Project - Devanagari Handwritten Character Classification

December 7, 2019

[0]:
```
from google.colab import drive
drive.mount('/content/drive')
```

[0]:
```
!unzip -q "/content/drive/My Drive/Colab Notebooks/
 ↪DevanagariHandwrittenCharacterDataset.zip"
```

[0]:
```
import numpy as np
import cv2
from google.colab.patches import cv2_imshow
import os
import random
import matplotlib.pyplot as plt
from PIL import Image
from sklearn.svm import SVC
from sklearn.metrics import classification_report
import shutil
from sklearn.metrics import accuracy_score
```

[0]:
```
#Extract names of classes from dataset folders
def fetchName(folderName):
    if folderName[0] == 'c':
        return str(folderName[(folderName.index('_',folderName.index('_')+1))+1:
 ↪])
    else:
        return str("d"+folderName[folderName.index('_')+1:])
```

[0]:
```
#Function to perform fetch the entire dataset
wholeDataset = []
def fetchDataset(rootPath):
    folders = os.listdir(rootPath)
    for fold in folders:
        files = os.listdir(rootPath+fold)
        for eachFile in files:
            openImage = cv2.imread(rootPath+fold+"/"+eachFile)
            grayscale = cv2.cvtColor(openImage, cv2.COLOR_BGR2GRAY)
            ret,thresh1 = cv2.threshold(grayscale,100,255,cv2.THRESH_BINARY)
            singleLineImage = np.array(thresh1).flatten()
```

```
            replacedValues = [1 if singleLineImage[i]==255 else␣
↪singleLineImage[i] for i in range(len(singleLineImage))]
            wholeDataset.append([replacedValues,fetchName(fold)])
```

[0]:
```
#Function defined in such a way that each fold can be performed seperately
startPoint = int(len(wholeDataset)/5)
def crossValidation(k,currentFold):
    for i in range(k):
        if i == (currentFold-1):
            print("("+str(startPoint*i)+","+str(startPoint*(i+1)-1)+")")
            performingFold([startPoint*i,startPoint*(i+1)-1],i)
```

[0]:
```
#Creating train and test Arrays and performing SVM
def performingFold(startEndPoints,k):
    trainDataset = []
    testDataset = []
    for j in range(len(wholeDataset)):
        if j >= startEndPoints[0] and j <= startEndPoints[1]:
            testDataset.append(wholeDataset[j])
        else:
            trainDataset.append(wholeDataset[j])
    performSVM(trainDataset, testDataset, k)
```

[0]:
```
#Collecting the Dataset into a Single Array
fetchDataset('DevanagariHandwrittenCharacterDataset/')
#Shuffling the Dataset
random.shuffle(wholeDataset)
```

[0]:
```
accuracyTrace = []
def performSVM(trainDataset, testDataset, k):
    print("Performing Fold-{}".format(k+1))
    print("-------------------")
    x_train,y_train,x_test,y_test = [],[],[],[]
    for p in range(len(trainDataset)):
        x_train.append(trainDataset[p][0])
        y_train.append(trainDataset[p][1])
    for q in range(len(testDataset)):
        x_test.append(testDataset[q][0])
        y_test.append(testDataset[q][1])
    print(np.shape(x_train),np.shape(y_train),np.shape(x_test),np.shape(y_test))

    #Building a Support Vector Machine having Gaussian Kernel
    model = SVC(kernel='rbf',decision_function_shape='ovo',gamma='scale')
    model.fit(x_train, y_train)

    #Predicting
    y_pred = model.predict(x_test)

    #Showing Entire Fold Accuracy as well as Class Wise
```

```
    print(classification_report(y_test, y_pred))

    #Storing Accuracy
    accuracyTrace.append(accuracy_score(y_test,y_pred))
    print("Fold Accuracy: {}".format(accuracy_score(y_test,y_pred)))
```

[0]:
```
#5-Fold Cross Validation
#Fold No. 1
crossValidation(5,1)
```

```
(0,18399)
Performing Fold-1
------------------
(73600, 1024) (73600,) (18400, 1024) (18400,)
              precision    recall  f1-score   support

        adna       0.89      0.96      0.92       368
          ba       0.89      0.84      0.87       404
         bha       0.91      0.91      0.91       405
         cha       0.91      0.95      0.93       454
        chha       0.91      0.91      0.91       383
       chhya       0.91      0.92      0.91       400
          d0       0.98      0.99      0.99       385
          d1       0.97      0.99      0.98       415
          d2       0.94      0.96      0.95       408
          d3       0.96      0.95      0.95       403
          d4       0.97      0.98      0.97       383
          d5       0.97      0.99      0.98       409
          d6       0.98      0.97      0.97       461
          d7       0.99      0.98      0.98       374
          d8       0.97      0.99      0.98       403
          d9       0.98      0.98      0.98       407
          da       0.90      0.88      0.89       394
         daa       0.92      0.92      0.92       368
         dha       0.90      0.90      0.90       397
        dhaa       0.95      0.93      0.94       374
          ga       0.94      0.90      0.92       386
         gha       0.86      0.92      0.89       409
         gya       0.94      0.94      0.94       415
          ha       0.93      0.92      0.93       377
          ja       0.94      0.95      0.95       371
         jha       0.98      0.95      0.97       427
          ka       0.96      0.96      0.96       385
         kha       0.94      0.90      0.92       444
         kna       0.94      0.90      0.92       397
          la       0.98      0.94      0.96       387
          ma       0.90      0.89      0.89       413
    motosaw       0.89      0.94      0.92       401
```

```
           na       0.93      0.91      0.92       439
           pa       0.86      0.92      0.89       413
   patalosaw        0.87      0.90      0.88       389
petchiryakha        0.91      0.91      0.91       384
          pha       0.96      0.97      0.96       423
           ra       0.96      0.95      0.96       413
    taamatar        0.94      0.95      0.94       385
      tabala        0.97      0.93      0.95       374
          tha       0.89      0.87      0.88       429
         thaa       0.95      0.94      0.94       407
          tra       0.92      0.90      0.91       378
          waw       0.90      0.91      0.90       379
          yaw       0.86      0.86      0.86       390
          yna       0.96      0.90      0.93       390

    accuracy                            0.93     18400
   macro avg        0.93      0.93      0.93     18400
weighted avg        0.93      0.93      0.93     18400

Fold Accuracy: 0.931141304347826
```

[0]:
```
#Fold No. 2
crossValidation(5,2)
```

```
(18400,36799)
Performing Fold-2
------------------
(73600, 1024) (73600,) (18400, 1024) (18400,)
              precision    recall  f1-score   support

        adna       0.90      0.95      0.93       432
          ba       0.88      0.86      0.87       416
         bha       0.93      0.90      0.91       413
         cha       0.90      0.93      0.92       372
        chha       0.93      0.89      0.91       404
       chhya       0.92      0.93      0.93       424
          d0       0.97      0.99      0.98       378
          d1       0.97      0.99      0.98       388
          d2       0.94      0.97      0.95       396
          d3       0.96      0.96      0.96       422
          d4       0.97      0.98      0.98       387
          d5       0.97      0.98      0.98       393
          d6       0.98      0.98      0.98       383
          d7       0.99      0.97      0.98       392
          d8       0.97      0.98      0.97       406
          d9       0.98      1.00      0.99       412
          da       0.86      0.88      0.87       405
```

```
          daa       0.88      0.89      0.89       390
          dha       0.88      0.89      0.89       421
         dhaa       0.92      0.92      0.92       398
           ga       0.94      0.93      0.94       396
          gha       0.85      0.87      0.86       393
          gya       0.92      0.94      0.93       385
           ha       0.92      0.94      0.93       407
           ja       0.94      0.93      0.94       381
          jha       0.98      0.95      0.97       374
           ka       0.97      0.97      0.97       402
          kha       0.91      0.94      0.93       397
          kna       0.92      0.86      0.89       370
           la       0.95      0.93      0.94       421
           ma       0.89      0.86      0.88       383
      motosaw       0.90      0.93      0.91       381
           na       0.89      0.92      0.91       384
           pa       0.84      0.90      0.87       378
    patalosaw       0.88      0.87      0.88       416
petchiryakha       0.90      0.94      0.92       405
          pha       0.97      0.95      0.96       377
           ra       0.95      0.94      0.95       394
     taamatar       0.97      0.92      0.95       426
       tabala       0.94      0.95      0.95       442
          tha       0.88      0.84      0.86       409
         thaa       0.96      0.94      0.95       417
          tra       0.89      0.90      0.89       393
          waw       0.91      0.88      0.89       423
          yaw       0.89      0.84      0.86       420
          yna       0.95      0.93      0.94       394

     accuracy                           0.93     18400
    macro avg       0.93      0.93      0.93     18400
weighted avg       0.93      0.93      0.93     18400

Fold Accuracy: 0.9265217391304348
```

[0]:
```
#Fold No. 3
crossValidation(5,3)
```

```
(36800,55199)
Performing Fold-3
------------------
(73600, 1024) (73600,) (18400, 1024) (18400,)
              precision    recall  f1-score   support

        adna       0.91      0.96      0.93       390
          ba       0.88      0.85      0.87       406
```

bha	0.93	0.93	0.93	368
cha	0.92	0.96	0.94	374
chha	0.92	0.88	0.90	392
chhya	0.94	0.91	0.93	438
d0	0.99	1.00	0.99	422
d1	0.97	0.99	0.98	406
d2	0.95	0.95	0.95	400
d3	0.94	0.95	0.95	387
d4	0.96	0.99	0.98	433
d5	0.98	0.99	0.99	387
d6	0.97	0.97	0.97	409
d7	0.98	0.99	0.99	410
d8	0.97	0.97	0.97	390
d9	0.99	0.99	0.99	403
da	0.92	0.89	0.91	404
daa	0.86	0.92	0.89	371
dha	0.90	0.89	0.90	379
dhaa	0.92	0.94	0.93	422
ga	0.95	0.92	0.94	423
gha	0.89	0.90	0.90	409
gya	0.95	0.92	0.93	433
ha	0.94	0.91	0.93	435
ja	0.93	0.95	0.94	408
jha	0.97	0.98	0.98	427
ka	0.96	0.97	0.96	377
kha	0.95	0.89	0.92	403
kna	0.92	0.88	0.90	399
la	0.97	0.94	0.96	391
ma	0.92	0.92	0.92	409
motosaw	0.86	0.93	0.89	352
na	0.93	0.90	0.91	375
pa	0.87	0.91	0.89	433
patalosaw	0.89	0.87	0.88	397
petchiryakha	0.88	0.93	0.91	393
pha	0.96	0.96	0.96	410
ra	0.97	0.95	0.96	382
taamatar	0.94	0.96	0.95	400
tabala	0.93	0.95	0.94	413
tha	0.87	0.85	0.86	356
thaa	0.94	0.94	0.94	377
tra	0.90	0.93	0.91	401
waw	0.90	0.87	0.88	391
yaw	0.89	0.88	0.89	403
yna	0.96	0.93	0.95	412
accuracy			0.93	18400
macro avg	0.93	0.93	0.93	18400
weighted avg	0.93	0.93	0.93	18400

```
Fold Accuracy: 0.9323369565217391
```

[30]:
```
#Fold No. 4
crossValidation(5,4)
```

```
(55200,73599)
Performing Fold-4
------------------
(73600, 1024) (73600,) (18400, 1024) (18400,)
              precision    recall  f1-score   support

        adna       0.92      0.94      0.93       422
          ba       0.89      0.88      0.88       392
         bha       0.91      0.91      0.91       411
         cha       0.91      0.94      0.92       409
        chha       0.93      0.93      0.93       404
       chhya       0.93      0.97      0.95       360
          d0       0.99      0.99      0.99       395
          d1       0.98      0.99      0.99       406
          d2       0.94      0.97      0.95       381
          d3       0.96      0.96      0.96       400
          d4       0.97      0.98      0.98       425
          d5       0.98      0.97      0.98       446
          d6       0.99      0.96      0.97       353
          d7       0.99      1.00      0.99       421
          d8       0.97      0.98      0.98       402
          d9       0.99      0.99      0.99       371
          da       0.91      0.89      0.90       377
         daa       0.90      0.93      0.91       427
         dha       0.87      0.87      0.87       389
        dhaa       0.95      0.94      0.95       414
          ga       0.95      0.92      0.93       428
         gha       0.88      0.86      0.87       413
         gya       0.94      0.93      0.94       375
          ha       0.91      0.94      0.93       378
          ja       0.94      0.93      0.94       414
         jha       0.97      0.96      0.97       364
          ka       0.97      0.97      0.97       408
         kha       0.91      0.93      0.92       373
         kna       0.94      0.91      0.92       398
          la       0.97      0.92      0.94       428
          ma       0.89      0.84      0.86       389
    motosaw       0.89      0.92      0.90       437
          na       0.93      0.91      0.92       413
          pa       0.83      0.93      0.87       405
  patalosaw       0.88      0.85      0.87       372
```

```
petchiryakha       0.89      0.91      0.90       406
         pha       0.96      0.94      0.95       358
          ra       0.97      0.95      0.96       402
    taamatar       0.97      0.95      0.96       408
      tabala       0.93      0.96      0.95       400
         tha       0.89      0.88      0.88       384
        thaa       0.96      0.96      0.96       403
         tra       0.91      0.90      0.90       415
         waw       0.88      0.87      0.88       391
         yaw       0.87      0.87      0.87       415
         yna       0.95      0.94      0.94       418

    accuracy                           0.93     18400
   macro avg       0.93      0.93      0.93     18400
weighted avg       0.93      0.93      0.93     18400

Fold Accuracy: 0.9309782608695653
```

[31]:
```
#Fold No. 5
crossValidation(5,5)
```

```
(73600,91999)
Performing Fold-5
------------------
(73600, 1024) (73600,) (18400, 1024) (18400,)
              precision    recall  f1-score   support

        adna       0.90      0.96      0.93       388
          ba       0.86      0.86      0.86       382
         bha       0.93      0.92      0.92       403
         cha       0.92      0.93      0.92       391
        chha       0.92      0.88      0.90       417
       chhya       0.95      0.94      0.95       378
          d0       0.98      1.00      0.99       420
          d1       0.96      0.99      0.97       385
          d2       0.94      0.98      0.96       415
          d3       0.96      0.96      0.96       388
          d4       0.97      0.97      0.97       372
          d5       0.96      0.98      0.97       365
          d6       0.98      0.97      0.97       394
          d7       0.98      0.99      0.99       403
          d8       0.97      0.98      0.98       399
          d9       0.98      0.99      0.98       407
          da       0.92      0.92      0.92       420
         daa       0.90      0.89      0.90       444
         dha       0.89      0.87      0.88       414
        dhaa       0.94      0.93      0.93       392
```

```
              ga       0.93      0.93      0.93       367
             gha       0.86      0.86      0.86       376
             gya       0.93      0.95      0.94       392
              ha       0.92      0.90      0.91       403
              ja       0.96      0.94      0.95       426
             jha       0.98      0.96      0.97       408
              ka       0.98      0.96      0.97       428
             kha       0.94      0.91      0.92       383
             kna       0.93      0.87      0.90       436
              la       0.96      0.95      0.96       373
              ma       0.89      0.89      0.89       406
         motosaw       0.90      0.94      0.92       429
              na       0.91      0.91      0.91       389
              pa       0.83      0.91      0.86       371
       patalosaw       0.90      0.87      0.88       426
    petchiryakha       0.88      0.94      0.91       412
             pha       0.97      0.95      0.96       432
              ra       0.96      0.91      0.94       409
        taamatar       0.96      0.97      0.96       381
          tabala       0.96      0.95      0.96       371
             tha       0.90      0.84      0.87       422
            thaa       0.95      0.96      0.95       396
             tra       0.91      0.91      0.91       413
             waw       0.90      0.89      0.89       416
             yaw       0.86      0.90      0.88       372
             yna       0.96      0.94      0.95       386

        accuracy                           0.93     18400
       macro avg       0.93      0.93      0.93     18400
    weighted avg       0.93      0.93      0.93     18400

    Fold Accuracy: 0.9302717391304348
```

[35]:
```
print("Average Accuracy: {}".format(np.mean(accuracyTrace)))
```

```
Average Accuracy: 0.93025
```