```

```
* CLASSIFICATION REPORT:
              precision    recall  f1-score   support

           0       1.00      1.00      1.00        17
           1       1.00      1.00      1.00        10
           2       1.00      1.00      1.00         7
           3       1.00      1.00      1.00         4
           4       1.00      1.00      1.00         6
           5       1.00      1.00      1.00        10
           6       1.00      1.00      1.00         8
           7       1.00      1.00      1.00         4
           8       1.00      1.00      1.00         8
           9       1.00      1.00      1.00        19
          10       1.00      1.00      1.00         9
          11       1.00      1.00      1.00         8
          12       1.00      1.00      1.00        11
          13       1.00      1.00      1.00        12
          14       1.00      1.00      1.00         8
          15       1.00      1.00      1.00         5
          16       1.00      1.00      1.00         8
          17       1.00      1.00      1.00        21
          18       1.00      1.00      1.00         7
          19       1.00      1.00      1.00        10

    accuracy                           1.00       192
   macro avg       1.00      1.00      1.00       192
weighted avg       1.00      1.00      1.00       192

------------------------END OF CLASSIFY RESULT WITH K = 3-------------------------
```

1.2.6. Faces (K = 5)

```
------------------------CLASSIFY RESULT WITH K = 5------------------------
* TRAIN FILE: data/faces/data.trn, WITH 768 SAMPLES
* TEST FILE: data/faces/data.tst, WITH 192 SAMPLES
* ACCURACY SCORE: 99%
* CONFUSION MATRIX:
 [[17  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0]
 [ 0 10  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0]
 [ 0  0  7  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0]
 [ 0  0  0  4  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0]
 [ 0  0  0  0  6  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0]
 [ 0  0  0  0  0 10  0  0  0  0  0  0  0  0  0  0  0  0  0  0]
 [ 0  0  0  0  0  0  8  0  0  0  0  0  0  0  0  0  0  0  0  0]
 [ 0  0  0  0  0  0  0  4  0  0  0  0  0  0  0  0  0  0  0  0]
 [ 0  0  0  0  0  0  0  0  8  0  0  0  0  0  0  0  0  0  0  0]
 [ 0  0  0  0  0  0  0  0  0 19  0  0  0  0  0  0  0  0  0  0]
 [ 0  0  0  0  0  0  0  0  1  0  8  0  0  0  0  0  0  0  0  0]
 [ 0  0  0  0  0  0  0  0  0  0  0  8  0  0  0  0  0  0  0  0]
 [ 0  0  0  0  0  0  0  0  0  0  0  0 11  0  0  0  0  0  0  0]
 [ 0  0  0  0  0  0  0  0  0  0  0  0  0 12  0  0  0  0  0  0]
 [ 0  0  0  0  0  0  0  0  0  0  0  0  0  0  8  0  0  0  0  0]
 [ 0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  5  0  0  0  0]
 [ 0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  8  0  0  0]
 [ 0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0 21  0  0]
 [ 0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  7  0]
 [ 0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0 10]]
```

```
* CLASSIFICATION REPORT:
              precision    recall  f1-score   support

           0       1.00      1.00      1.00        17
           1       1.00      1.00      1.00        10
           2       1.00      1.00      1.00         7
           3       1.00      1.00      1.00         4
           4       1.00      1.00      1.00         6
           5       1.00      1.00      1.00        10
           6       1.00      1.00      1.00         8
           7       1.00      1.00      1.00         4
           8       0.89      1.00      0.94         8
           9       1.00      1.00      1.00        19
          10       1.00      0.89      0.94         9
          11       1.00      1.00      1.00         8
          12       1.00      1.00      1.00        11
          13       1.00      1.00      1.00        12
          14       1.00      1.00      1.00         8
          15       1.00      1.00      1.00         5
          16       1.00      1.00      1.00         8
          17       1.00      1.00      1.00        21
          18       1.00      1.00      1.00         7
          19       1.00      1.00      1.00        10

    accuracy                           0.99       192
   macro avg       0.99      0.99      0.99       192
weighted avg       1.00      0.99      0.99       192

------------------------END OF CLASSIFY RESULT WITH K = 5-------------------------
```

1.2.7. Optics (K = 1)

```
-----------------------CLASSIFY RESULT WITH K = 1-----------------------
* TRAIN FILE: data/optics/opt.trn, WITH 3823 SAMPLES
* TEST FILE: data/optics/opt.tst, WITH 1797 SAMPLES
* ACCURACY SCORE: 98%
* CONFUSION MATRIX:
 [[178   0   0   0   0   0   0   0   0   0]
 [  0 182   0   0   0   0   0   0   0   0]
 [  0   2 175   0   0   0   0   0   0   0]
 [  0   0   0 180   0   0   0   2   0   1]
 [  0   2   0   0 178   0   0   0   1   0]
 [  0   0   0   0   1 179   0   0   0   2]
 [  0   0   0   0   0   0 181   0   0   0]
 [  0   0   0   0   0   0   0 177   0   2]
 [  0   5   0   0   0   0   0   0 168   1]
 [  0   0   0   3   2   1   0   0   3 171]]
* CLASSIFICATION REPORT:
              precision    recall  f1-score   support

           0       1.00      1.00      1.00       178
           1       0.95      1.00      0.98       182
           2       1.00      0.99      0.99       177
           3       0.98      0.98      0.98       183
           4       0.98      0.98      0.98       181
           5       0.99      0.98      0.99       182
           6       1.00      1.00      1.00       181
           7       0.99      0.99      0.99       179
           8       0.98      0.97      0.97       174
           9       0.97      0.95      0.96       180

    accuracy                           0.98      1797
   macro avg       0.98      0.98      0.98      1797
weighted avg       0.98      0.98      0.98      1797

-----------------------END OF CLASSIFY RESULT WITH K = 1-----------------------
```

1.2.8. Optics (K = 3)

```
------------------------CLASSIFY RESULT WITH K = 3------------------------
* TRAIN FILE: data/optics/opt.trn, WITH 3823 SAMPLES
* TEST FILE: data/optics/opt.tst, WITH 1797 SAMPLES
* ACCURACY SCORE: 97%
* CONFUSION MATRIX:
 [[178   0   0   0   0   0   0   0   0   0]
 [  0 180   0   0   0   0   1   0   1   0]
 [  0   4 173   0   0   0   0   0   0   0]
 [  0   0   0 181   0   0   0   1   1   0]
 [  0   2   0   0 178   0   0   0   1   0]
 [  0   0   0   1   1 179   0   0   0   1]
 [  0   0   0   0   0   0 181   0   0   0]
 [  0   0   0   0   0   0   0 172   1   6]
 [  0   8   0   1   0   0   0   0 163   2]
 [  0   0   0   3   0   1   0   0   1 175]]
* CLASSIFICATION REPORT:
              precision    recall  f1-score   support

           0       1.00      1.00      1.00       178
           1       0.93      0.99      0.96       182
           2       1.00      0.98      0.99       177
           3       0.97      0.99      0.98       183
           4       0.99      0.98      0.99       181
           5       0.99      0.98      0.99       182
           6       0.99      1.00      1.00       181
           7       0.99      0.96      0.98       179
           8       0.97      0.94      0.95       174
           9       0.95      0.97      0.96       180

    accuracy                           0.98      1797
   macro avg       0.98      0.98      0.98      1797
weighted avg       0.98      0.98      0.98      1797

------------------------END OF CLASSIFY RESULT WITH K = 3------------------------
```

1.2.9. Optics (K = 5)

```
------------------------CLASSIFY RESULT WITH K = 5------------------------
* TRAIN FILE: data/optics/opt.trn, WITH 3823 SAMPLES
* TEST FILE: data/optics/opt.tst, WITH 1797 SAMPLES
* ACCURACY SCORE: 98%
* CONFUSION MATRIX:
 [[178   0   0   0   0   0   0   0   0   0]
 [  0 181   0   0   0   0   1   0   0   0]
 [  0   3 174   0   0   0   0   0   0   0]
 [  0   1   1 178   0   1   0   1   1   0]
 [  0   1   0   0 179   0   0   0   1   0]
 [  0   0   0   0   1 180   0   0   0   1]
 [  0   0   0   0   0   0 181   0   0   0]
 [  0   0   0   0   0   0   0 173   1   5]
 [  0   8   0   1   0   1   0   0 163   1]
 [  0   0   0   2   1   1   0   0   1 175]]
* CLASSIFICATION REPORT:
              precision    recall  f1-score   support

           0       1.00      1.00      1.00       178
           1       0.93      0.99      0.96       182
           2       0.99      0.98      0.99       177
           3       0.98      0.97      0.98       183
           4       0.99      0.99      0.99       181
           5       0.98      0.99      0.99       182
           6       0.99      1.00      1.00       181
           7       0.99      0.97      0.98       179
           8       0.98      0.94      0.96       174
           9       0.96      0.97      0.97       180

    accuracy                           0.98      1797
   macro avg       0.98      0.98      0.98      1797
weighted avg       0.98      0.98      0.98      1797

------------------------END OF CLASSIFY RESULT WITH K = 5------------------------
```

1.2.10. Fp (K = 1)

```
-----------------------CLASSIFY RESULT WITH K = 1-----------------------
* TRAIN FILE: data/fp/fp.trn, WITH 320 SAMPLES
* TEST FILE: data/fp/fp.tst, WITH 160 SAMPLES
* ACCURACY SCORE: 91%
* CONFUSION MATRIX:
 [[29  0  0  0  0  0  0  0  0  0  0  0  0  0  0]
 [ 0  4  0  0  0  0  0  0  0  0  0  0  0  0  0]
 [ 1  0 10  0  0  0  1  0  0  0  0  0  0  0  0]
 [ 0  0  0  5  0  0  2  0  0  0  0  0  0  0  0]
 [ 0  0  0  0  9  0  0  0  0  0  0  0  0  0  0]
 [ 1  0  0  0  0 13  0  0  0  0  0  0  0  0  0]
 [ 0  0  0  0  0  0 10  0  0  0  0  0  0  0  0]
 [ 0  0  0  0  0  0  1 10  0  0  0  0  0  0  0]
 [ 0  0  0  0  0  0  0  0  7  0  0  0  0  0  0]
 [ 0  0  0  0  0  0  1  0  0  3  0  0  0  0  0]
 [ 0  0  0  0  0  0  0  0  0  0 10  0  0  0  0]
 [ 0  0  0  0  0  0  0  0  0  0  0  9  0  0  0]
 [ 0  0  0  0  0  0  2  0  0  0  0  1  7  0  0]
 [ 0  0  0  0  0  0  0  0  0  0  0  2  0  6  2]
 [ 0  0  0  0  0  0  0  0  0  0  0  0  0  0 14]]
```

```
* CLASSIFICATION REPORT:
              precision    recall  f1-score   support

           1       0.94      1.00      0.97        29
           2       1.00      1.00      1.00         4
           3       1.00      0.83      0.91        12
           4       1.00      0.71      0.83         7
           5       1.00      1.00      1.00         9
           6       1.00      0.93      0.96        14
           7       0.59      1.00      0.74        10
           8       1.00      0.91      0.95        11
           9       1.00      1.00      1.00         7
          10       1.00      0.75      0.86         4
          11       1.00      1.00      1.00        10
          12       0.75      1.00      0.86         9
          13       1.00      0.70      0.82        10
          14       1.00      0.60      0.75        10
          15       0.88      1.00      0.93        14

    accuracy                           0.91       160
   macro avg       0.94      0.90      0.91       160
weighted avg       0.94      0.91      0.91       160

-----------------------END OF CLASSIFY RESULT WITH K = 1-----------------------
```

1.2.11. Fp (K = 3)

```
----------------------CLASSIFY RESULT WITH K = 3----------------------
* TRAIN FILE: data/fp/fp.trn, WITH 320 SAMPLES
* TEST FILE: data/fp/fp.tst, WITH 160 SAMPLES
* ACCURACY SCORE: 90%
* CONFUSION MATRIX:
 [[29  0  0  0  0  0  0  0  0  0  0  0  0  0  0]
 [ 0  4  0  0  0  0  0  0  0  0  0  0  0  0  0]
 [ 2  0 10  0  0  0  0  0  0  0  0  0  0  0  0]
 [ 1  0  0  6  0  0  0  0  0  0  0  0  0  0  0]
 [ 0  0  0  0  9  0  0  0  0  0  0  0  0  0  0]
 [ 2  0  0  0  0 12  0  0  0  0  0  0  0  0  0]
 [ 0  0  0  0  0  0 10  0  0  0  0  0  0  0  0]
 [ 1  0  0  0  0  0  0 10  0  0  0  0  0  0  0]
 [ 0  0  0  0  0  0  0  0  7  0  0  0  0  0  0]
 [ 0  0  0  0  0  0  0  0  0  4  0  0  0  0  0]
 [ 0  0  0  0  0  0  0  0  0  0 10  0  0  0  0]
 [ 0  0  0  0  0  0  0  0  0  0  1  8  0  0  0]
 [ 0  0  0  0  0  0  4  0  0  0  0  0  6  0  0]
 [ 0  0  0  0  0  0  1  0  0  0  3  0  0  6  0]
 [ 0  0  0  0  0  0  0  0  0  0  0  0  0  0 14]]
```

```
* CLASSIFICATION REPORT:
              precision    recall  f1-score   support

           1       0.83      1.00      0.91        29
           2       1.00      1.00      1.00         4
           3       1.00      0.83      0.91        12
           4       1.00      0.86      0.92         7
           5       1.00      1.00      1.00         9
           6       1.00      0.86      0.92        14
           7       0.67      1.00      0.80        10
           8       1.00      0.91      0.95        11
           9       1.00      1.00      1.00         7
          10       1.00      1.00      1.00         4
          11       0.71      1.00      0.83        10
          12       1.00      0.89      0.94         9
          13       1.00      0.60      0.75        10
          14       1.00      0.60      0.75        10
          15       1.00      1.00      1.00        14

    accuracy                           0.91       160
   macro avg       0.95      0.90      0.91       160
weighted avg       0.93      0.91      0.91       160

----------------------END OF CLASSIFY RESULT WITH K = 3----------------------
```

1.2.12. Fp (K = 5)

```
------------------------CLASSIFY RESULT WITH K = 5------------------------
* TRAIN FILE: data/fp/fp.trn, WITH 320 SAMPLES
* TEST FILE: data/fp/fp.tst, WITH 160 SAMPLES
* ACCURACY SCORE: 88%
* CONFUSION MATRIX:
 [[29  0  0  0  0  0  0  0  0  0  0  0  0  0  0]
 [ 0  4  0  0  0  0  0  0  0  0  0  0  0  0  0]
 [ 2  0 10  0  0  0  0  0  0  0  0  0  0  0  0]
 [ 2  0  0  5  0  0  0  0  0  0  0  0  0  0  0]
 [ 0  0  0  0  9  0  0  0  0  0  0  0  0  0  0]
 [ 2  0  0  0  0 12  0  0  0  0  0  0  0  0  0]
 [ 0  0  0  0  0  0 10  0  0  0  0  0  0  0  0]
 [ 1  0  0  0  0  0  0 10  0  0  0  0  0  0  0]
 [ 0  0  0  0  0  0  0  0  6  0  1  0  0  0  0]
 [ 0  0  0  0  0  0  1  0  0  3  0  0  0  0  0]
 [ 0  0  0  0  0  0  0  0  0  0 10  0  0  0  0]
 [ 0  0  0  0  0  0  0  0  0  0  2  7  0  0  0]
 [ 0  0  0  0  0  0  2  0  0  0  2  0  6  0  0]
 [ 0  0  0  0  0  0  1  0  0  0  3  0  0  6  0]
 [ 0  0  0  0  0  0  0  0  0  0  0  0  0  0 14]]
```

```
* CLASSIFICATION REPORT:
              precision    recall  f1-score   support

           1       0.81      1.00      0.89        29
           2       1.00      1.00      1.00         4
           3       1.00      0.83      0.91        12
           4       1.00      0.71      0.83         7
           5       1.00      1.00      1.00         9
           6       1.00      0.86      0.92        14
           7       0.71      1.00      0.83        10
           8       1.00      0.91      0.95        11
           9       1.00      0.86      0.92         7
          10       1.00      0.75      0.86         4
          11       0.56      1.00      0.71        10
          12       1.00      0.78      0.88         9
          13       1.00      0.60      0.75        10
          14       1.00      0.60      0.75        10
          15       1.00      1.00      1.00        14

    accuracy                           0.88       160
   macro avg       0.94      0.86      0.88       160
weighted avg       0.92      0.88      0.88       160

------------------------END OF CLASSIFY RESULT WITH K = 5------------------------
```

1.2.13. Letter (K = 1)

```
------------------------CLASSIFY RESULT WITH K = 1------------------------
* TRAIN FILE: data/letter/let.trn, WITH 13334 SAMPLES
* TEST FILE: data/letter/let.tst, WITH 6666 SAMPLES
* ACCURACY SCORE: 99%
* CONFUSION MATRIX:
 [[271   3   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0
    0   0   0   0   0   0   0   0]
 [  0 240   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0
    0   0   0   0   0   0   0   0]
 [  0   0 226   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0
    0   0   0   0   0   0   0   0]
 [  0   0   0 277   0   0   0   0   0   0   0   0   0   0   0   0   0   0
    0   0   0   0   0   0   0   0]
 [  0   0   0   0 260   1   1   0   0   0   0   0   0   0   0   0   0   0
    0   0   0   0   0   0   0   0]
 [  0   0   0   0   1 267   1   0   0   0   0   0   0   0   0   0   0   0
    0   0   0   0   0   0   0   0]
 [  0   0   0   0   0   0 263   0   0   0   0   0   0   0   0   0   0   0
    0   0   0   0   0   0   0   0]
 [  0   0   0   0   0   1   2 225   0   0   2   0   0   0   0   0   0   0
    0   0   0   0   0   0   0   0]
 [  0   0   0   0   0   0   0   0 263   6   0   0   0   0   0   0   0   0
    0   0   0   0   0   0   0   0]
 [  0   0   0   0   0   0   0   1   7 231   0   0   0   0   0   0   0   0
    0   0   0   0   0   0   0   0]
 [  0   0   0   0   0   0   0   0   0   0 243   0   0   0   0   0   0   0
    0   0   0   0   0   0   0   0]
 [  0   0   0   0   0   0   0   0   0   1   0 269   0   0   0   0   0   0
    0   0   0   0   0   0   0   0]
 [  0   0   0   0   0   0   0   0   0   0   0   0 244   1   0   0   0   0
    0   0   0   0   0   0   0   0]
 [  0   0   0   0   0   0   0   0   0   0   0   0   4 267   1   0   0   0
    0   0   0   0   0   0   0   0]
 [  0   0   0   0   0   0   0   0   0   0   0   0   0   0 237   0   0   0
    0   0   0   0   0   0   0   0]
 [  0   0   0   0   0   0   0   0   0   0   0   0   0   0   0 264   2   0
    0   0   0   0   0   0   0   0]
 [  0   0   0   0   0   0   0   0   0   0   0   0   0   0   3   0 277   0
    0   0   0   0   0   0   0   0]
```

```
 [  0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0
    0   0   0   0   0   1 251   0]
 [  0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0
    0   0   0   0   0   0   0 259]]
* CLASSIFICATION REPORT:
              precision    recall  f1-score   support

           0       1.00      0.99      0.99       274
           1       0.99      1.00      0.99       240
           2       1.00      1.00      1.00       226
           3       1.00      1.00      1.00       277
           4       1.00      0.99      0.99       262
           5       0.99      0.99      0.99       269
           6       0.99      1.00      0.99       263
           7       1.00      0.98      0.99       230
           8       0.97      0.98      0.98       269
           9       0.97      0.97      0.97       239
          10       0.99      1.00      1.00       243
          11       1.00      1.00      1.00       270
          12       0.98      1.00      0.99       245
          13       1.00      0.98      0.99       272
          14       0.98      1.00      0.99       237
          15       1.00      0.99      1.00       266
          16       0.99      0.99      0.99       280
          17       1.00      1.00      1.00       271
          18       1.00      1.00      1.00       264
          19       1.00      1.00      1.00       244
          20       1.00      1.00      1.00       274
          21       1.00      1.00      1.00       238
          22       1.00      1.00      1.00       241
          23       1.00      1.00      1.00       261
          24       1.00      1.00      1.00       252
          25       1.00      1.00      1.00       259

    accuracy                           0.99      6666
   macro avg       0.99      0.99      0.99      6666
weighted avg       0.99      0.99      0.99      6666

-----------------------END OF CLASSIFY RESULT WITH K = 1-----------------------
```

1.2.14. Letter (K = 3)

```
-----------------------CLASSIFY RESULT WITH K = 3-----------------------
* TRAIN FILE: data/letter/let.trn, WITH 13334 SAMPLES
* TEST FILE: data/letter/let.tst, WITH 6666 SAMPLES
* ACCURACY SCORE: 99%
* CONFUSION MATRIX:
 [[273   1   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0
    0   0   0   0   0   0   0   0]
 [  0 240   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0
    0   0   0   0   0   0   0   0]
 [  1   0 225   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0
    0   0   0   0   0   0   0   0]
 [  0   0   0 277   0   0   0   0   0   0   0   0   0   0   0   0   0   0
    0   0   0   0   0   0   0   0]
 [  0   0   1   0 258   2   1   0   0   0   0   0   0   0   0   0   0   0
    0   0   0   0   0   0   0   0]
 [  0   0   0   0   0 267   1   1   0   0   0   0   0   0   0   0   0   0
    0   0   0   0   0   0   0   0]
 [  0   0   0   0   1   0 262   0   0   0   0   0   0   0   0   0   0   0
    0   0   0   0   0   0   0   0]
 [  0   0   0   0   0   0   2 227   0   0   1   0   0   0   0   0   0   0
    0   0   0   0   0   0   0   0]
 [  0   0   0   0   0   2   0   0 254  13   0   0   0   0   0   0   0   0
    0   0   0   0   0   0   0   0]
 [  0   0   0   0   0   0   0   0   7 232   0   0   0   0   0   0   0   0
    0   0   0   0   0   0   0   0]
 [  0   0   0   0   0   0   0   1   0   0 242   0   0   0   0   0   0   0
    0   0   0   0   0   0   0   0]
 [  0   0   0   0   0   0   0   0   0   1   0 269   0   0   0   0   0   0
    0   0   0   0   0   0   0   0]
 [  0   0   0   0   0   0   0   0   0   0   0   0 242   3   0   0   0   0
    0   0   0   0   0   0   0   0]
 [  0   0   0   0   0   0   0   0   0   0   0   0   4 264   4   0   0   0
    0   0   0   0   0   0   0   0]
 [  0   0   0   0   0   0   0   0   0   0   0   0   0   0 237   0   0   0
    0   0   0   0   0   0   0   0]
 [  0   0   0   0   0   0   0   0   0   0   0   0   0   0   0 264   2   0
    0   0   0   0   0   0   0   0]
 [  0   0   0   0   0   0   0   0   0   0   0   0   0   0   4   0 275   1
    0   0   0   0   0   0   0   0]
```

```
 [  0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0
    0   0   1   1   0   1 249   0]
 [  0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0
    0   0   0   0   0   0   0 259]]
* CLASSIFICATION REPORT:
              precision    recall  f1-score   support

           0       1.00      1.00      1.00       274
           1       1.00      1.00      1.00       240
           2       1.00      1.00      1.00       226
           3       1.00      1.00      1.00       277
           4       1.00      0.98      0.99       262
           5       0.99      0.99      0.99       269
           6       0.98      1.00      0.99       263
           7       0.99      0.99      0.99       230
           8       0.97      0.94      0.96       269
           9       0.94      0.97      0.96       239
          10       1.00      1.00      1.00       243
          11       1.00      1.00      1.00       270
          12       0.98      0.99      0.99       245
          13       0.99      0.97      0.98       272
          14       0.97      1.00      0.98       237
          15       1.00      0.99      1.00       266
          16       0.99      0.98      0.99       280
          17       1.00      1.00      1.00       271
          18       1.00      1.00      1.00       264
          19       1.00      1.00      1.00       244
          20       0.99      1.00      0.99       274
          21       0.99      1.00      0.99       238
          22       1.00      0.99      1.00       241
          23       1.00      1.00      1.00       261
          24       1.00      0.99      0.99       252
          25       1.00      1.00      1.00       259

    accuracy                           0.99      6666
   macro avg       0.99      0.99      0.99      6666
weighted avg       0.99      0.99      0.99      6666

------------------------END OF CLASSIFY RESULT WITH K = 3-----------------------
```

1.2.15. Letter (K = 5)

```
-----------------------CLASSIFY RESULT WITH K = 5-----------------------
* TRAIN FILE: data/letter/let.trn, WITH 13334 SAMPLES
* TEST FILE: data/letter/let.tst, WITH 6666 SAMPLES
* ACCURACY SCORE: 99%
* CONFUSION MATRIX:
 [[273   1   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0
    0   0   0   0   0   0   0   0]
 [  0 240   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0
    0   0   0   0   0   0   0   0]
 [  0   1 225   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0
    0   0   0   0   0   0   0   0]
 [  0   1   0 276   0   0   0   0   0   0   0   0   0   0   0   0   0   0
    0   0   0   0   0   0   0   0]
 [  0   0   1   0 259   1   1   0   0   0   0   0   0   0   0   0   0   0
    0   0   0   0   0   0   0   0]
 [  0   0   0   0   0 266   1   2   0   0   0   0   0   0   0   0   0   0
    0   0   0   0   0   0   0   0]
 [  0   0   0   0   1   0 261   1   0   0   0   0   0   0   0   0   0   0
    0   0   0   0   0   0   0   0]
 [  0   0   0   0   0   0   3 226   0   0   1   0   0   0   0   0   0   0
    0   0   0   0   0   0   0   0]
 [  0   0   0   0   0   2   0   0 254  13   0   0   0   0   0   0   0   0
    0   0   0   0   0   0   0   0]
 [  0   0   0   0   0   0   0   0   8 231   0   0   0   0   0   0   0   0
    0   0   0   0   0   0   0   0]
 [  0   0   0   0   0   0   0   0   0   0 243   0   0   0   0   0   0   0
    0   0   0   0   0   0   0   0]
 [  0   0   0   0   0   0   0   0   0   1   0 269   0   0   0   0   0   0
    0   0   0   0   0   0   0   0]
 [  0   0   0   0   0   0   0   0   0   0   0   0 244   1   0   0   0   0
    0   0   0   0   0   0   0   0]
 [  0   0   0   0   0   0   0   0   0   0   0   0   4 260   8   0   0   0
    0   0   0   0   0   0   0   0]
 [  0   0   0   0   0   0   0   0   0   0   0   0   0   0 237   0   0   0
    0   0   0   0   0   0   0   0]
 [  0   0   0   0   0   0   0   0   0   0   0   0   0   0   0 263   2   1
    0   0   0   0   0   0   0   0]
 [  0   0   0   0   0   0   0   0   0   0   0   0   0   0   1   0 278   1
    0   0   0   0   0   0   0   0]
```

```
 [  0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0
    0   0   0   0   1   1 250   0]
 [  0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0
    0   0   0   0   0   0   0 259]]
* CLASSIFICATION REPORT:
              precision    recall  f1-score   support

           0       1.00      1.00      1.00       274
           1       0.99      1.00      0.99       240
           2       1.00      1.00      1.00       226
           3       1.00      1.00      1.00       277
           4       1.00      0.99      0.99       262
           5       0.99      0.99      0.99       269
           6       0.98      0.99      0.99       263
           7       0.99      0.98      0.98       230
           8       0.97      0.94      0.96       269
           9       0.94      0.97      0.95       239
          10       1.00      1.00      1.00       243
          11       1.00      1.00      1.00       270
          12       0.98      1.00      0.99       245
          13       1.00      0.96      0.98       272
          14       0.96      1.00      0.98       237
          15       1.00      0.99      0.99       266
          16       0.99      0.99      0.99       280
          17       0.99      1.00      0.99       271
          18       1.00      0.99      1.00       264
          19       1.00      1.00      1.00       244
          20       1.00      1.00      1.00       274
          21       0.99      1.00      0.99       238
          22       0.99      0.99      0.99       241
          23       1.00      1.00      1.00       261
          24       1.00      0.99      1.00       252
          25       1.00      1.00      1.00       259

    accuracy                           0.99      6666
   macro avg       0.99      0.99      0.99      6666
weighted avg       0.99      0.99      0.99      6666

-----------------------END OF CLASSIFY RESULT WITH K = 5-----------------------
```

1.3. Proof of Cover-Hart's theorem:

Theorem: For sufficiently large training set size n, the error rate of the 1NN classifier is less than twice the Bayes error rate.

Proof: Let x be a query point and let r be its closest neighbor. The expected error rate of the 1NN classifier is

$$\sum_{i=1}^{c} p(i|x)[1 - p(i|r)]$$

where $p(i|x)$ is the probability that x has label i and $1 - p(i|r)$ is the probability that r has a different label. The critical fact is that if the number n of training examples is large enough, then the label probability distributions for all x and r will be essentially the same. In this case, the expected error rate of the 1NN classifier is

$$\sum_{i=1}^{c} p(i|x)[1 - p(i|x)]$$

To prove the theorem we need to show that

$$\sum_{i=1}^{c} p(i|x)[1 - p(i|r)] \leq 2\left[1 - \max_i p(i|x)\right]$$

Let $\max_i p(i|x) = r$ and let this maximum be attained with $i = j$. Then the lefthand side is

$$r(1 - r) + \sum_{i \neq j}^{c} p(i|x)[1 - p(i|r)]$$

and the righthand side is $2(1 - r)$. The summation above is maximized when all values $p(i|x)$ are equal for $i \neq j$. The value of the lefthand side is then

$$A = r(1 - r) + (c - 1)\frac{1 - r}{c - 1}\frac{(c - 1) - (1 - r)}{c - 1}$$

$$= r(1 - r) + (1 - r)\frac{c + r - 2}{m - 1}$$

Now $r \leq 1$ and $c - 2 + r < c - 1$ so $A < 2(1 - r)$ which is what we wanted to prove ■