A picture containing drawing

Description automatically generated

CMP466 – Machine Learning & Data Mining

Homework Assignment 3

6th of April 2021

|  |  |
| --- | --- |
| **Student Name** | **Student ID** |
| Zayed Alhosani | B00075279 |
| Saeed Alrafi | B00074839 |
| Mohammed Hamza | B00069499 |

**#Importing important libraries**

import graphviz

import numpy as np

from sklearn.model\_selection import cross\_validate

from time import time

import matplotlib.pyplot as plt

import sklearn

from sklearn.model\_selection import KFold

from sklearn import tree

from sklearn.model\_selection import GridSearchCV

from sklearn.datasets import fetch\_lfw\_people

from sklearn.metrics import classification\_report

from sklearn.metrics import confusion\_matrix

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

from sklearn.decomposition import PCA

from sklearn.svm import SVC

from sklearn.svm import LinearSVC

from sklearn.pipeline import make\_pipeline

from sklearn.preprocessing import StandardScaler

from sklearn.datasets import make\_classification

import os

import errno

import shutil

**#Copying the source directory**

def copy(src, dest):

try:

shutil.copytree(src, dest)

except OSError as e:

if e.errno == errno.ENOTDIR:

shutil.copy(src, dest)

else:

print('Directory not found! %s' % e)

src = '../input/'

dest = '../LFW/lfw\_home'

copy(src,dest)

path = '../LFW/'

**#Loading the dataset**

lfw\_dataset = sklearn.datasets.fetch\_lfw\_people(data\_home = path, min\_faces\_per\_person=100, download\_if\_missing = False)

t0 = time()

**#Performing K-Fold Splits, Printing the fold results, Printing the training and accuracy results for each SVC type with a confusion matrix and classification report.**

t0 = time()

C = 1.0 # Default C value

clf = clf.fit(X, y)

results = cross\_validate(clf, X, y, cv=5);

print(results)

results = cross\_validate(clf, X, y, cv=10);

print(results)

print("--------------------------------------")

print("LinearSVC")

for x in range(10):

X\_train, y\_train = make\_classification(n\_features=5, random\_state=0)

clf = make\_pipeline(StandardScaler(), LinearSVC(random\_state=0, tol=1e-5))

clf.fit(X\_train, y\_train)

y\_pred= clf.predict(X\_test)

print("--------------------------------------")

print("Training Accuracy in loop iteration " + str(x+1) + ": " + str(clf.score(X\_train, y\_train)))

print("Testing Accuracy in loop iteration " + str(x+1) + ": " + str(clf.score(X\_test, y\_test)))

print("--------------------------------------")

print(confusion\_matrix(y\_test,y\_pred))

print(classification\_report(y\_test,y\_pred))

print("Linear")

for x in range(10):

clf = SVC(kernel='linear', C=C).fit(X\_train, y\_train)

y\_pred= clf.predict(X\_test)

print("--------------------------------------")

print("Training Accuracy in loop iteration " + str(x+1) + ": " + str(clf.score(X\_train, y\_train)))

print("Testing Accuracy in loop iteration " + str(x+1) + ": " + str(clf.score(X\_test, y\_test)))

print("--------------------------------------")

print(confusion\_matrix(y\_test,y\_pred))

print(classification\_report(y\_test,y\_pred))

print("RBF")

for x in range(10):

clf = SVC(kernel='rbf', gamma=7, C=2).fit(X\_train, y\_train)

y\_pred= clf.predict(X\_test)

print("--------------------------------------")

print("Training Accuracy in loop iteration " + str(x+1) + ": " + str(clf.score(X\_train, y\_train)))

print("Testing Accuracy in loop iteration " + str(x+1) + ": " + str(clf.score(X\_test, y\_test)))

print("--------------------------------------")

print(confusion\_matrix(y\_test,y\_pred))

print(classification\_report(y\_test,y\_pred))

print("Sigmoid")

for x in range(10):

clf = SVC(kernel='sigmoid').fit(X\_train, y\_train)

y\_pred= clf.predict(X\_test)

print("--------------------------------------")

print("Training Accuracy in loop iteration " + str(x+1) + ": " + str(clf.score(X\_train, y\_train)))

print("Testing Accuracy in loop iteration " + str(x+1) + ": " + str(clf.score(X\_test, y\_test)))

print("--------------------------------------")

print(confusion\_matrix(y\_test,y\_pred))

print(classification\_report(y\_test,y\_pred))

print("Poly")

for x in range(10):

clf = SVC(kernel='poly', degree=0).fit(X\_train, y\_train)

y\_pred= clf.predict(X\_test)

print("--------------------------------------")

print("Training Accuracy in loop iteration " + str(x+1) + ": " + str(clf.score(X\_train, y\_train)))

print("Testing Accuracy in loop iteration " + str(x+1) + ": " + str(clf.score(X\_test, y\_test)))

print("--------------------------------------")

print(confusion\_matrix(y\_test,y\_pred))

print(classification\_report(y\_test,y\_pred))

**#OUTPUT (K-fold Cross Validation)**

{'fit\_time': array([0.00098634, 0.00087404, 0.00080824, 0.00089955, 0.00085568]), 'score\_time': array([0.00090122, 0.00042653, 0.0003407 , 0.00034213, 0.00034547]), 'test\_score': array([0.95, 0.9 , 0.9 , 0.9 , 0.95])}

{'fit\_time': array([0.00134826, 0.00098419, 0.00091434, 0.00076962, 0.00075722,

0.00094438, 0.00086164, 0.00075626, 0.00073361, 0.00083566]), 'score\_time': array([0.00033665, 0.00034475, 0.00036073, 0.00029707, 0.00031209,

0.00034046, 0.0003233 , 0.00028896, 0.00053263, 0.00033236]), 'test\_score': array([1. , 1. , 1. , 0.8, 0.8, 1.

, 1. , 0.8, 0.9, 0.9])}

**#OUTPUT (LinearSVC)**

LinearSVC

--------------------------------------

Training Accuracy in loop iteration 1: 0.94

Testing Accuracy in loop iteration 1: 0.9

--------------------------------------

--------------------------------------

Training Accuracy in loop iteration 2: 0.94

Testing Accuracy in loop iteration 2: 0.9

--------------------------------------

--------------------------------------

Training Accuracy in loop iteration 3: 0.94

Testing Accuracy in loop iteration 3: 0.9

--------------------------------------

--------------------------------------

Training Accuracy in loop iteration 4: 0.94

Testing Accuracy in loop iteration 4: 0.9

--------------------------------------

--------------------------------------

Training Accuracy in loop iteration 5: 0.94

Testing Accuracy in loop iteration 5: 0.9

--------------------------------------

--------------------------------------

Training Accuracy in loop iteration 6: 0.94

Testing Accuracy in loop iteration 6: 0.9

--------------------------------------

--------------------------------------

Training Accuracy in loop iteration 7: 0.94

Testing Accuracy in loop iteration 7: 0.9

--------------------------------------

--------------------------------------

Training Accuracy in loop iteration 8: 0.94

Testing Accuracy in loop iteration 8: 0.9

--------------------------------------

--------------------------------------

Training Accuracy in loop iteration 9: 0.94

Testing Accuracy in loop iteration 9: 0.9

--------------------------------------

--------------------------------------

Training Accuracy in loop iteration 10: 0.94

Testing Accuracy in loop iteration 10: 0.9

--------------------------------------

[[9 2]

[0 9]]

precision recall f1-score support

0 1.00 0.82 0.90 11

1 0.82 1.00 0.90 9

accuracy 0.90 20

macro avg 0.91 0.91 0.90 20

weighted avg 0.92 0.90 0.90 20

**#OUTPUT (Linear)**

Linear

--------------------------------------

Training Accuracy in loop iteration 1: 0.94

Testing Accuracy in loop iteration 1: 0.9

--------------------------------------

--------------------------------------

Training Accuracy in loop iteration 2: 0.94

Testing Accuracy in loop iteration 2: 0.9

--------------------------------------

--------------------------------------

Training Accuracy in loop iteration 3: 0.94

Testing Accuracy in loop iteration 3: 0.9

--------------------------------------

--------------------------------------

Training Accuracy in loop iteration 4: 0.94

Testing Accuracy in loop iteration 4: 0.9

--------------------------------------

--------------------------------------

Training Accuracy in loop iteration 5: 0.94

Testing Accuracy in loop iteration 5: 0.9

--------------------------------------

--------------------------------------

Training Accuracy in loop iteration 6: 0.94

Testing Accuracy in loop iteration 6: 0.9

--------------------------------------

--------------------------------------

Training Accuracy in loop iteration 7: 0.94

Testing Accuracy in loop iteration 7: 0.9

--------------------------------------

--------------------------------------

Training Accuracy in loop iteration 8: 0.94

Testing Accuracy in loop iteration 8: 0.9

--------------------------------------

--------------------------------------

Training Accuracy in loop iteration 9: 0.94

Testing Accuracy in loop iteration 9: 0.9

--------------------------------------

--------------------------------------

Training Accuracy in loop iteration 10: 0.94

Testing Accuracy in loop iteration 10: 0.9

--------------------------------------

[[9 2]

[0 9]]

precision recall f1-score support

0 1.00 0.82 0.90 11

1 0.82 1.00 0.90 9

accuracy 0.90 20

macro avg 0.91 0.91 0.90 20

weighted avg 0.92 0.90 0.90 20

**#OUTPUT (RBF gamma=7 C=1)**

RBF

--------------------------------------

Training Accuracy in loop iteration 1: 1.0

Testing Accuracy in loop iteration 1: 1.0

--------------------------------------

--------------------------------------

Training Accuracy in loop iteration 2: 1.0

Testing Accuracy in loop iteration 2: 1.0

--------------------------------------

--------------------------------------

Training Accuracy in loop iteration 3: 1.0

Testing Accuracy in loop iteration 3: 1.0

--------------------------------------

--------------------------------------

Training Accuracy in loop iteration 4: 1.0

Testing Accuracy in loop iteration 4: 1.0

--------------------------------------

--------------------------------------

Training Accuracy in loop iteration 5: 1.0

Testing Accuracy in loop iteration 5: 1.0

--------------------------------------

--------------------------------------

Training Accuracy in loop iteration 6: 1.0

Testing Accuracy in loop iteration 6: 1.0

--------------------------------------

--------------------------------------

Training Accuracy in loop iteration 7: 1.0

Testing Accuracy in loop iteration 7: 1.0

--------------------------------------

--------------------------------------

Training Accuracy in loop iteration 8: 1.0

Testing Accuracy in loop iteration 8: 1.0

--------------------------------------

--------------------------------------

Training Accuracy in loop iteration 9: 1.0

Testing Accuracy in loop iteration 9: 1.0

--------------------------------------

--------------------------------------

Training Accuracy in loop iteration 10: 1.0

Testing Accuracy in loop iteration 10: 1.0

--------------------------------------

[[11 0]

[ 0 9]]

precision recall f1-score support

0 1.00 1.00 1.00 11

1 1.00 1.00 1.00 9

accuracy 1.00 20

macro avg 1.00 1.00 1.00 20

weighted avg 1.00 1.00 1.00 20

**#OUTPUT (Sigmoid default)**

Sigmoid

--------------------------------------

Training Accuracy in loop iteration 1: 0.78

Testing Accuracy in loop iteration 1: 0.75

--------------------------------------

--------------------------------------

Training Accuracy in loop iteration 2: 0.78

Testing Accuracy in loop iteration 2: 0.75

--------------------------------------

--------------------------------------

Training Accuracy in loop iteration 3: 0.78

Testing Accuracy in loop iteration 3: 0.75

--------------------------------------

--------------------------------------

Training Accuracy in loop iteration 4: 0.78

Testing Accuracy in loop iteration 4: 0.75

--------------------------------------

--------------------------------------

Training Accuracy in loop iteration 5: 0.78

Testing Accuracy in loop iteration 5: 0.75

--------------------------------------

--------------------------------------

Training Accuracy in loop iteration 6: 0.78

Testing Accuracy in loop iteration 6: 0.75

--------------------------------------

--------------------------------------

Training Accuracy in loop iteration 7: 0.78

Testing Accuracy in loop iteration 7: 0.75

--------------------------------------

--------------------------------------

Training Accuracy in loop iteration 8: 0.78

Testing Accuracy in loop iteration 8: 0.75

--------------------------------------

--------------------------------------

Training Accuracy in loop iteration 9: 0.78

Testing Accuracy in loop iteration 9: 0.75

--------------------------------------

--------------------------------------

Training Accuracy in loop iteration 10: 0.78

Testing Accuracy in loop iteration 10: 0.75

--------------------------------------

[[8 3]

[2 7]]

precision recall f1-score support

0 0.80 0.73 0.76 11

1 0.70 0.78 0.74 9

accuracy 0.75 20

macro avg 0.75 0.75 0.75 20

weighted avg 0.76 0.75 0.75 20

**#OUTPUT (Poly 0 degrees)**

Poly

--------------------------------------

Training Accuracy in loop iteration 1: 0.5

Testing Accuracy in loop iteration 1: 0.45

--------------------------------------

--------------------------------------

Training Accuracy in loop iteration 2: 0.5

Testing Accuracy in loop iteration 2: 0.45

--------------------------------------

--------------------------------------

Training Accuracy in loop iteration 3: 0.5

Testing Accuracy in loop iteration 3: 0.45

--------------------------------------

--------------------------------------

Training Accuracy in loop iteration 4: 0.5

Testing Accuracy in loop iteration 4: 0.45

--------------------------------------

--------------------------------------

Training Accuracy in loop iteration 5: 0.5

Testing Accuracy in loop iteration 5: 0.45

--------------------------------------

--------------------------------------

Training Accuracy in loop iteration 6: 0.5

Testing Accuracy in loop iteration 6: 0.45

--------------------------------------

--------------------------------------

Training Accuracy in loop iteration 7: 0.5

Testing Accuracy in loop iteration 7: 0.45

--------------------------------------

--------------------------------------

Training Accuracy in loop iteration 8: 0.5

Testing Accuracy in loop iteration 8: 0.45

--------------------------------------

--------------------------------------

Training Accuracy in loop iteration 9: 0.5

Testing Accuracy in loop iteration 9: 0.45

--------------------------------------

--------------------------------------

Training Accuracy in loop iteration 10: 0.5

Testing Accuracy in loop iteration 10: 0.45

--------------------------------------

[[ 0 11]

[ 0 9]]

precision recall f1-score support

0 0.00 0.00 0.00 11

1 0.45 1.00 0.62 9

accuracy 0.45 20

macro avg 0.23 0.50 0.31 20

weighted avg 0.20 0.45 0.28 20

**#Detailed Table**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Classifier** | **Kernel** | **parameters** | **training accuracy each fold** | **avg training acc** | **testing accuracy each fold** | **avg testing acc** |
| LinearSVC | - |  | 0.94 | 0.94 | 0.9 | 0.9 |
| SVC | Linear | - | 0.94 | 0.94 | 0.9 | 0.9 |
|  | RBF | gamma=auto, C=default | 0.95 | 0.95 | 0.9 | 0.9 |
|  | RBF | tweak gamma=7 C=default | 1 | 1 | 1 | 1 |
|  | RBF | - | 0.95 | 0.95 | 0.9 | 0.9 |
|  |  | tweak C= 1 | 0.95 | 0.95 | 0.95 | 0.95 |
|  |  | tweak C= 2 | 0.94 | 0.94 | 0.9 | 0.9 |
|  |  | gamma=7 , C=2 | 1 | 1 | 1 | 1 |
|  | Sigmoid |  | 0.78 | 0.78 | 0.75 | 0.75 |
|  | Polynomial | degree=default | 0.92 | 0.92 | 0.9 | 0.9 |
|  |  | degree = 2 | 0.57 | 0.57 | 0.6 | 0.6 |
|  |  | degree= 1 | 0.94 | 0.94 | 0.9 | 0.9 |
|  |  | degree= 0 | 0.5 | 0.5 | 0.45 | 0.45 |