VIDYA PAYGUDE

Data Science Intern at LetsGrowMore Virtual Internship Program (APRIL-2022)

ADVANCED LEVEL TASK 9 - Handwritten equation solver using CNN

In [ ]:

```
import numpy as np
import cv2
from PIL import Image
from matplotlib import pyplot as plt
%matplotlib inline
import os
from os import listdir
from os.path import isfile, join
import pandas as pd
```

In [ ]:

```
def load_images_from_folder(folder):
    train_data=[]
    for filename in os.listdir(folder):
        img = cv2.imread(os.path.join(folder,filename),cv2.IMREAD_GRAYSCALE)
        img=~img
        if img is not None:
            ret,thresh=cv2.threshold(img,127,255,cv2.THRESH_BINARY)

            ctrs,ret=cv2.findContours(thresh,cv2.RETR_EXTERNAL,cv2.CHAIN_APPROX_NONE)
            cnt=sorted(ctrs, key=lambda ctr: cv2.boundingRect(ctr)[0])
            w=int(28)
            h=int(28)
            maxi=0
            for c in cnt:
                x,y,w,h=cv2.boundingRect(c)
                maxi=max(w*h,maxi)
                if maxi==w*h:
                    x_max=x
                    y_max=y
                    w_max=w
                    h_max=h
            im_crop= thresh[y_max:y_max+h_max+10, x_max:x_max+w_max+10]
```

```
            im_resize = cv2.resize(im_crop,(28,28))
            im_resize=np.reshape(im_resize,(784,1))
            train_data.append(im_resize)
    return train_data
```

In [ ]:
```
data=[]
```

In [ ]:
```
# Assign '-' = 10
data=load_images_from_folder('D://LGMVIP-DataScience//task 9 Handwritten equation solver using CNN//extracted_images//-//')
len(data)
for i in range(0,len(data)):
    data[i]=np.append(data[i],['10'])

print(len(data))
```

```
33997
```

In [ ]:
```
# Assign + = 11
data11=load_images_from_folder('D://LGMVIP-DataScience//task 9 Handwritten equation solver using CNN//extracted_images//+//')

for i in range(0,len(data11)):
    data11[i]=np.append(data11[i],['11'])
data=np.concatenate((data,data11))
print(len(data))
```

```
59109
```

In [ ]:
```
data0=load_images_from_folder('D://LGMVIP-DataScience//task 9 Handwritten equation solver using CNN//extracted_images//0//')

for i in range(0,len(data0)):
    data0[i]=np.append(data0[i],['0'])
data=np.concatenate((data,data0))
print(len(data))
```

```
66023
```

In [ ]:
```
data1=load_images_from_folder('D://LGMVIP-DataScience//task 9 Handwritten equation solver using CNN//extracted_images//1//')

for i in range(0,len(data1)):
```

```
    data1[i]=np.append(data1[i],['1'])
data=np.concatenate((data,data1))
print(len(data))
```

```
92543
```

In [ ]:

```
data2=load_images_from_folder('D://LGMVIP-DataScience//task 9 Handwritten equation solver using CNN//extracted_images//2//')

for i in range(0,len(data2)):
    data2[i]=np.append(data2[i],['2'])
data=np.concatenate((data,data2))
print(len(data))
```

```
118684
```

In [ ]:

```
data3=load_images_from_folder('D://LGMVIP-DataScience//task 9 Handwritten equation solver using CNN//extracted_images//3//')

for i in range(0,len(data3)):
    data3[i]=np.append(data3[i],['3'])
data=np.concatenate((data,data3))
print(len(data))
```

```
129593
```

In [ ]:

```
data4=load_images_from_folder('D://LGMVIP-DataScience//task 9 Handwritten equation solver using CNN//extracted_images//4//')

for i in range(0,len(data4)):
    data4[i]=np.append(data4[i],['4'])
data=np.concatenate((data,data4))
print(len(data))
```

```
136989
```

In [ ]:

```
data5=load_images_from_folder('D://LGMVIP-DataScience//task 9 Handwritten equation solver using CNN//extracted_images//5//')

for i in range(0,len(data5)):
    data5[i]=np.append(data5[i],['5'])
data=np.concatenate((data,data5))
print(len(data))
```

```
140534
```

In [ ]:
```
data6=load_images_from_folder('D://LGMVIP-DataScience//task 9 Handwritten equation solver using CNN//extracted_images//6//')

for i in range(0,len(data6)):
    data6[i]=np.append(data6[i],['6'])
data=np.concatenate((data,data6))
print(len(data))
```

```
143652
```

In [ ]:
```
data7=load_images_from_folder('D://LGMVIP-DataScience//task 9 Handwritten equation solver using CNN//extracted_images//7//')

for i in range(0,len(data7)):
    data7[i]=np.append(data7[i],['7'])
data=np.concatenate((data,data7))
print(len(data))
```

```
146561
```

In [ ]:
```
data8=load_images_from_folder('D://LGMVIP-DataScience//task 9 Handwritten equation solver using CNN//extracted_images//8//')

for i in range(0,len(data8)):
    data8[i]=np.append(data8[i],['8'])
data=np.concatenate((data,data8))
print(len(data))
```

```
149629
```

In [ ]:
```
data9=load_images_from_folder('D://LGMVIP-DataScience//task 9 Handwritten equation solver using CNN//extracted_images//9//')

for i in range(0,len(data9)):
    data9[i]=np.append(data9[i],['9'])
data=np.concatenate((data,data9))
print(len(data))
```

```
153366
```

In [ ]:
```
data12=load_images_from_folder('D://LGMVIP-DataScience//task 9 Handwritten equation solver using CNN//extracted_images//times//')

for i in range(0,len(data12)):
    data12[i]=np.append(data12[i],['12'])
```

```
data=np.concatenate((data,data12))
print(len(data))
```

```
156617
```

In [ ]:

```
df=pd.DataFrame(data,index=None)
df.to_csv('train_final.csv',index=False)
```

In [ ]: