

Microsoft Malware detection

1.Business/Real-world Problem

1.1. What is Malware?

The term malware is a contraction of malicious software. Put simply, malware is any piece of software that was written with the intent of doing harm to data, devices or to people.

Source: <https://www.avg.com/en/signal/what-is-malware> (<https://www.avg.com/en/signal/what-is-malware>).

1.2. Problem Statement

In the past few years, the malware industry has grown very rapidly that, the syndicates invest heavily in technologies to evade traditional protection, forcing the anti-malware groups/communities to build more robust softwares to detect and terminate these attacks. The major part of protecting a computer system from a malware attack is to **identify whether a given piece of file/software is a malware.**

1.3 Source/Useful Links

Microsoft has been very active in building anti-malware products over the years and it runs its anti-malware utilities over 150 million computers around the world. This generates tens of millions of daily data points to be analyzed as potential malware. In order to be effective in analyzing and classifying such large amounts of data, we need to be able to group them into groups and identify their respective families.

This dataset provided by Microsoft contains about 9 classes of malware. ,

Source: <https://www.kaggle.com/c/malware-classification>

1.4. Real-world/Business objectives and constraints.

1. Minimize multi-class error.
2. Multi-class probability estimates.
3. Malware detection should not take hours and block the user's computer. It should finish in a few seconds or a minute.

2. Machine Learning Problem

2.1. Data

2.1.1. Data Overview

- **Source :** <https://www.kaggle.com/c/malware-classification/data>
- **For every malware, we have two files**
 1. **.asm file** (read more: <https://www.reviversoft.com/file-extensions/asm>)
 2. **.bytes file** (the raw data contains the hexadecimal representation of the file's binary content, without the PE header)
- **Total train dataset consist of 200GB data out of which 50Gb of data is .bytes files and 150GB of data is .asm files:**
- **Lots of Data for a single-box/computer.**
- **There are total 10,868 .bytes files and 10,868 asm files total 21,736 files**
- **There are 9 types of malwares (9 classes) in our give data**
- **Types of Malware:**
 1. **Ramnit**
 2. **Lollipop**
 3. **Kelihos_ver3**
 4. **Vundo**
 5. **Simda**
 6. **Tracur**
 7. **Kelihos_ver1**
 8. **Obfuscator.ACY**
 9. **Gatak**

2.1.2. Example Data Point

.asm file

```
.text:00401000      assume es:nothing, ss:nothing, ds:_data, fs:nothing, gs:nothing
.text:00401000 56      push     esi
.text:00401001 8D 44 24 08      lea     eax, [esp+8]
.text:00401005 50      push     eax
.text:00401006 8B F1      mov     esi, ecx
.text:00401008 E8 1C 1B 00 00      call    ??0exception@std@@@QAE@ABQBD@Z ; std::exception::exception(char const * const &)
.text:0040100D C7 06 08 BB 42 00      mov     dword ptr [esi], offset off_42BB08
.text:00401013 8B C6      mov     eax, esi
.text:00401015 5E      pop     esi
.text:00401016 C2 04 00      retn    4
.text:00401016      ; -----
.text:00401019 CC CC CC CC CC CC CC      align 10h
.text:00401020 C7 01 08 BB 42 00      mov     dword ptr [ecx], offset off_42BB08
.text:00401026 E9 26 1C 00 00      jmp     sub_402C51
.text:00401026      ; -----
.text:0040102B CC CC CC CC CC      align 10h
.text:00401030 56      push     esi
.text:00401031 8B F1      mov     esi, ecx
.text:00401033 C7 06 08 BB 42 00      mov     dword ptr [esi], offset off_42BB08
.text:00401039 E8 13 1C 00 00      call    sub_402C51
.text:0040103E F6 44 24 08 01      test    byte ptr [esp+8], 1
.text:00401043 74 09      jz      short loc_40104E
.text:00401045 56      push     esi
.text:00401046 E8 6C 1E 00 00      call    ??3@YAXPAX@Z ; operator delete(void *)
.text:0040104B 83 C4 04      add     esp, 4
.text:0040104E
.text:0040104E      loc_40104E: ; CODE XREF: .text:00401043j
.text:0040104E 8B C6      mov     eax, esi
.text:00401050 5E      pop     esi
.text:00401051 C2 04 00      retn    4
.text:00401051      ; -----
```

.bytes file

00401000 00 00 80 40 40 28 00 1C 02 42 00 C4 00 20 04 20
00401010 00 00 20 09 2A 02 00 00 00 00 8E 10 41 0A 21 01
00401020 40 00 02 01 00 90 21 00 32 40 00 1C 01 40 C8 18
00401030 40 82 02 63 20 00 00 09 10 01 02 21 00 82 00 04
00401040 82 20 08 83 00 08 00 00 00 00 02 00 60 80 10 80
00401050 18 00 00 20 A9 00 00 00 00 04 04 78 01 02 70 90
00401060 00 02 00 08 20 12 00 00 00 40 10 00 80 00 40 19
00401070 00 00 00 00 11 20 80 04 80 10 00 20 00 00 25 00
00401080 00 00 01 00 00 04 00 10 02 C1 80 80 00 20 20 00
00401090 08 A0 01 01 44 28 00 00 08 10 20 00 02 08 00 00
004010A0 00 40 00 00 00 34 40 40 00 04 00 08 80 08 00 08
004010B0 10 00 40 00 68 02 40 04 E1 00 28 14 00 08 20 0A
004010C0 06 01 02 00 40 00 00 00 00 00 20 00 02 00 04
004010D0 80 18 90 00 00 10 A0 00 45 09 00 10 04 40 44 82
004010E0 90 00 26 10 00 00 04 00 82 00 00 00 20 40 00 00
004010F0 B4 00 00 40 00 02 20 25 08 00 00 00 00 00 00
00401100 08 00 00 50 00 08 40 50 00 02 06 22 08 85 30 00
00401110 00 80 00 80 60 00 09 00 04 20 00 00 00 00 00
00401120 00 82 40 02 00 11 46 01 4A 01 8C 01 E6 00 86 10
00401130 4C 01 22 00 64 00 AE 01 EA 01 2A 11 E8 10 26 11
00401140 4E 11 8E 11 C2 00 6C 00 0C 11 60 01 CA 00 62 10
00401150 6C 01 A0 11 CE 10 2C 11 4E 10 8C 00 CE 01 AE 01
00401160 6C 10 6C 11 A2 01 AE 00 46 11 EE 10 22 00 A8 00
00401170 EC 01 08 11 A2 01 AE 10 6C 00 6E 00 AC 11 8C 00
00401180 EC 01 2A 10 2A 01 AE 00 40 00 C8 10 48 01 4E 11
00401190 0E 00 EC 11 24 10 4A 10 04 01 C8 11 E6 01 C2 00

2.2. Mapping the real-world problem to an ML problem

2.2.1. Type of Machine Learning Problem

There are nine different classes of malware that we need to classify a given a data point => Multi class classification problem

2.2.2. Performance Metric

Source: <https://www.kaggle.com/c/malware-classification#evaluation> (<https://www.kaggle.com/c/malware-classification#evaluation>)

Metric(s):

- Multi class log-loss
- Confusion matrix

2.2.3. Machine Learning Objectives and Constraints

Objective: Predict the probability of each data-point belonging to each of the nine classes.

Constraints:

- Class probabilities are needed.
- Penalize the errors in class probabilities => Metric is Log-loss.
- Some Latency constraints.

2.3. Train and Test Dataset

Split the dataset randomly into three parts train, cross validation and test with 64%,16%, 20% of data respectively

2.4. Useful blogs, videos and reference papers

<http://blog.kaggle.com/2015/05/26/microsoft-malware-winners-interview-1st-place-no-to-overfitting/>

<https://arxiv.org/pdf/1511.04317.pdf>

First place solution in Kaggle competition: <https://www.youtube.com/watch?v=VLQTRILGz5Y>

<https://github.com/dchad/malware-detection>

<http://vizsec.org/files/2011/Nataraj.pdf>

https://www.dropbox.com/sh/gfqzv0ckgs4l1bf/AAB6EeInEjvvuQg2nu_pIB6ua?dl=0

" Cross validation is more trustworthy than domain knowledge."

3. Exploratory Data Analysis

```
In [ ]: import warnings
warnings.filterwarnings("ignore")
import shutil
import os
import pandas as pd
import matplotlib
matplotlib.use('nbAgg')
import matplotlib.pyplot as plt
import seaborn as sns
import numpy as np
import pickle
from sklearn.manifold import TSNE
from sklearn import preprocessing
import pandas as pd
from multiprocessing import Process# this is used for multithreading
import multiprocessing
import codecs# this is used for file operations
import random as r
from xgboost import XGBClassifier
from sklearn.model_selection import RandomizedSearchCV
from sklearn.tree import DecisionTreeClassifier
from sklearn.calibration import CalibratedClassifierCV
from sklearn.neighbors import KNeighborsClassifier
from sklearn.metrics import log_loss
from sklearn.metrics import confusion_matrix
from sklearn.model_selection import train_test_split
from sklearn.linear_model import LogisticRegression
from sklearn.ensemble import RandomForestClassifier
```



```
In [ ]: #separating byte files and asm files
```

```
source = 'train'
destination_1 = 'byteFiles'
destination_2 = 'asmFiles'

# we will check if the folder 'byteFiles' exists if it not there we will create a folder with the same name
if not os.path.isdir(destination_1):
    os.makedirs(destination_1)
if not os.path.isdir(destination_2):
    os.makedirs(destination_2)

# if we have folder called 'train' (train folder contains both .asm files and .bytes files) we will rename it 'asmFiles'
# for every file that we have in our 'asmFiles' directory we check if it is ending with .bytes, if yes we will move it to
# 'byteFiles' folder

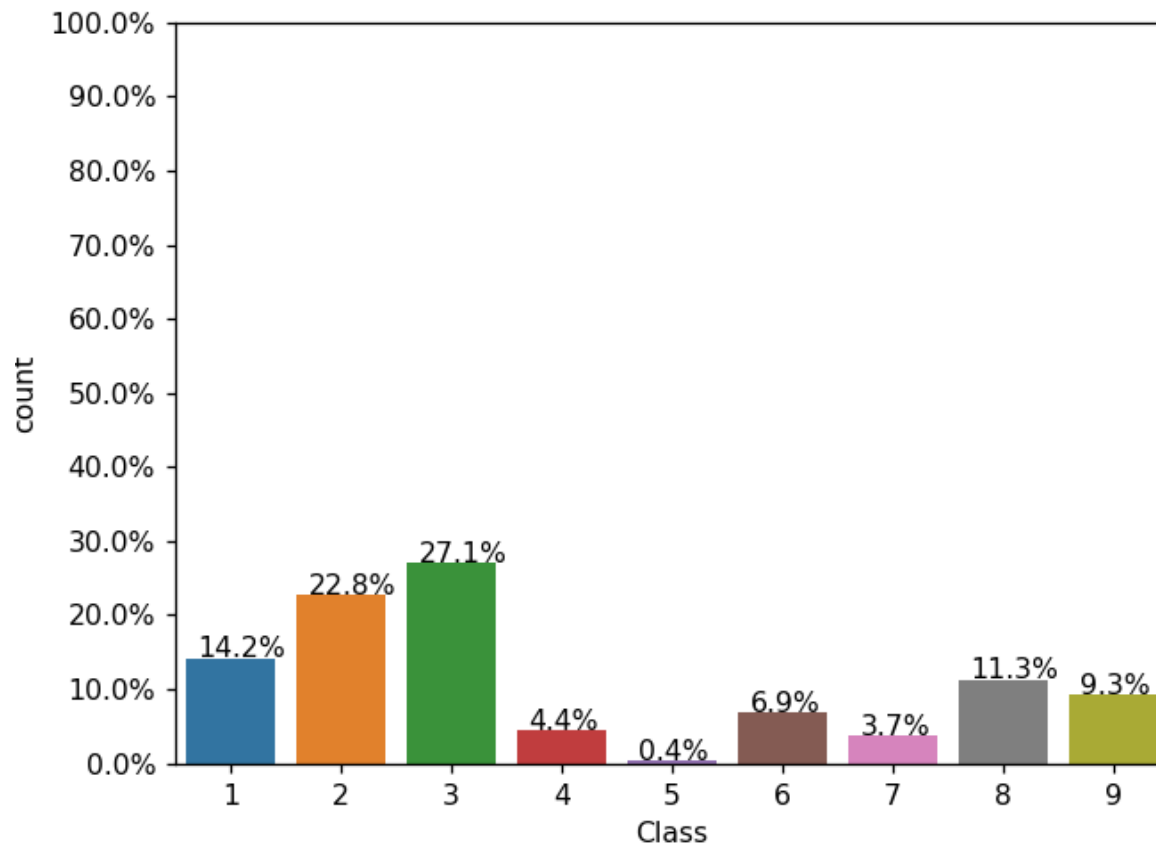
# so by the end of this snippet we will separate all the .byte files and .asm files
if os.path.isdir(source):
    data_files = os.listdir(source)
    for file in data_files:
        print(file)
        if (file.endswith("bytes")):
            shutil.move(source+'\\'+file,destination_1)
        if (file.endswith("asm")):
            shutil.move(source+'\\'+file,destination_2)
```

3.1. Distribution of malware classes in whole data set

```
In [ ]: Y=pd.read_csv("trainLabels.csv")
total = len(Y)*1.
ax=sns.countplot(x="Class", data=Y)
for p in ax.patches:
    ax.annotate('{:.1f}%'.format(100*p.get_height()/total), (p.get_x()+0.1, p.get_height()+5))

#put 11 ticks (therefore 10 steps), from 0 to the total number of rows in the dataframe
ax.yaxis.set_ticks(np.linspace(0, total, 11))

#adjust the ticklabel to the desired format, without changing the position of the ticks.
ax.set_yticklabels(map('{:.1f}%'.format, 100*ax.yaxis.get_majorticklocs()/total))
plt.show()
```



3.2. Feature extraction

3.2.1 File size of byte files as a feature

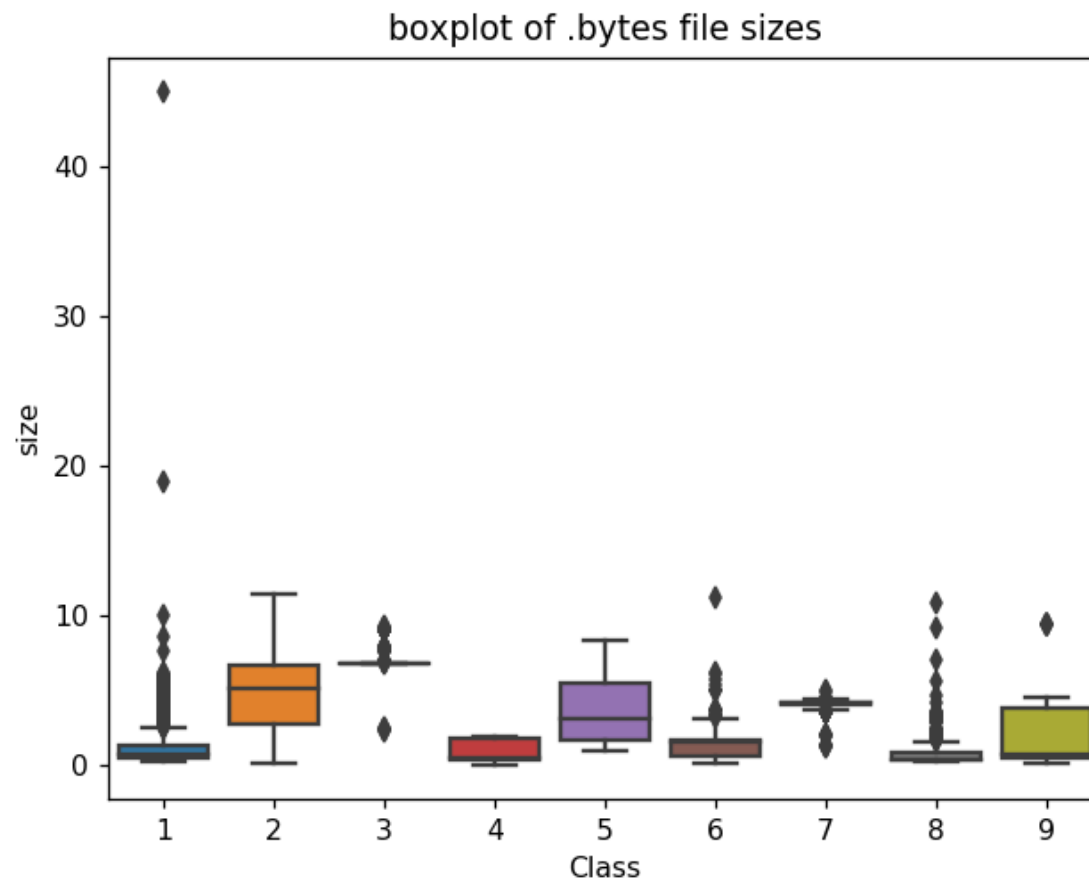
```
In [ ]: #file sizes of byte files

files=os.listdir('byteFiles')
filenames=Y['Id'].tolist()
class_y=Y['Class'].tolist()
class_bytes=[]
sizebytes=[]
fnames=[]
for file in files:
    # print(os.stat('byteFiles/0A32eTdBKayjCWhZqDOQ.txt'))
    # os.stat_result(st_mode=33206, st_ino=1125899906874507, st_dev=3561571700, st_nlink=1, st_uid=0, st_gid=0,
    # st_size=3680109, st_atime=1519638522, st_mtime=1519638522, st_ctime=1519638522)
    # read more about os.stat: here https://www.tutorialspoint.com/python/os\_stat.htm
    statinfo=os.stat('byteFiles/'+file)
    # split the file name at '.' and take the first part of it i.e the file name
    file=file.split('.')[0]
    if any(file == filename for filename in filenames):
        i=filenames.index(file)
        class_bytes.append(class_y[i])
        # converting into Mb's
        sizebytes.append(statinfo.st_size/(1024.0*1024.0))
        fnames.append(file)
data_size_byte=pd.DataFrame({'ID':fnames, 'size':sizebytes, 'Class':class_bytes})
print (data_size_byte.head())
```

	ID	size	Class
0	01azqd4InC7m9JpocGv5	4.234863	9
1	01IsoiSMh5gxyDYL14CB	5.538818	2
2	01jsnpXSA1gw6aPeDxrU	3.887939	9
3	01kcPWA9K2B0xQeS5Rju	0.574219	1
4	01SuzwMJEIXsK7A8dQb1	0.370850	8

3.2.2 box plots of file size (.byte files) feature

```
In [ ]: #boxplot of byte files
ax = sns.boxplot(x="Class", y="size", data=data_size_byte)
plt.title("boxplot of .bytes file sizes")
plt.show()
```



3.2.3 feature extraction from byte files

```

In [ ]: #removal of addres from byte files
# contents of .byte files
# -----
#00401000 56 8D 44 24 08 50 8B F1 E8 1C 1B 00 00 C7 06 08
#-----
#we remove the starting address 00401000

files = os.listdir('byteFiles')
filenames=[]
array=[]
for file in files:
    if(file.endswith("bytes")):
        file=file.split('.')[0]
        text_file = open('byteFiles/'+file+".txt", 'w+')
        with open('byteFiles/'+file+".bytes","r") as fp:
            lines=""
            for line in fp:
                a=line.rstrip().split(" ")[1:]
                b=' '.join(a)
                b=b+"\n"
                text_file.write(b)
            fp.close()
            os.remove('byteFiles/'+file+".bytes")
        text_file.close()

files = os.listdir('byteFiles')
filenames2=[]
feature_matrix = np.zeros((len(files),257),dtype=int)
k=0

#program to convert into bag of words of bytefiles
#this is custom-built bag of words this is unigram bag of words
byte_feature_file=open('result.csv','w+')
byte_feature_file.write("ID,0,1,2,3,4,5,6,7,8,9,0a,0b,0c,0d,0e,0f,10,11,12,13,14,15,16,17,18,19,1a,1b,1c,1d,1e,1f,20,21,22,23,24,25,26,27,28,29,2a,2b,2c,2d,2e,2f,30,31,32,33,34,35,36,37,38,39,3a,3b,3c,3d,3e,3f,40,41,42,43,44,45,46,47,48,49,4a,4b,4c,4d,4e,4f,50,51,52,53,54,55,56,57,58,59,5a,5b,5c,5d,5e,5f,60,61,62,63,64,65,66,67,68,69,6a,6b,6c,6d,6e,6f,70,71,72,73,74,75,76,77,78,79,7a,7b,7c,7d,7e,7f,80,81,82,83,84,85,86,87,88,89,8a,8b,8c,8d,8e,8f,90,91,92,93,94,95,96,97,98,99,9a,9b,9c,9d,9e,9f,a0,a1,a2,a3,a4,a5,a6,a7,a8,a9,aa,ab,ac,ad,ae,af,b0,b1,b2,b3,b4,b5,b6,b7,b8,b9,ba,bb,bc,bd,be,bf,c0,c1,c2,c3,c4,c5,c6,c7,c8,c9,ca,cb,cc,cd,ce,cf,d0,d1,d2,d3,d4,d5,d6,d7,d8,d9,da,db,dc,dd,de,df,e0,e1,e2,e3,e4,e5,e6,e7,e8,e9,ea,eb,ec,ed,ee,ef,f0,f1,f2,f3,f4,f5,f6,f7,f8,f9,fa,fb,fc,fd,fe,ff,??")
byte_feature_file.write("\n")
for file in files:
    filenames2.append(file)
    byte_feature_file.write(file+",")

```

```

if(file.endswith(".txt")):
    with open('byteFiles/'+file,"r") as byte_flie:
        for lines in byte_flie:
            line=line.rstrip().split(" ")
            for hex_code in line:
                if hex_code=='??':
                    feature_matrix[k][256]+=1
                else:
                    feature_matrix[k][int(hex_code,16)]+=1
            byte_flie.close()
for i, row in enumerate(feature_matrix[k]):
    if i!=len(feature_matrix[k])-1:
        byte_feature_file.write(str(row)+",")
    else:
        byte_feature_file.write(str(row))
byte_feature_file.write("\n")

k += 1

byte_feature_file.close()

```

```

In [ ]: byte_features=pd.read_csv("result.csv")
byte_features['ID'] = byte_features['ID'].str.split('.').str[0]
byte_features.head(2)

```

```

Out[ ]:

```

	ID	0	1	2	3	4	5	6	7	8	...	f7	f8	f9	fa	fb	fc	fd	fe	ff	??
0	01azqd4lnC7m9JpocGv5	601905	3905	2816	3832	3345	3242	3650	3201	2965	...	2804	3687	3101	3211	3097	2758	3099	2759	5753	1824
1	01lsoiSMh5gxyDYTI4CB	39755	8337	7249	7186	8663	6844	8420	7589	9291	...	451	6536	439	281	302	7639	518	17001	54902	8588

2 rows × 258 columns

```

In [ ]: data_size_byte.head(2)

```

```

Out[ ]:

```

	ID	size	Class
0	01azqd4lnC7m9JpocGv5	4.234863	9
1	01lsoiSMh5gxyDYTI4CB	5.538818	2

```
In [ ]: byte_features_with_size = byte_features.merge(data_size_byte, on='ID')
byte_features_with_size.to_csv("result_with_size.csv")
byte_features_with_size.head(2)
```

Out[]:

	ID	0	1	2	3	4	5	6	7	8	...	f9	fa	fb	fc	fd	fe	ff	??	size	Class
0	01azqd4lnC7m9JpocGv5	601905	3905	2816	3832	3345	3242	3650	3201	2965	...	3101	3211	3097	2758	3099	2759	5753	1824	4.234863	
1	01IsoiSMh5gxyDYTI4CB	39755	8337	7249	7186	8663	6844	8420	7589	9291	...	439	281	302	7639	518	17001	54902	8588	5.538818	

2 rows × 260 columns

```
In [ ]: # https://stackoverflow.com/a/29651514
def normalize(df):
    result1 = df.copy()
    for feature_name in df.columns:
        if (str(feature_name) != str('ID') and str(feature_name) != str('Class')):
            max_value = df[feature_name].max()
            min_value = df[feature_name].min()
            result1[feature_name] = (df[feature_name] - min_value) / (max_value - min_value)
    return result1
result = normalize(byte_features_with_size)
```

In []:

result.head(2)

Out[]:

	ID	0	1	2	3	4	5	6	7	8	...	f9	fa	fb	
0	01azqd4lnC7m9JpocGv5	0.262806	0.005498	0.001567	0.002067	0.002048	0.001835	0.002058	0.002946	0.002638	...	0.01356	0.013107	0.013634	0.031
1	01IsoiSMh5gxyDYTI4CB	0.017358	0.011737	0.004033	0.003876	0.005303	0.003873	0.004747	0.006984	0.008267	...	0.00192	0.001147	0.001329	0.087

2 rows × 260 columns

```
In [ ]: data_y = result['Class']
result.head()
```

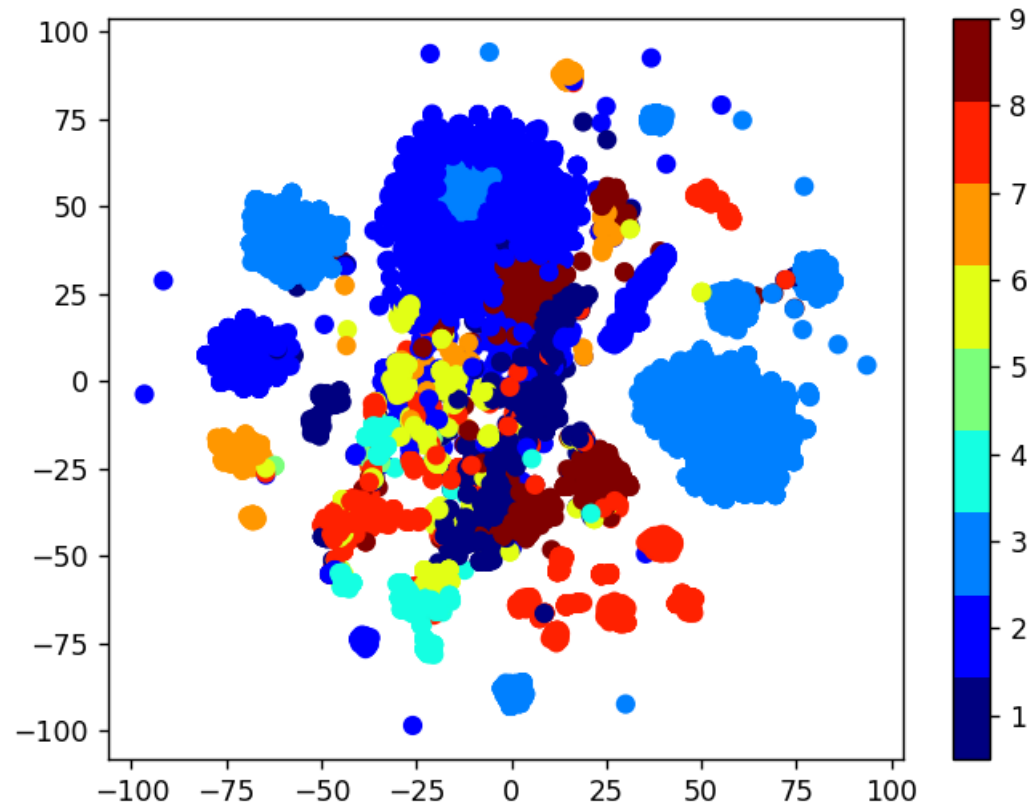
Out[]:

	ID	0	1	2	3	4	5	6	7	8	...	f9	fa	fb	
0	01azqd4lnC7m9JpocGv5	0.262806	0.005498	0.001567	0.002067	0.002048	0.001835	0.002058	0.002946	0.002638	...	0.013560	0.013107	0.013634	0.0
1	01lsoiSMh5gxyDYTI4CB	0.017358	0.011737	0.004033	0.003876	0.005303	0.003873	0.004747	0.006984	0.008267	...	0.001920	0.001147	0.001329	0.0
2	01jsnpXSAlgW6aPeDxrU	0.040827	0.013434	0.001429	0.001315	0.005464	0.005280	0.005078	0.002155	0.008104	...	0.009804	0.011777	0.012604	0.0
3	01kcPWA9K2BOxQeS5Rju	0.009209	0.001708	0.000404	0.000441	0.000770	0.000354	0.000310	0.000481	0.000959	...	0.002121	0.001886	0.002272	0.0
4	01SuzwMJEIXsK7A8dQbl	0.008629	0.001000	0.000168	0.000234	0.000342	0.000232	0.000148	0.000229	0.000376	...	0.001530	0.000853	0.001052	0.0

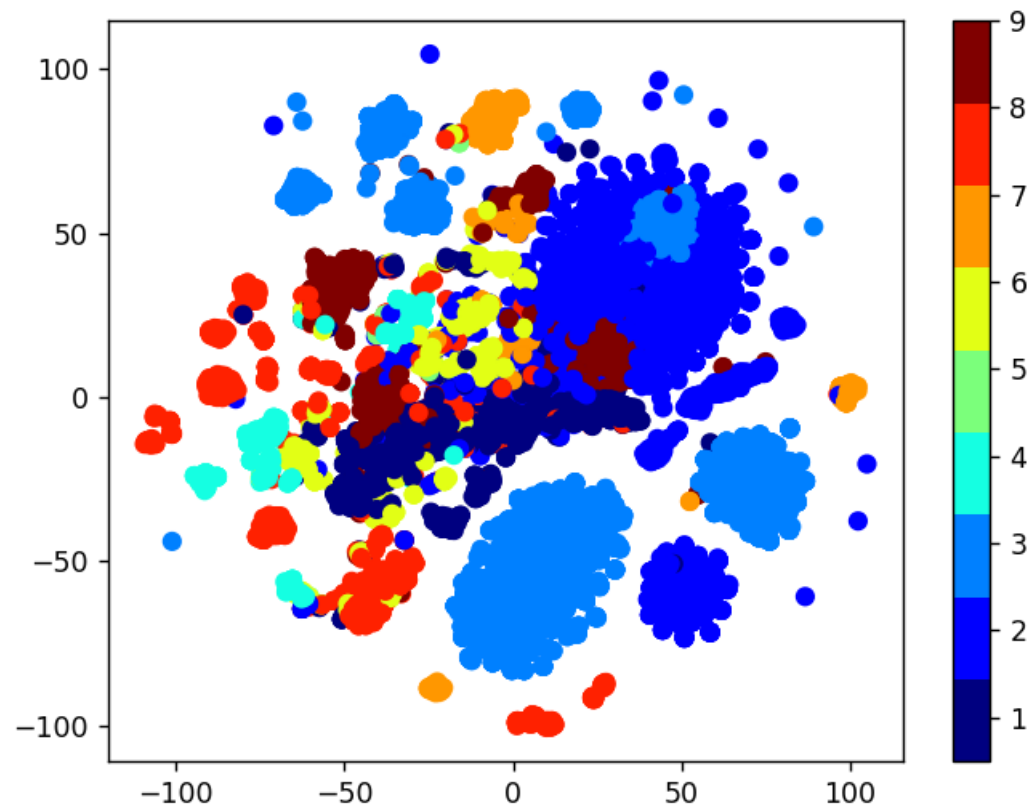
5 rows × 260 columns

3.2.4 Multivariate Analysis


```
In [ ]: #multivariate analysis on byte files
#this is with perplexity 50
xtsne=TSNE(perplexity=50)
results=xtsne.fit_transform(result.drop(['ID','Class'], axis=1))
vis_x = results[:, 0]
vis_y = results[:, 1]
plt.scatter(vis_x, vis_y, c=data_y, cmap=plt.cm.get_cmap("jet", 9))
plt.colorbar(ticks=range(10))
plt.clim(0.5, 9)
plt.show()
```



```
In [ ]: #this is with perplexity 30
xtsne=TSNE(perplexity=30)
results=xtsne.fit_transform(result.drop(['ID', 'Class'], axis=1))
vis_x = results[:, 0]
vis_y = results[:, 1]
plt.scatter(vis_x, vis_y, c=data_y, cmap=plt.cm.get_cmap("jet", 9))
plt.colorbar(ticks=range(10))
plt.clim(0.5, 9)
plt.show()
```



Train Test split

```
In [ ]: data_y = result['Class']
# split the data into test and train by maintaining same distribution of output variable 'y_true' [stratify=y_true]
X_train, X_test, y_train, y_test = train_test_split(result.drop(['ID', 'Class'], axis=1), data_y, stratify=data_y, test_size=0.20)
# split the train data into train and cross validation by maintaining same distribution of output variable 'y_train' [stratify=y_train]
X_train, X_cv, y_train, y_cv = train_test_split(X_train, y_train, stratify=y_train, test_size=0.20)
```

```
In [ ]: print('Number of data points in train data:', X_train.shape[0])
print('Number of data points in test data:', X_test.shape[0])
print('Number of data points in cross validation data:', X_cv.shape[0])
```

Number of data points in train data: 6955

Number of data points in test data: 2174

Number of data points in cross validation data: 1739

```

In [ ]: # it returns a dict, keys as class labels and values as the number of data points in that class
train_class_distribution = y_train.value_counts().sortlevel()
test_class_distribution = y_test.value_counts().sortlevel()
cv_class_distribution = y_cv.value_counts().sortlevel()

my_colors = 'rgbkymc'
train_class_distribution.plot(kind='bar', color=my_colors)
plt.xlabel('Class')
plt.ylabel('Data points per Class')
plt.title('Distribution of yi in train data')
plt.grid()
plt.show()

# ref: argsort https://docs.scipy.org/doc/numpy/reference/generated/numpy.argsort.html
# -(train_class_distribution.values): the minus sign will give us in decreasing order
sorted_yi = np.argsort(-train_class_distribution.values)
for i in sorted_yi:
    print('Number of data points in class', i+1, ':', train_class_distribution.values[i], '(', np.round((train_class_distribution.values[i]/y_train.shape[0]*100), 3), '%)')

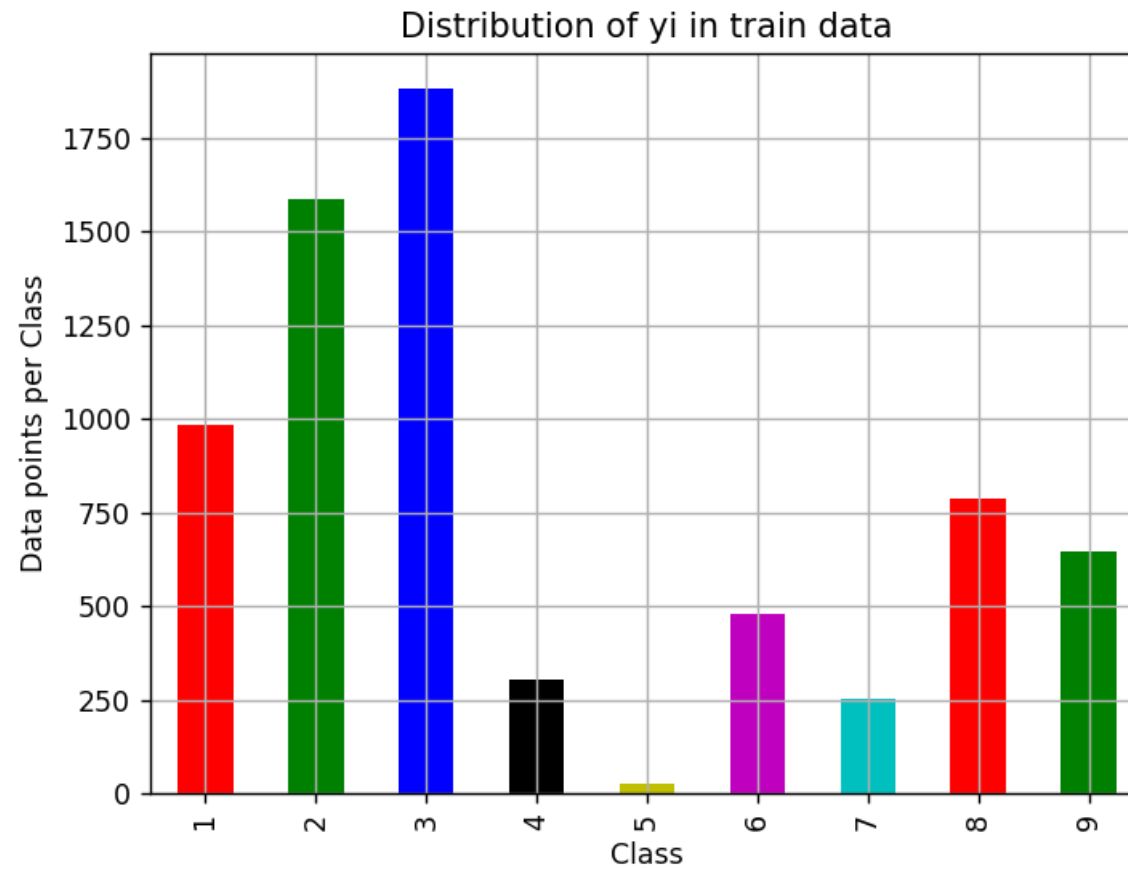
print('-'*80)
my_colors = 'rgbkymc'
test_class_distribution.plot(kind='bar', color=my_colors)
plt.xlabel('Class')
plt.ylabel('Data points per Class')
plt.title('Distribution of yi in test data')
plt.grid()
plt.show()

# ref: argsort https://docs.scipy.org/doc/numpy/reference/generated/numpy.argsort.html
# -(train_class_distribution.values): the minus sign will give us in decreasing order
sorted_yi = np.argsort(-test_class_distribution.values)
for i in sorted_yi:
    print('Number of data points in class', i+1, ':', test_class_distribution.values[i], '(', np.round((test_class_distribution.values[i]/y_test.shape[0]*100), 3), '%)')

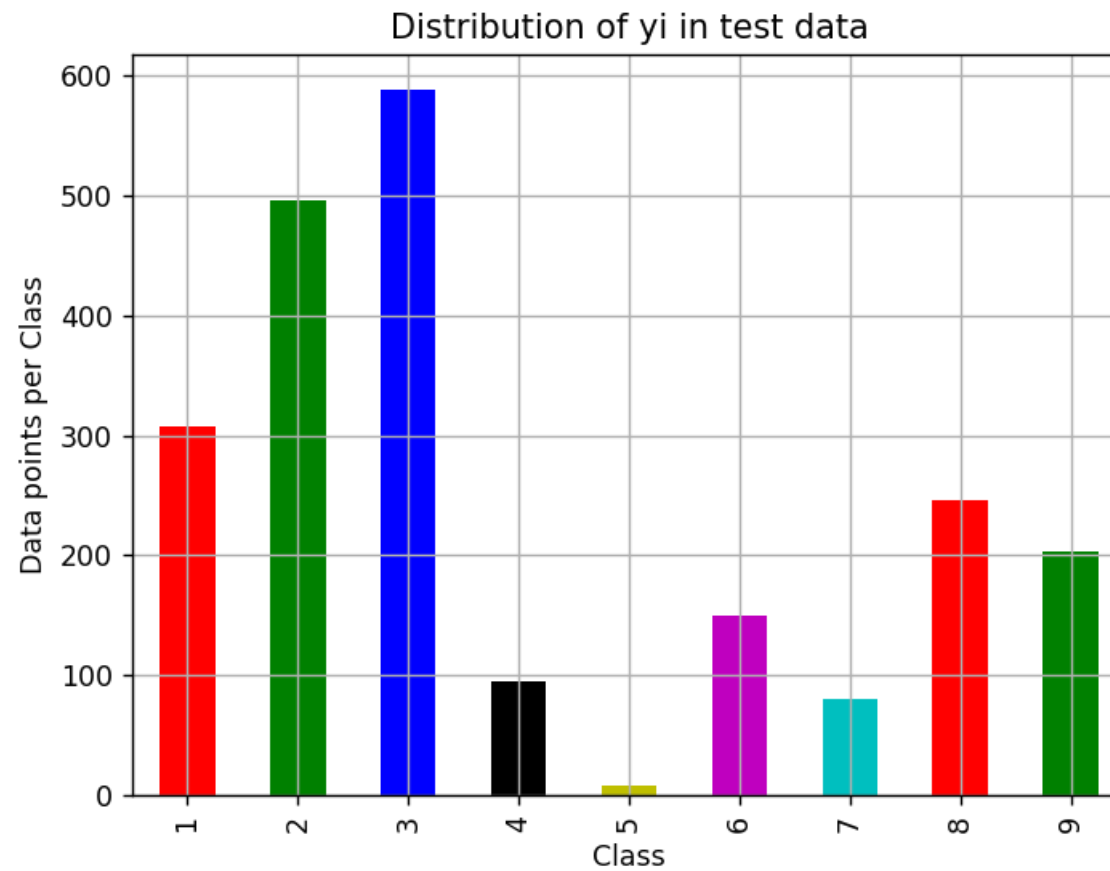
print('-'*80)
my_colors = 'rgbkymc'
cv_class_distribution.plot(kind='bar', color=my_colors)
plt.xlabel('Class')
plt.ylabel('Data points per Class')
plt.title('Distribution of yi in cross validation data')
plt.grid()
plt.show()

```

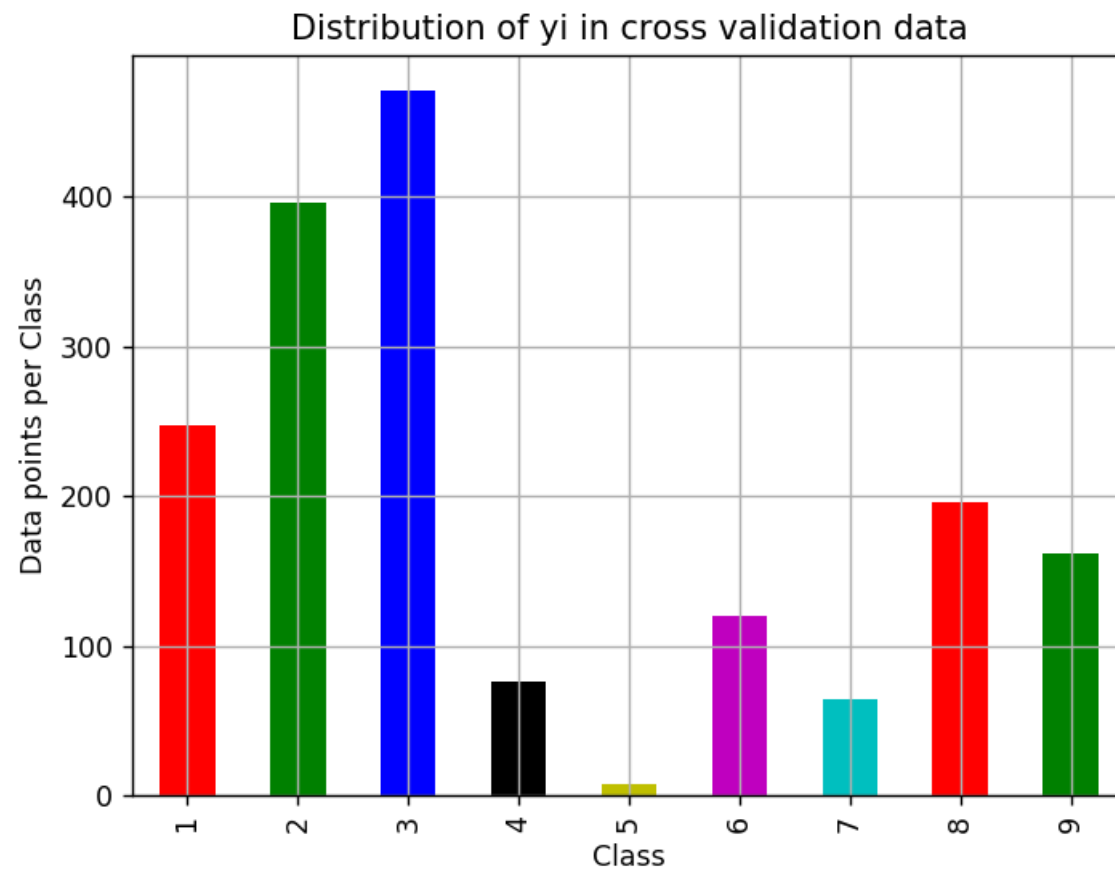
```
# ref: argsort https://docs.scipy.org/doc/numpy/reference/generated/numpy.argsort.html
# -(train_class_distribution.values): the minus sign will give us in decreasing order
sorted_yi = np.argsort(-train_class_distribution.values)
for i in sorted_yi:
    print('Number of data points in class', i+1, ':', cv_class_distribution.values[i], '(', np.round((cv_class_distribution.values[i]/y_cv.shape[0]*100), 3), '%')
```



Number of data points in class 3 : 1883 (27.074 %)
Number of data points in class 2 : 1586 (22.804 %)
Number of data points in class 1 : 986 (14.177 %)
Number of data points in class 8 : 786 (11.301 %)
Number of data points in class 9 : 648 (9.317 %)
Number of data points in class 6 : 481 (6.916 %)
Number of data points in class 4 : 304 (4.371 %)
Number of data points in class 7 : 254 (3.652 %)
Number of data points in class 5 : 27 (0.388 %)



Number of data points in class 3 : 588 (27.047 %)
Number of data points in class 2 : 496 (22.815 %)
Number of data points in class 1 : 308 (14.167 %)
Number of data points in class 8 : 246 (11.316 %)
Number of data points in class 9 : 203 (9.338 %)
Number of data points in class 6 : 150 (6.9 %)
Number of data points in class 4 : 95 (4.37 %)
Number of data points in class 7 : 80 (3.68 %)
Number of data points in class 5 : 8 (0.368 %)



Number of data points in class 3 : 471 (27.085 %)
Number of data points in class 2 : 396 (22.772 %)
Number of data points in class 1 : 247 (14.204 %)
Number of data points in class 8 : 196 (11.271 %)
Number of data points in class 9 : 162 (9.316 %)
Number of data points in class 6 : 120 (6.901 %)
Number of data points in class 4 : 76 (4.37 %)
Number of data points in class 7 : 64 (3.68 %)
Number of data points in class 5 : 7 (0.403 %)


```

In [ ]: def plot_confusion_matrix(test_y, predict_y):
    C = confusion_matrix(test_y, predict_y)
    print("Number of misclassified points ", (len(test_y)-np.trace(C))/len(test_y)*100)
    # C = 9,9 matrix, each cell (i,j) represents number of points of class i are predicted class j

    A = (((C.T)/(C.sum(axis=1))).T)
    #divid each element of the confusion matrix with the sum of elements in that column

    # C = [[1, 2],
    #      [3, 4]]
    # C.T = [[1, 3],
    #        [2, 4]]
    # C.sum(axis = 1)  axis=0 corresonds to columns and axis=1 corresponds to rows in two dimensional array
    # C.sum(axix =1) = [[3, 7]]
    # ((C.T)/(C.sum(axis=1))) = [[1/3, 3/7]
    #                             [2/3, 4/7]]

    # ((C.T)/(C.sum(axis=1))).T = [[1/3, 2/3]
    #                               [3/7, 4/7]]
    # sum of row elements = 1

    B = (C/C.sum(axis=0))
    #divid each element of the confusion matrix with the sum of elements in that row
    # C = [[1, 2],
    #      [3, 4]]
    # C.sum(axis = 0)  axis=0 corresonds to columns and axis=1 corresponds to rows in two dimensional array
    # C.sum(axix =0) = [[4, 6]]
    # (C/C.sum(axis=0)) = [[1/4, 2/6],
    #                       [3/4, 4/6]]

    labels = [1,2,3,4,5,6,7,8,9]
    cmap=sns.light_palette("green")
    # representing A in heatmap format
    print("-"*50, "Confusion matrix", "-"*50)
    plt.figure(figsize=(10,5))
    sns.heatmap(C, annot=True, cmap=cmap, fmt=".3f", xticklabels=labels, yticklabels=labels)
    plt.xlabel('Predicted Class')
    plt.ylabel('Original Class')
    plt.show()

    print("-"*50, "Precision matrix", "-"*50)
    plt.figure(figsize=(10,5))
    sns.heatmap(B, annot=True, cmap=cmap, fmt=".3f", xticklabels=labels, yticklabels=labels)
    plt.xlabel('Predicted Class')
    plt.ylabel('Original Class')

```

```
plt.show()
print("Sum of columns in precision matrix",B.sum(axis=0))

# representing B in heatmap format
print("-"*50, "Recall matrix"      , "-"*50)
plt.figure(figsize=(10,5))
sns.heatmap(A, annot=True, cmap=cmap, fmt=".3f", xticklabels=labels, yticklabels=labels)
plt.xlabel('Predicted Class')
plt.ylabel('Original Class')
plt.show()
print("Sum of rows in precision matrix",A.sum(axis=1))
```

4. Machine Learning Models

4.1. Machine Leaning Models on bytes files

4.1.1. Random Model

```
In [ ]: # we need to generate 9 numbers and the sum of numbers should be 1
# one solution is to generate 9 numbers and divide each of the numbers by their sum
# ref: https://stackoverflow.com/a/18662466/4084039

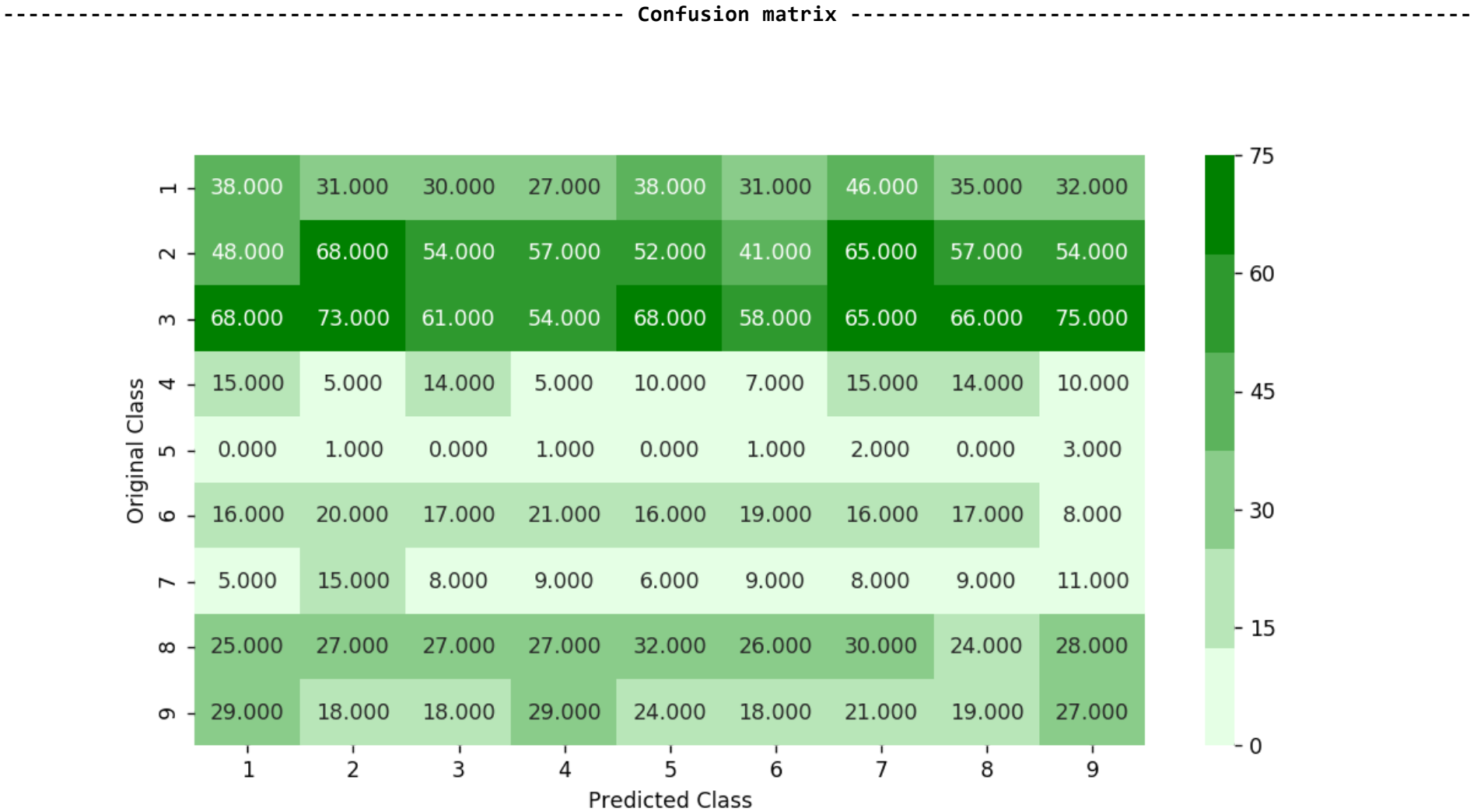
test_data_len = X_test.shape[0]
cv_data_len = X_cv.shape[0]

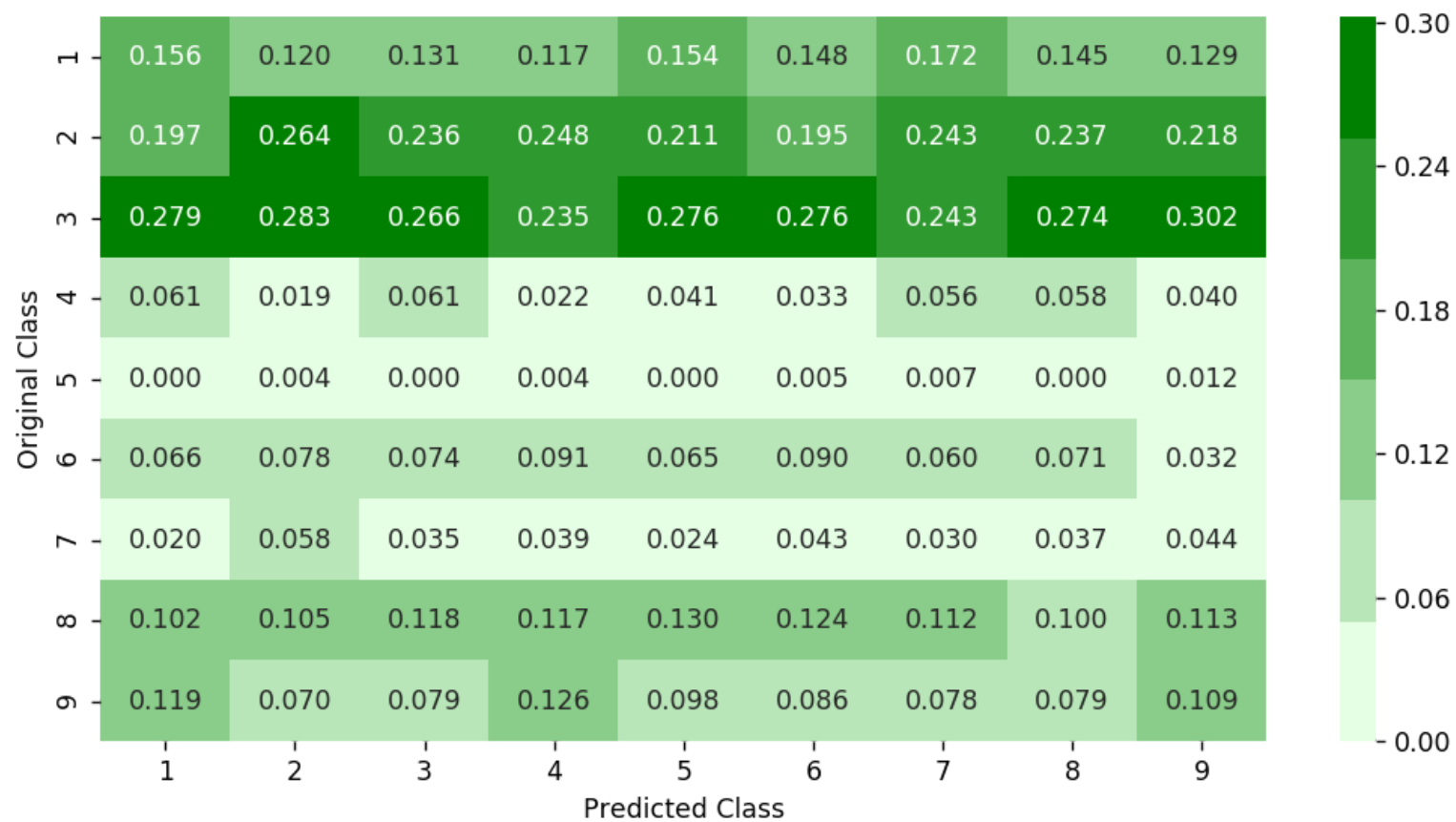
# we create a output array that has exactly same size as the CV data
cv_predicted_y = np.zeros((cv_data_len,9))
for i in range(cv_data_len):
    rand_probs = np.random.rand(1,9)
    cv_predicted_y[i] = ((rand_probs/sum(sum(rand_probs))))[0])
print("Log loss on Cross Validation Data using Random Model",log_loss(y_cv,cv_predicted_y, eps=1e-15))

# Test-Set error.
#we create a output array that has exactly same as the test data
test_predicted_y = np.zeros((test_data_len,9))
for i in range(test_data_len):
    rand_probs = np.random.rand(1,9)
    test_predicted_y[i] = ((rand_probs/sum(sum(rand_probs))))[0])
print("Log loss on Test Data using Random Model",log_loss(y_test,test_predicted_y, eps=1e-15))

predicted_y =np.argmax(test_predicted_y, axis=1)
plot_confusion_matrix(y_test, predicted_y+1)
```

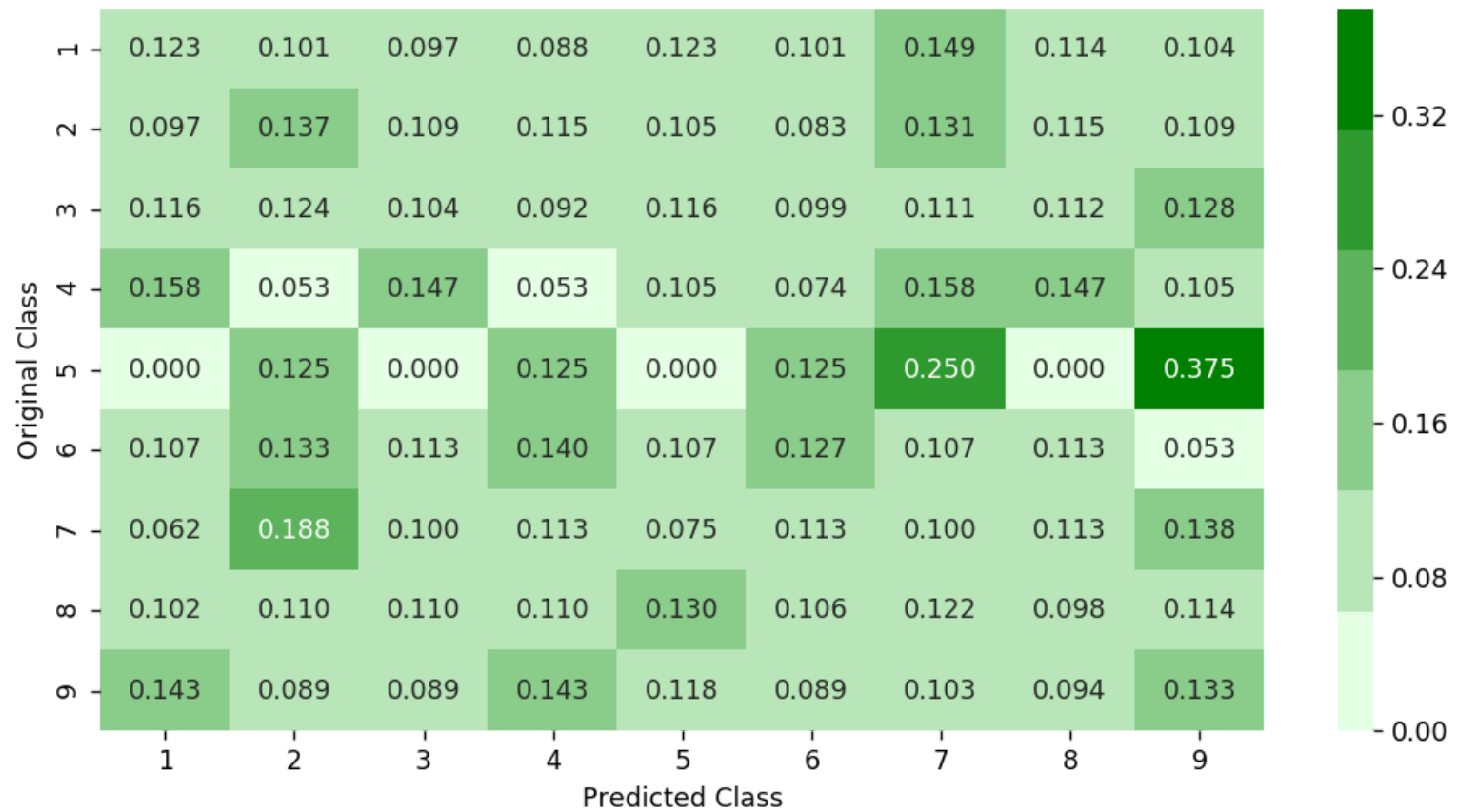
Log loss on Cross Validation Data using Random Model 2.45615644965
Log loss on Test Data using Random Model 2.48503905509
Number of misclassified points 88.5004599816





Sum of columns in precision matrix [1. 1. 1. 1. 1. 1. 1. 1. 1.]

----- Recall matrix -----



Sum of rows in precision matrix [1. 1. 1. 1. 1. 1. 1. 1. 1.]

4.1.2. K Nearest Neighbour Classification

```

In [ ]: # find more about KNeighborsClassifier() here http://scikit-learn.org/stable/modules/generated/sklearn.neighbors.KNeighborsClassifier.html
# -----
# default parameter
# KNeighborsClassifier(n_neighbors=5, weights='uniform', algorithm='auto', leaf_size=30, p=2,
# metric='minkowski', metric_params=None, n_jobs=1, **kwargs)

# methods of
# fit(X, y) : Fit the model using X as training data and y as target values
# predict(X):Predict the class labels for the provided data
# predict_proba(X):Return probability estimates for the test data X.
#-----
# video link: https://www.appliedaicourse.com/course/applied-ai-course-online/lessons/k-nearest-neighbors-geometric-intuition-with-a-toy-example-1/
#-----

# find more about CalibratedClassifierCV here at http://scikit-learn.org/stable/modules/generated/sklearn.calibration.CalibratedClassifierCV.html
# -----
# default paramters
# sklearn.calibration.CalibratedClassifierCV(base_estimator=None, method='sigmoid', cv=3)
#
# some of the methods of CalibratedClassifierCV()
# fit(X, y[, sample_weight]) Fit the calibrated model
# get_params([deep]) Get parameters for this estimator.
# predict(X) Predict the target of new samples.
# predict_proba(X) Posterior probabilities of classification
#-----
# video link:
#-----

alpha = [x for x in range(1, 15, 2)]
cv_log_error_array=[]
for i in alpha:
    k_cfl=KNeighborsClassifier(n_neighbors=i)
    k_cfl.fit(X_train,y_train)
    sig_clf = CalibratedClassifierCV(k_cfl, method="sigmoid")
    sig_clf.fit(X_train, y_train)
    predict_y = sig_clf.predict_proba(X_cv)
    cv_log_error_array.append(log_loss(y_cv, predict_y, labels=k_cfl.classes_, eps=1e-15))

for i in range(len(cv_log_error_array)):
    print ('log_loss for k = ',alpha[i],'is',cv_log_error_array[i])

```

```

best_alpha = np.argmin(cv_log_error_array)

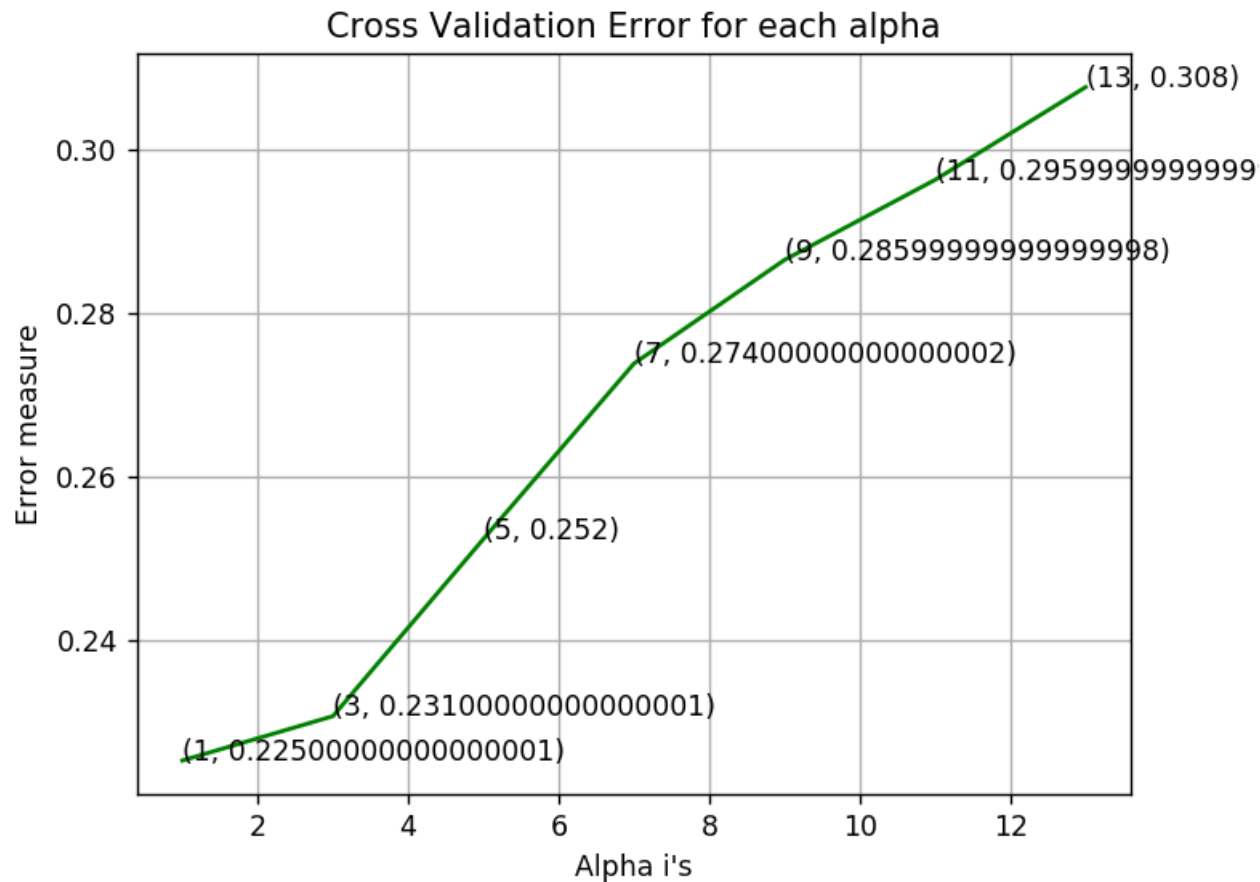
fig, ax = plt.subplots()
ax.plot(alpha, cv_log_error_array, c='g')
for i, txt in enumerate(np.round(cv_log_error_array, 3)):
    ax.annotate((alpha[i], np.round(txt, 3)), (alpha[i], cv_log_error_array[i]))
plt.grid()
plt.title("Cross Validation Error for each alpha")
plt.xlabel("Alpha i's")
plt.ylabel("Error measure")
plt.show()

k_cfl = KNeighborsClassifier(n_neighbors=alpha[best_alpha])
k_cfl.fit(X_train, y_train)
sig_clf = CalibratedClassifierCV(k_cfl, method="sigmoid")
sig_clf.fit(X_train, y_train)

predict_y = sig_clf.predict_proba(X_train)
print('For values of best alpha = ', alpha[best_alpha], "The train log loss is:", log_loss(y_train, predict_y))
predict_y = sig_clf.predict_proba(X_cv)
print('For values of best alpha = ', alpha[best_alpha], "The cross validation log loss is:", log_loss(y_cv, predict_y))
predict_y = sig_clf.predict_proba(X_test)
print('For values of best alpha = ', alpha[best_alpha], "The test log loss is:", log_loss(y_test, predict_y))
plot_confusion_matrix(y_test, sig_clf.predict(X_test))

```


log_loss for k = 1 is 0.225386237304
log_loss for k = 3 is 0.230795229168
log_loss for k = 5 is 0.252421408646
log_loss for k = 7 is 0.273827486888
log_loss for k = 9 is 0.286469181555
log_loss for k = 11 is 0.29623391147
log_loss for k = 13 is 0.307551203154

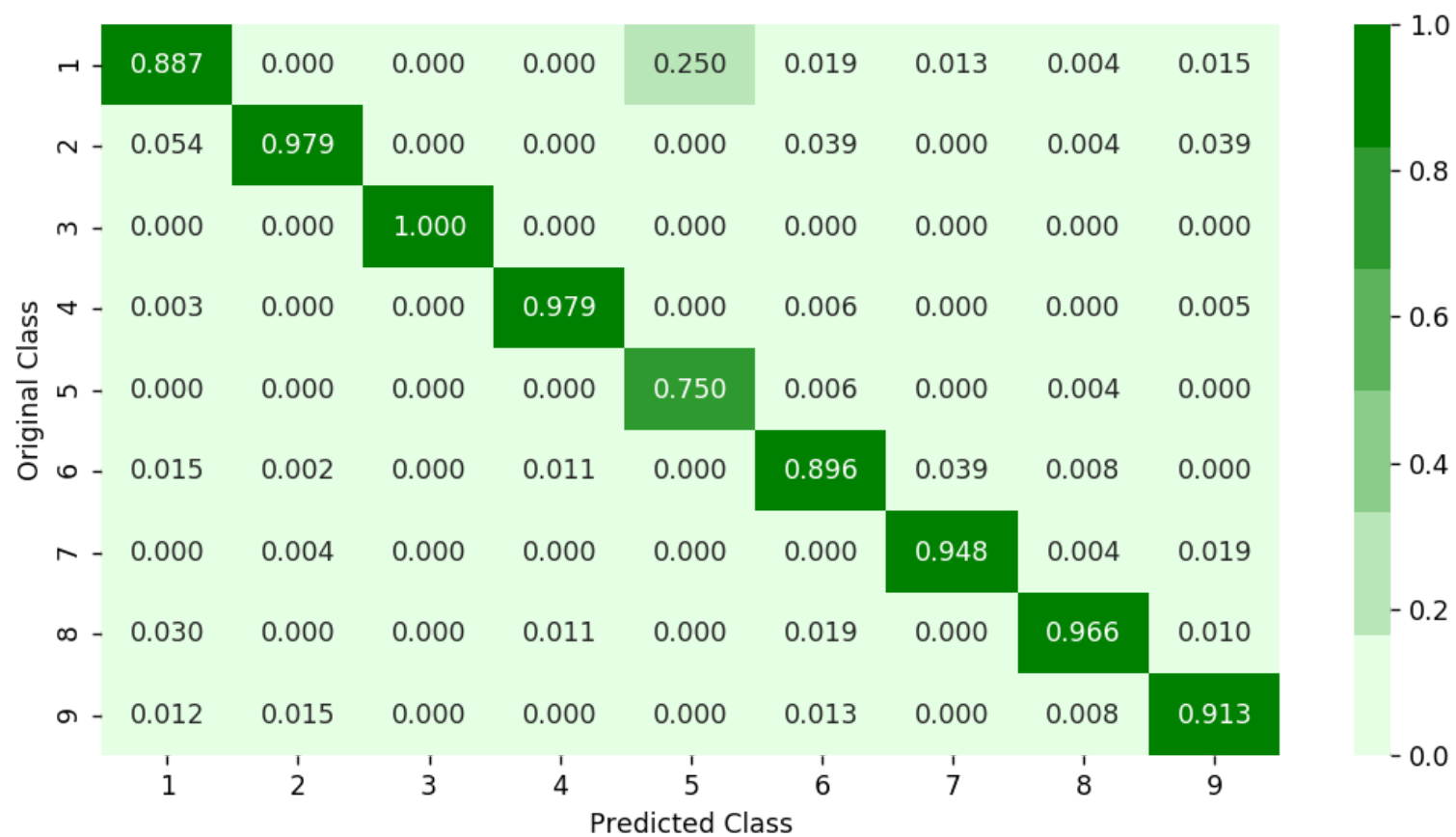


For values of best alpha = 1 The train log loss is: 0.0782947669247
For values of best alpha = 1 The cross validation log loss is: 0.225386237304
For values of best alpha = 1 The test log loss is: 0.241508604195
Number of misclassified points 4.50781968721

----- Confusion matrix -----

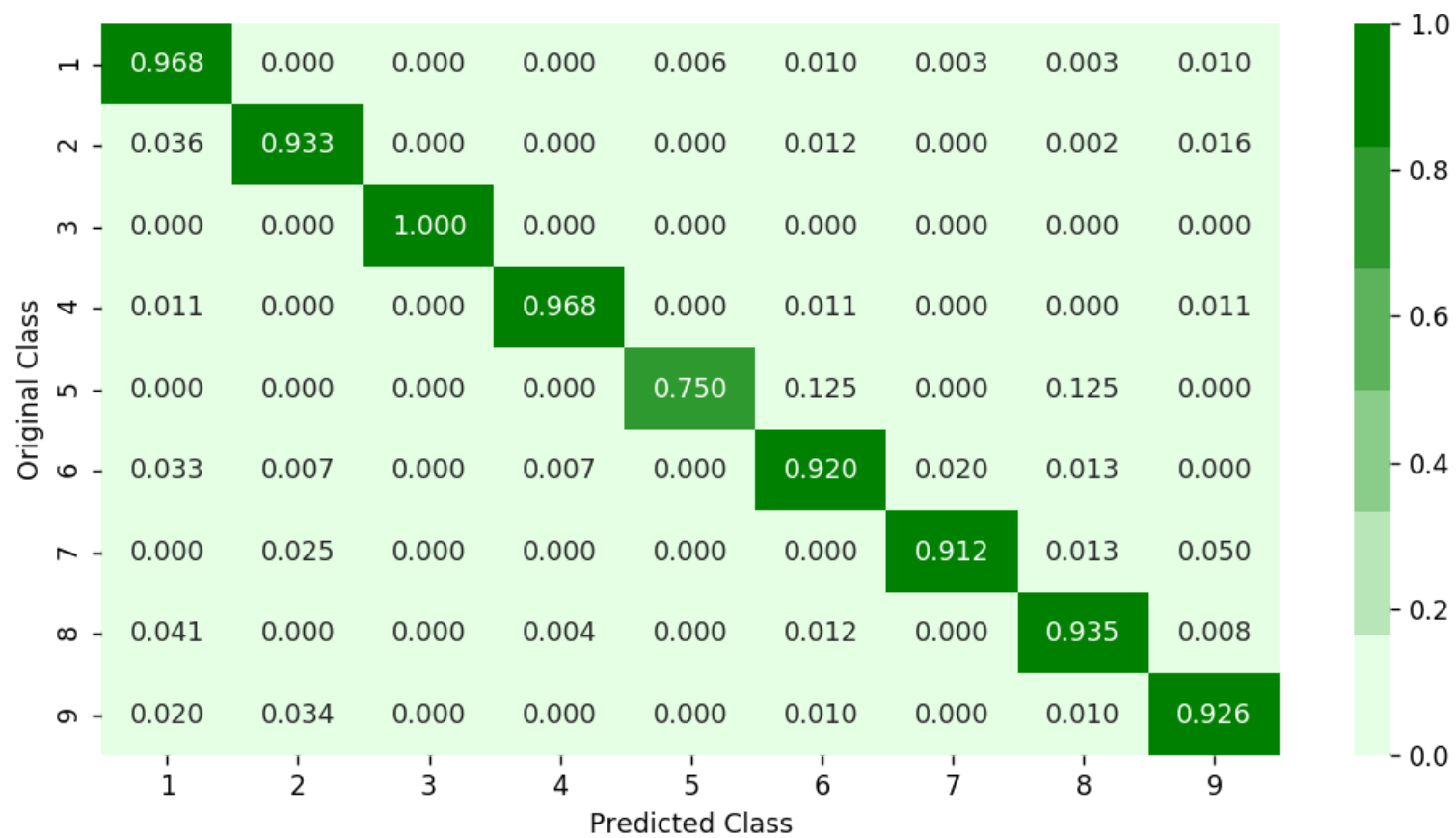


Precision matrix



Sum of columns in precision matrix [1. 1. 1. 1. 1. 1. 1. 1. 1.]

----- Recall matrix -----



Sum of rows in precision matrix [1. 1. 1. 1. 1. 1. 1. 1. 1.]

4.1.3. Logistic Regression

```

In [ ]: # read more about SGDClassifier() at http://scikit-learn.org/stable/modules/generated/sklearn.linear_model.SGDClassifier.html
# -----
# default parameters
# SGDClassifier(loss='hinge', penalty='l2', alpha=0.0001, l1_ratio=0.15, fit_intercept=True, max_iter=None, tol=None,
# shuffle=True, verbose=0, epsilon=0.1, n_jobs=1, random_state=None, learning_rate='optimal', eta0=0.0, power_t=0.5,
# class_weight=None, warm_start=False, average=False, n_iter=None)

# some of methods
# fit(X, y[, coef_init, intercept_init, ...])      Fit linear model with Stochastic Gradient Descent.
# predict(X)      Predict class labels for samples in X.

#-----
# video link: https://www.appliedaicourse.com/course/applied-ai-course-online/lessons/geometric-intuition-1/
#-----

alpha = [10 ** x for x in range(-5, 4)]
cv_log_error_array=[]
for i in alpha:
    logisticR=LogisticRegression(penalty='l2',C=i,class_weight='balanced')
    logisticR.fit(X_train,y_train)
    sig_clf = CalibratedClassifierCV(logisticR, method="sigmoid")
    sig_clf.fit(X_train, y_train)
    predict_y = sig_clf.predict_proba(X_cv)
    cv_log_error_array.append(log_loss(y_cv, predict_y, labels=logisticR.classes_, eps=1e-15))

for i in range(len(cv_log_error_array)):
    print ('log_loss for c = ',alpha[i],'is',cv_log_error_array[i])

best_alpha = np.argmin(cv_log_error_array)

fig, ax = plt.subplots()
ax.plot(alpha, cv_log_error_array,c='g')
for i, txt in enumerate(np.round(cv_log_error_array,3)):
    ax.annotate((alpha[i],np.round(txt,3)), (alpha[i],cv_log_error_array[i]))
plt.grid()
plt.title("Cross Validation Error for each alpha")
plt.xlabel("Alpha i's")
plt.ylabel("Error measure")
plt.show()

logisticR=LogisticRegression(penalty='l2',C=alpha[best_alpha],class_weight='balanced')
logisticR.fit(X_train,y_train)
sig_clf = CalibratedClassifierCV(logisticR, method="sigmoid")
sig_clf.fit(X_train, y_train)

```

```
pred_y=sig_clf.predict(X_test)
```

```
predict_y = sig_clf.predict_proba(X_train)
```

```
print ('log loss for train data',log_loss(y_train, predict_y, labels=logisticR.classes_, eps=1e-15))
```

```
predict_y = sig_clf.predict_proba(X_cv)
```

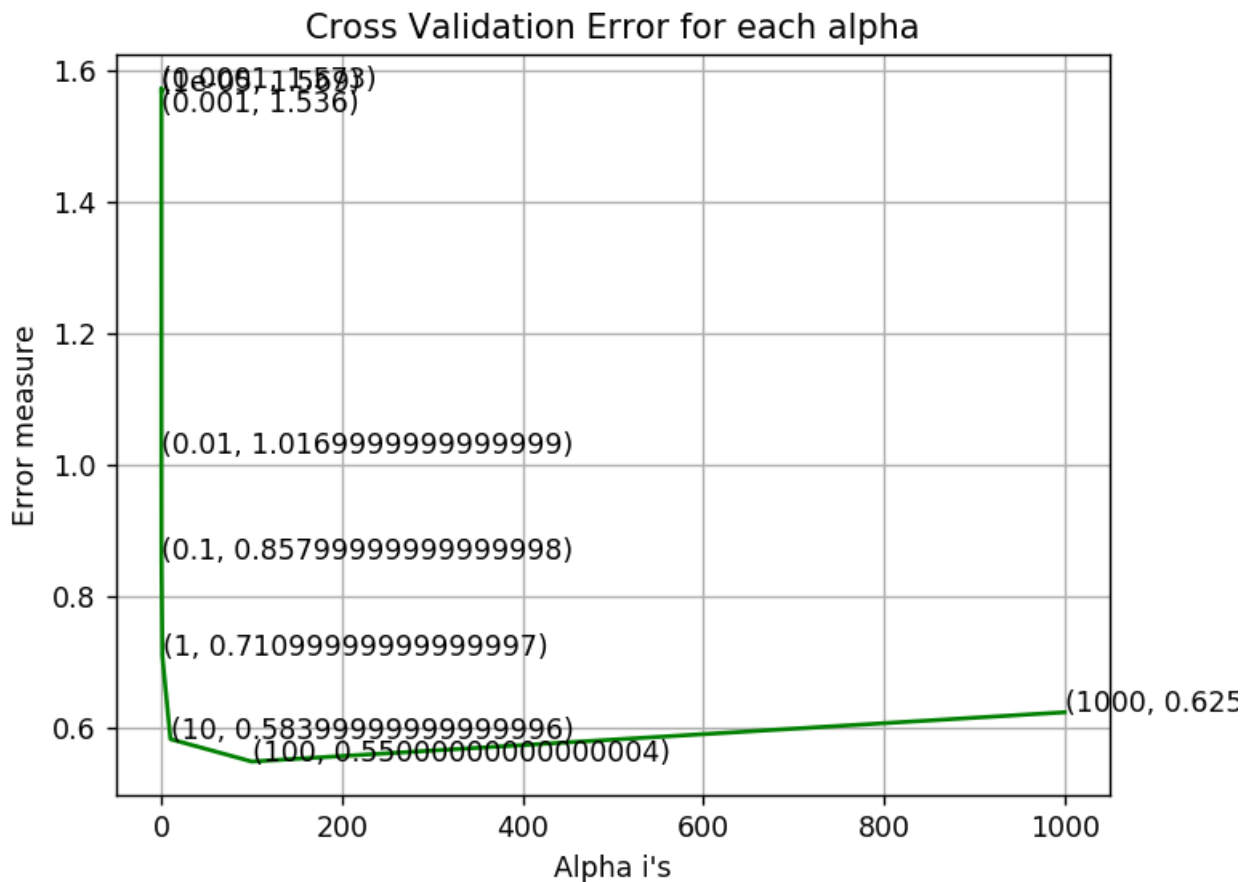
```
print ('log loss for cv data',log_loss(y_cv, predict_y, labels=logisticR.classes_, eps=1e-15))
```

```
predict_y = sig_clf.predict_proba(X_test)
```

```
print ('log loss for test data',log_loss(y_test, predict_y, labels=logisticR.classes_, eps=1e-15))
```

```
plot_confusion_matrix(y_test, sig_clf.predict(X_test))
```

log_loss for c = 1e-05 is 1.56916911178
log_loss for c = 0.0001 is 1.57336384417
log_loss for c = 0.001 is 1.53598598273
log_loss for c = 0.01 is 1.01720972418
log_loss for c = 0.1 is 0.857766083873
log_loss for c = 1 is 0.711154393309
log_loss for c = 10 is 0.583929522635
log_loss for c = 100 is 0.549929846589
log_loss for c = 1000 is 0.624746769121

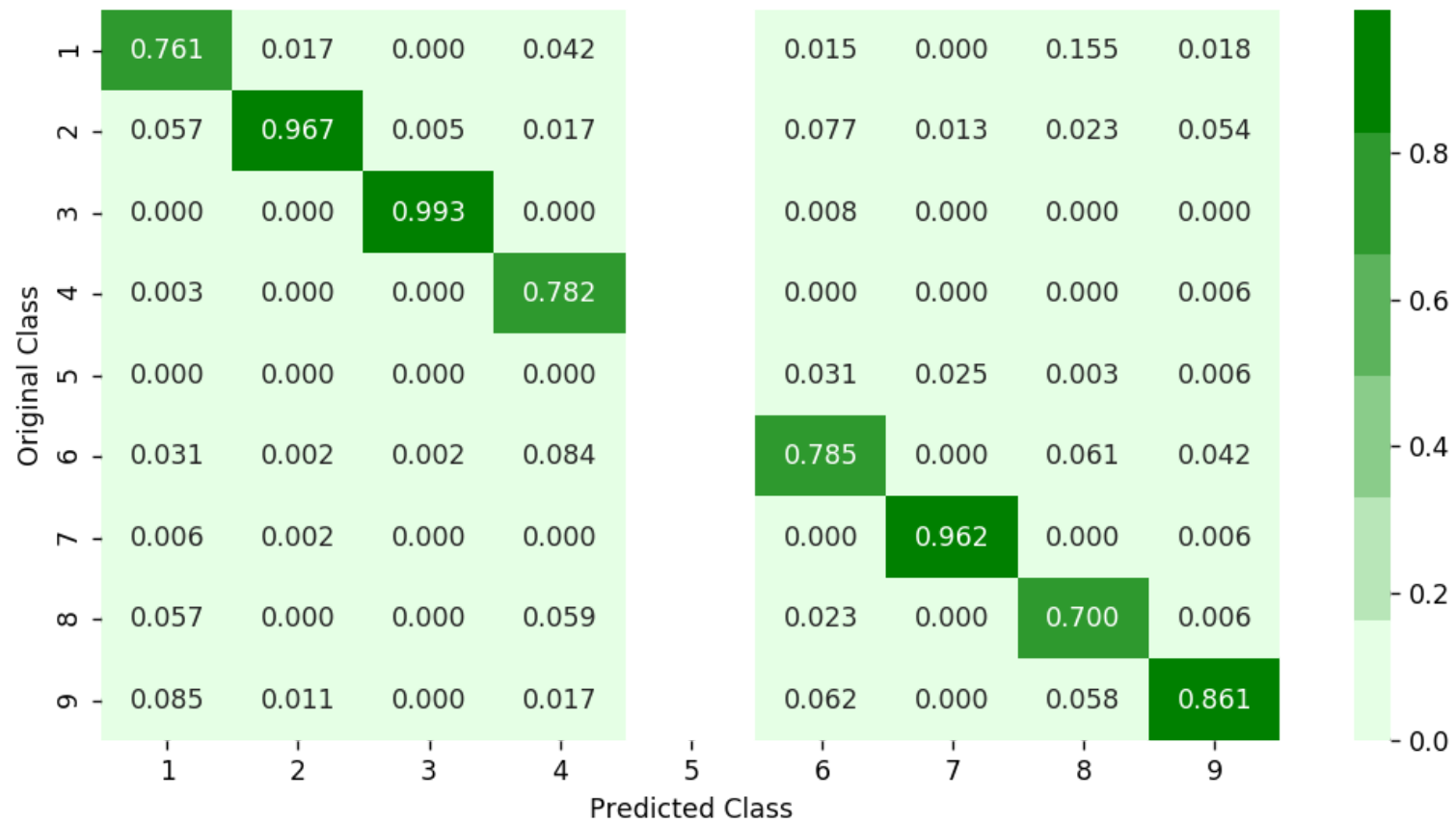


log loss for train data 0.498923428696
log loss for cv data 0.549929846589
log loss for test data 0.528347316704
Number of misclassified points 12.3275068997

----- Confusion matrix -----

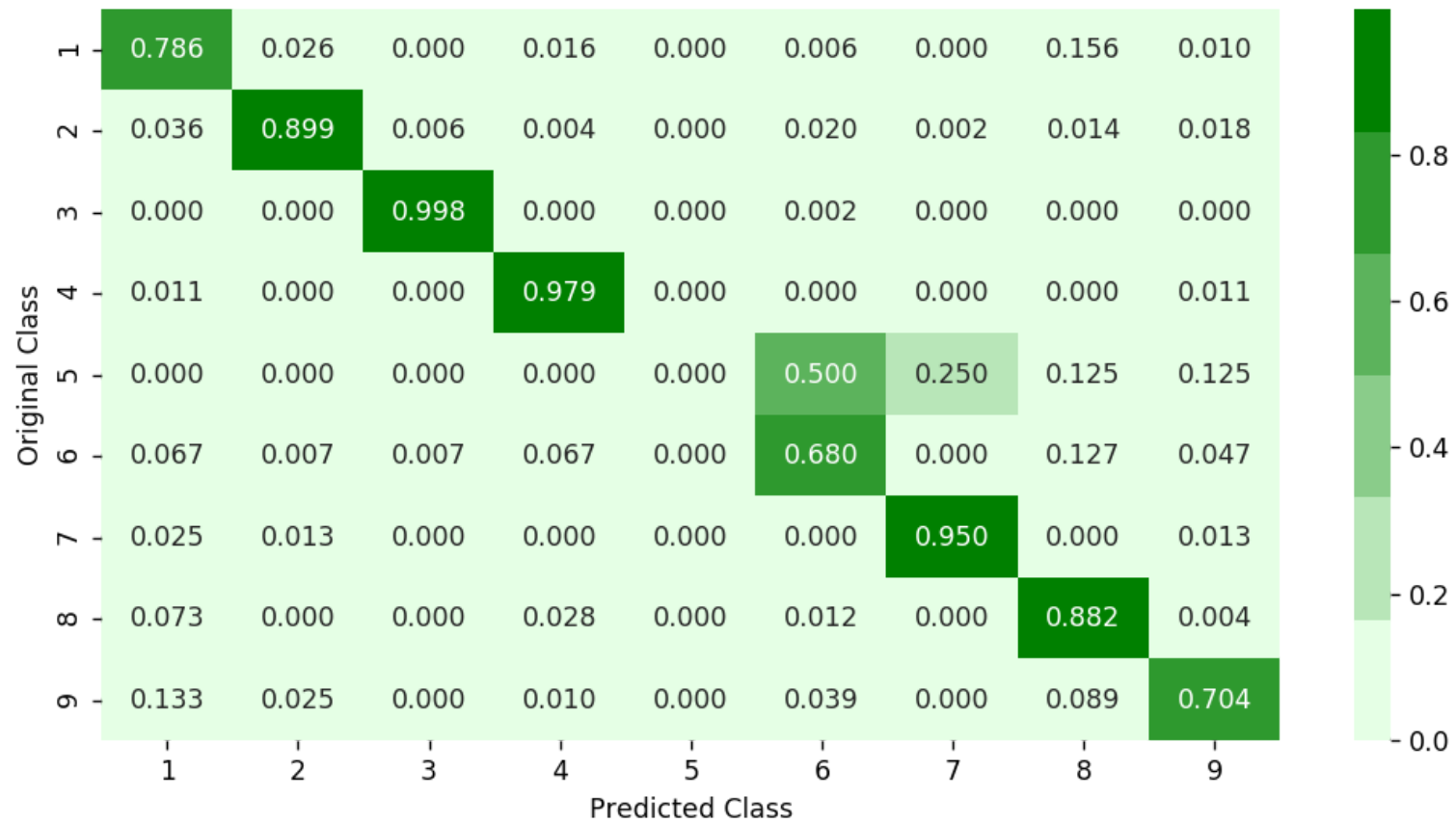


----- Precision matrix -----



Sum of columns in precision matrix [1. 1. 1. 1. nan 1. 1. 1. 1.]

----- Recall matrix -----



Sum of rows in precision matrix [1. 1. 1. 1. 1. 1. 1. 1. 1.]

4.1.4. Random Forest Classifier

```

In [ ]: # -----
# default parameters
# sklearn.ensemble.RandomForestClassifier(n_estimators=10, criterion='gini', max_depth=None, min_samples_split=2,
# min_samples_leaf=1, min_weight_fraction_leaf=0.0, max_features='auto', max_leaf_nodes=None, min_impurity_decrease=0.0,
# min_impurity_split=None, bootstrap=True, oob_score=False, n_jobs=1, random_state=None, verbose=0, warm_start=False,
# class_weight=None)

# Some of methods of RandomForestClassifier()
# fit(X, y, [sample_weight])    Fit the SVM model according to the given training data.
# predict(X)    Perform classification on samples in X.
# predict_proba (X)    Perform classification on samples in X.

# some of attributes of RandomForestClassifier()
# feature_importances_ : array of shape = [n_features]
# The feature importances (the higher, the more important the feature).

# -----
# video link: https://www.appliedaicourse.com/course/applied-ai-course-online/lessons/random-forest-and-their-construction-2/
# -----

alpha=[10,50,100,500,1000,2000,3000]
cv_log_error_array=[]
train_log_error_array=[]
from sklearn.ensemble import RandomForestClassifier
for i in alpha:
    r_cfl=RandomForestClassifier(n_estimators=i,random_state=42,n_jobs=-1)
    r_cfl.fit(X_train,y_train)
    sig_clf = CalibratedClassifierCV(r_cfl, method="sigmoid")
    sig_clf.fit(X_train, y_train)
    predict_y = sig_clf.predict_proba(X_cv)
    cv_log_error_array.append(log_loss(y_cv, predict_y, labels=r_cfl.classes_, eps=1e-15))

for i in range(len(cv_log_error_array)):
    print ('log_loss for c = ',alpha[i],'is',cv_log_error_array[i])

best_alpha = np.argmin(cv_log_error_array)

fig, ax = plt.subplots()
ax.plot(alpha, cv_log_error_array,c='g')
for i, txt in enumerate(np.round(cv_log_error_array,3)):
    ax.annotate((alpha[i],np.round(txt,3)), (alpha[i],cv_log_error_array[i]))
plt.grid()
plt.title("Cross Validation Error for each alpha")

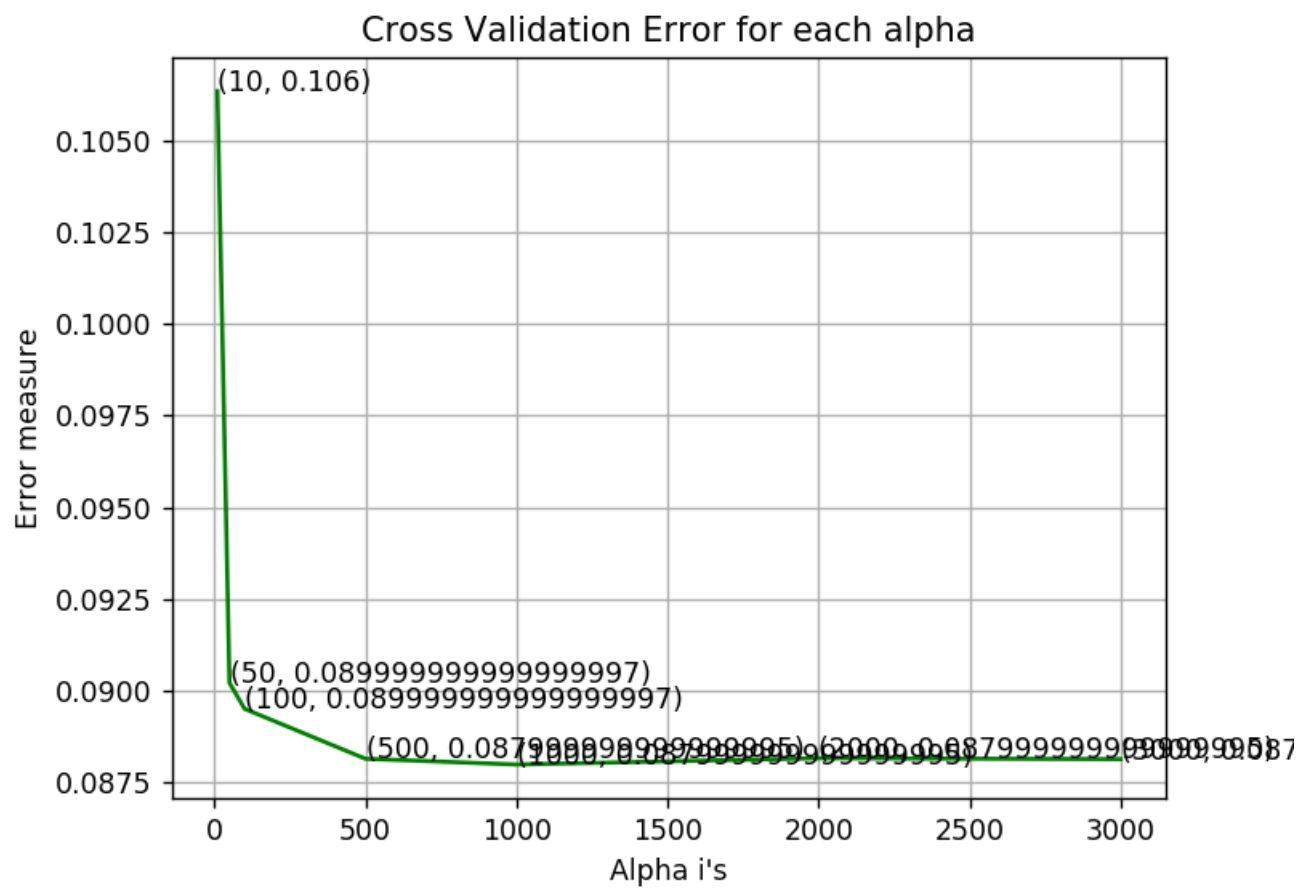
```

```
plt.xlabel("Alpha i's")
plt.ylabel("Error measure")
plt.show()

r_cfl=RandomForestClassifier(n_estimators=alpha[best_alpha],random_state=42,n_jobs=-1)
r_cfl.fit(X_train,y_train)
sig_clf = CalibratedClassifierCV(r_cfl, method="sigmoid")
sig_clf.fit(X_train, y_train)

predict_y = sig_clf.predict_proba(X_train)
print('For values of best alpha = ', alpha[best_alpha], "The train log loss is:",log_loss(y_train, predict_y))
predict_y = sig_clf.predict_proba(X_cv)
print('For values of best alpha = ', alpha[best_alpha], "The cross validation log loss is:",log_loss(y_cv, predict_y))
predict_y = sig_clf.predict_proba(X_test)
print('For values of best alpha = ', alpha[best_alpha], "The test log loss is:",log_loss(y_test, predict_y))
plot_confusion_matrix(y_test, sig_clf.predict(X_test))
```

log_loss for c = 10 is 0.106357709164
log_loss for c = 50 is 0.0902124124145
log_loss for c = 100 is 0.0895043339776
log_loss for c = 500 is 0.0881420869288
log_loss for c = 1000 is 0.0879849524621
log_loss for c = 2000 is 0.0881566647295
log_loss for c = 3000 is 0.0881318948443

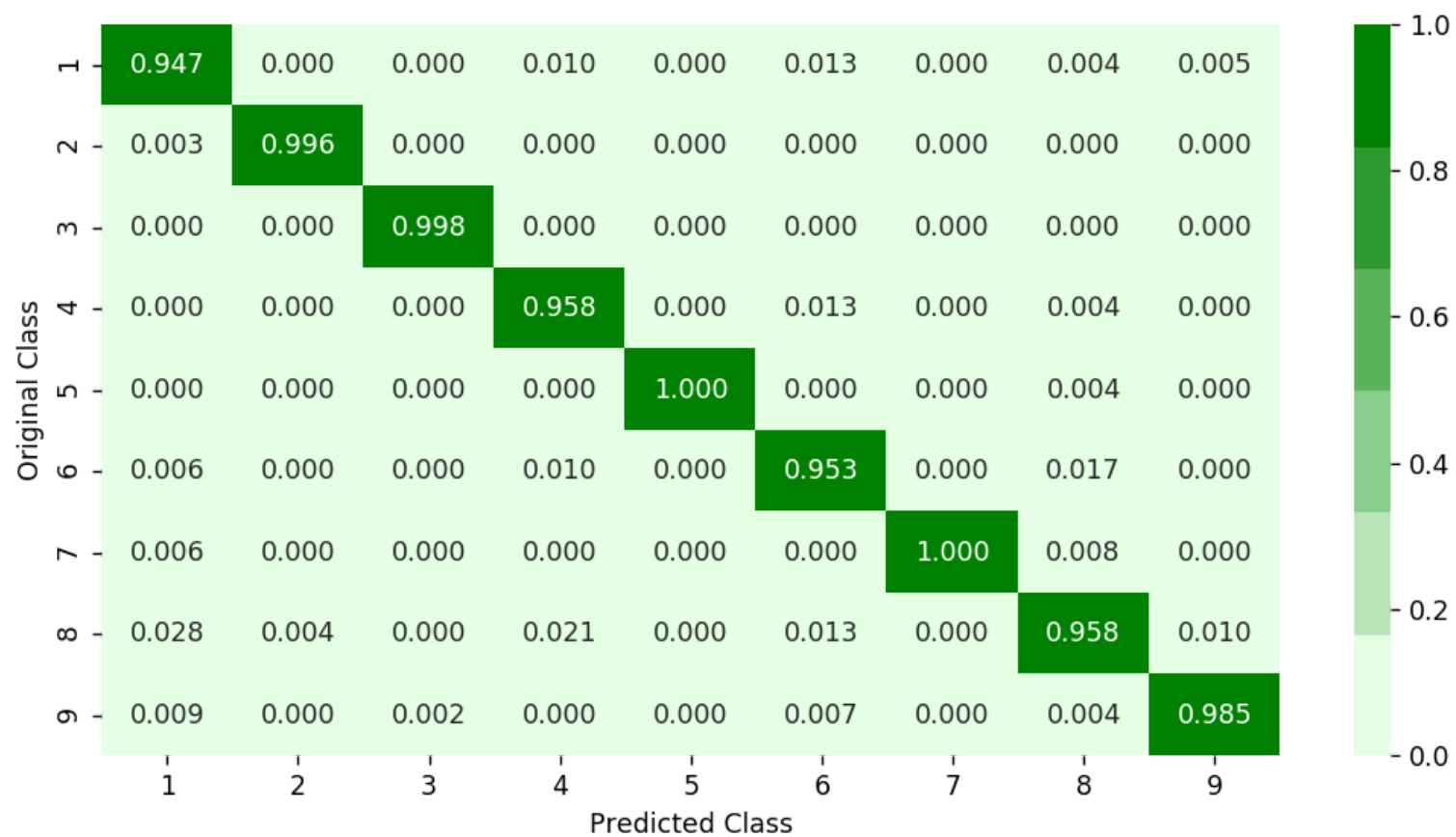


For values of best alpha = 1000 The train log loss is: 0.0266476291801
For values of best alpha = 1000 The cross validation log loss is: 0.0879849524621
For values of best alpha = 1000 The test log loss is: 0.0858346961407
Number of misclassified points 2.02391904324

----- Confusion matrix -----

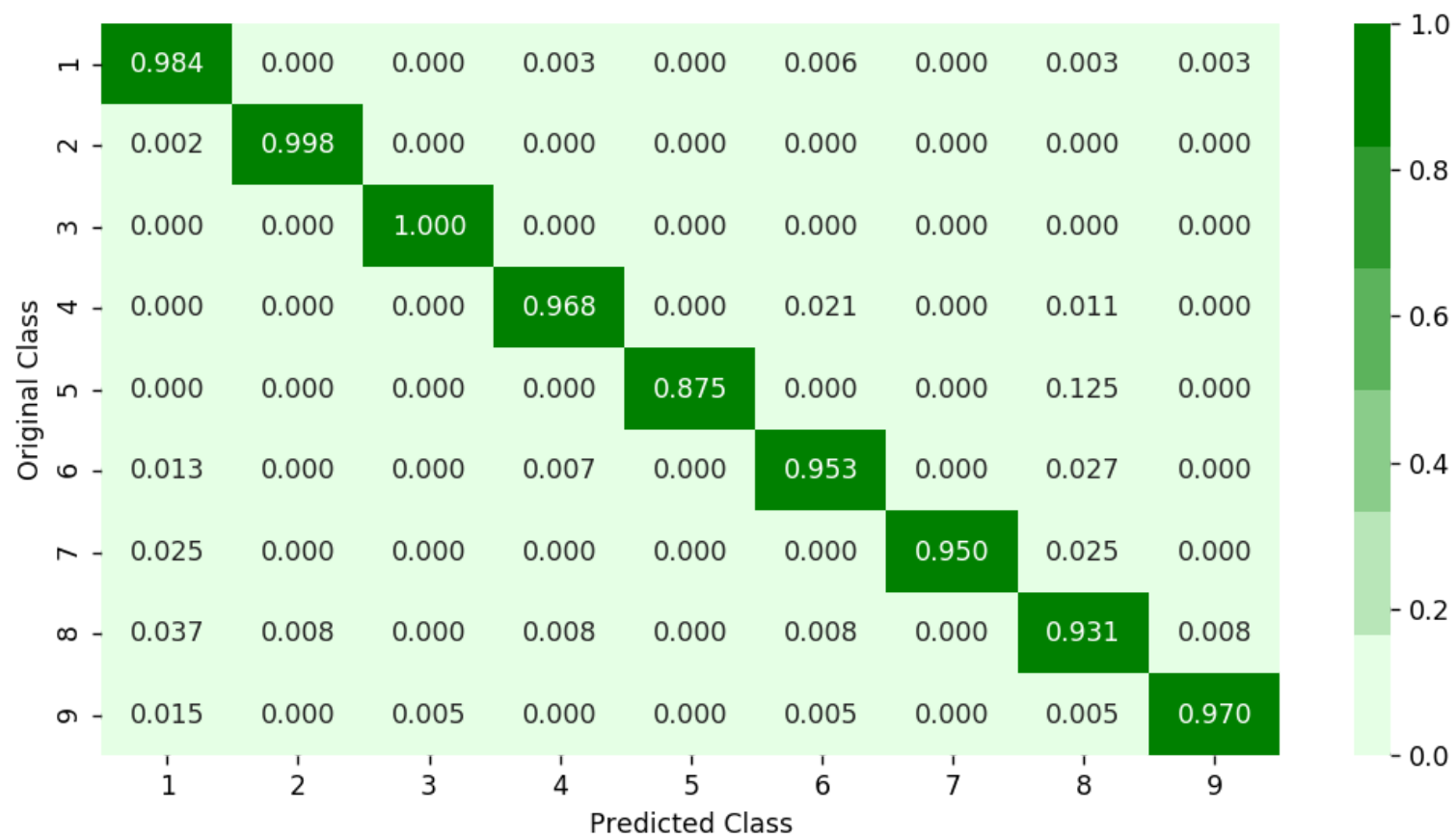


----- Precision matrix -----



Sum of columns in precision matrix [1. 1. 1. 1. 1. 1. 1. 1. 1.]

----- Recall matrix -----



Sum of rows in precision matrix [1. 1. 1. 1. 1. 1. 1. 1. 1.]

4.1.5. XgBoost Classification


```

In [ ]: # Training a hyper-parameter tuned Xg-Boost regressor on our train data

# find more about XGBClassifier function here http://xgboost.readthedocs.io/en/latest/python/python\_api.html?#xgboost.XGBClassifier
# -----
# default paramters
# class xgboost.XGBClassifier(max_depth=3, learning_rate=0.1, n_estimators=100, silent=True,
# objective='binary:logistic', booster='gbtree', n_jobs=1, nthread=None, gamma=0, min_child_weight=1,
# max_delta_step=0, subsample=1, colsample_bytree=1, colsample_bylevel=1, reg_alpha=0, reg_lambda=1,
# scale_pos_weight=1, base_score=0.5, random_state=0, seed=None, missing=None, **kwargs)

# some of methods of RandomForestRegressor()
# fit(X, y, sample_weight=None, eval_set=None, eval_metric=None, early_stopping_rounds=None, verbose=True, xgb_model=None)
# get_params([deep]) Get parameters for this estimator.
# predict(data, output_margin=False, ntree_limit=0) : Predict with data. NOTE: This function is not thread safe.
# get_score(importance_type='weight') -> get the feature importance
# -----
# video link1: https://www.applidaicourse.com/course/applied-ai-course-online/lessons/regression-using-decision-trees-2/
# video link2: https://www.applidaicourse.com/course/applied-ai-course-online/lessons/what-are-ensembles/
# -----

alpha=[10,50,100,500,1000,2000]
cv_log_error_array=[]
for i in alpha:
    x_cfl=XGBClassifier(n_estimators=i,nthread=-1)
    x_cfl.fit(X_train,y_train)
    sig_clf = CalibratedClassifierCV(x_cfl, method="sigmoid")
    sig_clf.fit(X_train, y_train)
    predict_y = sig_clf.predict_proba(X_cv)
    cv_log_error_array.append(log_loss(y_cv, predict_y, labels=x_cfl.classes_, eps=1e-15))

for i in range(len(cv_log_error_array)):
    print ('log_loss for c = ',alpha[i],'is',cv_log_error_array[i])

best_alpha = np.argmin(cv_log_error_array)

fig, ax = plt.subplots()
ax.plot(alpha, cv_log_error_array,c='g')
for i, txt in enumerate(np.round(cv_log_error_array,3)):
    ax.annotate((alpha[i],np.round(txt,3)), (alpha[i],cv_log_error_array[i]))
plt.grid()
plt.title("Cross Validation Error for each alpha")
plt.xlabel("Alpha i's")
plt.ylabel("Error measure")

```

```
plt.show()
```

```
x_cfl=XGBClassifier(n_estimators=alpha[best_alpha],nthread=-1)
```

```
x_cfl.fit(X_train,y_train)
```

```
sig_clf = CalibratedClassifierCV(x_cfl, method="sigmoid")
```

```
sig_clf.fit(X_train, y_train)
```

```
predict_y = sig_clf.predict_proba(X_train)
```

```
print ('For values of best alpha = ', alpha[best_alpha], "The train log loss is:",log_loss(y_train, predict_y))
```

```
predict_y = sig_clf.predict_proba(X_cv)
```

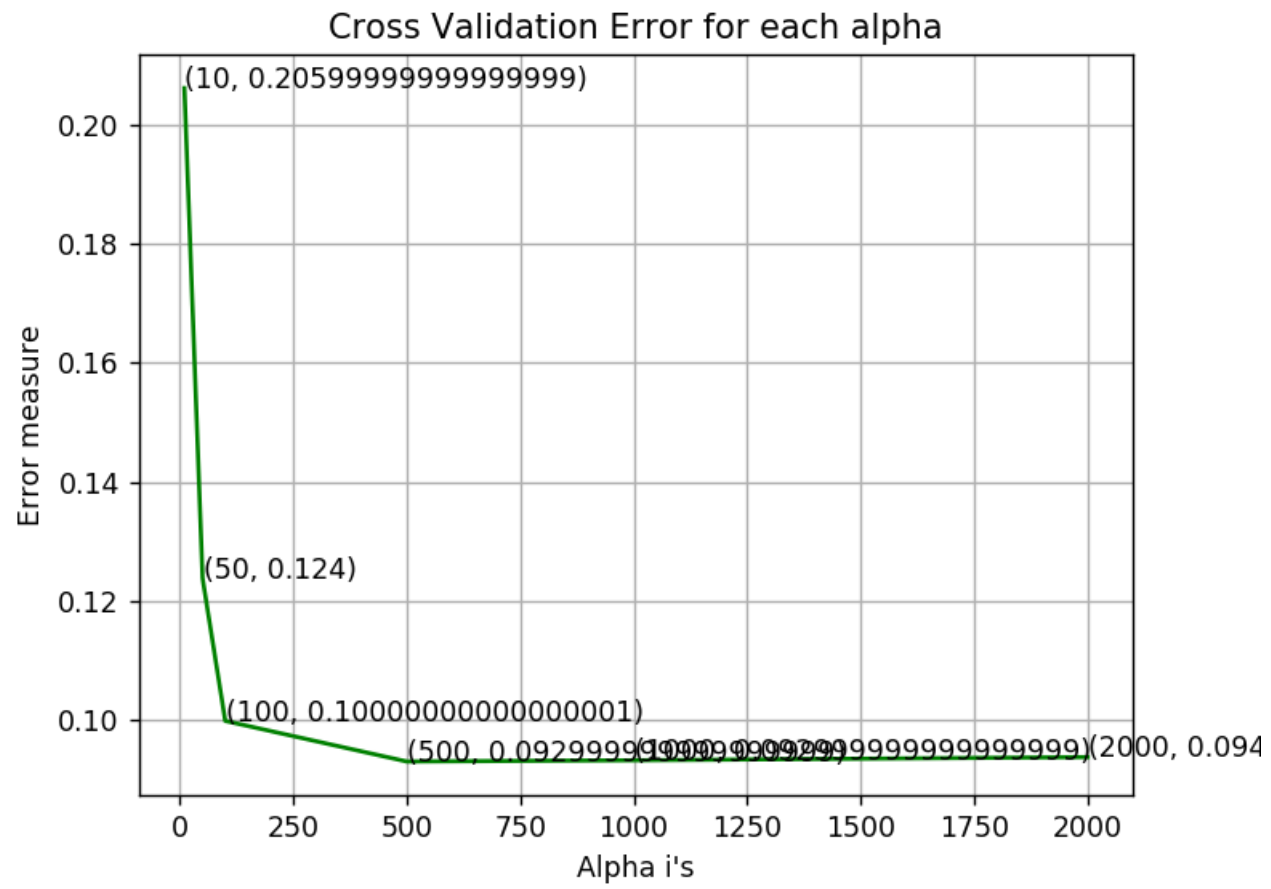
```
print('For values of best alpha = ', alpha[best_alpha], "The cross validation log loss is:",log_loss(y_cv, predict_y))
```

```
predict_y = sig_clf.predict_proba(X_test)
```

```
print('For values of best alpha = ', alpha[best_alpha], "The test log loss is:",log_loss(y_test, predict_y))
```

```
plot_confusion_matrix(y_test, sig_clf.predict(X_test))
```

log_loss for c = 10 is 0.20615980494
log_loss for c = 50 is 0.123888382365
log_loss for c = 100 is 0.099919437112
log_loss for c = 500 is 0.0931035681289
log_loss for c = 1000 is 0.0933084876012
log_loss for c = 2000 is 0.0938395690309

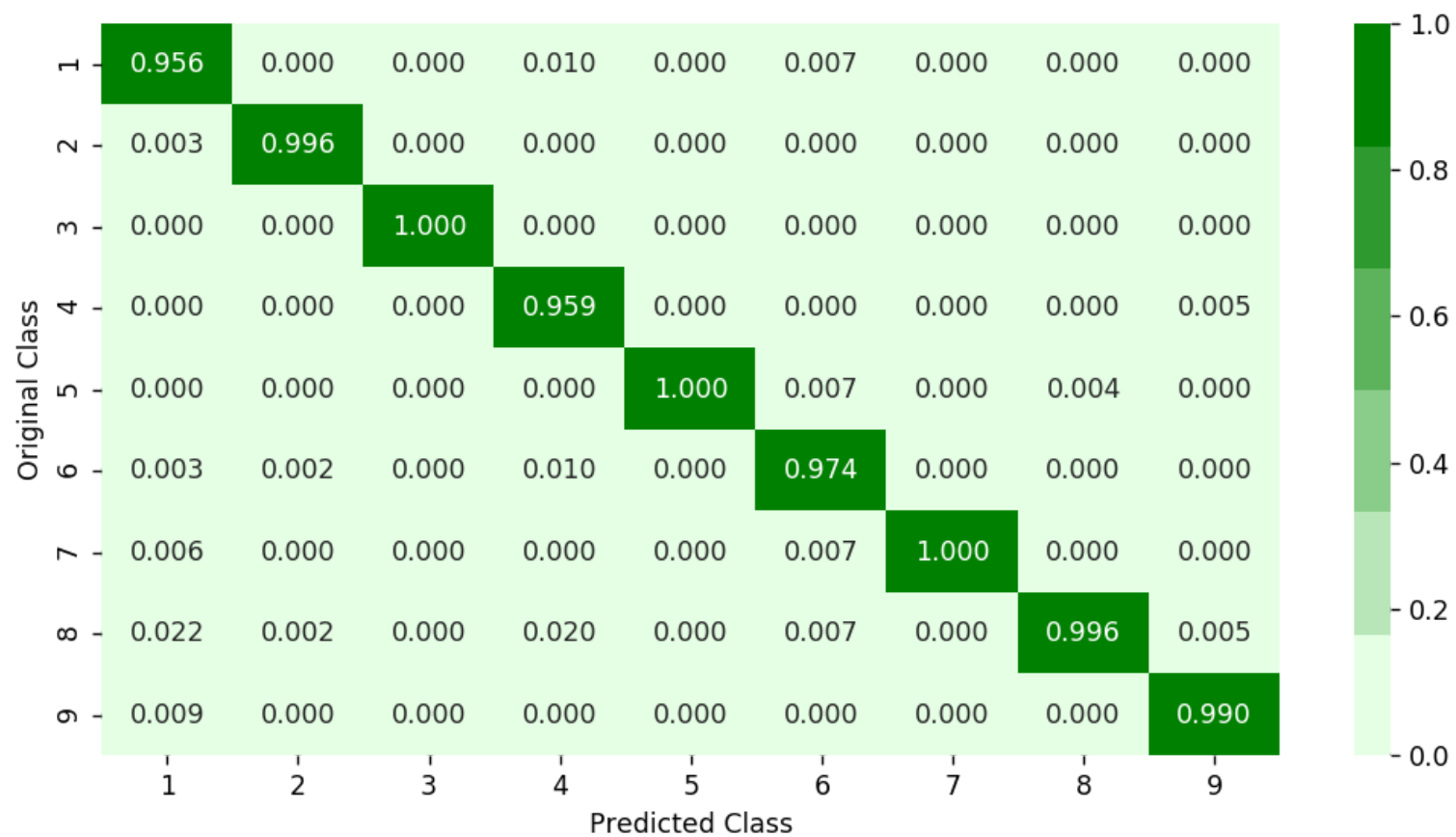


For values of best alpha = 500 The train log loss is: 0.0225231805824
For values of best alpha = 500 The cross validation log loss is: 0.0931035681289
For values of best alpha = 500 The test log loss is: 0.0792067651731
Number of misclassified points 1.24195032199

----- Confusion matrix -----

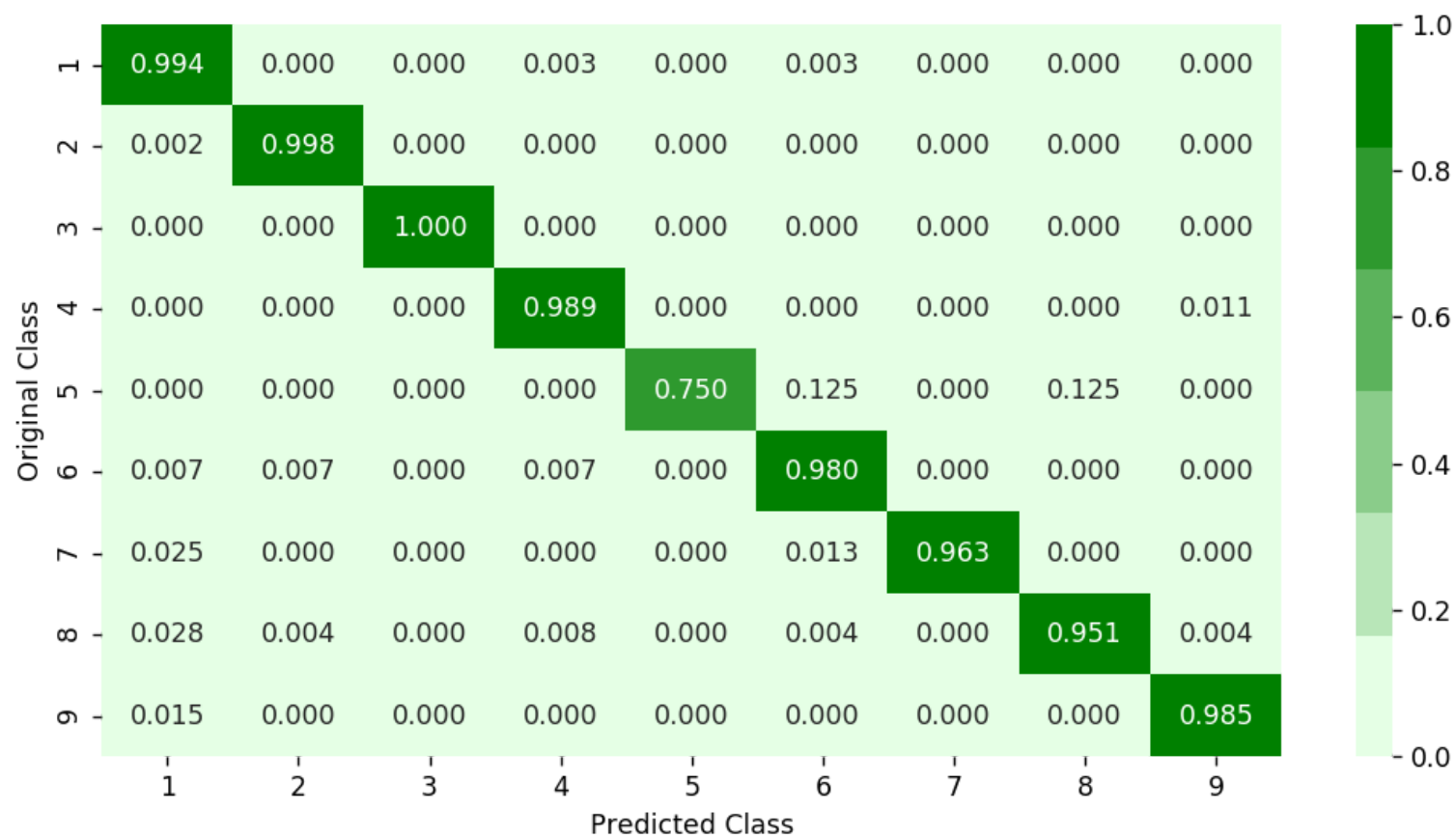


Precision matrix



Sum of columns in precision matrix [1. 1. 1. 1. 1. 1. 1. 1. 1.]

----- Recall matrix -----



Sum of rows in precision matrix [1. 1. 1. 1. 1. 1. 1. 1. 1.]

4.1.5. XgBoost Classification with best hyper parameters using RandomSearch

```
In [ ]: # https://www.analyticsvidhya.com/blog/2016/03/complete-guide-parameter-tuning-xgboost-with-codes-python/
x_cfl=XGBClassifier()
```

```
prams={
    'learning_rate':[0.01,0.03,0.05,0.1,0.15,0.2],
    'n_estimators':[100,200,500,1000,2000],
    'max_depth':[3,5,10],
    'colsample_bytree':[0.1,0.3,0.5,1],
    'subsample':[0.1,0.3,0.5,1]
}
random_cfl1=RandomizedSearchCV(x_cfl,param_distributions=prams,verbose=10,n_jobs=-1,)
random_cfl1.fit(X_train,y_train)
```

Fitting 3 folds for each of 10 candidates, totalling 30 fits

```
[Parallel(n_jobs=-1)]: Done    2 tasks      | elapsed:    26.5s
[Parallel(n_jobs=-1)]: Done    9 tasks      | elapsed:    5.8min
[Parallel(n_jobs=-1)]: Done   19 out of  30 | elapsed:    9.3min remaining:  5.4min
[Parallel(n_jobs=-1)]: Done   23 out of  30 | elapsed:   10.1min remaining:  3.1min
[Parallel(n_jobs=-1)]: Done   27 out of  30 | elapsed:   14.0min remaining:  1.6min
[Parallel(n_jobs=-1)]: Done   30 out of  30 | elapsed:   14.2min finished
```

```
Out[ ]: RandomizedSearchCV(cv=None, error_score='raise',
    estimator=XGBClassifier(base_score=0.5, colsample_bylevel=1, colsample_bytree=1,
    gamma=0, learning_rate=0.1, max_delta_step=0, max_depth=3,
    min_child_weight=1, missing=None, n_estimators=100, nthread=-1,
    objective='binary:logistic', reg_alpha=0, reg_lambda=1,
    scale_pos_weight=1, seed=0, silent=True, subsample=1),
    fit_params=None, iid=True, n_iter=10, n_jobs=-1,
    param_distributions={'learning_rate': [0.01, 0.03, 0.05, 0.1, 0.15, 0.2], 'n_estimators': [100, 200, 500, 1000, 2
    000], 'max_depth': [3, 5, 10], 'colsample_bytree': [0.1, 0.3, 0.5, 1], 'subsample': [0.1, 0.3, 0.5, 1]},
    pre_dispatch='2*n_jobs', random_state=None, refit=True,
    return_train_score=True, scoring=None, verbose=10)
```

```
In [ ]: print (random_cfl1.best_params_)
```

```
{'subsample': 1, 'n_estimators': 500, 'max_depth': 5, 'learning_rate': 0.05, 'colsample_bytree': 0.5}
```

```

In [ ]: # Training a hyper-parameter tuned Xg-Boost regressor on our train data

# find more about XGBClassifier function here http://xgboost.readthedocs.io/en/latest/python/python\_api.html?#xgboost.XGBClassifier
# -----
# default paramters
# class xgboost.XGBClassifier(max_depth=3, learning_rate=0.1, n_estimators=100, silent=True,
# objective='binary:logistic', booster='gbtree', n_jobs=1, nthread=None, gamma=0, min_child_weight=1,
# max_delta_step=0, subsample=1, colsample_bytree=1, colsample_bylevel=1, reg_alpha=0, reg_lambda=1,
# scale_pos_weight=1, base_score=0.5, random_state=0, seed=None, missing=None, **kwargs)

# some of methods of RandomForestRegressor()
# fit(X, y, sample_weight=None, eval_set=None, eval_metric=None, early_stopping_rounds=None, verbose=True, xgb_model=None)
# get_params([deep]) Get parameters for this estimator.
# predict(data, output_margin=False, ntree_limit=0) : Predict with data. NOTE: This function is not thread safe.
# get_score(importance_type='weight') -> get the feature importance
# -----
# video link2: https://www.appliedaicourse.com/course/applied-ai-course-online/lessons/what-are-ensembles/
# -----

x_cfl=XGBClassifier(n_estimators=2000, learning_rate=0.05, colsample_bytree=1, max_depth=3)
x_cfl.fit(X_train,y_train)
c_cfl=CalibratedClassifierCV(x_cfl,method='sigmoid')
c_cfl.fit(X_train,y_train)

predict_y = c_cfl.predict_proba(X_train)
print ('train loss',log_loss(y_train, predict_y))
predict_y = c_cfl.predict_proba(X_cv)
print ('cv loss',log_loss(y_cv, predict_y))
predict_y = c_cfl.predict_proba(X_test)
print ('test loss',log_loss(y_test, predict_y))

train loss 0.022540976086
cv loss 0.0928710624158
test loss 0.0782688587098

```

4.2 Modeling with .asm files

There are 10868 files of asm

All the files make up about 150 GB

The asm files contains :

1. Address
2. Segments
3. Opcodes
4. Registers
5. function calls
6. APIs

With the help of parallel processing we extracted all the features. In parallel we can use all the cores that are present in our computer.

Here we extracted 52 features from all the asm files which are important.

We read the top solutions and handpicked the features from those papers/videos/blogs.

Refer: <https://www.kaggle.com/c/malware-classification/discussion>

4.2.1 Feature extraction from asm files

- To extract the unigram features from the .asm files we need to process ~150GB of data
- **Note: Below two cells will take lot of time (over 48 hours to complete)**
- We will provide you the output file of these two cells, which you can directly use it

```
In [ ]: #initially create five folders
#first
#second
#thrid
#fourth
#fifth
#this code tells us about random split of files into five folders
folder_1 = 'first'
folder_2 = 'second'
folder_3 = 'third'
folder_4 = 'fourth'
folder_5 = 'fifth'
folder_6 = 'output'
for i in [folder_1, folder_2, folder_3, folder_4, folder_5, folder_6]:
    if not os.path.isdir(i):
        os.makedirs(i)

source='train/'
files = os.listdir('train')
#ID=df['Id'].tolist()
data=range(0,10868)
r.shuffle(data)
count=0
for i in range(0,10868):
    if i % 5==0:
        shutil.move(source+files[data[i]], 'first')
    elif i%5==1:
        shutil.move(source+files[data[i]], 'second')
    elif i%5 ==2:
        shutil.move(source+files[data[i]], 'thrid')
    elif i%5 ==3:
        shutil.move(source+files[data[i]], 'fourth')
    elif i%5==4:
        shutil.move(source+files[data[i]], 'fifth')
```

```
In [ ]: #http://flint.cs.yale.edu/cs421/papers/x86-asm/asm.html
```

```
def firstprocess():
    #The prefixes tells about the segments that are present in the asm files
    #There are 450 segments(approx) present in all asm files.
    #this prefixes are best segments that gives us best values.
    #https://en.wikipedia.org/wiki/Data_segment

    prefixes = ['HEADER:', '.text:', '.Pav:', '.idata:', '.data:', '.bss:', '.rdata:', '.edata:', '.rsrc:', '.tls:', '.reloc:', '.BS
S:', '.CODE']
    #this are opcodes that are used to get best results
    #https://en.wikipedia.org/wiki/X86_instruction_listings

    opcodes = ['jmp', 'mov', 'retf', 'push', 'pop', 'xor', 'retn', 'nop', 'sub', 'inc', 'dec', 'add', 'imul', 'xchg', 'or',
'shr', 'cmp', 'call', 'shl', 'ror', 'rol', 'jnb', 'jz', 'rtn', 'lea', 'movzx']
    #best keywords that are taken from different blogs
    keywords = ['.dll', 'std:', ':dword']
    #Below taken registers are general purpose registers and special registers
    #All the registers which are taken are best
    registers=['edx', 'esi', 'eax', 'ebx', 'ecx', 'edi', 'ebp', 'esp', 'eip']
    file1=open("output\asmsmallfile.txt", "w+")
    files = os.listdir('first')
    for f in files:
        #filling the values with zeros into the arrays
        prefixescount=np.zeros(len(prefixes), dtype=int)
        opcodescount=np.zeros(len(opcodes), dtype=int)
        keywordcount=np.zeros(len(keywords), dtype=int)
        registerscount=np.zeros(len(registers), dtype=int)
        features=[]
        f2=f.split('.')[0]
        file1.write(f2+",")
        opcodefile.write(f2+" ")
        # https://docs.python.org/3/library/codecs.html#codecs.ignore_errors
        # https://docs.python.org/3/library/codecs.html#codecs.Codec.encode
        with codecs.open('first/'+f, encoding='cp1252', errors = 'replace') as fli:
            for lines in fli:
                # https://www.tutorialspoint.com/python3/string_rstrip.htm
                line=lines.rstrip().split()
                l=line[0]
                #counting the prefixes in each and every line
                for i in range(len(prefixes)):
                    if prefixes[i] in line[0]:
                        prefixescount[i]+=1
                line=line[1:]
                #counting the opcodes in each and every line
```

```

    for i in range(len(opcodes)):
        if any(opcodes[i]==li for li in line):
            features.append(opcodes[i])
            opcodescount[i]+=1
    #counting registers in the line
    for i in range(len(registers)):
        for li in line:
            # we will use registers only in 'text' and 'CODE' segments
            if registers[i] in li and ('text' in l or 'CODE' in l):
                registerscount[i]+=1
    #counting keywords in the line
    for i in range(len(keywords)):
        for li in line:
            if keywords[i] in li:
                keywordcount[i]+=1
    #pushing the values into the file after reading whole file
    for prefix in prefixescount:
        file1.write(str(prefix)+",")
    for opcode in opcodescount:
        file1.write(str(opcode)+",")
    for register in registerscount:
        file1.write(str(register)+",")
    for key in keywordcount:
        file1.write(str(key)+",")
    file1.write("\n")
file1.close()

```

#same as above

```

def secondprocess():
    prefixes = ['HEADER:', '.text:', '.Pav:', '.idata:', '.data:', '.bss:', '.rdata:', '.edata:', '.rsrc:', '.tls:', '.reloc:', '.BS
S:', '.CODE']
    opcodes = ['jmp', 'mov', 'retf', 'push', 'pop', 'xor', 'retn', 'nop', 'sub', 'inc', 'dec', 'add', 'imul', 'xchg', 'or',
'shr', 'cmp', 'call', 'shl', 'ror', 'rol', 'jnb', 'jz', 'rtn', 'lea', 'movzx']
    keywords = ['.dll', 'std::', ':dword']
    registers=['edx', 'esi', 'eax', 'ebx', 'ecx', 'edi', 'ebp', 'esp', 'eip']
    file1=open("output\mediumasmfile.txt", "w+")
    files = os.listdir('second')
    for f in files:
        prefixescount=np.zeros(len(prefixes), dtype=int)
        opcodescount=np.zeros(len(opcodes), dtype=int)
        keywordcount=np.zeros(len(keywords), dtype=int)
        registerscount=np.zeros(len(registers), dtype=int)
        features=[]
        f2=f.split('.')[0]
        file1.write(f2+",")

```

```

opcodefile.write(f2+" ")
with codecs.open('second/'+f,encoding='cp1252',errors='replace') as fli:
    for lines in fli:
        line=line.rstrip().split()
        l=line[0]
        for i in range(len(prefixes)):
            if prefixes[i] in line[0]:
                prefixescount[i]+=1
        line=line[1:]
        for i in range(len(opcodes)):
            if any(opcodes[i]==li for li in line):
                features.append(opcodes[i])
                opcodecount[i]+=1
        for i in range(len(registers)):
            for li in line:
                if registers[i] in li and ('text' in l or 'CODE' in l):
                    registerscount[i]+=1
        for i in range(len(keywords)):
            for li in line:
                if keywords[i] in li:
                    keywordcount[i]+=1
    for prefix in prefixescount:
        file1.write(str(prefix)+",")
    for opcode in opcodecount:
        file1.write(str(opcode)+",")
    for register in registerscount:
        file1.write(str(register)+",")
    for key in keywordcount:
        file1.write(str(key)+",")
    file1.write("\n")
file1.close()

```

same as smallprocess() functions

```

def thirdprocess():
    prefixes = ['HEADER:', '.text:', '.Pav:', '.idata:', '.data:', '.bss:', '.rdata:', '.edata:', '.rsrc:', '.tls:', '.reloc:', '.BS
S:', '.CODE']
    opcodes = ['jmp', 'mov', 'retf', 'push', 'pop', 'xor', 'retn', 'nop', 'sub', 'inc', 'dec', 'add', 'imul', 'xchg', 'or',
'shr', 'cmp', 'call', 'shl', 'ror', 'rol', 'jnb', 'jz', 'rtn', 'lea', 'movzx']
    keywords = ['.dll', 'std::', ':dword']
    registers=['edx', 'esi', 'eax', 'ebx', 'ecx', 'edi', 'ebp', 'esp', 'eip']
    file1=open("output\largeasmfile.txt", "w+")
    files = os.listdir('thrid')
    for f in files:
        prefixescount=np.zeros(len(prefixes),dtype=int)
        opcodecount=np.zeros(len(opcodes),dtype=int)
        keywordcount=np.zeros(len(keywords),dtype=int)

```

```

registerscount=np.zeros(len(registers),dtype=int)
features=[]
f2=f.split('.')[0]
file1.write(f2+",")
opcodefile.write(f2+" ")
with codecs.open('thrid/'+f,encoding='cp1252',errors='replace') as fli:
    for lines in fli:
        line=lines.rstrip().split()
        l=line[0]
        for i in range(len(prefixes)):
            if prefixes[i] in line[0]:
                prefixescount[i]+=1
        line=line[1:]
        for i in range(len(opcodes)):
            if any(opcodes[i]==li for li in line):
                features.append(opcodes[i])
                opcodecount[i]+=1
        for i in range(len(registers)):
            for li in line:
                if registers[i] in li and ('text' in l or 'CODE' in l):
                    registerscount[i]+=1
        for i in range(len(keywords)):
            for li in line:
                if keywords[i] in li:
                    keywordcount[i]+=1
    for prefix in prefixescount:
        file1.write(str(prefix)+",")
    for opcode in opcodecount:
        file1.write(str(opcode)+",")
    for register in registerscount:
        file1.write(str(register)+",")
    for key in keywordcount:
        file1.write(str(key)+",")
    file1.write("\n")
file1.close()

```

```

def fourthprocess():
    prefixes = ['HEADER:', '.text:', '.Pav:', '.idata:', '.data:', '.bss:', '.rdata:', '.edata:', '.rsrc:', '.tls:', '.reloc:', '.BS
S:', '.CODE']
    opcodes = ['jmp', 'mov', 'retf', 'push', 'pop', 'xor', 'retn', 'nop', 'sub', 'inc', 'dec', 'add', 'imul', 'xchg', 'or',
'shr', 'cmp', 'call', 'shl', 'ror', 'rol', 'jnb', 'jz', 'rtn', 'lea', 'movzx']
    keywords = ['.dll', 'std:', ':dword']
    registers=['edx', 'esi', 'eax', 'ebx', 'ecx', 'edi', 'ebp', 'esp', 'eip']
    file1=open("output\hugeasmfile.txt", "w+")
    files = os.listdir('fourth/')

```

```

for f in files:
    prefixescount=np.zeros(len(prefixes),dtype=int)
    opcodescount=np.zeros(len(opcodes),dtype=int)
    keywordcount=np.zeros(len(keywords),dtype=int)
    registerscount=np.zeros(len(registers),dtype=int)
    features=[]
    f2=f.split('.')[0]
    file1.write(f2+",")
    opcodefile.write(f2+" ")
    with codecs.open('fourth/'+f,encoding='cp1252',errors='replace') as fli:
        for lines in fli:
            line=lines.rstrip().split()
            l=line[0]
            for i in range(len(prefixes)):
                if prefixes[i] in line[0]:
                    prefixescount[i]+=1
            line=line[1:]
            for i in range(len(opcodes)):
                if any(opcodes[i]==li for li in line):
                    features.append(opcodes[i])
                    opcodescount[i]+=1
            for i in range(len(registers)):
                for li in line:
                    if registers[i] in li and ('text' in l or 'CODE' in l):
                        registerscount[i]+=1
            for i in range(len(keywords)):
                for li in line:
                    if keywords[i] in li:
                        keywordcount[i]+=1
        for prefix in prefixescount:
            file1.write(str(prefix)+",")
        for opcode in opcodescount:
            file1.write(str(opcode)+",")
        for register in registerscount:
            file1.write(str(register)+",")
        for key in keywordcount:
            file1.write(str(key)+",")
        file1.write("\n")
    file1.close()

```

```

def fifthprocess():
    prefixes = ['HEADER:', '.text:', '.Pav:', '.idata:', '.data:', '.bss:', '.rdata:', '.edata:', '.rsrc:', '.tls:', '.reloc:', '.BS
S:', '.CODE']
    opcodes = ['jmp', 'mov', 'retf', 'push', 'pop', 'xor', 'retn', 'nop', 'sub', 'inc', 'dec', 'add', 'imul', 'xchg', 'or',
'shr', 'cmp', 'call', 'shl', 'ror', 'rol', 'jnb', 'jz', 'rtn', 'lea', 'movzx']

```

```

keywords = ['.dll', 'std::', ':dword']
registers=['edx', 'esi', 'eax', 'ebx', 'ecx', 'edi', 'ebp', 'esp', 'eip']
file1=open("output\trainasmfile.txt", "w+")
files = os.listdir('fifth/')
for f in files:
    prefixcount=np.zeros(len(prefixes),dtype=int)
    opcodecount=np.zeros(len(opcodes),dtype=int)
    keywordcount=np.zeros(len(keywords),dtype=int)
    registerscount=np.zeros(len(registers),dtype=int)
    features=[]
    f2=f.split('.')[0]
    file1.write(f2+",")
    opcodefile.write(f2+" ")
    with codecs.open('fifth/'+f,encoding='cp1252',errors='replace') as fli:
        for lines in fli:
            line=lines.rstrip().split()
            l=line[0]
            for i in range(len(prefixes)):
                if prefixes[i] in line[0]:
                    prefixcount[i]+=1
            line=line[1:]
            for i in range(len(opcodes)):
                if any(opcodes[i]==li for li in line):
                    features.append(opcodes[i])
                    opcodecount[i]+=1
            for i in range(len(registers)):
                for li in line:
                    if registers[i] in li and ('text' in li or 'CODE' in li):
                        registerscount[i]+=1
            for i in range(len(keywords)):
                for li in line:
                    if keywords[i] in li:
                        keywordcount[i]+=1
    for prefix in prefixcount:
        file1.write(str(prefix)+",")
    for opcode in opcodecount:
        file1.write(str(opcode)+",")
    for register in registerscount:
        file1.write(str(register)+",")
    for key in keywordcount:
        file1.write(str(key)+",")
    file1.write("\n")
file1.close()

```

```
def main():
```



```

#the below code is used for multiprocessing
#the number of process depends upon the number of cores present System
#process is used to call multiprocessing
manager=multiprocessing.Manager()
p1=Process(target=firstprocess)
p2=Process(target=secondprocess)
p3=Process(target=thirdprocess)
p4=Process(target=fourthprocess)
p5=Process(target=fifthprocess)
#p1.start() is used to start the thread execution
p1.start()
p2.start()
p3.start()
p4.start()
p5.start()
#After completion all the threads are joined
p1.join()
p2.join()
p3.join()
p4.join()
p5.join()

if __name__=="__main__":
    main()

```

```

In [ ]: # asmoutputfile.csv(output generated from the above two cells) will contain all the extracted features from .asm files
# this file will be uploaded in the drive, you can directly use this
dfasm=pd.read_csv("asmoutputfile.csv")
Y.columns = ['ID', 'Class']
result_asm = pd.merge(dfasm, Y,on='ID', how='left')
result_asm.head()

```

```

Out[ ]:

```

	ID	HEADER:	.text:	.Pav:	.ldata:	.data:	.bss:	.rdata:	.edata:	.rsrc:	...	edx	esi	eax	ebx	ecx	edi	ebp	esp	eip	Class
0	01kcPWA9K2BOxQeS5Rju	19	744	0	127	57	0	323	0	3	...	18	66	15	43	83	0	17	48	29	1
1	1E93CpP60RHFNiT5Qfvn	17	838	0	103	49	0	0	0	3	...	18	29	48	82	12	0	14	0	20	1
2	3ekVow2ajZHbTnBcsDfX	17	427	0	50	43	0	145	0	3	...	13	42	10	67	14	0	11	0	9	1
3	3X2nY7iQaPBIWDrAZqJe	17	227	0	43	19	0	0	0	3	...	6	8	14	7	2	0	8	0	6	1
4	46OZzdsSKDCFV8h7XWxf	17	402	0	59	170	0	0	0	3	...	12	9	18	29	5	0	11	0	11	1

5 rows × 53 columns

4.2.1.1 Files sizes of each .asm file

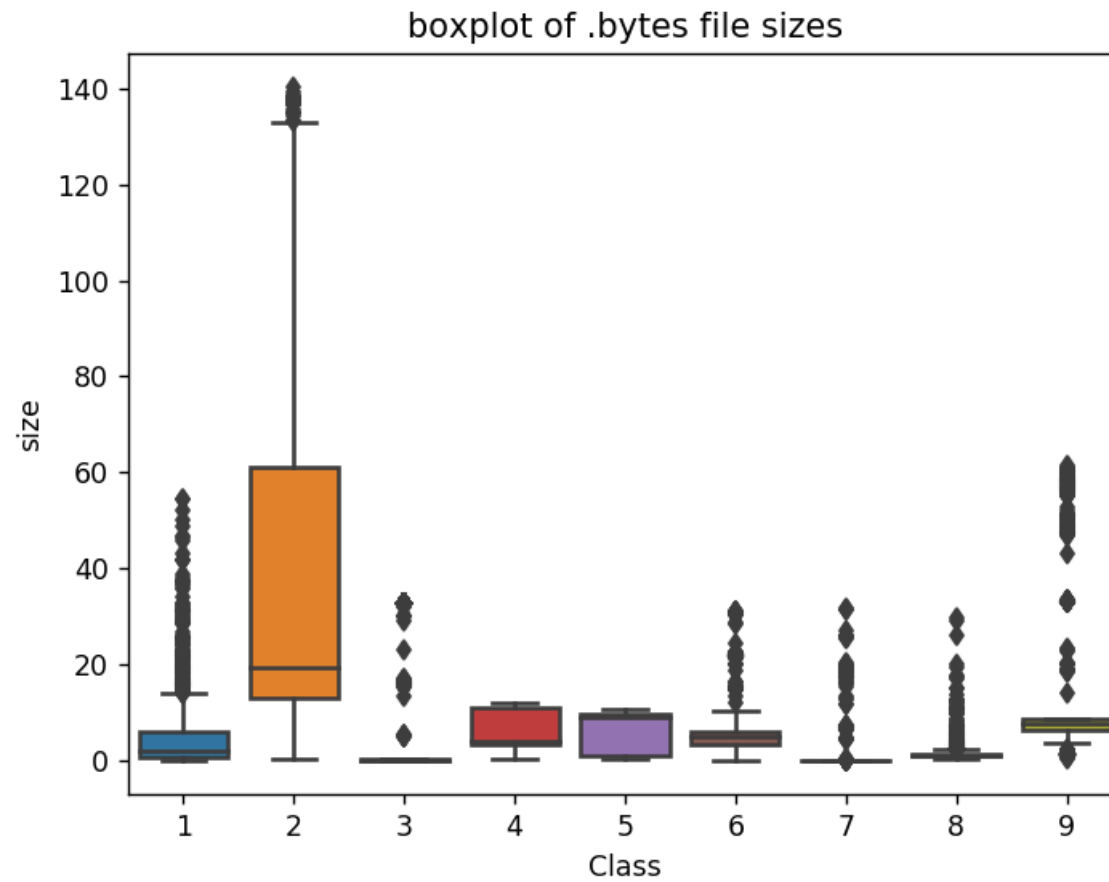
```
In [ ]: #file sizes of byte files

files=os.listdir('asmFiles')
filenames=Y['ID'].tolist()
class_y=Y['Class'].tolist()
class_bytes=[]
sizebytes=[]
fnames=[]
for file in files:
    # print(os.stat('byteFiles/0A32eTdBKayjCWhZqDOQ.txt'))
    # os.stat_result(st_mode=33206, st_ino=1125899906874507, st_dev=3561571700, st_nlink=1, st_uid=0, st_gid=0,
    # st_size=3680109, st_atime=1519638522, st_mtime=1519638522, st_ctime=1519638522)
    # read more about os.stat: here https://www.tutorialspoint.com/python/os\_stat.htm
    statinfo=os.stat('asmFiles/'+file)
    # split the file name at '.' and take the first part of it i.e the file name
    file=file.split('.')[0]
    if any(file == filename for filename in filenames):
        i=filenames.index(file)
        class_bytes.append(class_y[i])
        # converting into Mb's
        sizebytes.append(statinfo.st_size/(1024.0*1024.0))
        fnames.append(file)
asm_size_byte=pd.DataFrame({'ID':fnames,'size':sizebytes,'Class':class_bytes})
print (asm_size_byte.head())
```

	Class	ID	size
0	9	01azqd4InC7m9JpocGv5	56.229886
1	2	01IsoiSMh5gxyDYT14CB	13.999378
2	9	01jsnpXSA1gw6aPeDxrU	8.507785
3	1	01kcPWA9K2B0xQeS5Rju	0.078190
4	8	01SuzwMJEIXsK7A8dQb1	0.996723

4.2.1.2 Distribution of .asm file sizes

```
In [ ]: #boxplot of asm files
ax = sns.boxplot(x="Class", y="size", data=asm_size_byte)
plt.title("boxplot of .bytes file sizes")
plt.show()
```



```
In [ ]: # add the file size feature to previous extracted features
print(result_asm.shape)
print(asm_size_byte.shape)
result_asm = pd.merge(result_asm, asm_size_byte.drop(['Class'], axis=1), on='ID', how='left')
result_asm.head()
```

```
(10868, 53)
(10868, 3)
```

```
Out[ ]:
```

	ID	HEADER:	.text:	.Pav:	.idata:	.data:	.bss:	.rdata:	.edata:	.rsrc:	...	esi	eax	ebx	ecx	edi	ebp	esp	eip	Class	size
0	01kcPWA9K2BOxQeS5Rju	19	744	0	127	57	0	323	0	3	...	66	15	43	83	0	17	48	29	1	0.078190
1	1E93CpP60RHFNiT5Qfvn	17	838	0	103	49	0	0	0	3	...	29	48	82	12	0	14	0	20	1	0.063400
2	3ekVow2ajZHbTnBcsDfX	17	427	0	50	43	0	145	0	3	...	42	10	67	14	0	11	0	9	1	0.041695
3	3X2nY7iQaPBIWDrAZqJe	17	227	0	43	19	0	0	0	3	...	8	14	7	2	0	8	0	6	1	0.018757
4	46OZzdsSKDCFV8h7XWxf	17	402	0	59	170	0	0	0	3	...	9	18	29	5	0	11	0	11	1	0.037567

5 rows × 54 columns

```
In [ ]: # we normalize the data each column
result_asm = normalize(result_asm)
result_asm.head()
```

```
Out[ ]:
```

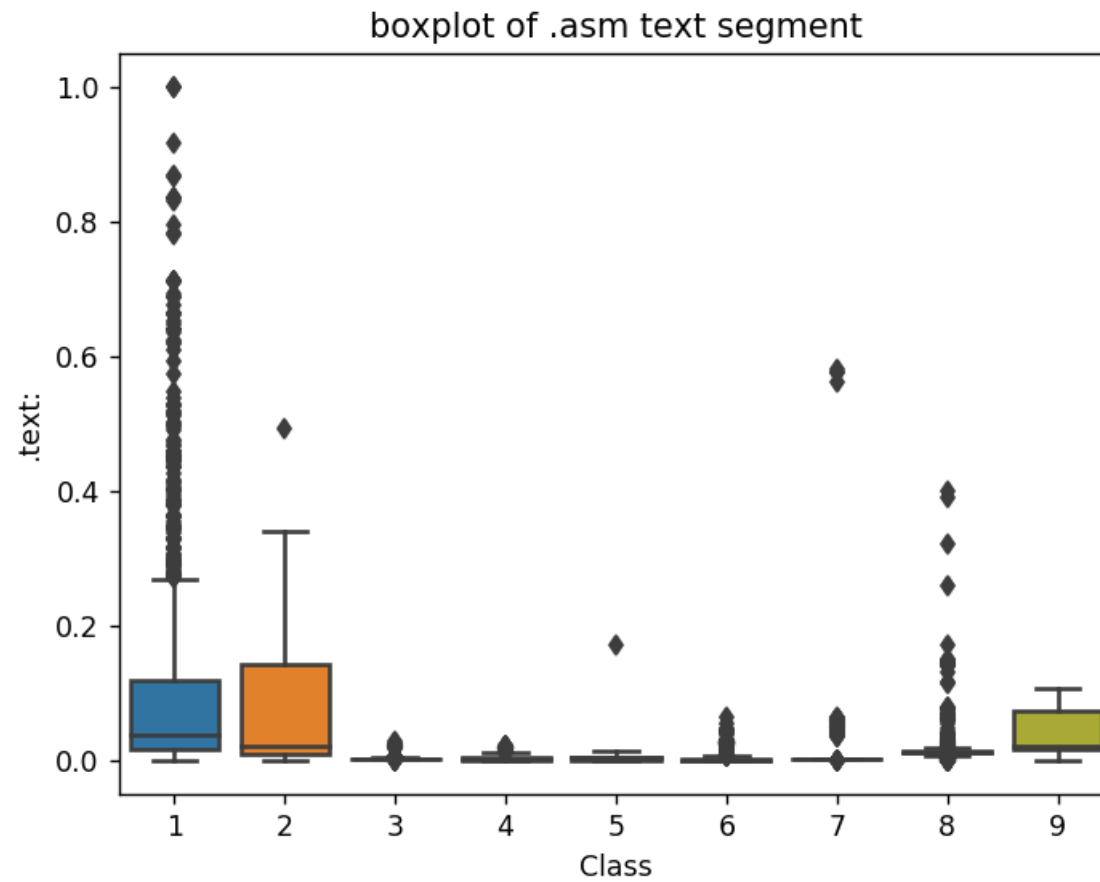
	ID	HEADER:	.text:	.Pav:	.idata:	.data:	.bss:	.rdata:	.edata:	.rsrc:	...	esi	eax	ebx	ecx	ed
0	01kcPWA9K2BOxQeS5Rju	0.107345	0.001092	0.0	0.000761	0.000023	0.0	0.000084	0.0	0.000072	...	0.000746	0.000301	0.000360	0.001057	0.0
1	1E93CpP60RHFNiT5Qfvn	0.096045	0.001230	0.0	0.000617	0.000019	0.0	0.000000	0.0	0.000072	...	0.000328	0.000965	0.000686	0.000153	0.0
2	3ekVow2ajZHbTnBcsDfX	0.096045	0.000627	0.0	0.000300	0.000017	0.0	0.000038	0.0	0.000072	...	0.000475	0.000201	0.000560	0.000178	0.0
3	3X2nY7iQaPBIWDrAZqJe	0.096045	0.000333	0.0	0.000258	0.000008	0.0	0.000000	0.0	0.000072	...	0.000090	0.000281	0.000059	0.000025	0.0
4	46OZzdsSKDCFV8h7XWxf	0.096045	0.000590	0.0	0.000353	0.000068	0.0	0.000000	0.0	0.000072	...	0.000102	0.000362	0.000243	0.000064	0.0

5 rows × 54 columns



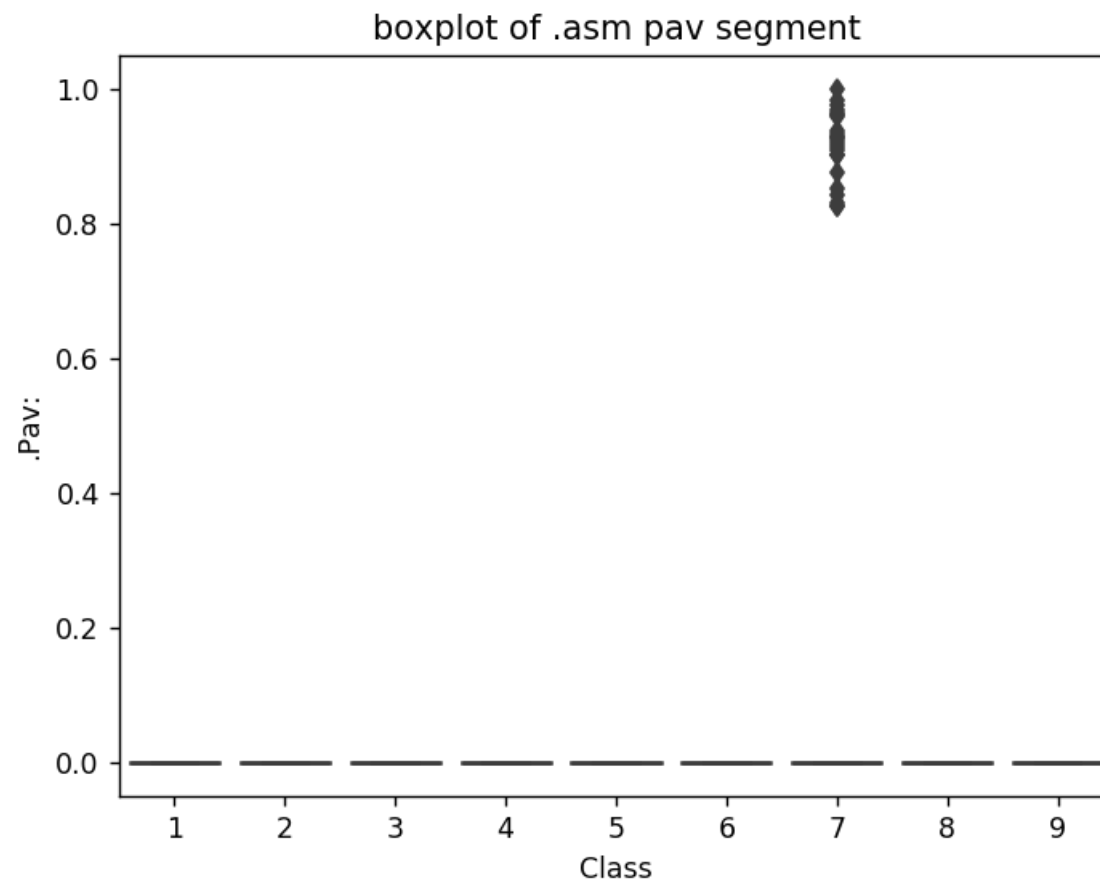
4.2.2 Univariate analysis on asm file features

```
In [ ]: ax = sns.boxplot(x="Class", y=".text:", data=result_asm)
plt.title("boxplot of .asm text segment")
plt.show()
```

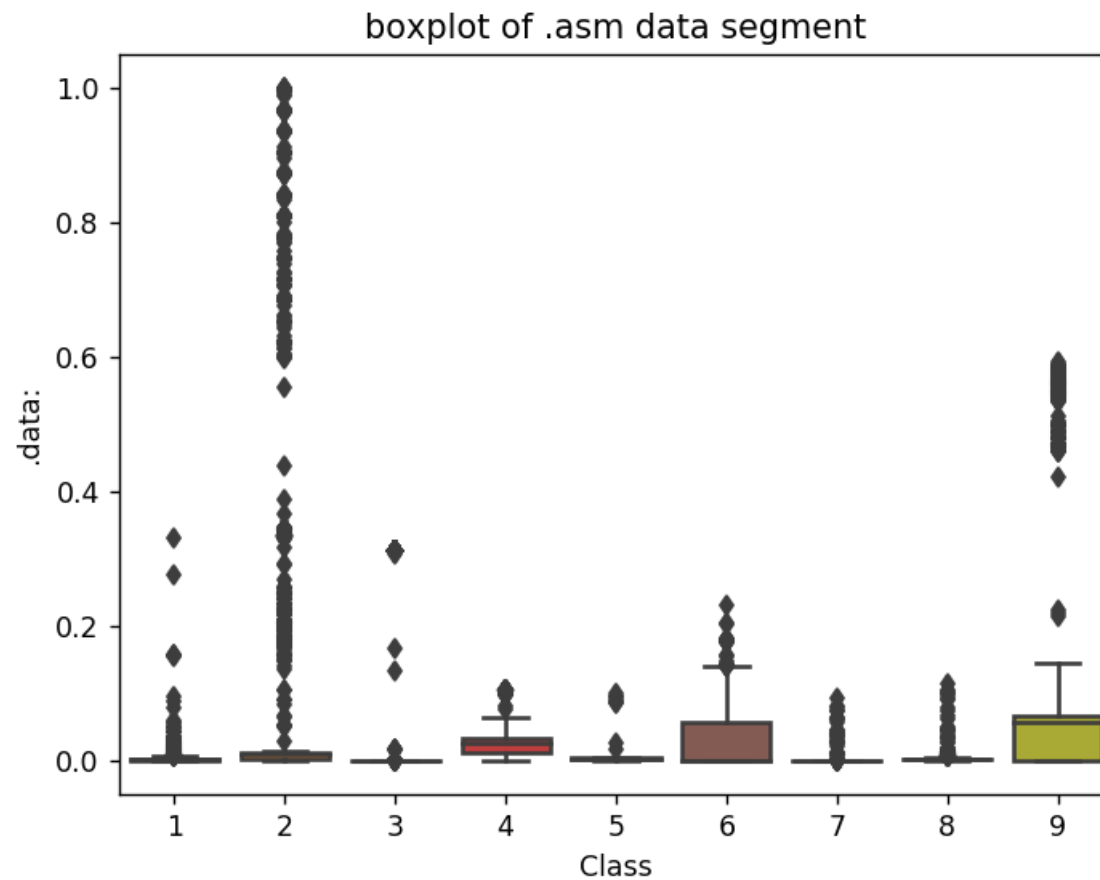


The plot is between Text and class
Class 1,2 and 9 can be easily separated

```
In [ ]: ax = sns.boxplot(x="Class", y=".Pav:", data=result_asm)
plt.title("boxplot of .asm pav segment")
plt.show()
```

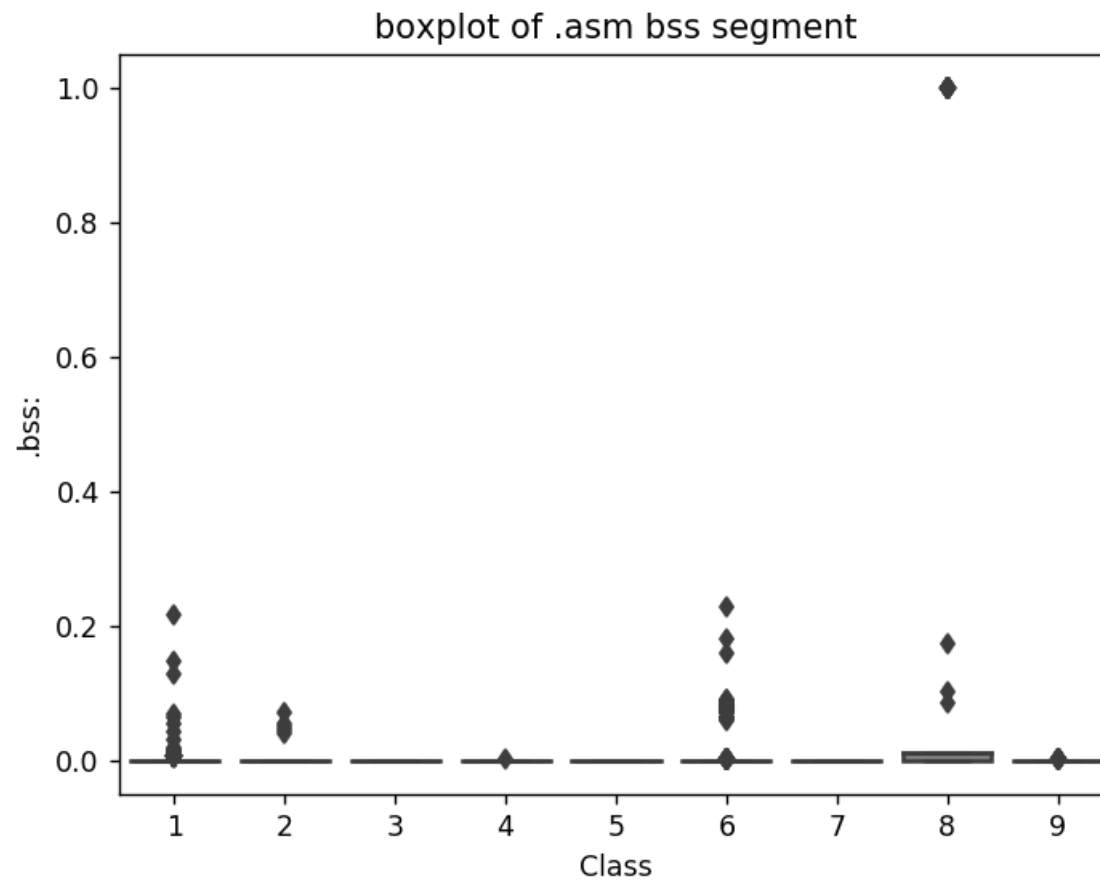


```
In [ ]: ax = sns.boxplot(x="Class", y=".data:", data=result_asm)
plt.title("boxplot of .asm data segment")
plt.show()
```



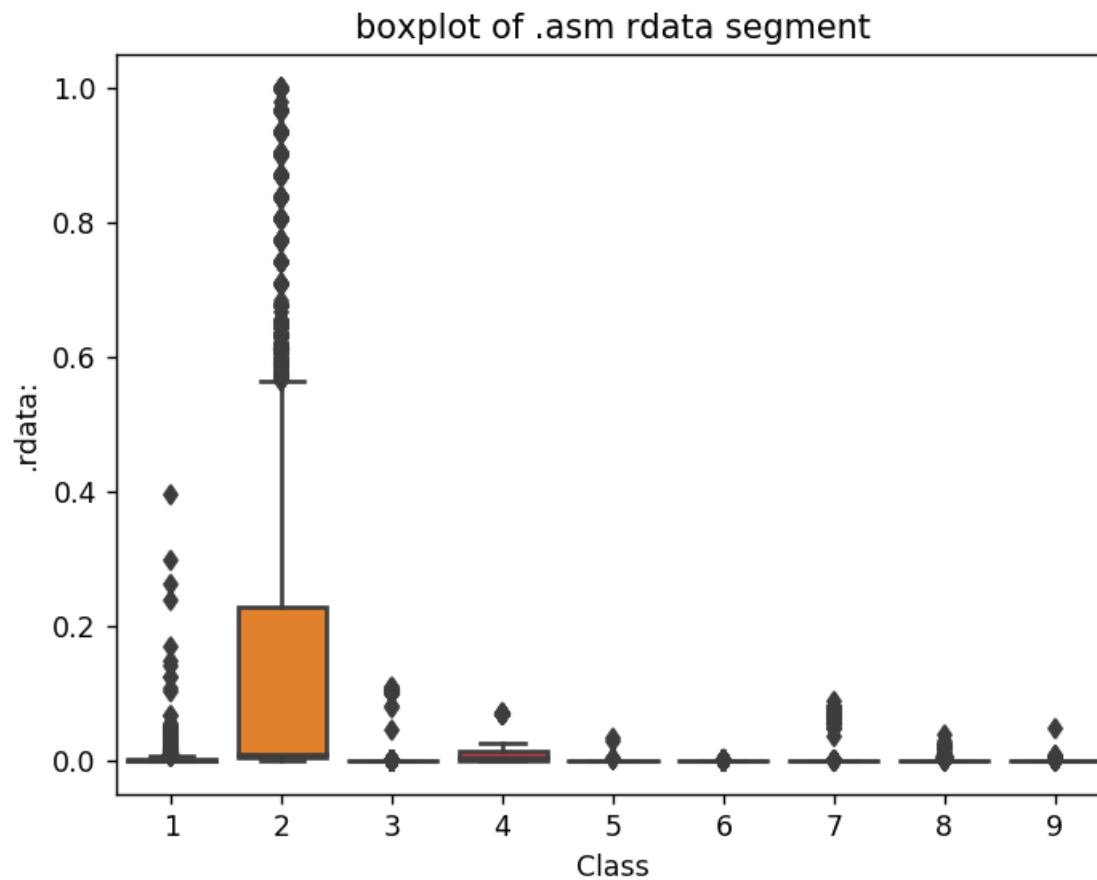
The plot is between data segment and class label
class 6 and class 9 can be easily separated from given points

```
In [ ]: ax = sns.boxplot(x="Class", y=".bss:", data=result_asm)
plt.title("boxplot of .asm bss segment")
plt.show()
```



plot between bss segment and class label
very less number of files are having bss segment

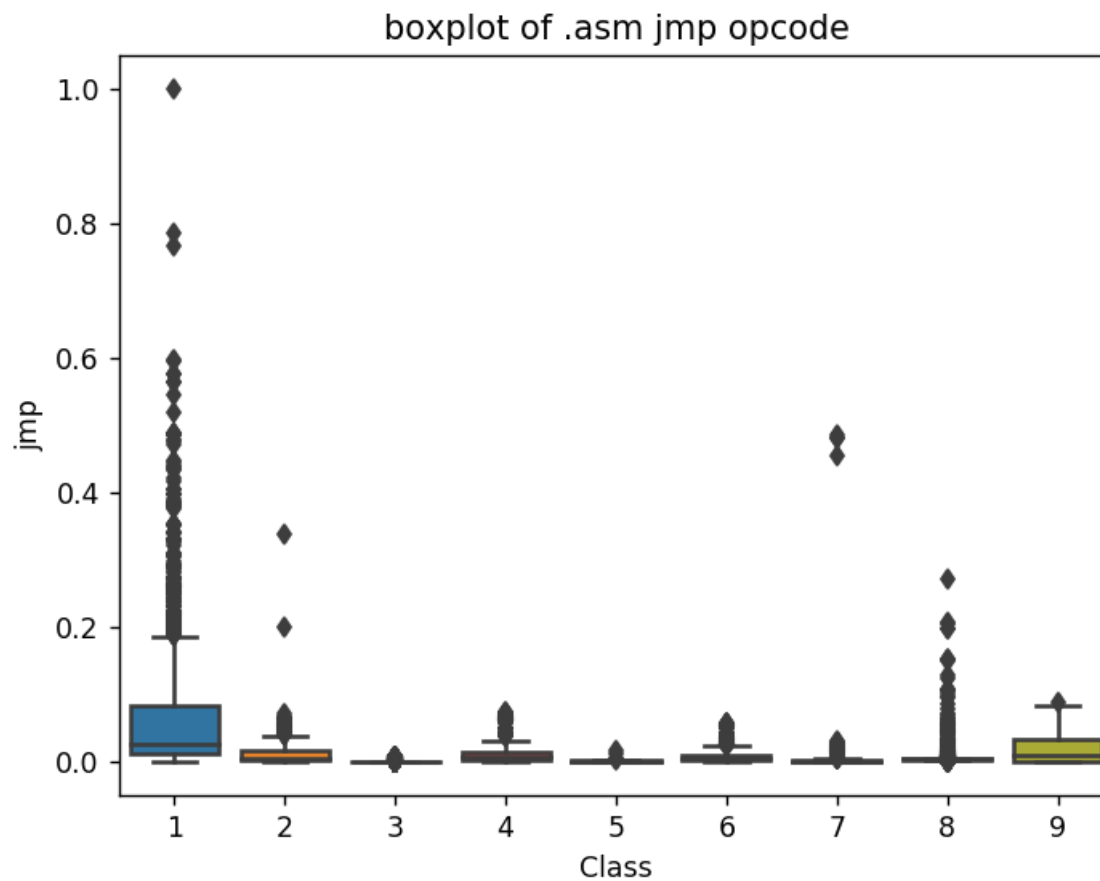

```
In [ ]: ax = sns.boxplot(x="Class", y=".rdata:", data=result_asm)
plt.title("boxplot of .asm rdata segment")
plt.show()
```



Plot between rdata segment and Class segment

Class 2 can be easily separated 75 percentile files are having 1M rdata lines

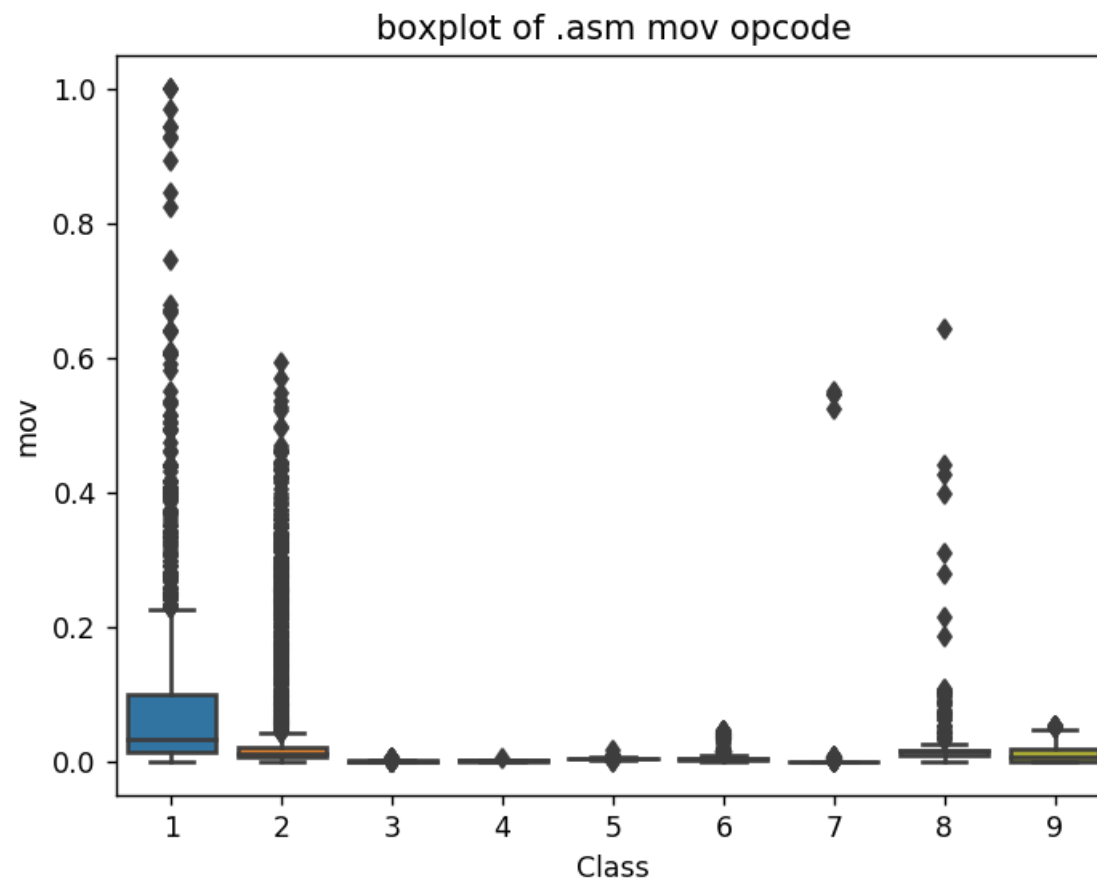
```
In [ ]: ax = sns.boxplot(x="Class", y="jmp", data=result_asm)
plt.title("boxplot of .asm jmp opcode")
plt.show()
```



plot between jmp and Class label

Class 1 is having frequency of 2000 approx in 75 percentile of files

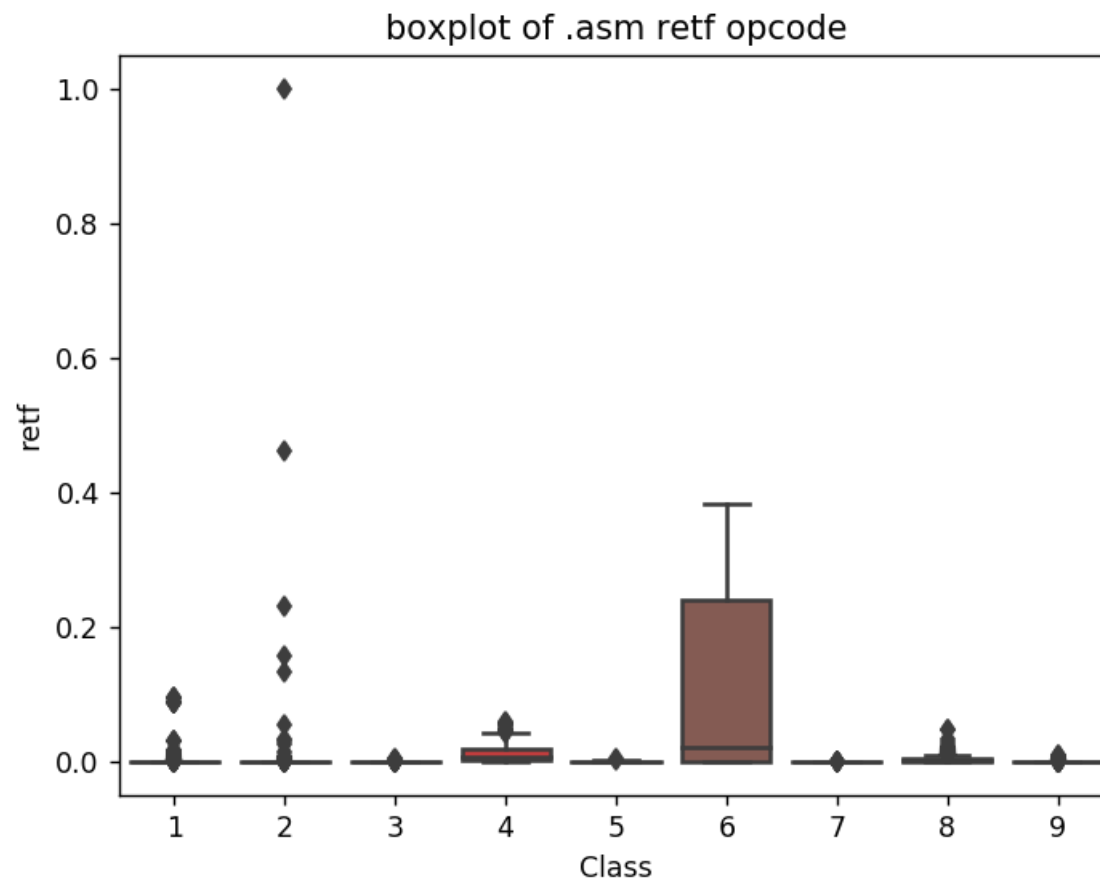
```
In [ ]: ax = sns.boxplot(x="Class", y="mov", data=result_asm)
plt.title("boxplot of .asm mov opcode")
plt.show()
```



plot between Class label and mov opcode

Class 1 is having frequency of 2000 approx in 75 percentile of files

```
In [ ]: ax = sns.boxplot(x="Class", y="retf", data=result_asm)
plt.title("boxplot of .asm retf opcode")
plt.show()
```

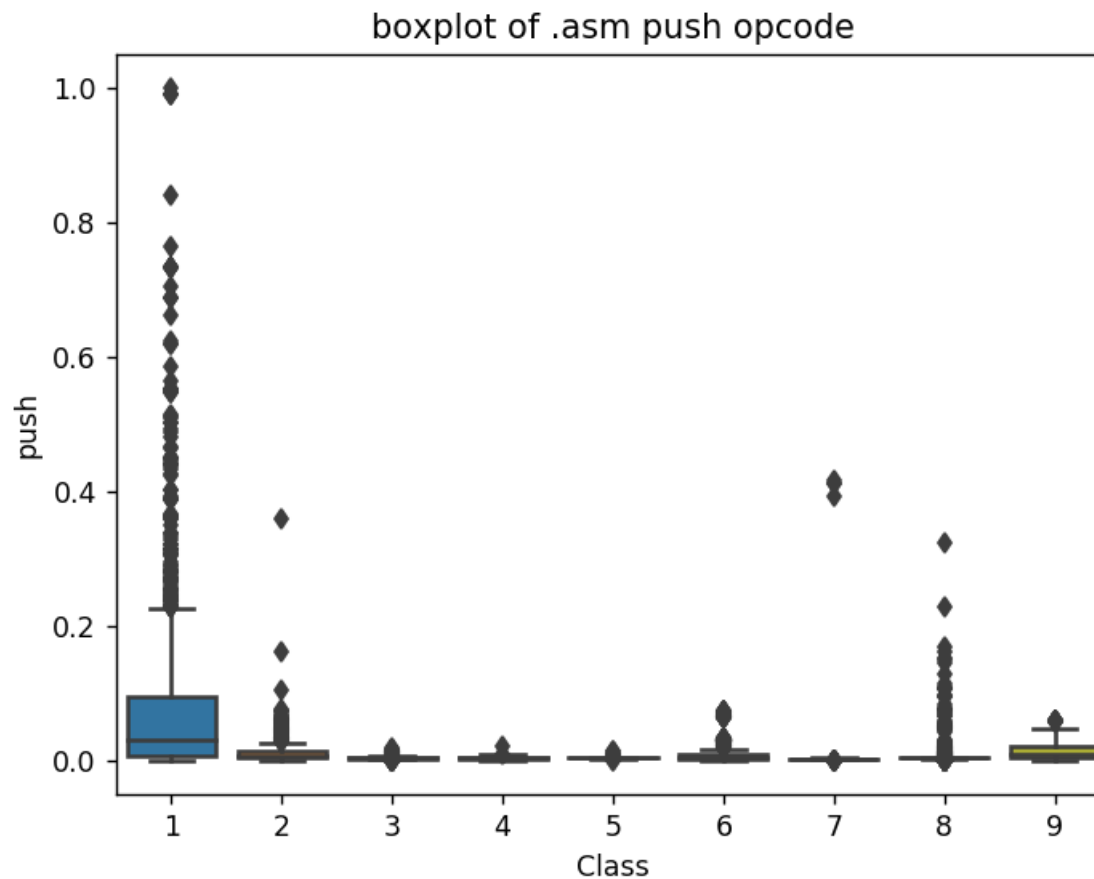


plot between Class label and retf

Class 6 can be easily separated with opcode retf

The frequency of retf is approx of 250.

```
In [ ]: ax = sns.boxplot(x="Class", y="push", data=result_asm)
plt.title("boxplot of .asm push opcode")
plt.show()
```



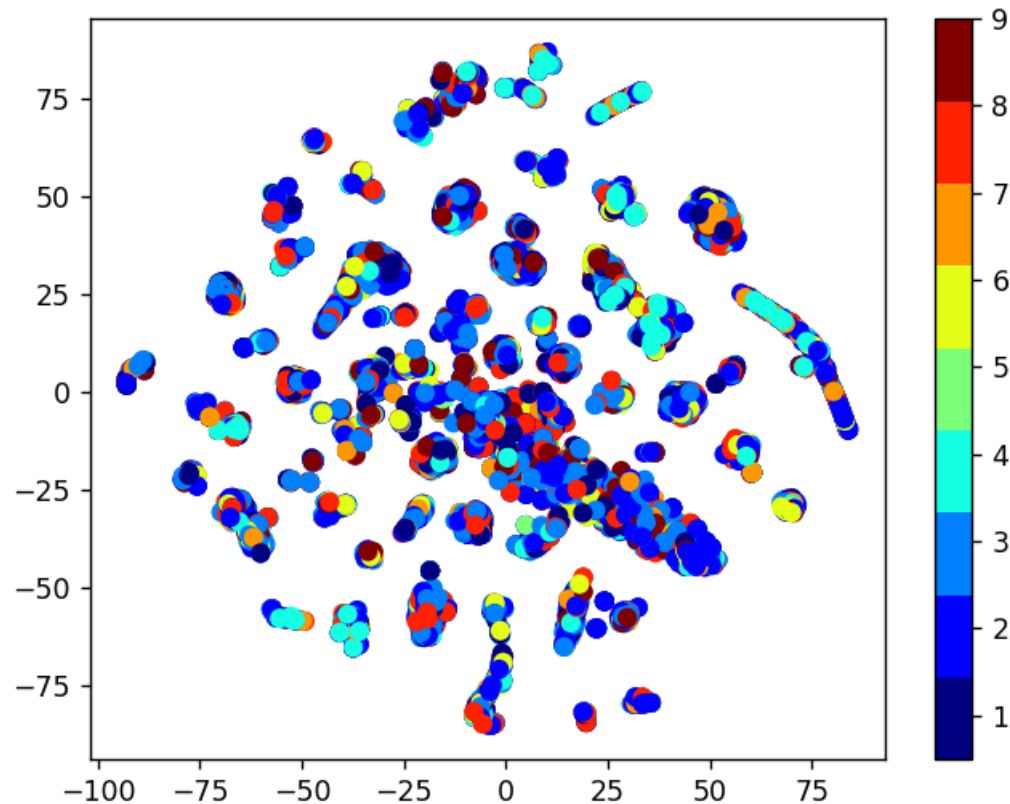
plot between push opcode and Class label

Class 1 is having 75 precentile files with push opcodes of frequency 1000

4.2.2 Multivariate Analysis on .asm file features

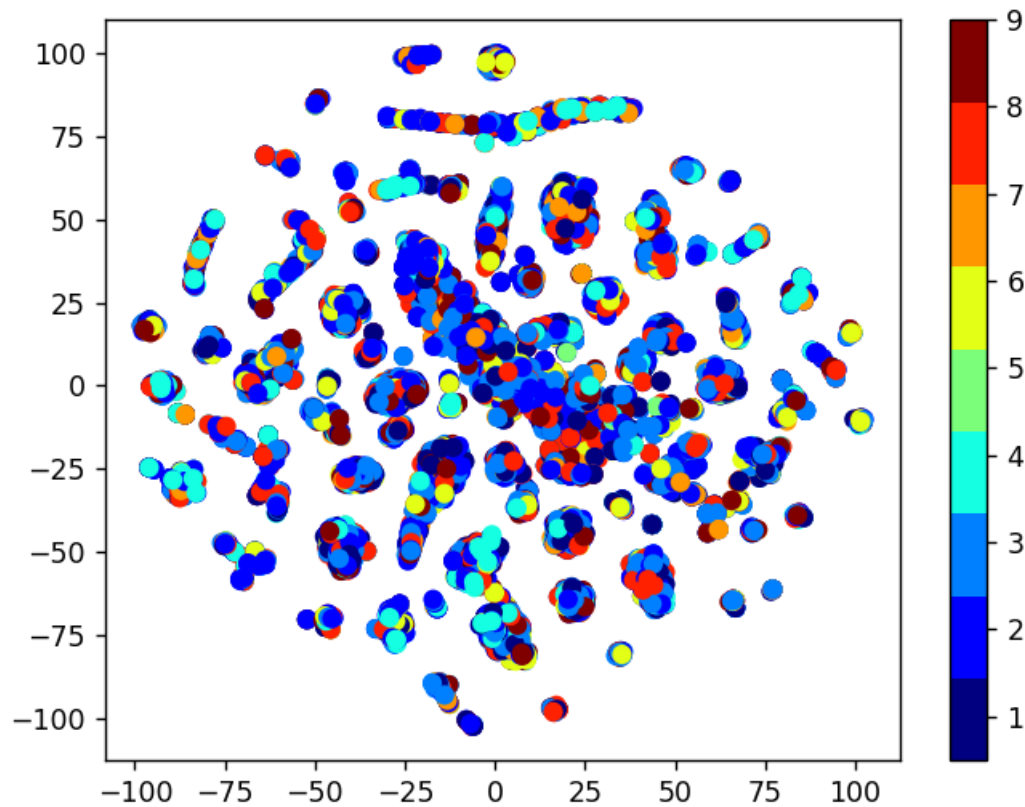
```
In [ ]: # check out the course content for more explantion on tsne algorithm
# https://www.applidaicourse.com/course/applied-ai-course-online/lessons/t-distributed-stochastic-neighbourhood-embedding
t-sne-part-1/

#multivariate analysis on byte files
#this is with perplexity 50
xtsne=TSNE(perplexity=50)
results=xtsne.fit_transform(result_asm.drop(['ID','Class'], axis=1).fillna(0))
vis_x = results[:, 0]
vis_y = results[:, 1]
plt.scatter(vis_x, vis_y, c=data_y, cmap=plt.cm.get_cmap("jet", 9))
plt.colorbar(ticks=range(10))
plt.clim(0.5, 9)
plt.show()
```



```
In [ ]: # by univariate analysis on the .asm file features we are getting very negligible information from
# 'rtn', '.BSS:', '.CODE' features, so here we are trying multivariate analysis after removing those features
# the plot looks very messy
```

```
xtsne=TSNE(perplexity=30)
results=xtsne.fit_transform(result_asm.drop(['ID','Class', 'rtn', '.BSS:', '.CODE','size'], axis=1))
vis_x = results[:, 0]
vis_y = results[:, 1]
plt.scatter(vis_x, vis_y, c=data_y, cmap=plt.cm.get_cmap("jet", 9))
plt.colorbar(ticks=range(10))
plt.clim(0.5, 9)
plt.show()
```



TSNE for asm data with perplexity 50

4.2.3 Conclusion on EDA

- We have taken only 52 features from asm files (after reading through many blogs and research papers)
- The univariate analysis was done only on few important features.
- Take-aways
 - 1. Class 3 can be easily separated because of the frequency of segments, opcodes and keywords being less
 - 2. Each feature has its unique importance in separating the Class labels.

4.3 Train and test split

```
In [ ]: asm_y = result_asm['Class']  
asm_x = result_asm.drop(['ID', 'Class', '.BSS:', 'rtn', '.CODE'], axis=1)
```

```
In [ ]: X_train_asm, X_test_asm, y_train_asm, y_test_asm = train_test_split(asm_x, asm_y, stratify=asm_y, test_size=0.20)  
X_train_asm, X_cv_asm, y_train_asm, y_cv_asm = train_test_split(X_train_asm, y_train_asm, stratify=y_train_asm, test_size=0.20)
```



```
In [ ]: print( X_cv_asm.isnull().all())
```

HEADER:	False
.text:	False
.Pav:	False
.idata:	False
.data:	False
.bss:	False
.rdata:	False
.edata:	False
.rsrc:	False
.tls:	False
.reloc:	False
jmp	False
mov	False
retf	False
push	False
pop	False
xor	False
retn	False
nop	False
sub	False
inc	False
dec	False
add	False
imul	False
xchg	False
or	False
shr	False
cmp	False
call	False
shl	False
ror	False
rol	False
jnb	False
jz	False
lea	False
movzx	False
.dll	False
std::	False
:dword	False
edx	False
esi	False
eax	False
ebx	False
ecx	False
edi	False
ebp	False

```
esp      False
eip      False
size     False
dtype: bool
```

4.4. Machine Learning models on features of .asm files

4.4.1 K-Nearest Neighbors

```

In [ ]: # find more about KNeighborsClassifier() here http://scikit-learn.org/stable/modules/generated/sklearn.neighbors.KNeighborsClassifier.html
# -----
# default parameter
# KNeighborsClassifier(n_neighbors=5, weights='uniform', algorithm='auto', leaf_size=30, p=2,
# metric='minkowski', metric_params=None, n_jobs=1, **kwargs)

# methods of
# fit(X, y) : Fit the model using X as training data and y as target values
# predict(X):Predict the class labels for the provided data
# predict_proba(X):Return probability estimates for the test data X.
#-----
# video link: https://www.appliedaicourse.com/course/applied-ai-course-online/lessons/k-nearest-neighbors-geometric-intuition-with-a-toy-example-1/
#-----

# find more about CalibratedClassifierCV here at http://scikit-learn.org/stable/modules/generated/sklearn.calibration.CalibratedClassifierCV.html
# -----
# default paramters
# sklearn.calibration.CalibratedClassifierCV(base_estimator=None, method='sigmoid', cv=3)
#
# some of the methods of CalibratedClassifierCV()
# fit(X, y[, sample_weight]) Fit the calibrated model
# get_params([deep]) Get parameters for this estimator.
# predict(X) Predict the target of new samples.
# predict_proba(X) Posterior probabilities of classification
#-----
# video link:
#-----

alpha = [x for x in range(1, 21,2)]
cv_log_error_array=[]
for i in alpha:
    k_cfl=KNeighborsClassifier(n_neighbors=i)
    k_cfl.fit(X_train_asm,y_train_asm)
    sig_clf = CalibratedClassifierCV(k_cfl, method="sigmoid")
    sig_clf.fit(X_train_asm, y_train_asm)
    predict_y = sig_clf.predict_proba(X_cv_asm)
    cv_log_error_array.append(log_loss(y_cv_asm, predict_y, labels=k_cfl.classes_, eps=1e-15))

for i in range(len(cv_log_error_array)):
    print ('log_loss for k = ',alpha[i],'is',cv_log_error_array[i])

```

```

best_alpha = np.argmin(cv_log_error_array)

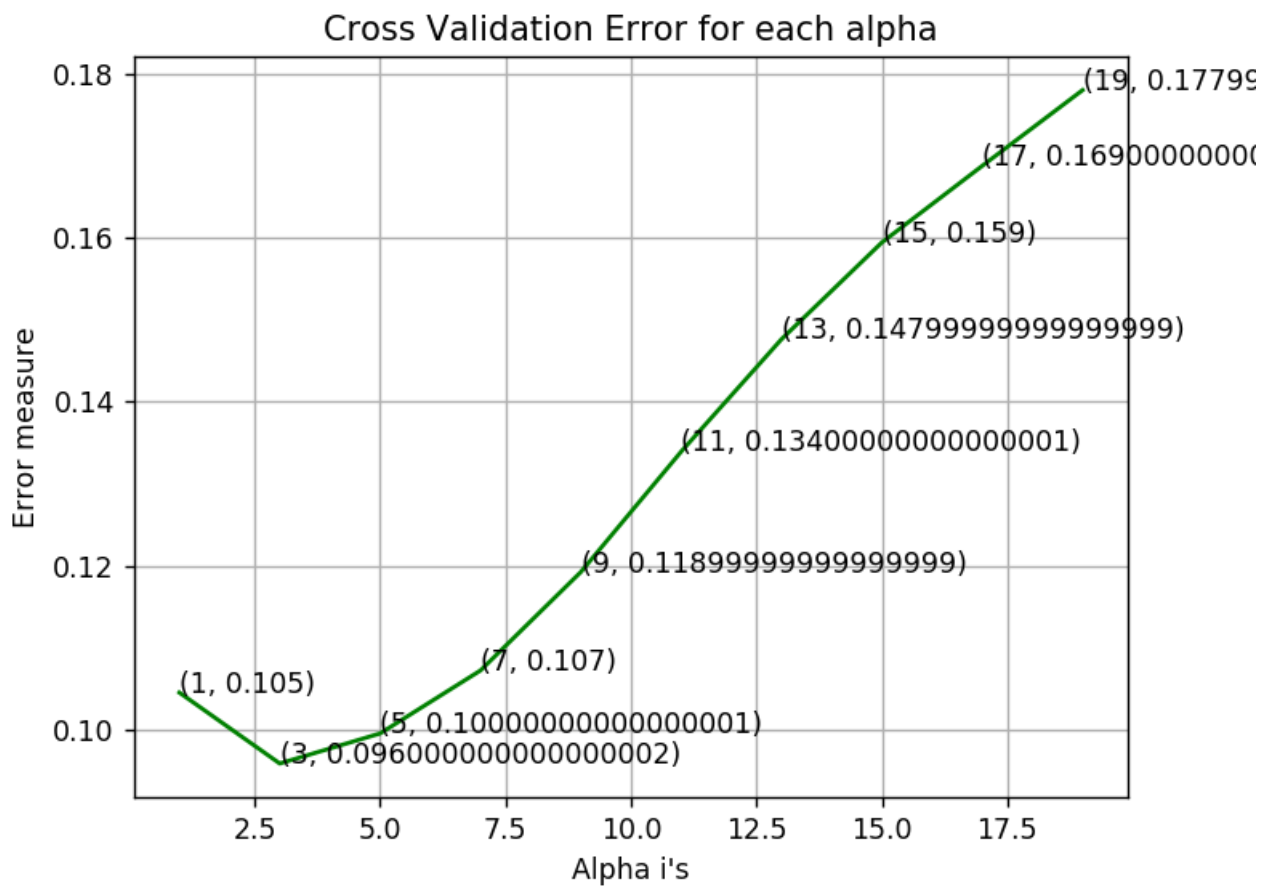
fig, ax = plt.subplots()
ax.plot(alpha, cv_log_error_array,c='g')
for i, txt in enumerate(np.round(cv_log_error_array,3)):
    ax.annotate((alpha[i],np.round(txt,3)), (alpha[i],cv_log_error_array[i]))
plt.grid()
plt.title("Cross Validation Error for each alpha")
plt.xlabel("Alpha i's")
plt.ylabel("Error measure")
plt.show()

k_cfl=KNeighborsClassifier(n_neighbors=alpha[best_alpha])
k_cfl.fit(X_train_asm,y_train_asm)
sig_clf = CalibratedClassifierCV(k_cfl, method="sigmoid")
sig_clf.fit(X_train_asm, y_train_asm)
pred_y=sig_clf.predict(X_test_asm)

predict_y = sig_clf.predict_proba(X_train_asm)
print ('log loss for train data',log_loss(y_train_asm, predict_y))
predict_y = sig_clf.predict_proba(X_cv_asm)
print ('log loss for cv data',log_loss(y_cv_asm, predict_y))
predict_y = sig_clf.predict_proba(X_test_asm)
print ('log loss for test data',log_loss(y_test_asm, predict_y))
plot_confusion_matrix(y_test_asm,sig_clf.predict(X_test_asm))

```

log_loss for k = 1 is 0.104531321344
log_loss for k = 3 is 0.0958800580948
log_loss for k = 5 is 0.0995466557335
log_loss for k = 7 is 0.107227274345
log_loss for k = 9 is 0.119239543547
log_loss for k = 11 is 0.133926642781
log_loss for k = 13 is 0.147643793967
log_loss for k = 15 is 0.159439699615
log_loss for k = 17 is 0.16878376444
log_loss for k = 19 is 0.178020728839

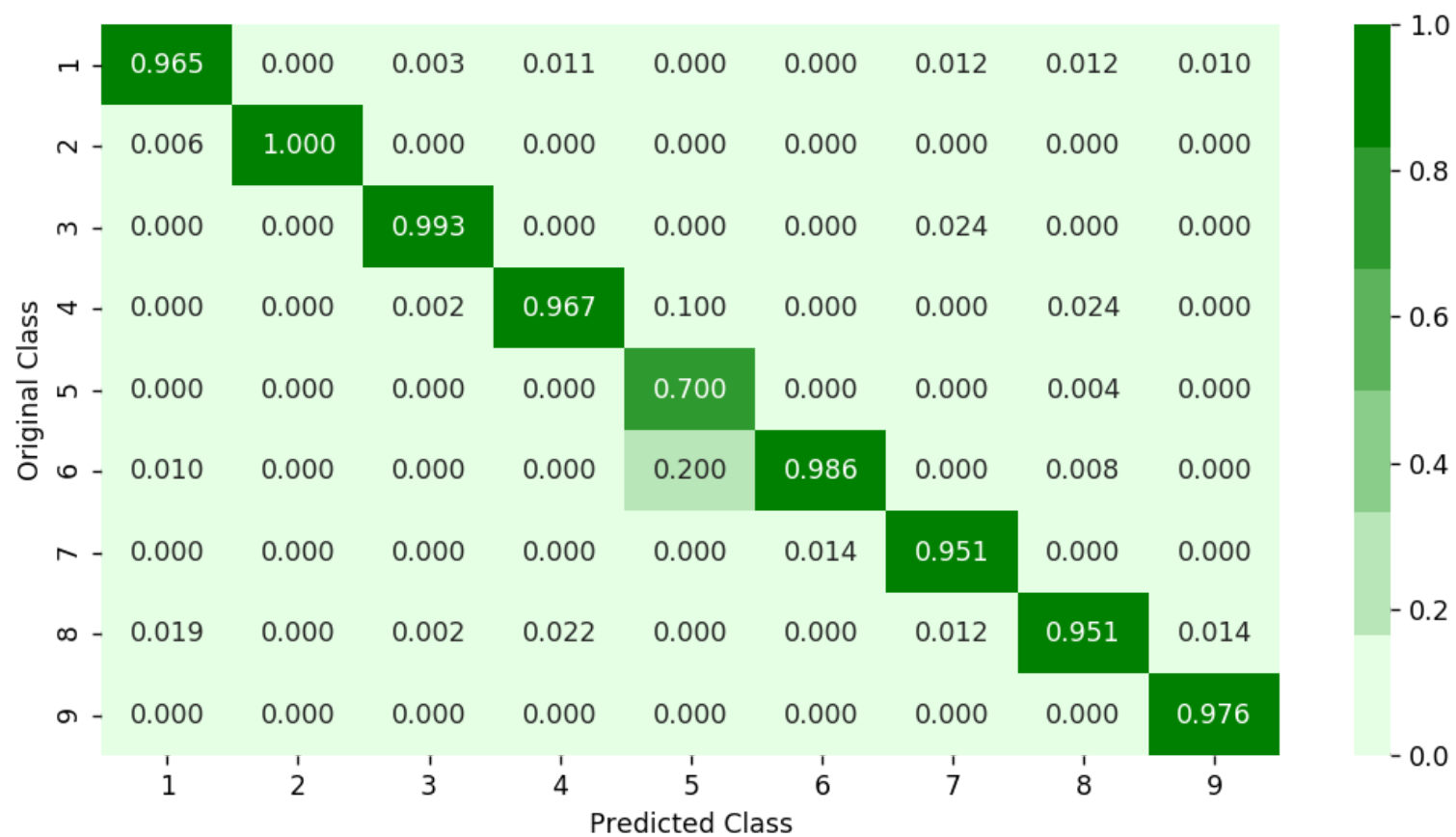


log_loss for train data 0.0476773462198
log_loss for cv data 0.0958800580948
log_loss for test data 0.0894810720832
Number of misclassified points 2.02391904324

----- Confusion matrix -----

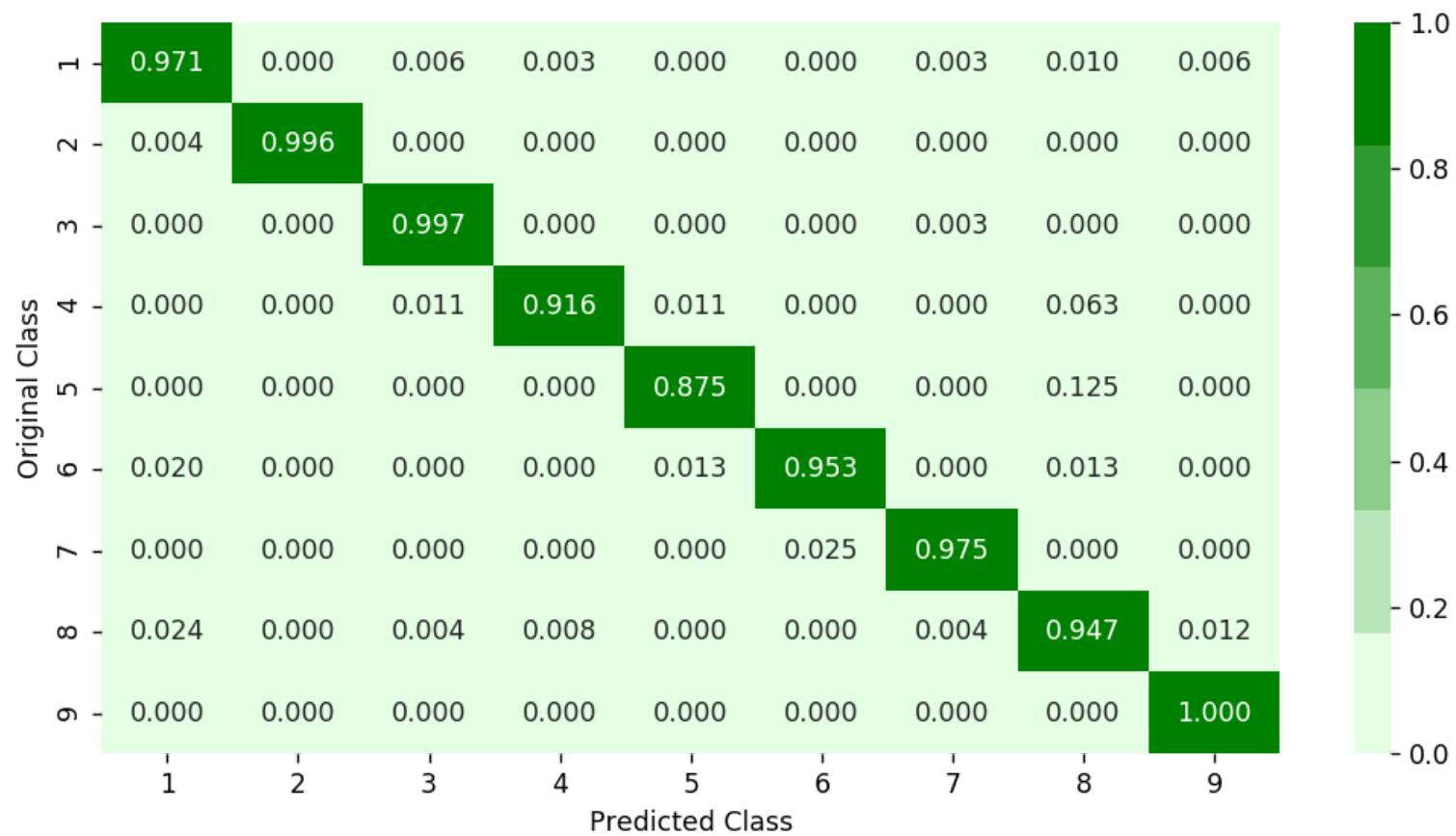


----- Precision matrix -----



Sum of columns in precision matrix [1. 1. 1. 1. 1. 1. 1. 1. 1.]

----- Recall matrix -----



Sum of rows in precision matrix [1. 1. 1. 1. 1. 1. 1. 1. 1.]

4.4.2 Logistic Regression

```

In [ ]: # read more about SGDClassifier() at http://scikit-learn.org/stable/modules/generated/sklearn.linear\_model.SGDClassifier.h
        tml
        # -----
        # default parameters
        # SGDClassifier(loss='hinge', penalty='l2', alpha=0.0001, l1_ratio=0.15, fit_intercept=True, max_iter=None, tol=None,
        # shuffle=True, verbose=0, epsilon=0.1, n_jobs=1, random_state=None, learning_rate='optimal', eta0=0.0, power_t=0.5,
        # class_weight=None, warm_start=False, average=False, n_iter=None)

        # some of methods
        # fit(X, y[, coef_init, intercept_init, ...])      Fit linear model with Stochastic Gradient Descent.
        # predict(X)      Predict class labels for samples in X.

        #-----
        # video link: https://www.appliedaicourse.com/course/applied-ai-course-online/lessons/geometric-intuition-1/
        #-----

alpha = [10 ** x for x in range(-5, 4)]
cv_log_error_array=[]
for i in alpha:
    logisticR=LogisticRegression(penalty='l2',C=i,class_weight='balanced')
    logisticR.fit(X_train_asm,y_train_asm)
    sig_clf = CalibratedClassifierCV(logisticR, method="sigmoid")
    sig_clf.fit(X_train_asm, y_train_asm)
    predict_y = sig_clf.predict_proba(X_cv_asm)
    cv_log_error_array.append(log_loss(y_cv_asm, predict_y, labels=logisticR.classes_, eps=1e-15))

for i in range(len(cv_log_error_array)):
    print ('log_loss for c = ',alpha[i],'is',cv_log_error_array[i])

best_alpha = np.argmin(cv_log_error_array)

fig, ax = plt.subplots()
ax.plot(alpha, cv_log_error_array,c='g')
for i, txt in enumerate(np.round(cv_log_error_array,3)):
    ax.annotate((alpha[i],np.round(txt,3)), (alpha[i],cv_log_error_array[i]))
plt.grid()
plt.title("Cross Validation Error for each alpha")
plt.xlabel("Alpha i's")
plt.ylabel("Error measure")
plt.show()

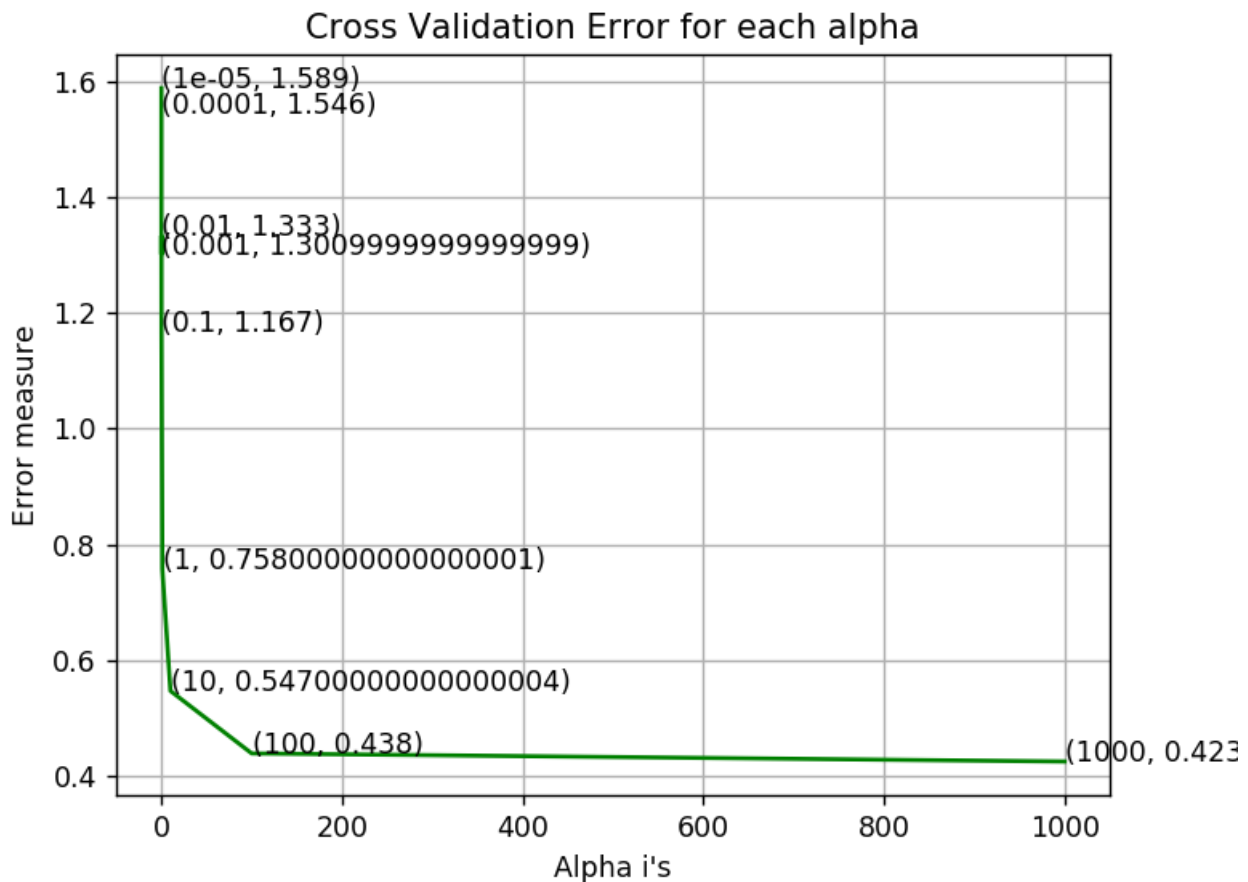
logisticR=LogisticRegression(penalty='l2',C=alpha[best_alpha],class_weight='balanced')
logisticR.fit(X_train_asm,y_train_asm)
sig_clf = CalibratedClassifierCV(logisticR, method="sigmoid")

```

```
sig_clf.fit(X_train_asm, y_train_asm)

predict_y = sig_clf.predict_proba(X_train_asm)
print ('log loss for train data',(log_loss(y_train_asm, predict_y, labels=logisticR.classes_, eps=1e-15)))
predict_y = sig_clf.predict_proba(X_cv_asm)
print ('log loss for cv data',(log_loss(y_cv_asm, predict_y, labels=logisticR.classes_, eps=1e-15)))
predict_y = sig_clf.predict_proba(X_test_asm)
print ('log loss for test data',(log_loss(y_test_asm, predict_y, labels=logisticR.classes_, eps=1e-15)))
plot_confusion_matrix(y_test_asm,sig_clf.predict(X_test_asm))
```

log_loss for c = 1e-05 is 1.58867274165
log_loss for c = 0.0001 is 1.54560797884
log_loss for c = 0.001 is 1.30137786807
log_loss for c = 0.01 is 1.33317456931
log_loss for c = 0.1 is 1.16705751378
log_loss for c = 1 is 0.757667807779
log_loss for c = 10 is 0.546533939819
log_loss for c = 100 is 0.438414998062
log_loss for c = 1000 is 0.424423536526

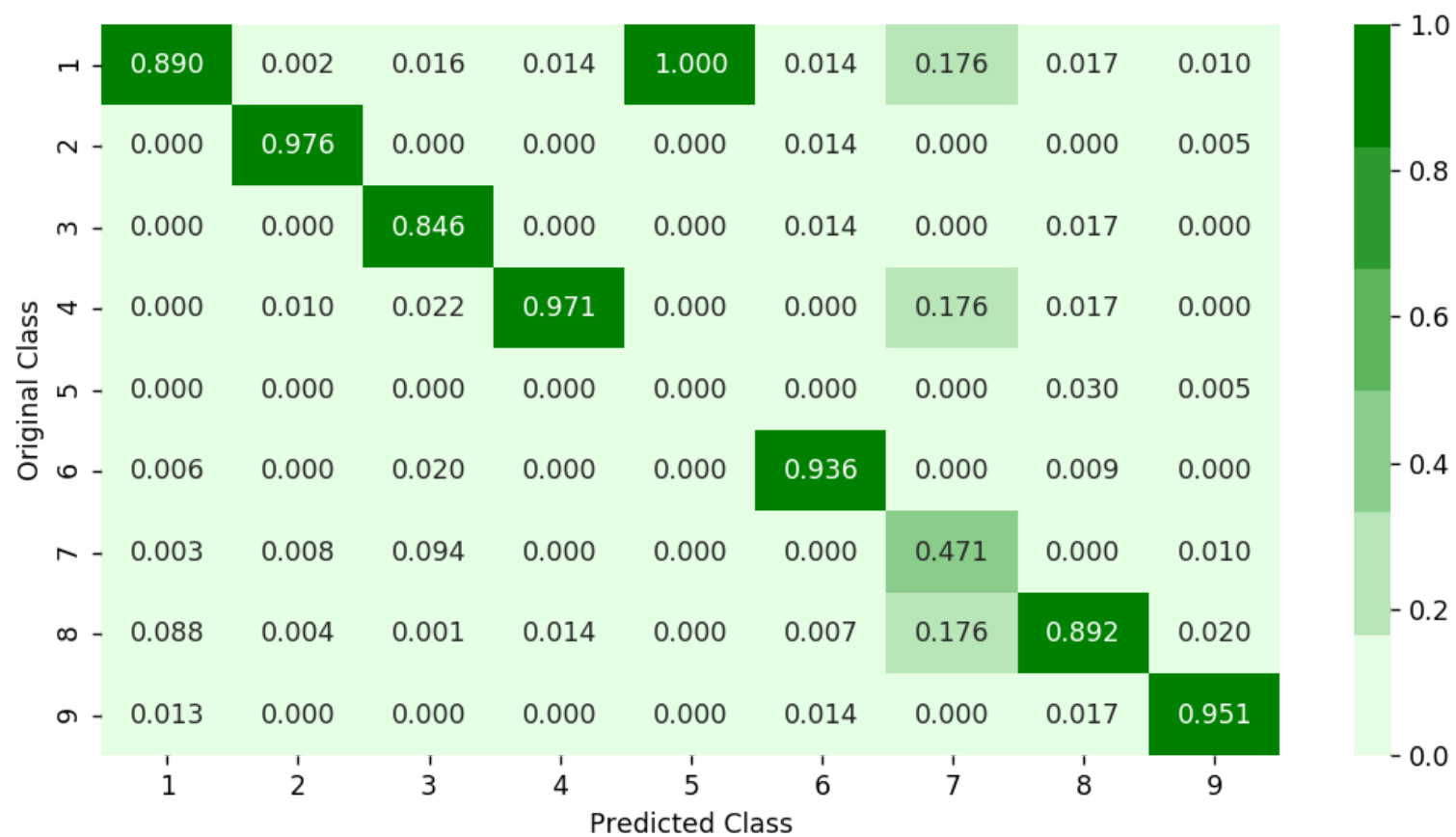


log loss for train data 0.396219394701
log loss for cv data 0.424423536526
log loss for test data 0.415685592517
Number of misclassified points 9.61361545538

----- Confusion matrix -----

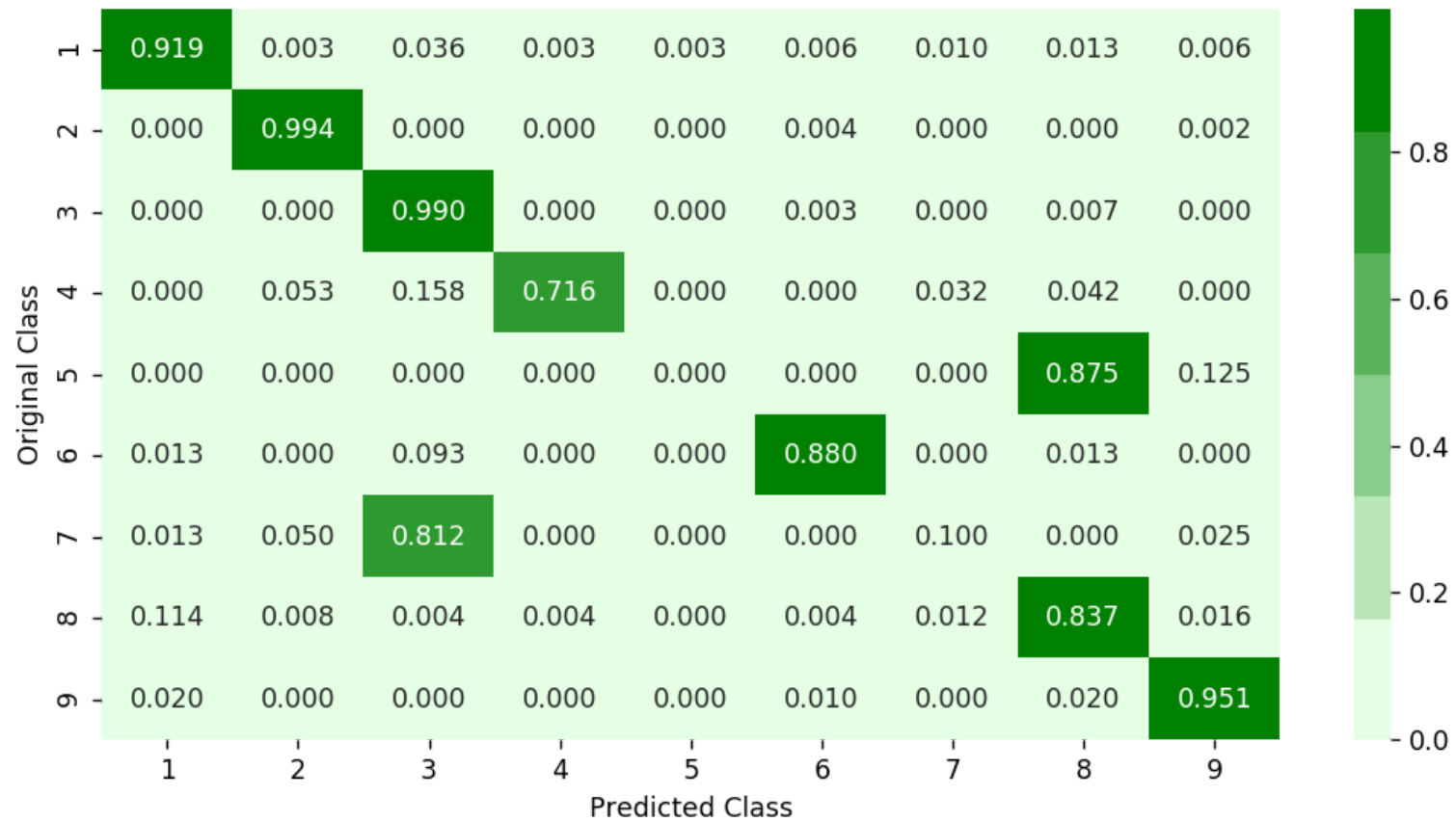


----- Precision matrix -----



Sum of columns in precision matrix [1. 1. 1. 1. 1. 1. 1. 1. 1.]

----- Recall matrix -----



Sum of rows in precision matrix [1. 1. 1. 1. 1. 1. 1. 1. 1.]

4.4.3 Random Forest Classifier

```

In [ ]: # -----
# default parameters
# sklearn.ensemble.RandomForestClassifier(n_estimators=10, criterion='gini', max_depth=None, min_samples_split=2,
# min_samples_leaf=1, min_weight_fraction_leaf=0.0, max_features='auto', max_leaf_nodes=None, min_impurity_decrease=0.0,
# min_impurity_split=None, bootstrap=True, oob_score=False, n_jobs=1, random_state=None, verbose=0, warm_start=False,
# class_weight=None)

# Some of methods of RandomForestClassifier()
# fit(X, y, [sample_weight])    Fit the SVM model according to the given training data.
# predict(X)    Perform classification on samples in X.
# predict_proba (X)    Perform classification on samples in X.

# some of attributes of RandomForestClassifier()
# feature_importances_ : array of shape = [n_features]
# The feature importances (the higher, the more important the feature).

# -----
# video link: https://www.appliedaicourse.com/course/applied-ai-course-online/lessons/random-forest-and-their-construction-2/
# -----

alpha=[10,50,100,500,1000,2000,3000]
cv_log_error_array=[]
for i in alpha:
    r_cfl=RandomForestClassifier(n_estimators=i,random_state=42,n_jobs=-1)
    r_cfl.fit(X_train_asm,y_train_asm)
    sig_clf = CalibratedClassifierCV(r_cfl, method="sigmoid")
    sig_clf.fit(X_train_asm, y_train_asm)
    predict_y = sig_clf.predict_proba(X_cv_asm)
    cv_log_error_array.append(log_loss(y_cv_asm, predict_y, labels=r_cfl.classes_, eps=1e-15))

for i in range(len(cv_log_error_array)):
    print ('log_loss for c = ',alpha[i],'is',cv_log_error_array[i])

best_alpha = np.argmin(cv_log_error_array)

fig, ax = plt.subplots()
ax.plot(alpha, cv_log_error_array,c='g')
for i, txt in enumerate(np.round(cv_log_error_array,3)):
    ax.annotate((alpha[i],np.round(txt,3)), (alpha[i],cv_log_error_array[i]))
plt.grid()
plt.title("Cross Validation Error for each alpha")
plt.xlabel("Alpha i's")
plt.ylabel("Error measure")

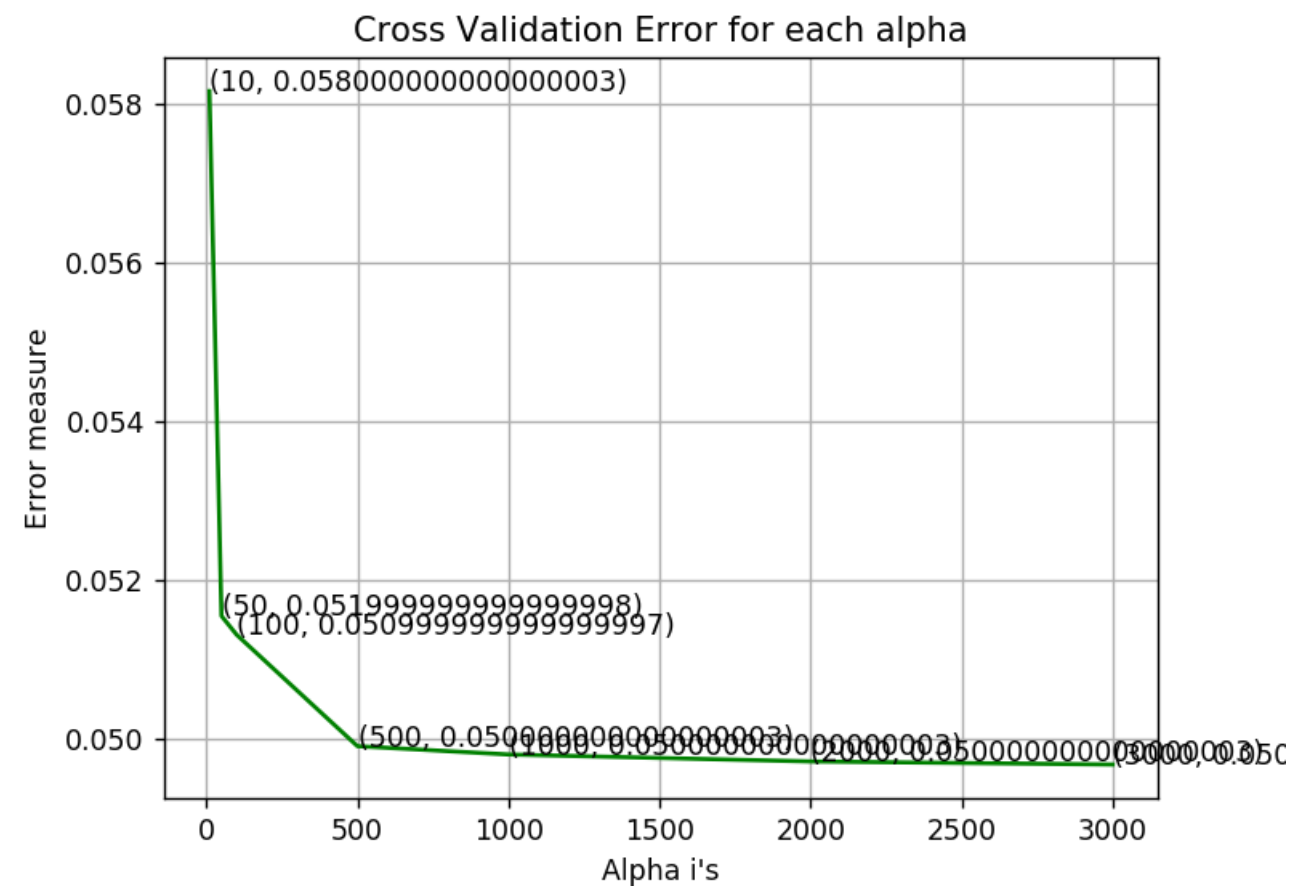
```



```
plt.show()
```

```
r_cfl=RandomForestClassifier(n_estimators=alpha[best_alpha],random_state=42,n_jobs=-1)
r_cfl.fit(X_train_asm,y_train_asm)
sig_clf = CalibratedClassifierCV(r_cfl, method="sigmoid")
sig_clf.fit(X_train_asm, y_train_asm)
predict_y = sig_clf.predict_proba(X_train_asm)
print ('log loss for train data',(log_loss(y_train_asm, predict_y, labels=sig_clf.classes_, eps=1e-15)))
predict_y = sig_clf.predict_proba(X_cv_asm)
print ('log loss for cv data',(log_loss(y_cv_asm, predict_y, labels=sig_clf.classes_, eps=1e-15)))
predict_y = sig_clf.predict_proba(X_test_asm)
print ('log loss for test data',(log_loss(y_test_asm, predict_y, labels=sig_clf.classes_, eps=1e-15)))
plot_confusion_matrix(y_test_asm,sig_clf.predict(X_test_asm))
```

log_loss for c = 10 is 0.0581657906023
log_loss for c = 50 is 0.0515443148419
log_loss for c = 100 is 0.0513084973231
log_loss for c = 500 is 0.0499021761479
log_loss for c = 1000 is 0.0497972474298
log_loss for c = 2000 is 0.0497091690815
log_loss for c = 3000 is 0.0496706817633

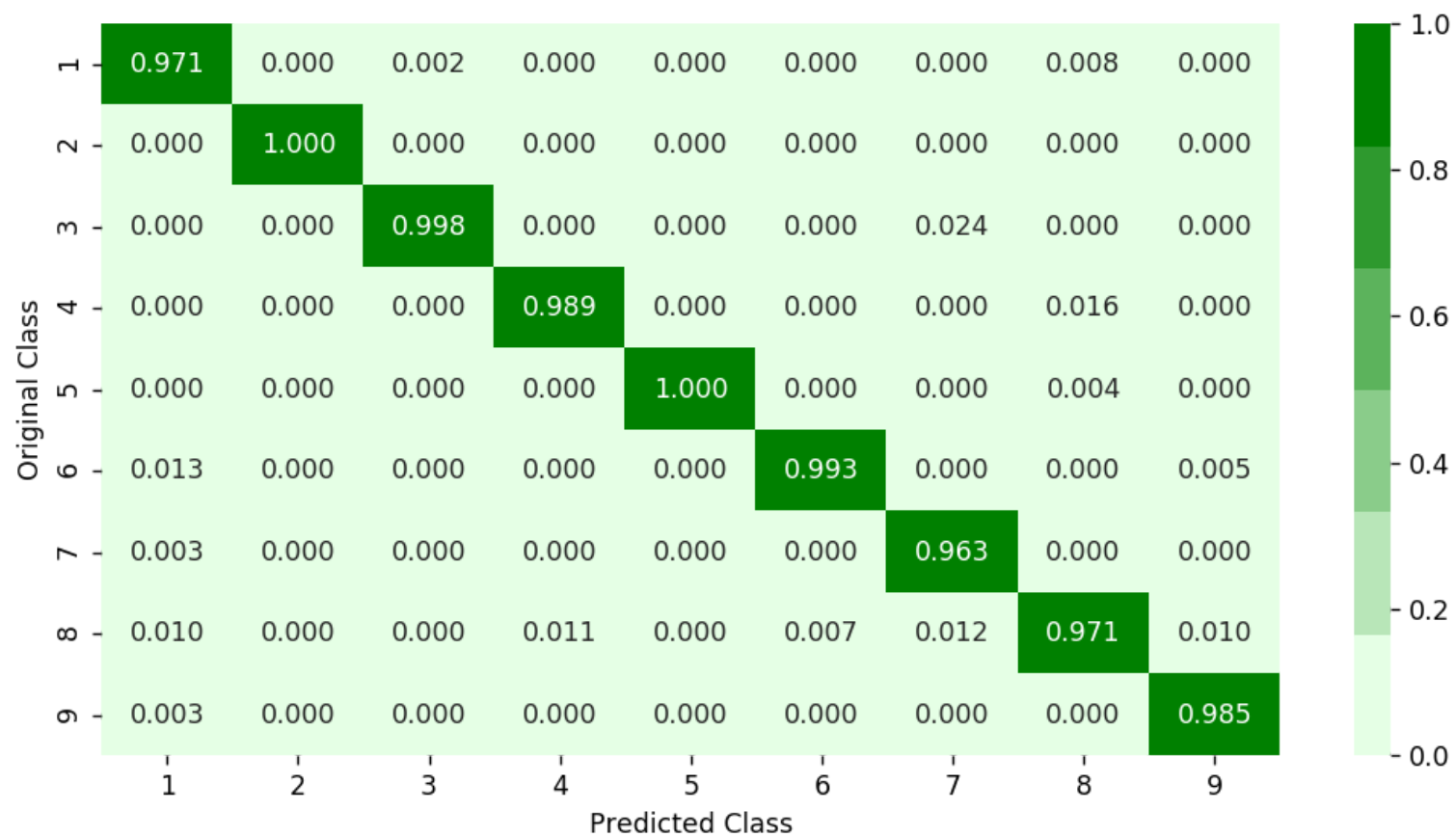


log loss for train data 0.0116517052676
log loss for cv data 0.0496706817633
log loss for test data 0.0571239496453
Number of misclassified points 1.14995400184

----- Confusion matrix -----

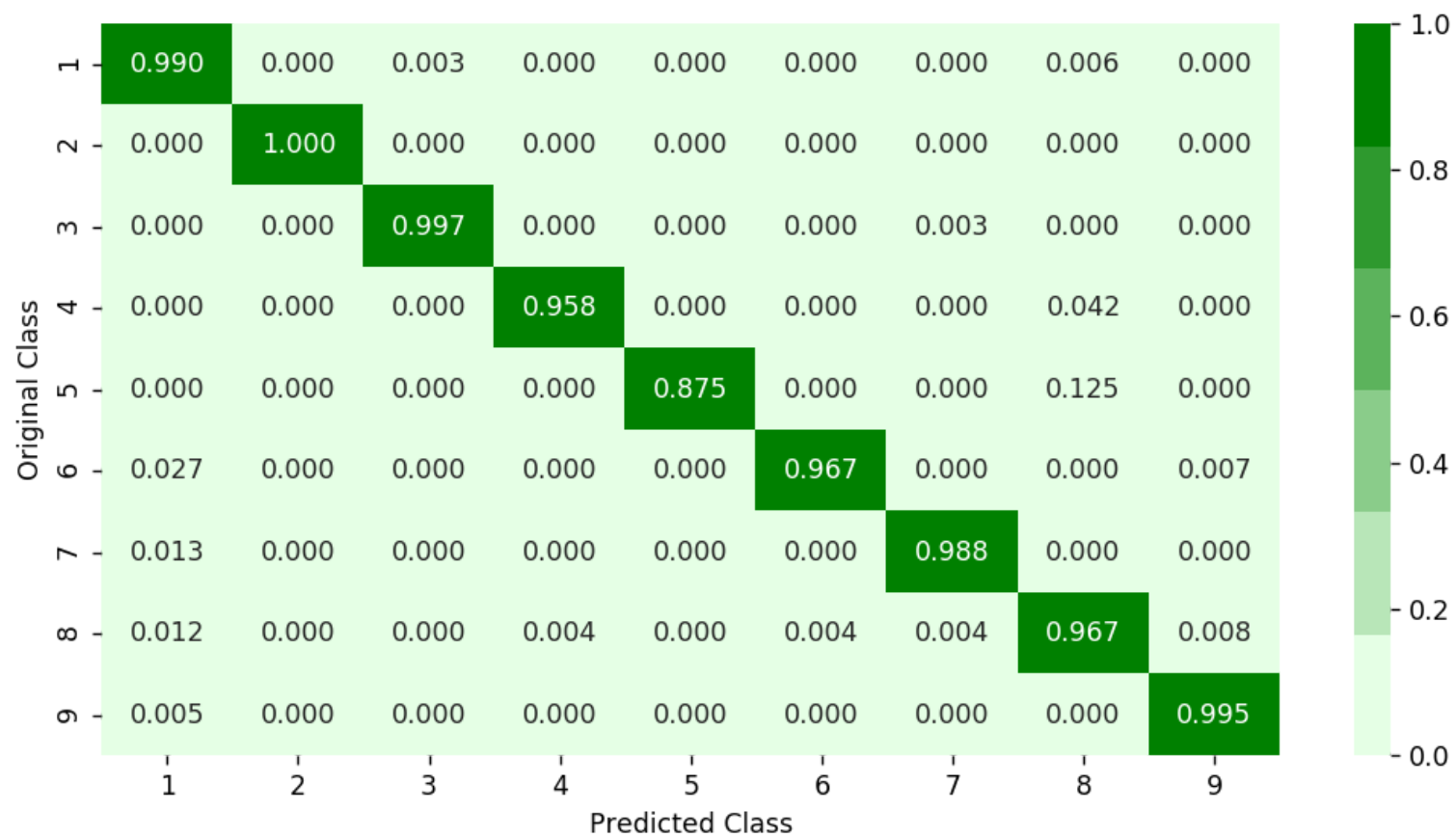


Precision matrix



Sum of columns in precision matrix [1. 1. 1. 1. 1. 1. 1. 1. 1.]

----- Recall matrix -----



Sum of rows in precision matrix [1. 1. 1. 1. 1. 1. 1. 1. 1.]

4.4.4 XgBoost Classifier

```

In [ ]: # Training a hyper-parameter tuned Xg-Boost regressor on our train data

# find more about XGBClassifier function here http://xgboost.readthedocs.io/en/latest/python/python\_api.html?#xgboost.XGBClassifier
# -----
# default paramters
# class xgboost.XGBClassifier(max_depth=3, learning_rate=0.1, n_estimators=100, silent=True,
# objective='binary:logistic', booster='gbtree', n_jobs=1, nthread=None, gamma=0, min_child_weight=1,
# max_delta_step=0, subsample=1, colsample_bytree=1, colsample_bylevel=1, reg_alpha=0, reg_lambda=1,
# scale_pos_weight=1, base_score=0.5, random_state=0, seed=None, missing=None, **kwargs)

# some of methods of RandomForestRegressor()
# fit(X, y, sample_weight=None, eval_set=None, eval_metric=None, early_stopping_rounds=None, verbose=True, xgb_model=None)
# get_params([deep]) Get parameters for this estimator.
# predict(data, output_margin=False, ntree_limit=0) : Predict with data. NOTE: This function is not thread safe.
# get_score(importance_type='weight') -> get the feature importance
# -----
# video link2: https://www.appliedaicourse.com/course/applied-ai-course-online/lessons/what-are-ensembles/
# -----

alpha=[10,50,100,500,1000,2000,3000]
cv_log_error_array=[]
for i in alpha:
    x_cfl=XGBClassifier(n_estimators=i,nthread=-1)
    x_cfl.fit(X_train_asm,y_train_asm)
    sig_clf = CalibratedClassifierCV(x_cfl, method="sigmoid")
    sig_clf.fit(X_train_asm, y_train_asm)
    predict_y = sig_clf.predict_proba(X_cv_asm)
    cv_log_error_array.append(log_loss(y_cv_asm, predict_y, labels=x_cfl.classes_, eps=1e-15))

for i in range(len(cv_log_error_array)):
    print ('log_loss for c = ',alpha[i],'is',cv_log_error_array[i])

best_alpha = np.argmin(cv_log_error_array)

fig, ax = plt.subplots()
ax.plot(alpha, cv_log_error_array,c='g')
for i, txt in enumerate(np.round(cv_log_error_array,3)):
    ax.annotate((alpha[i],np.round(txt,3)), (alpha[i],cv_log_error_array[i]))
plt.grid()
plt.title("Cross Validation Error for each alpha")
plt.xlabel("Alpha i's")
plt.ylabel("Error measure")
plt.show()

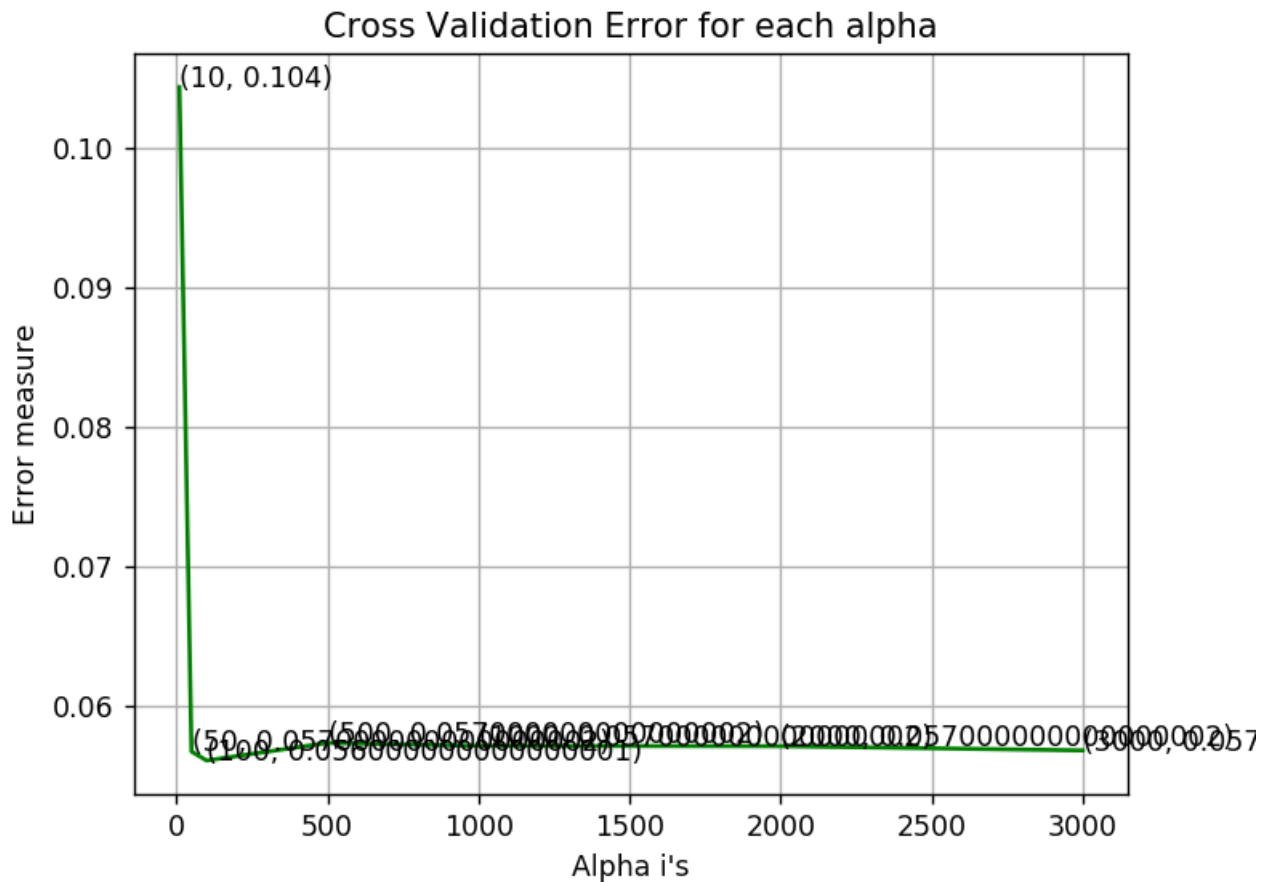
```

```
x_cfl=XGBClassifier(n_estimators=alpha[best_alpha],nthread=-1)
x_cfl.fit(X_train_asm,y_train_asm)
sig_clf = CalibratedClassifierCV(x_cfl, method="sigmoid")
sig_clf.fit(X_train_asm, y_train_asm)

predict_y = sig_clf.predict_proba(X_train_asm)

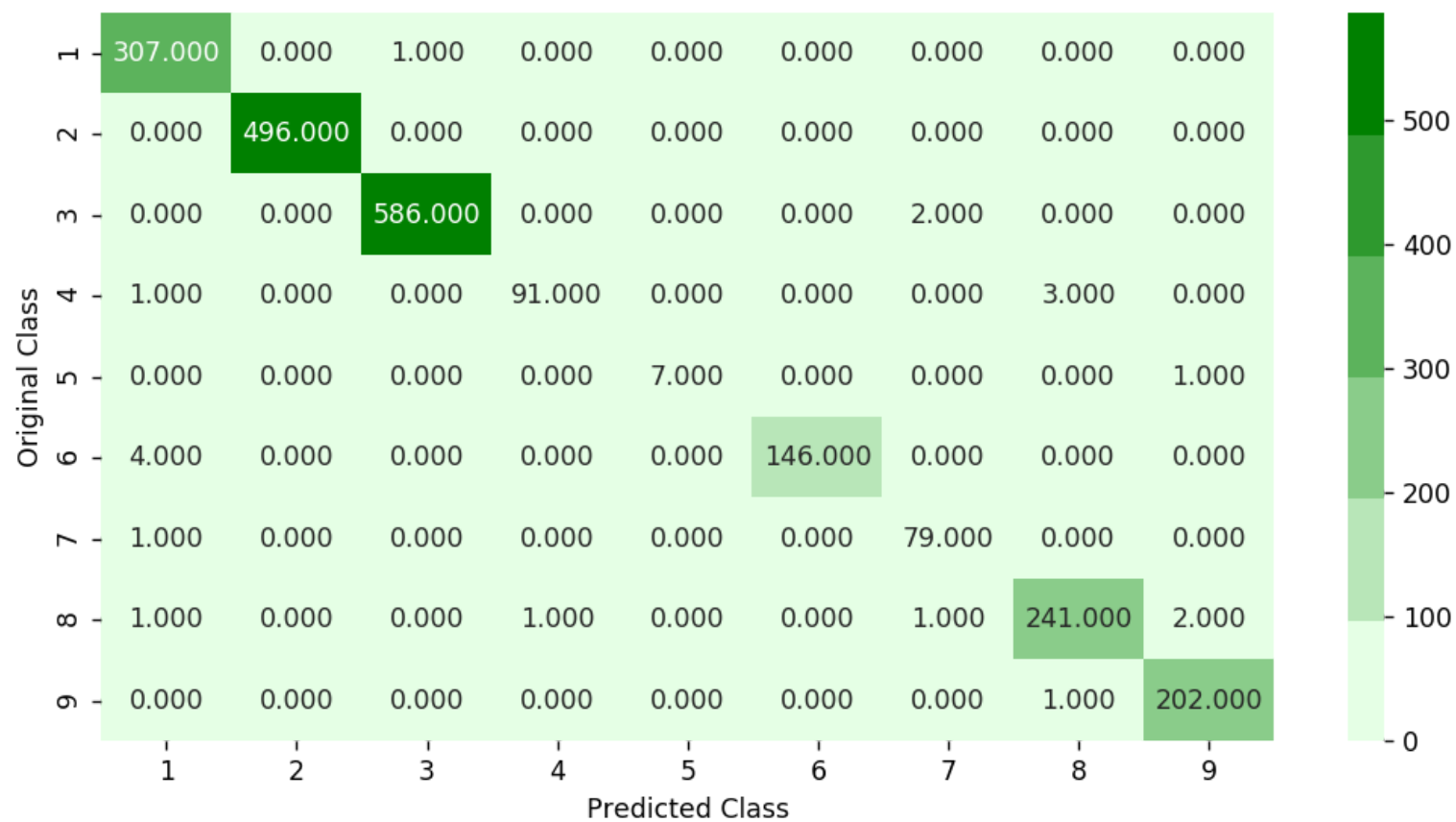
print ('For values of best alpha = ', alpha[best_alpha], "The train log loss is:",log_loss(y_train_asm, predict_y))
predict_y = sig_clf.predict_proba(X_cv_asm)
print('For values of best alpha = ', alpha[best_alpha], "The cross validation log loss is:",log_loss(y_cv_asm, predict_y))
predict_y = sig_clf.predict_proba(X_test_asm)
print('For values of best alpha = ', alpha[best_alpha], "The test log loss is:",log_loss(y_test_asm, predict_y))
plot_confusion_matrix(y_test_asm,sig_clf.predict(X_test_asm))
```

log_loss for c = 10 is 0.104344888454
log_loss for c = 50 is 0.0567190635611
log_loss for c = 100 is 0.056075038646
log_loss for c = 500 is 0.057336051683
log_loss for c = 1000 is 0.0571265109903
log_loss for c = 2000 is 0.057103406781
log_loss for c = 3000 is 0.0567993215778

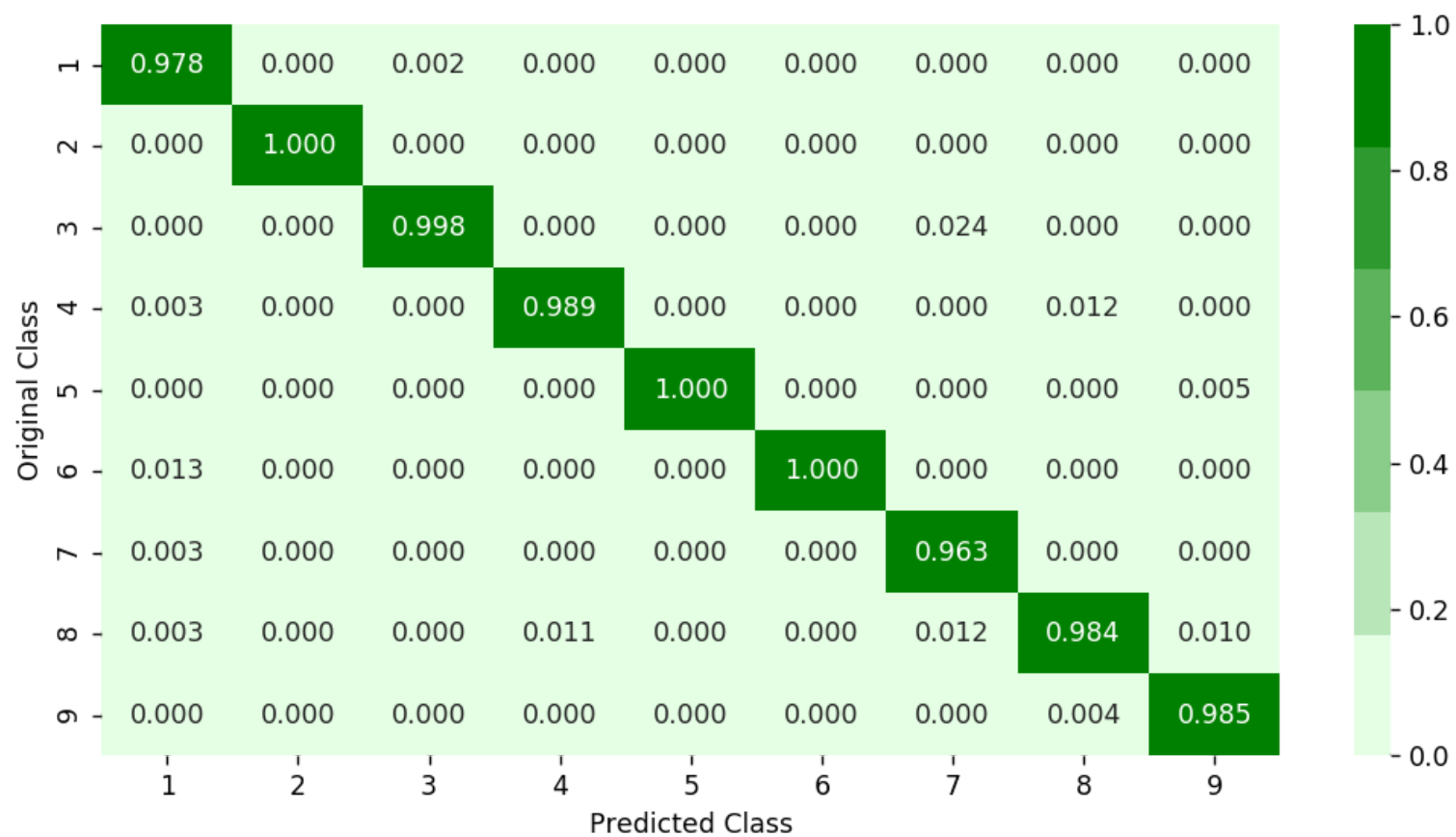


For values of best alpha = 100 The train log loss is: 0.0117883742574
For values of best alpha = 100 The cross validation log loss is: 0.056075038646
For values of best alpha = 100 The test log loss is: 0.0491647763845
Number of misclassified points 0.873965041398

----- Confusion matrix -----

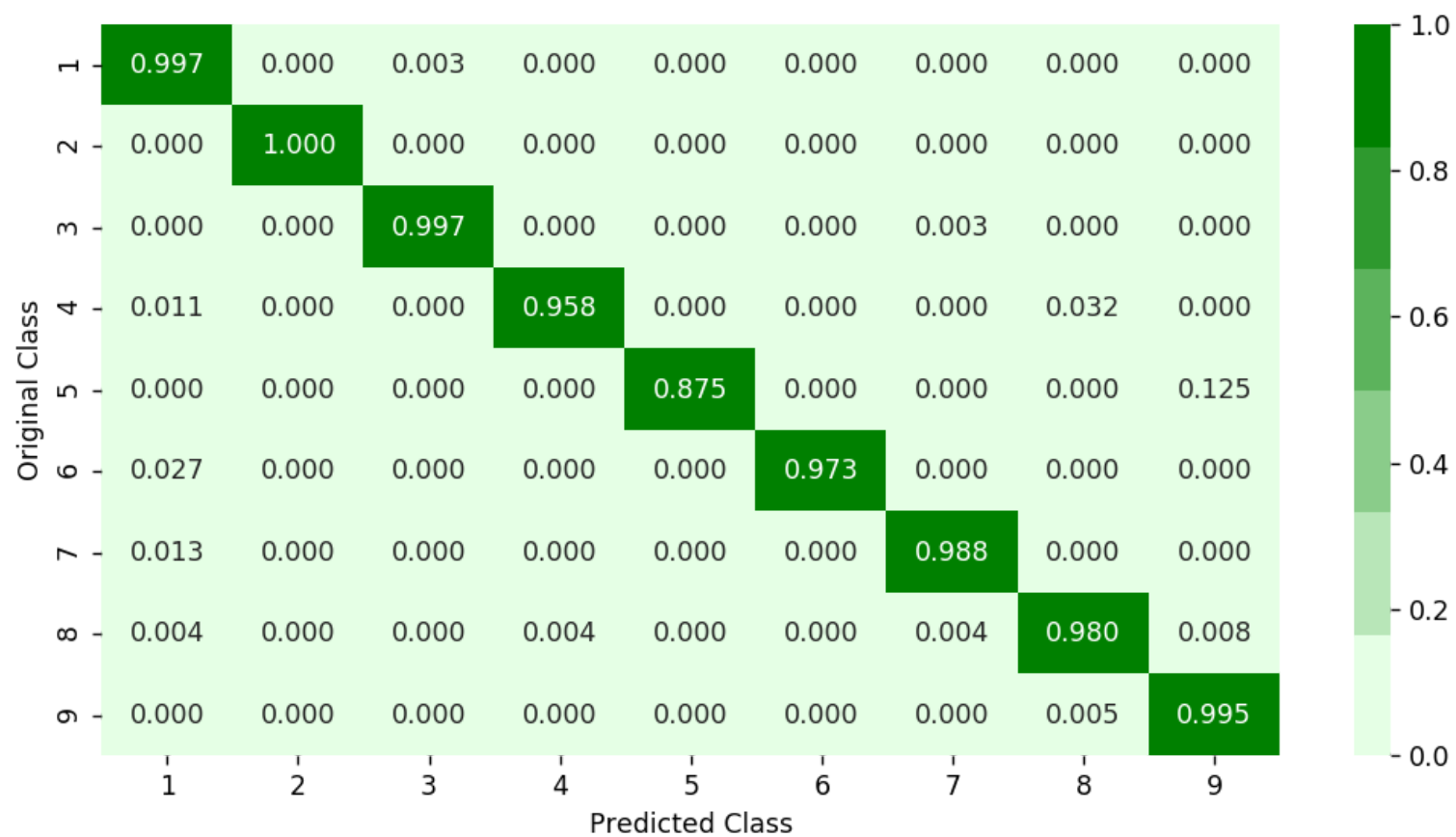


Precision matrix



Sum of columns in precision matrix [1. 1. 1. 1. 1. 1. 1. 1. 1.]

----- Recall matrix -----



Sum of rows in precision matrix [1. 1. 1. 1. 1. 1. 1. 1. 1.]

4.4.5 Xgboost Classifier with best hyperparameters

```
In [ ]: x_cfl=XGBClassifier()

prams={
    'learning_rate':[0.01,0.03,0.05,0.1,0.15,0.2],
    'n_estimators':[100,200,500,1000,2000],
    'max_depth':[3,5,10],
    'colsample_bytree':[0.1,0.3,0.5,1],
    'subsample':[0.1,0.3,0.5,1]
}
random_cfl=RandomizedSearchCV(x_cfl,param_distributions=prams,verbose=10,n_jobs=-1,)
random_cfl.fit(X_train_asm,y_train_asm)
```

Fitting 3 folds for each of 10 candidates, totalling 30 fits

```
[Parallel(n_jobs=-1)]: Done    2 tasks      | elapsed:    8.1s
[Parallel(n_jobs=-1)]: Done    9 tasks      | elapsed:   32.8s
[Parallel(n_jobs=-1)]: Done   19 out of  30 | elapsed:   1.1min remaining:   39.3s
[Parallel(n_jobs=-1)]: Done   23 out of  30 | elapsed:   1.3min remaining:   23.0s
[Parallel(n_jobs=-1)]: Done   27 out of  30 | elapsed:   1.4min remaining:    9.2s
[Parallel(n_jobs=-1)]: Done   30 out of  30 | elapsed:   2.3min finished
```

```
Out[ ]: RandomizedSearchCV(cv=None, error_score='raise',
    estimator=XGBClassifier(base_score=0.5, colsample_bylevel=1, colsample_bytree=1,
    gamma=0, learning_rate=0.1, max_delta_step=0, max_depth=3,
    min_child_weight=1, missing=None, n_estimators=100, nthread=-1,
    objective='binary:logistic', reg_alpha=0, reg_lambda=1,
    scale_pos_weight=1, seed=0, silent=True, subsample=1),
    fit_params=None, iid=True, n_iter=10, n_jobs=-1,
    param_distributions={'learning_rate': [0.01, 0.03, 0.05, 0.1, 0.15, 0.2], 'n_estimators': [100, 200, 500, 1000, 2000], 'max_depth': [3, 5, 10], 'colsample_bytree': [0.1, 0.3, 0.5, 1], 'subsample': [0.1, 0.3, 0.5, 1]},
    pre_dispatch='2*n_jobs', random_state=None, refit=True,
    return_train_score=True, scoring=None, verbose=10)
```

```
In [ ]: print (random_cfl.best_params_)

{'subsample': 1, 'n_estimators': 200, 'max_depth': 5, 'learning_rate': 0.15, 'colsample_bytree': 0.5}
```

```

In [ ]: # Training a hyper-parameter tuned Xg-Boost regressor on our train data

# find more about XGBClassifier function here http://xgboost.readthedocs.io/en/latest/python/python\_api.html?#xgboost.XGBClassifier
# -----
# default paramters
# class xgboost.XGBClassifier(max_depth=3, learning_rate=0.1, n_estimators=100, silent=True,
# objective='binary:logistic', booster='gbtree', n_jobs=1, nthread=None, gamma=0, min_child_weight=1,
# max_delta_step=0, subsample=1, colsample_bytree=1, colsample_bylevel=1, reg_alpha=0, reg_lambda=1,
# scale_pos_weight=1, base_score=0.5, random_state=0, seed=None, missing=None, **kwargs)

# some of methods of RandomForestRegressor()
# fit(X, y, sample_weight=None, eval_set=None, eval_metric=None, early_stopping_rounds=None, verbose=True, xgb_model=None)
# get_params([deep]) Get parameters for this estimator.
# predict(data, output_margin=False, ntree_limit=0) : Predict with data. NOTE: This function is not thread safe.
# get_score(importance_type='weight') -> get the feature importance
# -----
# video link2: https://www.appliedaicourse.com/course/applied-ai-course-online/lessons/what-are-ensembles/
# -----

x_cfl=XGBClassifier(n_estimators=200,subsample=0.5,learning_rate=0.15,colsample_bytree=0.5,max_depth=3)
x_cfl.fit(X_train_asm,y_train_asm)
c_cfl=CalibratedClassifierCV(x_cfl,method='sigmoid')
c_cfl.fit(X_train_asm,y_train_asm)

predict_y = c_cfl.predict_proba(X_train_asm)
print ('train loss',log_loss(y_train_asm, predict_y))
predict_y = c_cfl.predict_proba(X_cv_asm)
print ('cv loss',log_loss(y_cv_asm, predict_y))
predict_y = c_cfl.predict_proba(X_test_asm)
print ('test loss',log_loss(y_test_asm, predict_y))

train loss 0.0102661325822
cv loss 0.0501201796687
test loss 0.0483908764397

```

4.5. Machine Learning models on features of both .asm and .bytes files

4.5.1. Merging both asm and byte file features

```
In [ ]: result.head()
```

Out[]:

	ID	0	1	2	3	4	5	6	7	8	...	f9	fa	fb
0	01azqd4InC7m9JpocGv5	0.262806	0.005498	0.001567	0.002067	0.002048	0.001835	0.002058	0.002946	0.002638	...	0.013560	0.013107	0.013634
1	01IsoiSMh5gxyDYTI4CB	0.017358	0.011737	0.004033	0.003876	0.005303	0.003873	0.004747	0.006984	0.008267	...	0.001920	0.001147	0.001329
2	01jsnpXSAIgw6aPeDxrU	0.040827	0.013434	0.001429	0.001315	0.005464	0.005280	0.005078	0.002155	0.008104	...	0.009804	0.011777	0.012604
3	01kcPWA9K2BOxQeS5Rju	0.009209	0.001708	0.000404	0.000441	0.000770	0.000354	0.000310	0.000481	0.000959	...	0.002121	0.001886	0.002272
4	01SuzwMJEIXsK7A8dQbl	0.008629	0.001000	0.000168	0.000234	0.000342	0.000232	0.000148	0.000229	0.000376	...	0.001530	0.000853	0.001052

5 rows × 260 columns

```
In [ ]: result_asm.head()
```

Out[]:

	ID	HEADER:	.text:	.Pav:	.idata:	.data:	.bss:	.rdata:	.edata:	.rsrc:	...	esi	eax	ebx	ecx	ed
0	01kcPWA9K2BOxQeS5Rju	0.107345	0.001092	0.0	0.000761	0.000023	0.0	0.000084	0.0	0.000072	...	0.000746	0.000301	0.000360	0.001057	0.0
1	1E93CpP60RHFNiT5Qfvn	0.096045	0.001230	0.0	0.000617	0.000019	0.0	0.000000	0.0	0.000072	...	0.000328	0.000965	0.000686	0.000153	0.0
2	3ekVow2ajZHbTnBcsDfX	0.096045	0.000627	0.0	0.000300	0.000017	0.0	0.000038	0.0	0.000072	...	0.000475	0.000201	0.000560	0.000178	0.0
3	3X2nY7iQaPBIWDrAZqJe	0.096045	0.000333	0.0	0.000258	0.000008	0.0	0.000000	0.0	0.000072	...	0.000090	0.000281	0.000059	0.000025	0.0
4	46OZzdsSKDCFV8h7XWxf	0.096045	0.000590	0.0	0.000353	0.000068	0.0	0.000000	0.0	0.000072	...	0.000102	0.000362	0.000243	0.000064	0.0

5 rows × 54 columns

```
In [ ]: print(result.shape)
print(result_asm.shape)

(10868, 260)
(10868, 54)
```

```
In [ ]: result_x = pd.merge(result,result_asm.drop(['Class'], axis=1),on='ID', how='left')
result_y = result_x['Class']
result_x = result_x.drop(['ID','rtn','.BSS:', '.CODE', 'Class'], axis=1)
result_x.head()
```

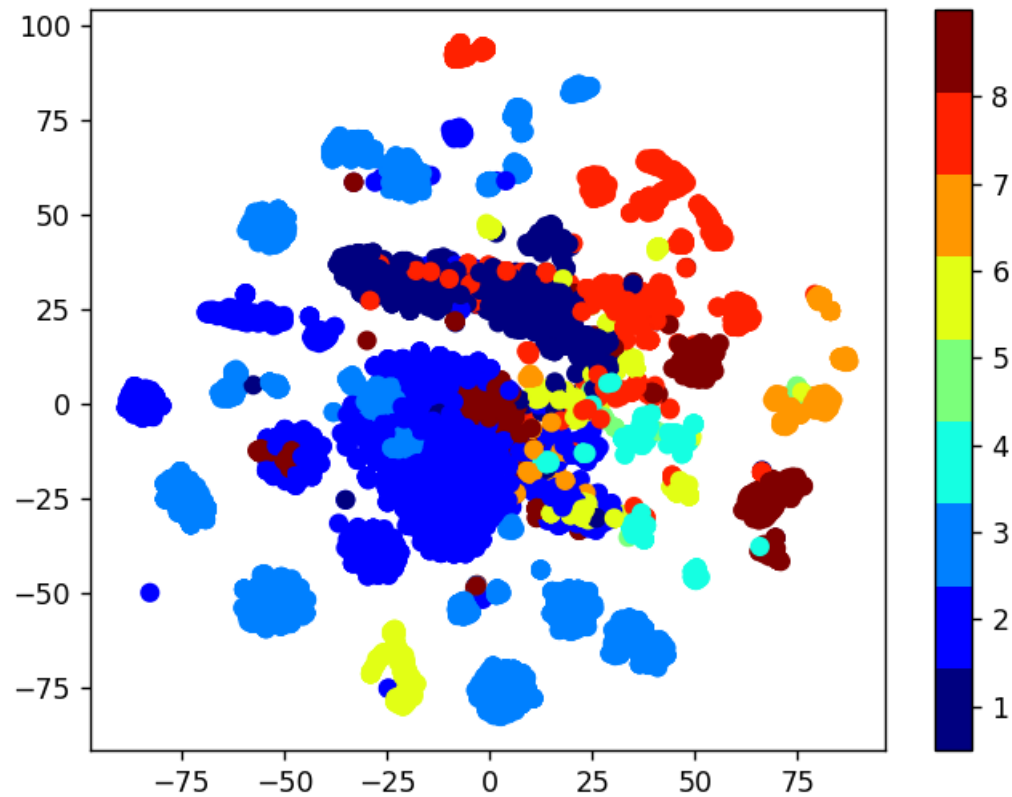
Out[]:

	0	1	2	3	4	5	6	7	8	9	...	edx	esi	eax	ebx	ecx
0	0.262806	0.005498	0.001567	0.002067	0.002048	0.001835	0.002058	0.002946	0.002638	0.003531	...	0.015418	0.025875	0.025744	0.004910	0.008930
1	0.017358	0.011737	0.004033	0.003876	0.005303	0.003873	0.004747	0.006984	0.008267	0.000394	...	0.004961	0.012316	0.007858	0.007570	0.005350
2	0.040827	0.013434	0.001429	0.001315	0.005464	0.005280	0.005078	0.002155	0.008104	0.002707	...	0.000095	0.006181	0.000100	0.003773	0.000713
3	0.009209	0.001708	0.000404	0.000441	0.000770	0.000354	0.000310	0.000481	0.000959	0.000521	...	0.000343	0.000746	0.000301	0.000360	0.001057
4	0.008629	0.001000	0.000168	0.000234	0.000342	0.000232	0.000148	0.000229	0.000376	0.000246	...	0.000343	0.013875	0.000482	0.012932	0.001363

5 rows × 307 columns

4.5.2. Multivariate Analysis on final fearures

```
In [ ]: xtsne=TSNE(perplexity=50)
results=xtsne.fit_transform(result_x, axis=1))
vis_x = results[:, 0]
vis_y = results[:, 1]
plt.scatter(vis_x, vis_y, c=result_y, cmap=plt.cm.get_cmap("jet", 9))
plt.colorbar(ticks=range(9))
plt.clim(0.5, 9)
plt.show()
```



4.5.3. Train and Test split


```
In [ ]: X_train, X_test_merge, y_train, y_test_merge = train_test_split(result_x, result_y, stratify=result_y, test_size=0.20)
X_train_merge, X_cv_merge, y_train_merge, y_cv_merge = train_test_split(X_train, y_train, stratify=y_train, test_size=0.20)
```

4.5.4. Random Forest Classifier on final features

```

In [ ]: # -----
# default parameters
# sklearn.ensemble.RandomForestClassifier(n_estimators=10, criterion='gini', max_depth=None, min_samples_split=2,
# min_samples_leaf=1, min_weight_fraction_leaf=0.0, max_features='auto', max_leaf_nodes=None, min_impurity_decrease=0.0,
# min_impurity_split=None, bootstrap=True, oob_score=False, n_jobs=1, random_state=None, verbose=0, warm_start=False,
# class_weight=None)

# Some of methods of RandomForestClassifier()
# fit(X, y, [sample_weight])    Fit the SVM model according to the given training data.
# predict(X)    Perform classification on samples in X.
# predict_proba (X)    Perform classification on samples in X.

# some of attributes of RandomForestClassifier()
# feature_importances_ : array of shape = [n_features]
# The feature importances (the higher, the more important the feature).

# -----
# video link: https://www.appliedaicourse.com/course/applied-ai-course-online/lessons/random-forest-and-their-construction-2/
# -----

alpha=[10,50,100,500,1000,2000,3000]
cv_log_error_array=[]
from sklearn.ensemble import RandomForestClassifier
for i in alpha:
    r_cfl=RandomForestClassifier(n_estimators=i,random_state=42,n_jobs=-1)
    r_cfl.fit(X_train_merge,y_train_merge)
    sig_clf = CalibratedClassifierCV(r_cfl, method="sigmoid")
    sig_clf.fit(X_train_merge, y_train_merge)
    predict_y = sig_clf.predict_proba(X_cv_merge)
    cv_log_error_array.append(log_loss(y_cv_merge, predict_y, labels=r_cfl.classes_, eps=1e-15))

for i in range(len(cv_log_error_array)):
    print ('log_loss for c = ',alpha[i],'is',cv_log_error_array[i])

best_alpha = np.argmin(cv_log_error_array)

fig, ax = plt.subplots()
ax.plot(alpha, cv_log_error_array,c='g')
for i, txt in enumerate(np.round(cv_log_error_array,3)):
    ax.annotate((alpha[i],np.round(txt,3)), (alpha[i],cv_log_error_array[i]))
plt.grid()
plt.title("Cross Validation Error for each alpha")
plt.xlabel("Alpha i's")

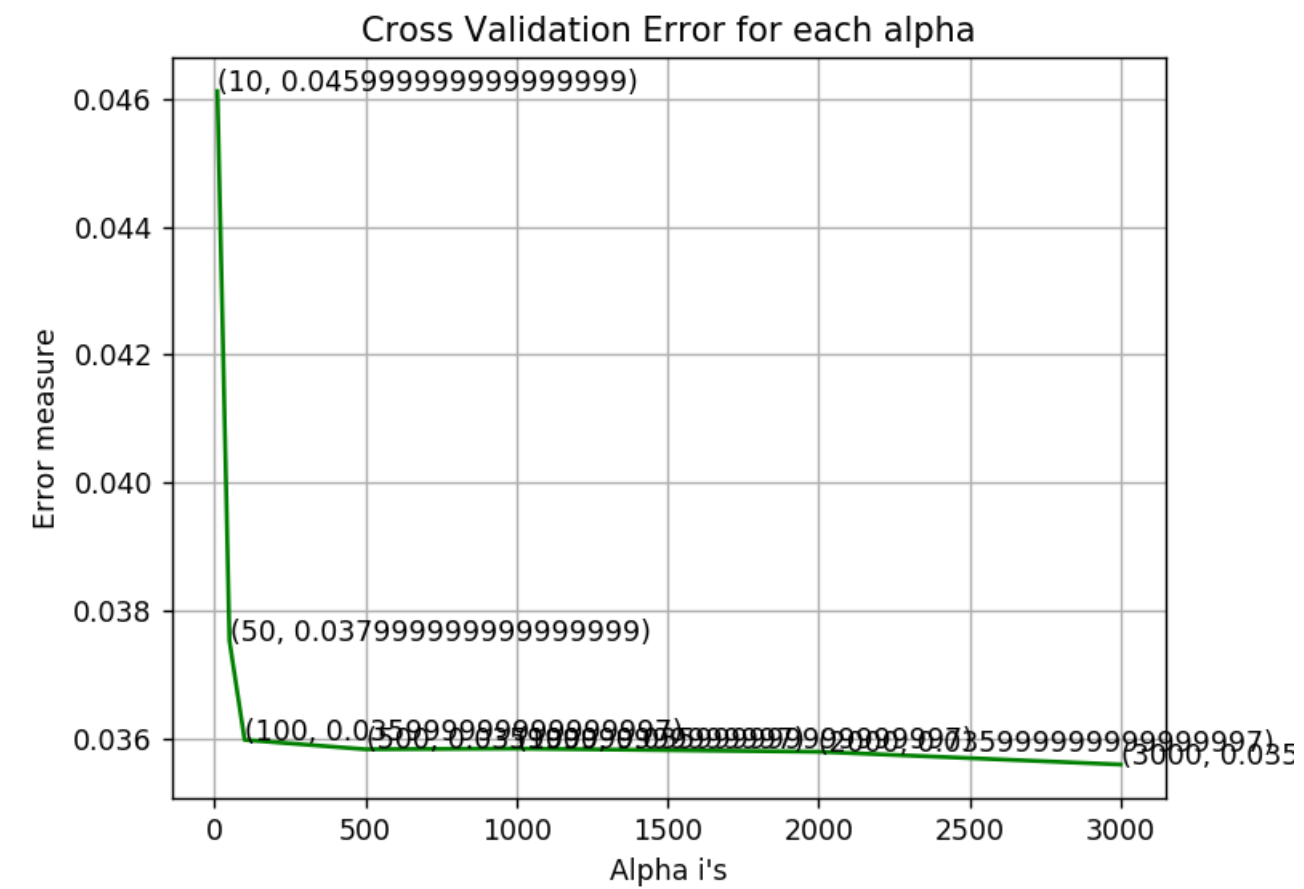
```

```
plt.ylabel("Error measure")
plt.show()

r_cfl=RandomForestClassifier(n_estimators=alpha[best_alpha],random_state=42,n_jobs=-1)
r_cfl.fit(X_train_merge,y_train_merge)
sig_clf = CalibratedClassifierCV(r_cfl, method="sigmoid")
sig_clf.fit(X_train_merge, y_train_merge)

predict_y = sig_clf.predict_proba(X_train_merge)
print ('For values of best alpha = ', alpha[best_alpha], "The train log loss is:",log_loss(y_train_merge, predict_y))
predict_y = sig_clf.predict_proba(X_cv_merge)
print('For values of best alpha = ', alpha[best_alpha], "The cross validation log loss is:",log_loss(y_cv_merge, predict_y
))
predict_y = sig_clf.predict_proba(X_test_merge)
print('For values of best alpha = ', alpha[best_alpha], "The test log loss is:",log_loss(y_test_merge, predict_y))
```

log_loss for c = 10 is 0.0461221662017
log_loss for c = 50 is 0.0375229563452
log_loss for c = 100 is 0.0359765822455
log_loss for c = 500 is 0.0358291883873
log_loss for c = 1000 is 0.0358403093496
log_loss for c = 2000 is 0.0357908022178
log_loss for c = 3000 is 0.0355909487962



For values of best alpha = 3000 The train log loss is: 0.0166267614753
For values of best alpha = 3000 The cross validation log loss is: 0.0355909487962
For values of best alpha = 3000 The test log loss is: 0.0401141303589

4.5.5. XgBoost Classifier on final features

```

In [ ]: # Training a hyper-parameter tuned Xg-Boost regressor on our train data

# find more about XGBClassifier function here http://xgboost.readthedocs.io/en/latest/python/python\_api.html?#xgboost.XGBClassifier
# -----
# default paramters
# class xgboost.XGBClassifier(max_depth=3, learning_rate=0.1, n_estimators=100, silent=True,
# objective='binary:logistic', booster='gbtree', n_jobs=1, nthread=None, gamma=0, min_child_weight=1,
# max_delta_step=0, subsample=1, colsample_bytree=1, colsample_bylevel=1, reg_alpha=0, reg_lambda=1,
# scale_pos_weight=1, base_score=0.5, random_state=0, seed=None, missing=None, **kwargs)

# some of methods of RandomForestRegressor()
# fit(X, y, sample_weight=None, eval_set=None, eval_metric=None, early_stopping_rounds=None, verbose=True, xgb_model=None)
# get_params([deep]) Get parameters for this estimator.
# predict(data, output_margin=False, ntree_limit=0) : Predict with data. NOTE: This function is not thread safe.
# get_score(importance_type='weight') -> get the feature importance
# -----
# video link2: https://www.appliedaicourse.com/course/applied-ai-course-online/lessons/what-are-ensembles/
# -----

alpha=[10,50,100,500,1000,2000,3000]
cv_log_error_array=[]
for i in alpha:
    x_cfl=XGBClassifier(n_estimators=i)
    x_cfl.fit(X_train_merge,y_train_merge)
    sig_clf = CalibratedClassifierCV(x_cfl, method="sigmoid")
    sig_clf.fit(X_train_merge, y_train_merge)
    predict_y = sig_clf.predict_proba(X_cv_merge)
    cv_log_error_array.append(log_loss(y_cv_merge, predict_y, labels=x_cfl.classes_, eps=1e-15))

for i in range(len(cv_log_error_array)):
    print ('log_loss for c = ',alpha[i],'is',cv_log_error_array[i])

best_alpha = np.argmin(cv_log_error_array)

fig, ax = plt.subplots()
ax.plot(alpha, cv_log_error_array,c='g')
for i, txt in enumerate(np.round(cv_log_error_array,3)):
    ax.annotate((alpha[i],np.round(txt,3)), (alpha[i],cv_log_error_array[i]))
plt.grid()
plt.title("Cross Validation Error for each alpha")
plt.xlabel("Alpha i's")
plt.ylabel("Error measure")
plt.show()

```

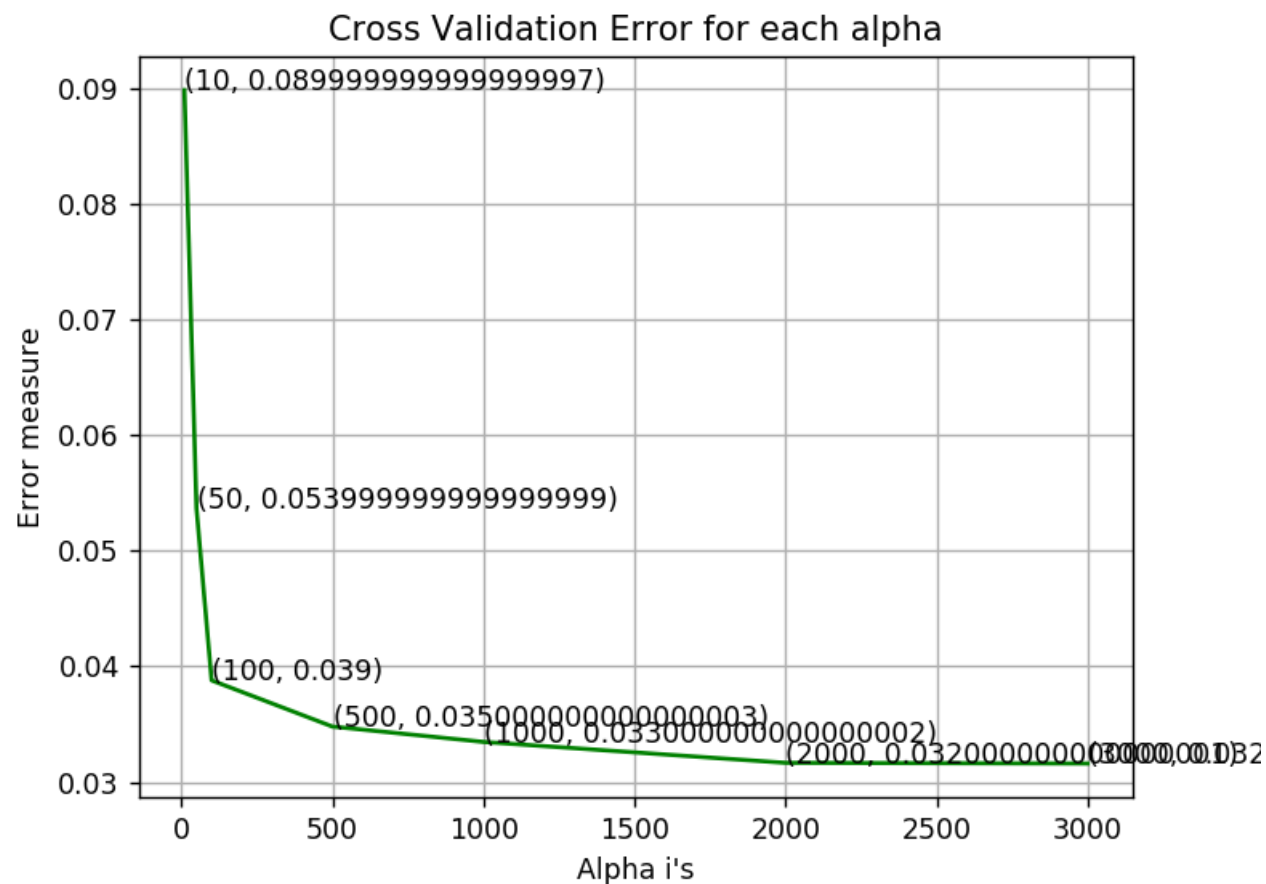
```
x_cfl=XGBClassifier(n_estimators=3000,nthread=-1)
x_cfl.fit(X_train_merge,y_train_merge,verbose=True)
sig_clf = CalibratedClassifierCV(x_cfl, method="sigmoid")
sig_clf.fit(X_train_merge, y_train_merge)

predict_y = sig_clf.predict_proba(X_train_merge)
print('For values of best alpha = ', alpha[best_alpha], "The train log loss is:",log_loss(y_train_merge, predict_y))
predict_y = sig_clf.predict_proba(X_cv_merge)
print('For values of best alpha = ', alpha[best_alpha], "The cross validation log loss is:",log_loss(y_cv_merge, predict_y
))
predict_y = sig_clf.predict_proba(X_test_merge)
print('For values of best alpha = ', alpha[best_alpha], "The test log loss is:",log_loss(y_test_merge, predict_y))
```

```

log_loss for c = 10 is 0.0898979446265
log_loss for c = 50 is 0.0536946658041
log_loss for c = 100 is 0.0387968186177
log_loss for c = 500 is 0.0347960327293
log_loss for c = 1000 is 0.0334668083237
log_loss for c = 2000 is 0.0316569078846
log_loss for c = 3000 is 0.0315972694477

```



```

For values of best alpha = 3000 The train log loss is: 0.0111918809342
For values of best alpha = 3000 The cross validation log loss is: 0.0315972694477
For values of best alpha = 3000 The test log loss is: 0.0323978515915

```

4.5.5. XgBoost Classifier on final features with best hyper parameters using Random search

```
In [ ]: x_cfl=XGBClassifier()

prams={
    'learning_rate':[0.01,0.03,0.05,0.1,0.15,0.2],
    'n_estimators':[100,200,500,1000,2000],
    'max_depth':[3,5,10],
    'colsample_bytree':[0.1,0.3,0.5,1],
    'subsample':[0.1,0.3,0.5,1]
}
random_cfl=RandomizedSearchCV(x_cfl,param_distributions=prams,verbose=10,n_jobs=-1,)
random_cfl.fit(X_train_merge, y_train_merge)
```

Fitting 3 folds for each of 10 candidates, totalling 30 fits

```
[Parallel(n_jobs=-1)]: Done    2 tasks      | elapsed:  1.1min
[Parallel(n_jobs=-1)]: Done    9 tasks      | elapsed:  2.2min
[Parallel(n_jobs=-1)]: Done   19 out of  30 | elapsed:  4.5min remaining:  2.6min
[Parallel(n_jobs=-1)]: Done   23 out of  30 | elapsed:  5.8min remaining:  1.8min
[Parallel(n_jobs=-1)]: Done   27 out of  30 | elapsed:  6.7min remaining:   44.5s
[Parallel(n_jobs=-1)]: Done   30 out of  30 | elapsed:  7.4min finished
```

```
Out[ ]: RandomizedSearchCV(cv=None, error_score='raise',
    estimator=XGBClassifier(base_score=0.5, colsample_bylevel=1, colsample_bytree=1,
    gamma=0, learning_rate=0.1, max_delta_step=0, max_depth=3,
    min_child_weight=1, missing=None, n_estimators=100, nthread=-1,
    objective='binary:logistic', reg_alpha=0, reg_lambda=1,
    scale_pos_weight=1, seed=0, silent=True, subsample=1),
    fit_params=None, iid=True, n_iter=10, n_jobs=-1,
    param_distributions={'learning_rate': [0.01, 0.03, 0.05, 0.1, 0.15, 0.2], 'n_estimators': [100, 200, 500, 1000, 2
    000], 'max_depth': [3, 5, 10], 'colsample_bytree': [0.1, 0.3, 0.5, 1], 'subsample': [0.1, 0.3, 0.5, 1]},
    pre_dispatch='2*n_jobs', random_state=None, refit=True,
    return_train_score=True, scoring=None, verbose=10)
```

```
In [ ]: print (random_cfl.best_params_)

{'subsample': 1, 'n_estimators': 1000, 'max_depth': 10, 'learning_rate': 0.15, 'colsample_bytree': 0.3}
```



```
In [ ]: # find more about XGBClassifier function here http://xgboost.readthedocs.io/en/latest/python/python\_api.html?#xgboost.XGBClassifier
# -----
# default paramters
# class xgboost.XGBClassifier(max_depth=3, learning_rate=0.1, n_estimators=100, silent=True,
# objective='binary:logistic', booster='gbtree', n_jobs=1, nthread=None, gamma=0, min_child_weight=1,
# max_delta_step=0, subsample=1, colsample_bytree=1, colsample_bylevel=1, reg_alpha=0, reg_lambda=1,
# scale_pos_weight=1, base_score=0.5, random_state=0, seed=None, missing=None, **kwargs)

# some of methods of RandomForestRegressor()
# fit(X, y, sample_weight=None, eval_set=None, eval_metric=None, early_stopping_rounds=None, verbose=True, xgb_model=None)
# get_params([deep]) Get parameters for this estimator.
# predict(data, output_margin=False, ntree_limit=0) : Predict with data. NOTE: This function is not thread safe.
# get_score(importance_type='weight') -> get the feature importance
# -----
# video Link2: https://www.appliedaicourse.com/course/applied-ai-course-online/lessons/what-are-ensembles/
# -----

x_cfl=XGBClassifier(n_estimators=1000,max_depth=10,learning_rate=0.15,colsample_bytree=0.3,subsample=1,nthread=-1)
x_cfl.fit(X_train_merge,y_train_merge,verbose=True)
sig_clf = CalibratedClassifierCV(x_cfl, method="sigmoid")
sig_clf.fit(X_train_merge, y_train_merge)

predict_y = sig_clf.predict_proba(X_train_merge)
print('For values of best alpha = ', alpha[best_alpha], "The train log loss is:",log_loss(y_train_merge, predict_y))
predict_y = sig_clf.predict_proba(X_cv_merge)
print('For values of best alpha = ', alpha[best_alpha], "The cross validation log loss is:",log_loss(y_cv_merge, predict_y
))
predict_y = sig_clf.predict_proba(X_test_merge)
print('For values of best alpha = ', alpha[best_alpha], "The test log loss is:",log_loss(y_test_merge, predict_y))
plot_confusion_matrix(y_test_asm,sig_clf.predict(X_test_merge))

For values of best alpha = 3000 The train log loss is: 0.0121922832297
For values of best alpha = 3000 The cross validation log loss is: 0.0344955487471
For values of best alpha = 3000 The test log loss is: 0.0317041132442
```

5. Assignments

1. Add bi-grams on byte files and improve the log-loss
2. Watch the video ([video \(https://www.youtube.com/watch?v=VLQTRILGz5Y#t=13m11s\)](https://www.youtube.com/watch?v=VLQTRILGz5Y#t=13m11s)) and include pixel intensity features to improve the logloss

1. you need to download the train from kaggle, which is of size ~17GB, after extracting it will occupy ~128GB data your drive
2. if you are having computation power limitations, you can try using google colab, with GPU option enabled (you can search for how to enable GPU in colab) or you can work with the Google Cloud, check this tutorials by one of our student: https://www.youtube.com/channel/UCRH_z-oM0LR0vHPe_KYR4Wg (we suggest you to use GCP over Colab)
3. To Extract the .7z file in google cloud, once after you upload the file into server, in your ipython notebook create a new cell and write these commands
 - a. `!sudo apt-get install p7zip`
 - b. `!7z x file_name.7z -o path/where/you/want/to/extract`

<https://askubuntu.com/a/341637>

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

Mounted at /content/drive

```
In [ ]: #!mkdir ~/.kaggle
        ! cp kaggle.json ~/.kaggle/
        !chmod 600 ~/.kaggle/kaggle.json
```

```
In [ ]: !kaggle competitions download -c malware-classification
```

Warning: Looks like you're using an outdated API Version, please consider updating (server 1.5.6 / client 1.5.4)

Downloading sampleSubmission.csv.zip to /content

0% 0.00/185k [00:00<?, ?B/s]

100% 185k/185k [00:00<00:00, 70.3MB/s]

Downloading trainLabels.csv to /content

0% 0.00/265k [00:00<?, ?B/s]

100% 265k/265k [00:00<00:00, 88.0MB/s]

Downloading dataSample.7z to /content

100% 4.06M/4.06M [00:00<00:00, 16.3MB/s]

Downloading train.7z to /content

100% 17.5G/17.5G [07:25<00:00, 27.7MB/s]

100% 17.5G/17.5G [07:25<00:00, 42.2MB/s]

Downloading test.7z to /content

100% 17.8G/17.8G [07:52<00:00, 40.2MB/s]

100% 17.8G/17.8G [07:52<00:00, 40.4MB/s]

```
In [ ]: !rm -rf test.7z  
        !7z e train.7z -o/content/bytes *.bytes -r
```

```
p7zip Version 16.02 (locale=en_US.UTF-8,Utf16=on,HugeFiles=on,64 bits,2 CPUs Intel(R) Xeon(R) CPU @ 2.20GHz (406F0),ASM,AES-NI)
```

```
0M Scan????????? 1 file, 18810691091 bytes (18 GiB)
```

— —

Blocks = 94

[illegible]

s 1% 82 - train/05aiMRw13bYwQz80Hvj1.byte
s
 1% 85 - train/05Kps4iFw8mOLJZQrb1H.byte
s
 1% 89 - train/06aLOj8EUXMByS423sum.byte
s
 1% 90 - train/06arUi9q3wHS2C8RZxeB.byte
s
 1% 92 - train/06osXqPUVM1HbvBGNncT.byte
s
 1% 94 - train/07ECKjDTyQLnabNoxrIH.byte
s
 1% 96 - train/07nrG1cLKUPxj0lWMFiV.byte
s
 1% 99 - train/09CPNMYyQjSguFrE8UOf.byte
s
 1% 103 - train/0ACDbR5M3ZhBJajygTuf.byte
s
 1% 108 - train/0ASH2csN7k8jZyoRaqtN.byte
s
 2% 110 - train/0aU7XWsr8RtN94jvo3lG.byte
s
 2% 112 - train/0aVNj3qFgEZI6Akf4Kuv.byte
s
 2% 118 - train/0BFIPv1r083whtpMYyAs.byte
s
 2% 121 - train/0BKcmNv4iGY2hsVSaXJ6.byte
s
 2% 123 - train/0bN60DYWw2xeCQBn3tEg.byte
s
 2% 125 - train/0BZQIJak6Pu2tyAXfrzR.byte
s
 2% 128 - train/0cfGJLYgE6R0aZH7KT1h.byte
s
 2% 130 - train/0cGWK6VvCkm702AxDjtw.byte
s
 2% 134 - train/0Cq4wfhLrKBJiut1lYAZ.byte
s
 2% 136 - train/0cTu2bkef0AJqIhYUWFK.byte
s
 2% 138 - train/0czUXKSCiGY2j5mxLdWa.byte
s
 2% 144 - train/0dhL8Jvcswa7U1qHiDS5.byte
s
 2% 147 - train/0DM3hS6Gg2QVKb1fZydv.byte
s

s	2%	148	-	train/0dnTixlMYzDUpsvEVrGc.byte
s				
	2%	152	-	train/0DTs2PhZfCwEv7q8349K.byte
s				
	2%	154	-	train/0eaNKwluUmkYdIvZ923c.byte
s				
	2%	157	-	train/0Eo9qT6idXHDMebwmvPA.byte
s				
	2%	159	-	train/0F4qIHaR7xOrm19Set3o.byte
s				
	2%	161	-	train/0fGuCWgTraQ6nEmLPN8q.byte
s				
	2%	165	-	train/0FOXjzmnD9CUMVcSlEqh.byte
s				
	2%	167	-	train/0fvnGU7dkbr8iEhZuMcP.byte
s				
	2%	168	-	train/0fxgjYEC1PL1BDbcshzJ.byte
s				
	2%	171	-	train/0GbMkYlNyt720zBjIcVh.byte
s				
	2%	174	-	train/0gDsIvrylX5fPbG7cSBn.byte
s				
	2%	179	-	train/0GKzFQ81IYXqUWkmfv26.byte
s				
	2%	181	-	train/0glscKoNakWL84EpunPe.byte
s				
	2%	184	-	train/0GUIi7xA10DwZ4YBenNM.byte
s				
	2%	186	-	train/0GuYe4J7oLwQ82xr3pWS.byte
s				
	2%	187	-	train/0GVcTdBQXWUJ2t7vjphN.byte
s				
	2%	189	-	train/0gWUIudhwovMYb3NSnZA.byte
s				
	2%	193	-	train/0hAlkjTR1Q6PewMczavb.byte
s				
	2%	195	-	train/0HcZRmLi9VTpuQJCoXny.byte
s				
	2%	200	-	train/0Hlm4XgE1cQhC6BkMays.byte
s				
	2%	203	-	train/0HVAnMrp1LjKDMuoOJFY.byte
s				
	3%	206	-	train/0hZqVRKkw7GfMdpa1LiN.byte
s				
	3%	211	-	train/0IA1cuEiP9G6epb710om.byte
s				

s	3%	212	-	train/0iBaz3krsQ8HuA7cGDSt.byte
s				
	3%	216	-	train/0iS3pwlGJco8XORD4TLq.byte
s				
	3%	218	-	train/0itbI5mjJF28ocTkrUf9.byte
s				
	3%	222	-	train/0IYZ1tU7uMpaco85PfKr.byte
s				
	3%	225	-	train/0J61YGoWjV25TzxeSluf.byte
s				
	3%	228	-	train/0JAzwGUKORhFQWr3o1dN.byte
s				
	3%	231	-	train/0JfwyrEcBqaRzN9TgFMh.byte
s				
	3%	235	-	train/0JOb8TyN6VBGCrjAkzfP.byte
s				
	3%	236	-	train/0JPAX13cjxewaTh6tRCi.byte
s				
	3%	239	-	train/0KgE6ksUeytoHf12cT4r.byte
s				
	3%	241	-	train/0KLUAMqmPJhOwaYrbSCE.byte
s				
	3%	242	-	train/0KyDiQb1whgaSrm1x58J.byte
s				
	3%	248	-	train/0LQSi5wnRZ3muIs6Mx9E.byte
s				
	3%	250	-	train/0LZkc7qeS39TUtVHuJB1.byte
s				
	3%	253	-	train/0mcWyK6unLRGV8Hfr97Y.byte
s				
	3%	255	-	train/0mfwTlekXE1poYAnqMRO.byte
s				
	3%	258	-	train/0MmZ8j5pn2R3VG9w1xYi.byte
s				
	3%	261	-	train/0MPV9Y8WNcFoyRZqQ76G.byte
s				
	3%	263	-	train/0MsQ16zDg9XVerbmfcDU.byte
s				
	3%	268	-	train/0NmoAE0tIDdwiVr9PCBf.byte
s				
	3%	275	-	train/00Q9MEykugoYZJrnj64d.byte
s				
	3%	276	-	train/0otC2dSxuGDzY6XHLkq3.byte
s				
	3%	278	-	train/0P6tEopzr2mIhkuOKiCD.byte
s				

s	3%	281	-	train/0PGo0fXAjVDBytKRH4Um.byte
s				
	3%	283	-	train/0PlfqyKM1JtYZx2me5FU.byte
s				
	3%	286	-	train/0pQ1231KqiLJtFTwjBfG.byte
s				
	3%	290	-	train/0PVAd2041MGUbawzfXBT.byte
s				
	3%	293	-	train/0q6eABpEHSbX1IzGh7jc.byte
s				
	3%	295	-	train/0qeE63wvCGzSapRuW4TI.byte
s				
	3%	298	-	train/0qjuDC7Rhx9rHkLlItAp.byte
s				
	4%	300	-	train/0qN2U7Ayorv9IbiTW4Gd.byte
s				
	4%	305	-	train/0QqWlUsiyj3AaPwJD81S.byte
s				
	4%	307	-	train/0QWUChwNSFt10jr45JpT.byte
s				
	4%	310	-	train/0RCcpJ1DuKNMBsaQUgmA.byte
s				
	4%	315	-	train/0r1keuBLpWm15TyFRHUJ.byte
s				
	4%	320	-	train/0SaYqdFRA5G7rfktjT63.byte
s				
	4%	322	-	train/0sdr2cfHLzN09Ep4RSjT.byte
s				
	4%	326	-	train/0SHsZKUKI2iN1Rw4be6E.byte
s				
	4%	330	-	train/0sM2atqTo48GgnFwhQXu.byte
s				
	4%	333	-	train/0SUAYQWoTeZb8qmIPMK3.byte
s				
	4%	334	-	train/0sYDNzwVMaWE2TIZeA81.byte
s				
	4%	337	-	train/0TpWNF7ULhd4Yk5bXQyP.byte
s				
	4%	340	-	train/0tZy7jgOKCIJs1PRvrop.byte
s				
	4%	344	-	train/0UcDbq2yojiThfdLRNS4.byte
s				
	4%	348	-	train/0uNkt6sirCnUWw175pjl.byte
s				
	4%	351	-	train/0UwHZzuXy7GFStRKEgdx.byte
s				

s	4%	354	-	train/0V6UmdMg7wQGxs1Az0RJ.byte
s				
	4%	356	-	train/0Vi6RsKd1HBzySe5Zp4Y.byte
s				
	4%	362	-	train/0wkFafuhqSr1Z5J7K6C9.byte
s				
	4%	369	-	train/0X8fojtR6QIblypvHLsM.byte
s				
	4%	372	-	train/0xfjbVBPYX1pSt4rdW0e.byte
s				
	4%	376	-	train/0XnPCouGZdhIKli0FQHR.byte
s				
	4%	380	-	train/0XsnKFjqPQxzCf81DvWp.byte
s				
	4%	383	-	train/0yBQ5dAc2gN4qx9YWuiS.byte
s				
	4%	386	-	train/0yjkG4BHeCbTimIslgr7.byte
s				
	4%	391	-	train/0YwidX9hOD5sPrtTvc2M.byte
s				
	4%	392	-	train/0Z3nQDPiU7LMEesmG4TB.byte
s				
	4%	398	-	train/0ZiQmgtxzHe9v508Lf2k.byte
s				
	4%	402	-	train/10Esk16DUSGugoWxvtY0.byte
s				
	4%	405	-	train/125y4VsArzkCNOGZfu6o.byte
s				
	4%	407	-	train/12Jd4qp0zTtQC3E6PXDb.byte
s				
	4%	409	-	train/12TEseaJ4jnbyORqKxhf.byte
s				
	4%	411	-	train/13FZ9D0hywk78sJiVULu.byte
s				
	4%	414	-	train/13YpdP5vTL0azSQFRgJn.byte
s				
	4%	418	-	train/15ADIikydx7eNq90WfCB.byte
s				
	5%	421	-	train/15mfGKRS6aVevjAoYL8w.byte
s				
	5%	423	-	train/15tQNEdpb0mi63kf0WM8.byte
s				
	5%	427	-	train/16gDhwbZTRjLM09zI5cX.byte
s				
	5%	430	-	train/17aN50JFZXyrcAk2iCtG.byte
s				

s	5%	432	-	train/17kCoALmxaXpcNzd5Kne.byte
s				
	5%	435	-	train/18A9fzqbeyripCSuLko4.byte
s				
	5%	440	-	train/18T9jLsd0DZQc2aRrSpb.byte
s				
	5%	443	-	train/198PdXRpGVZAzYxCrF0e.byte
s				
	5%	445	-	train/19B6wSHOLmjhGPoquTbW.byte
s				
	5%	447	-	train/19Emlw3WgC204iJknVtP.byte
s				
	5%	451	-	train/19UExqPS8Go3MTNci4Hh.byte
s				
	5%	454	-	train/1A7UqoIMrxHgVuW3FS9c.byte
s				
	5%	459	-	train/1AhLOsNdHzDPfJrSXmQC.byte
s				
	5%	460	-	train/1AhQif5kMPb24RnCmjrg.byte
s				
	5%	463	-	train/1AvCEkV76rjuYp9aqPbK.byte
s				
	5%	467	-	train/1AXEyUDVt42T6ZFx8COo.byte
s				
	5%	470	-	train/1bgW3y0fFDHBjdps5Jhe.byte
s				
	5%	472	-	train/1bL4yiwCUvS0g7tBJudf.byte
s				
	5%	476	-	train/1cB6EdrmebFjAX3QHkS9.byte
s				
	5%	477	-	train/1CbXmgAIJtiNLR0So9ul.byte
s				
	5%	480	-	train/1cEKCiXn0hLveSk2dDyW.byte
s				
	5%	485	-	train/1cP8XiEvMd7tg3rJmnAj.byte
s				
	5%	487	-	train/1Cub9LvOgMWkQqs3R2ef.byte
s				
	5%	490	-	train/1d485UYC6qhXKrAcDJPZ.byte
s				
	5%	493	-	train/1dFAk8Yzck04RVmsaS9Z.byte
s				
	5%	494	-	train/1dhuKsA0t2Dnf9qNQLjl.byte
s				
	5%	497	-	train/1DNWriJEg56S03yvjtBu.byte
s				

s	5%	499	-	train/1e4vEcCMmyGJ3wrPXTq8.byte
s				
	5%	502	-	train/1eAC0zqkIGHRBvm97P6i.byte
s				
	5%	505	-	train/1EisIfzG0TX8hKaFH6uy.byte
s				
	5%	510	-	train/1esMpo6SAPHiTUGEb3Bt.byte
s				
	5%	516	-	train/1FAxSngsym35phwROGl0.byte
s				
	5%	518	-	train/1fJqS7HUCEa6PYx0ZLc9.byte
s				
	6%	522	-	train/1FsfvPlaKG3TpImqOejM.byte
s				
	6%	530	-	train/1GRsOmJXetEkqIraBuSd.byte
s				
	6%	532	-	train/1GvV8Jsz09wLnje4kSyN.byte
s				
	6%	539	-	train/1her6tuVWjBkqZPUf8KR.byte
s				
	6%	543	-	train/1HJhGL4zxVvcI2se07in.byte
s				
	6%	546	-	train/1hToAGU8YgHQxfVONLqJ.byte
s				
	6%	549	-	train/1i2SThtkCGEp1HeoJxqI.byte
s				
	6%	551	-	train/1I3EGeyOpWHNtYjLvnk5.byte
s				
	6%	552	-	train/1I8aJRE7T0igPZHWcvVQ.byte
s				
	6%	557	-	train/1ik3ncdBTYIJAgfmjpwz.byte
s				
	6%	562	-	train/1iWpGIDVnJEaTHR8QSPL.byte
s				
	6%	565	-	train/1Jd0PTkEWtKcyCSbBD6H.byte
s				
	6%	572	-	train/1JuwbkFyUg7T6KfoNEjp.byte
s				
	6%	576	-	train/1KB3Z7gd5aN4Xmx8W0sf.byte
s				
	6%	580	-	train/1KjwODdhC3VxA4MSLbFI.byte
s				
	6%	583	-	train/1KozfUh5ri3ngNCGH098.byte
s				
	6%	585	-	train/1l5V2AwZS9ixLMmecCkp.byte
s				

s	6%	587	-	train/1LE6uK9BdgNXIcSTnrMJ.byte
s				
	6%	591	-	train/1lvtjJ0rfDVkE6IxaUsc.byte
s				
	6%	593	-	train/1LZcKGIkna9HU7vXl3YN.byte
s				
	6%	595	-	train/1mCvIc0K8Gd3P5suXTRM.byte
s				
	6%	599	-	train/1MQI2rYz86Sbxa5WhCFG.byte
s				
	6%	601	-	train/1mStzMaQ8RUVBx7TkF4o.byte
s				
	6%	605	-	train/1NiArstKkEmMf0vjDd7l.byte
s				
	6%	609	-	train/1nsIO2yMfHuVcFxY7Tlb.byte
s				
	6%	611	-	train/1nv9wldWsg0fU2hEcKb3.byte
s				
	6%	614	-	train/1o2t0KwOIqHdhYCcPAfp.byte
s				
	6%	615	-	train/10ayZHFbz3tPBE5XNkcS.byte
s				
	6%	619	-	train/1otbAhISV0yQ7vF6zndx.byte
s				
	6%	621	-	train/1P6o3IJetcjyL4pTaXqi.byte
s				
	6%	624	-	train/1pCDZ6lLytxzQ0KYXqe3.byte
s				
	6%	625	-	train/1Pcq2sTQEV7r3gBSwFyU.byte
s				
	6%	629	-	train/1pjpg7bvn9oxqzCOPN2kZ.byte
s				
	6%	633	-	train/1Q7DXjLweH09EokGNWiI.byte
s				
	7%	636	-	train/1QHbFGR3qcur6aWZw0Jk.byte
s				
	7%	639	-	train/1QuMnTw8zc3oWLSZ4KXt.byte
s				
	7%	642	-	train/1r6oKe4cpzuBj0YIhEsL.byte
s				
	7%	643	-	train/1rbXZBI9R3EleQDH74Sk.byte
s				
	7%	645	-	train/1rCGTyze8MqbtX49fQOu.byte
s				
	7%	648	-	train/1rpACyIiMJjE8mxc3NK4.byte
s				

s	7%	651	-	train/1RX4GxMFoCJUDvs1qBaP.byte
s				
	7%	652	-	train/1RyBhpeVSJd96rZkEI7w.byte
s				
	7%	657	-	train/1SFK4QHeDoOz3AMJ6bfh.byte
s				
	7%	663	-	train/1sqBw5DpJaRkx4S1QdLY.byte
s				
	7%	666	-	train/1T0BWPoJVtNFzsycKh1r.byte
s				
	7%	673	-	train/1t03IYLhvDx6CKoHamQb.byte
s				
	7%	675	-	train/1txR5rhb4MwNvXuPAFpz.byte
s				
	7%	679	-	train/1u3qTGiRvckQZW7dBY58.byte
s				
	7%	682	-	train/1UE8oOGSuImQZxeXLYK0.byte
s				
	7%	684	-	train/1uHZ0dS3zTnrk8jXUQ2i.byte
s				
	7%	687	-	train/1uNTL24yH13EVFidSzUA.byte
s				
	7%	690	-	train/1v5SAIfct9srqh7p8JHL.byte
s				
	7%	692	-	train/1Vd3xZnQ187fbhPoSYJ6.byte
s				
	7%	695	-	train/1VMhvfCS30zbeFK4UZxL.byte
s				
	7%	697	-	train/1vQKJIkiBaTVxc5uYCZW.byte
s				
	7%	701	-	train/1wf4Y2Xa0mlyK8VNTDBn.byte
s				
	7%	702	-	train/1WJXH29t4FmdcLBK5DNR.byte
s				
	7%	705	-	train/1x2u5Ws7tzFRAGyqoJBV.byte
s				
	7%	707	-	train/1XgEYavLi9R0fV8BxHS5.byte
s				
	7%	711	-	train/1XKY6TxM2eICFymS4AsN.byte
s				
	7%	713	-	train/1xSW8f3MoVAKsjTCmE79.byte
s				
	7%	715	-	train/1xVIAa0rkQePG1BLpMgs.byte
s				
	7%	720	-	train/1YhJFKP0mIC5e6kD7iqS.byte
s				

s	7%	723	-	train/1YkSNHcbq4y0ofMpR93Z.byte
s				
	7%	724	-	train/1ym2nFJUv1Z4zAYBbwa7.byte
s				
	7%	727	-	train/1yx5iQYgClzaUEIPdLJF.byte
s				
	7%	730	-	train/1zb0rkEtweXH7ABuqoM2.byte
s				
	7%	733	-	train/1zP8Fh4ZWpELUrXB7Say.byte
s				
	7%	735	-	train/1ZYtDkrdm9yVLpeU5cSJ.byte
s				
	7%	737	-	train/20loY7fRhmcSbSKctgqT.byte
s				
	7%	740	-	train/21l6hcFemsiIapSqw8X0.byte
s				
	8%	744	-	train/21S5IzZvPWd0MQr9yoq4.byte
s				
	8%	745	-	train/21XVtgk6bMRdLWP4xoKs.byte
s				
	8%	748	-	train/23iLdjsJ6c01QNkrIRXD.byte
s				
	8%	751	-	train/24DYHUbwxidKLV05Genl.byte
s				
	8%	752	-	train/24NSkH5K03grqedBQM8T.byte
s				
	8%	758	-	train/25MVSa0YETk0ueIroJ1L.byte
s				
	8%	762	-	train/26GkX7dIghv0w38A9ZzF.byte
s				
	8%	763	-	train/26IcTMH970BwKz5nFbrU.byte
s				
	8%	764	-	train/26mSoP1NGRYSyvqA5fbE.byte
s				
	8%	766	-	train/26xuQtiVm9qcW0ErByXF.byte
s				
	8%	770	-	train/27fiQdcEXxeyWoV9qu8r.byte
s				
	8%	771	-	train/27o93DUCHcGPXdTx1Nwg.byte
s				
	8%	779	-	train/28ndZemTJQNpV5rWl9Mq.byte
s				
	8%	782	-	train/295nPpgmDUkQw4TScqfh.byte
s				
	8%	787	-	train/29WyrbfmNGFBVLoQhg8q.byte
s				

s	8%	791	-	train/2aHfrLhcPTj5GnFZXUCN.byte
s				
	8%	794	-	train/2AkLpJEONXZSefoC0BgQ.byte
s				
	8%	797	-	train/2aqbztXhp0QDeABEGLrV.byte
s				
	8%	801	-	train/2BcVLYAwvCqM37U6bkj9.byte
s				
	8%	806	-	train/2bf65CdKZTUGXjzq9KNJ.byte
s				
	8%	808	-	train/2bwDkqs87r5IES13WLxM.byte
s				
	8%	810	-	train/2bZlmksaBHeNDVnf9PRx.byte
s				
	8%	818	-	train/2d1DA6k5BE0H3MzRXv7w.byte
s				
	8%	821	-	train/2dEMyoGL3sI08NVu1r4b.byte
s				
	8%	823	-	train/2dHolcmQKIbU1hf43jF9.byte
s				
	8%	827	-	train/2eA8avWyDsI5h0Zrc1lg.byte
s				
	8%	829	-	train/2ECqL8au6iS3jNJDo0dn.byte
s				
	8%	830	-	train/2Ecxb0XKTPYwm0MfkJUz.byte
s				
	8%	833	-	train/2efsFypAKd1kBN9q8V6t.byte
s				
	8%	836	-	train/2EkHfwNj5o7JcDq6UzuG.byte
s				
	8%	838	-	train/2Ep7f84CvJDn9NwMHBPK.byte
s				
	8%	840	-	train/2EQ7bhH8pDCsw45djNeR.byte
s				
	8%	845	-	train/2fBi0gkxcmrHRInXwVWP.byte
s				
	8%	848	-	train/2FfmJdWv6RrIjzYhaKZ1.byte
s				
	9%	851	-	train/2fl8q7dsoC1PwApx6QNg.byte
s				
	9%	855	-	train/2FpVtbaHJWAG0ImwTzxU.byte
s				
	9%	859	-	train/2GIqt8Rpx7uEHrWCdoNg.byte
s				
	9%	862	-	train/2GPmRJEgDjHqoT7YN9M5.byte
s				

s	9%	865	-	train/2gwSoct3BzRUvajVeCLu.byte
s				
	9%	866	-	train/2gXZWDa3VeTPr1I7pjYa.byte
s				
	9%	868	-	train/2H4xSOLEzJhKATZ5orni.byte
s				
	9%	870	-	train/2HAv14QhqKBGu8cWedtY.byte
s				
	9%	873	-	train/2HNmJG5tcI7aXQsnuvpk.byte
s				
	9%	876	-	train/2HRQ1j0Dx3fv8Ad7n9rg.byte
s				
	9%	884	-	train/2Ig4GdWcj10LQ9YEz6aV.byte
s				
	9%	888	-	train/2ipXFC47Umax9kQfgcTe.byte
s				
	9%	891	-	train/2IUJrWpZ8tboCh607aqv.byte
s				
	9%	892	-	train/2iXg1178rZEW5YVRJyv6.byte
s				
	9%	895	-	train/2j6bn1rCgNW9cvm0KIMH.byte
s				
	9%	896	-	train/2jA0Zpk5aetMqrz6nswK.byte
s				
	9%	899	-	train/2JfpcM9ACq4XatFLho8z.byte
s				
	9%	901	-	train/2JpSgz3vxQqm9H4EIRhV.byte
s				
	9%	905	-	train/2kKae4J5IY6LWtTv3bc1.byte
s				
	9%	908	-	train/2L5fIAXjYka3yrmJHswR.byte
s				
	9%	912	-	train/2LeQVofH8Jl4da69AUxk.byte
s				
	9%	916	-	train/2lv0CPzQVLwtcAKZT1FB.byte
s				
	9%	917	-	train/2M4c9art0HkgiC6dRjfx.byte
s				
	9%	921	-	train/2mNqMtbkL87i14gXBPHY.byte
s				
	9%	923	-	train/2mQdWY1BRCsIZTPGA1Sc.byte
s				
	9%	925	-	train/2mxSawkp0F5oK4gJGqTV.byte
s				
	9%	927	-	train/2mypfAUu04n7R1PBMLZq.byte
s				

s	9%	929	-	train/2mZqzjspRt5Xy4B9f03L.byte
s				
	9%	931	-	train/2n8uBGd0wsYvgz4UfH59.byte
s				
	9%	934	-	train/2nBdD7s8KKlrtvm3zSow.byte
s				
	9%	936	-	train/2N1bHE3LTPkmvV1COSJ4.byte
s				
	9%	938	-	train/2nqWjb58uFIORHeyCrkY.byte
s				
	9%	942	-	train/2NYAUVgtXPpbJzRQ07C1.byte
s				
	9%	943	-	train/208w7LMDfCgRhmvNuetV.byte
s				
	9%	944	-	train/2o9qPQI3SL0ZV7wirTA0.byte
s				
	10%	947	-	train/2oH0w4StzEV0hiKq3ByD.byte
s				
	10%	950	-	train/2ondxDIwf4si7URbH1ZJ.byte
s				
	10%	952	-	train/2oPaLFU6SQDdZ45Wu0EB.byte
s				
	10%	959	-	train/2pdHWrsICS6VPXhfyEw1.byte
s				
	10%	964	-	train/2POsRoLQDjVz89qgSpnY.byte
s				
	10%	965	-	train/2pwjzv6eGEb8QmHPfxSc.byte
s				
	10%	969	-	train/2Q300EUydTbhDIxVAztg.byte
s				
	10%	971	-	train/2Q941EwIimV0z13xXNLY.byte
s				
	10%	973	-	train/2QCuzf8vh6JL41dZmSNg.byte
s				
	10%	976	-	train/2qJP9B5Q1xeDECIVzLo4.byte
s				
	10%	979	-	train/2qpZmcvFs4LCBNi9IX6H.byte
s				
	10%	982	-	train/2qSURceVYgQEh0FAdPXC.byte
s				
	10%	987	-	train/2qZrQoxyYDeadjIp3mSH.byte
s				
	10%	989	-	train/2r0EQuyUfz1k9BhPgpv0.byte
s				
	10%	990	-	train/2r0UYLcKCxnRovpj5d1Q.byte
s				

10% 994 - train/2Rn8XiSKtxyT6Cv1juQ0.byte
s
10% 999 - train/2Ryi1HVUPdwNxuMebW0X.byte
s
10% 1002 - train/2sf3AQd9MjP1xcLFRYyo.byte
s
10% 1003 - train/2sHzcQk5T1UExobShRmF.byte
s
10% 1006 - train/2SofB197xJ3CKhuiIwVQ.byte
s
10% 1008 - train/2SuEocCLvdekMN1jJFzI.byte
s
10% 1011 - train/2T9zBqN60tJEQmOblyM8.byte
s
10% 1014 - train/2Te7QKFCEJPy8wjB5Y6l.byte
s
10% 1016 - train/2Tlr63msziCjLuBtvZDa.byte
s
10% 1017 - train/2TqmRPjBwf36enxUDGKQ.byte
s
10% 1018 - train/2tqQkycVPSxIRGuW3wnh.byte
s
10% 1020 - train/2u4RQPDJzrf1TwqI0taY.byte
s
10% 1024 - train/2UcBWpt6mf8dohDsYxOR.byte
s
10% 1026 - train/2UfwYHR03ZMokQgFAzqt.byte
s
10% 1030 - train/2uwHk0ITbfVelsUgRt7o.byte
s
10% 1034 - train/2vAZ5dpB7hzci0V9EmnP.byte
s
10% 1036 - train/2VEsIiQxy5a60hwGAo1T.byte
s
10% 1039 - train/2VHna141dt9Ar0EhQSjZ.byte
s
10% 1041 - train/2vQgFHCYq9wfb8G0V70B.byte
s
10% 1043 - train/2VY0sEQDxLRHn0jIUzF9.byte
s
10% 1046 - train/2WabDISz1fTxFnXvekm9.byte
s
10% 1049 - train/2WHevmqLEZAb5ug3I7DN.byte
s
10% 1051 - train/2WkGMV9mKJrBAOZL5zdI.byte
s

10% 1053 - train/2WVtFRnNabH41C9dqzDB.byte
s
10% 1055 - train/2wxRyP0giXLMC4daDunr.byte
s
11% 1059 - train/2X3tMeVL1QZ4WJmdqkUz.byte
s
11% 1064 - train/2xkoYzhi0RMSad95G7Pg.byte
s
11% 1065 - train/2XL3TUHZNYkuvS0n5KpC.byte
s
11% 1068 - train/2XnTALlRBbSxQjo4py58.byte
s
11% 1069 - train/2XoqdSMTgrZKACjPynfW.byte
s
11% 1070 - train/2xQa1ME3fIsHzjc7Zo6w.byte
s
11% 1073 - train/2Y8yNd7pfDX4l3qacwEk.byte
s
11% 1075 - train/2ydtPK1XCYeBqwSIcRGZ.byte
s
11% 1078 - train/2YiILGOvJNDh18Acw1r3.byte
s
11% 1083 - train/2z8s73BXN0wqj0UdtJFD.byte
s
11% 1086 - train/2ZI9SdMvc7WAXK03nPyy.byte
s
11% 1092 - train/2ZoTE3iUtwFfuNAYlyOk.byte
s
11% 1094 - train/2zrDU8t1pNjoWaisTeK9.byte
s
11% 1097 - train/2zUdQEyv5oKARaYGxhXk.byte
s
11% 1099 - train/2zyuIDkXNCwY0tjdTqfH.byte
s
11% 1100 - train/30259LfMxFdATBcqN0i1.byte
s
11% 1104 - train/30ryX9b4e1IiQcPaVx2W.byte
s
11% 1105 - train/30Sj9r1HUZveV8FN1k54.byte
s
11% 1107 - train/31KjQiNWATfPaozHtY0l.byte
s
11% 1109 - train/31q6gcPKtCp8EQGj7yTL.byte
s
11% 1111 - train/31xrLF5GIWsnJ9p1tCBE.byte
s

11% 1117 - train/32WwkMHD8dafQApiKm1v.byte
s
11% 1118 - train/34anOKthJ6CXbZs2MeI5.byte
s
11% 1125 - train/35i2FSK0EnV0mNHpd8QU.byte
s
11% 1127 - train/365qzFMSdZmIU1YbfoAx.byte
s
11% 1129 - train/36eMEj40inf8r5vVQ1Bx.byte
s
11% 1135 - train/36zPKsYmJc8wOiEqStFN.byte
s
11% 1138 - train/37kuvmtW8EGoCeFTgBKH.byte
s
11% 1142 - train/37YaF9jLifHNxQPbq8hm.byte
s
11% 1145 - train/38IdJg2Se7po9rGZaNAw.byte
s
11% 1147 - train/38MtDCdfSEF7BWhizVyL.byte
s
11% 1150 - train/39SjamCOL05zq1Eni4Qf.bytes
11% 1152 - train/39VK5R6xNtwkBZqDY7Lr.byte
s
11% 1155 - train/3a01JWnNRb8QIxVHALGd.byte
s
12% 1157 - train/3aC1EztfpUXIbuyi8sGP.byte
s
12% 1166 - train/3b7cK2nPtTHIQ658wsYk.byte
s
12% 1167 - train/3bapyeGYC5Zot4Sm1N0d.byte
s
12% 1168 - train/3boTYd1KBg0PIL5pVvkN.byte
s
12% 1171 - train/3BwvQWhe4kXgPqYrKLCT.byte
s
12% 1175 - train/3Ce9LP6ngx2uBtSyczRG.byte
s
12% 1179 - train/3cOQWb04rxmgsYS8nfNK.byte
s
12% 1182 - train/3CwcNOpIWJsBERGPHVkr.byte
s
12% 1184 - train/3CyxQdJhW5trsfwIz4ik.byte
s
12% 1189 - train/3DfvjbgBnePmskyhpCT2.byte
s
12% 1192 - train/3dstJNaJYRkxEV4g0nA6.byte

s	12%	1196	-	train/3e80boJq70FQ9Ccnaugp.byte
s	12%	1200	-	train/3eHwdBo1FKi4PahtXSyg.byte
s	12%	1204	-	train/3EMZ7J5djsBm1cwHhtSF.byte
s	12%	1206	-	train/3ESs1rk0uyp9w1GgFaKj.byte
s	12%	1210	-	train/3f8ms4TX9ax0B2zjCQuS.byte
s	12%	1215	-	train/3fN5b0wAxSQG2M1hWV4m.byte
s	12%	1219	-	train/3Fq0zG8mAv7H609nSyuj.byte
s	12%	1222	-	train/3FRuYdNiL4ye96gBoWK2.byte
s	12%	1224	-	train/3Fv8npJNEqWBSmi60xXw.byte
s	12%	1227	-	train/3G6jJKRVLz1poimkvfhA.byte
s	12%	1231	-	train/3gJDz9PFyZxd2KQmSutG.byte
s	12%	1232	-	train/3gkZ4wV0vUx8haYiMuq1.byte
s	12%	1237	-	train/3GZL4dcysYpv2MSVOPb5.byte
s	12%	1238	-	train/3h0t6wQAFR10HK9nfkVY.byte
s	12%	1242	-	train/3H1jRUIAsgc25iJNBeqo.byte
s	12%	1245	-	train/3I8UmNjSvVu5sb1AYfXB.byte
s	12%	1247	-	train/3IBKi0NtaWxXpQmufSZs.byte
s	12%	1249	-	train/3iHvsmdx7DauhSPKfec4.byte
s	12%	1252	-	train/3IZHa2xXVu6KrgNoPd9W.byte
s	12%	1254	-	train/3JCnTRxD1a0UX1qr854L.byte
s	12%	1258	-	train/3JI42LUuZET9tgv1YDqx.byte
s	12%	1260	-	train/3jmcq4dRMkXIaL70EWCT.byte
s	12%	1264	-	train/3jt6Y8zf1DH0vgKsTPqy.byte

s	12%	1268	-	train/3KCHUXujYosV6ywDkZPA.byte
s	13%	1273	-	train/3Kn9Ai6EwNaemtS70Qf1.byte
s	13%	1275	-	train/3kqaIlVuPtXQJevjvNZ0.byte
s	13%	1278	-	train/3KWDrvPXBaAxfGYsoN8.byte
s	13%	1279	-	train/3kXqmsjc6Jx00g8uQZWB.byte
s	13%	1284	-	train/3Lrcb9Md4Isugz5a1UK7.byte
s	13%	1287	-	train/3m8kb5ILPrHcMC1o9Nht.byte
s	13%	1289	-	train/3MbT7ePSqZuEWgXpiKdM.byte
s	13%	1291	-	train/3MG6tJEmkC1b5pB0aOQT.byte
s	13%	1294	-	train/3mpsxg2FODqCAXHbMKPi.byte
s	13%	1297	-	train/3MZri0FjTzUcpWf14nah.byte
s	13%	1300	-	train/3nD402vc1wpMazdUobEB.byte
s	13%	1305	-	train/3NPVdenAHU8cyZ1sQBv1.byte
s	13%	1308	-	train/3nxujXtip8hGHb6ZQdA4.byte
s	13%	1311	-	train/3NZJCsnHq1zpoF72dt6y.byte
s	13%	1315	-	train/3oA2emBxRQTkvDrVqPfL.byte
s	13%	1317	-	train/3oMQAPEXqKpiB1Hv0JYz.byte
s	13%	1320	-	train/30QNh7iEaut1HUSSZyRW.byte
s	13%	1324	-	train/30y7QgdVxkjuEsfXT8B5.byte
s	13%	1327	-	train/3PFdwRiY9seVIa7DBkly.byte
s	13%	1329	-	train/3PmaDd1EYWrhG6e480xp.byte
s	13%	1334	-	train/3Q8Wvm1ceyOhSajNdCub.byte
s	13%	1337	-	train/3qi6bX91BAxoCPapGzmR.byte

s	13%	1341	-	train/3QYwp0Uha1uvGBfbmg86.byte
s	13%	1344	-	train/3rf7KJiPd8hIZ645wBVN.byte
s	13%	1347	-	train/3RpLgGMCVIXuZcnAS98K.byte
s	13%	1350	-	train/3RrLNMzxy8g26wFXeYS.byte
s	13%	1353	-	train/3SdTQDpt7X4C0NkgamMA.byte
s	13%	1356	-	train/3sKk00hxBNy6MILz8gia.byte
s	13%	1358	-	train/3sw1FG5VqW06nEUR1yej.byte
s	13%	1359	-	train/3T271W9cVsDRatMYLFqp.byte
s	13%	1362	-	train/3t8F2RIgnQAGxe1jrdSl.byte
s	13%	1364	-	train/3TBSm5I1CVEyQY17kJPG.byte
s	13%	1366	-	train/3ThmZXnQYN5ApG61k4dL.byte
s	13%	1368	-	train/3tpMAGVr740gkCde51zL.byte
s	14%	1371	-	train/3TYS1o5Qt1IHCg284whV.byte
s	14%	1373	-	train/3u6fqdFzHRhgZLC4AJa9.byte
s	14%	1376	-	train/3uCF0Mmy1oSeHxbtTVnY.byte
s	14%	1378	-	train/3UGJfgV58TXmWAnh64De.byte
s	14%	1381	-	train/3UXfgqQ0mhcrFZv4jzDn.byte
s	14%	1383	-	train/3v7fe0Diw9c0u6dAzsaE.byte
s	14%	1386	-	train/3VdhPHBJMc1nzEyqARWj.byte
s	14%	1391	-	train/3VrMnozUtP519ThSG4iy.byte
s	14%	1392	-	train/3vs19KYdAo1tibP5DRSX.byte
s	14%	1395	-	train/3wBXsM1xyZQodCTzbDtf.byte
s	14%	1402	-	train/3wNfTntvMkp18LuY0WPm.byte

s	14%	1406	-	train/3WZjh7BUpwRVDtkmtISi.byte
s	14%	1408	-	train/3X1RZPbneqjVdBU5St4x.byte
s	14%	1411	-	train/3X9o8zrqkZThUFMBJjyL.byte
s	14%	1412	-	train/3x9R8E6Gv5JhDda2eoq4.byte
s	14%	1417	-	train/3XwBkUPfuGrKvoisljnt.byte
s	14%	1418	-	train/3y5HQzqXhowfrkn0CwF6.byte
s	14%	1423	-	train/3yHaDhq1kp67fzMQAgxc.byte
s	14%	1425	-	train/3ym0zq0n8cWVLhx4ZGvJ.byte
s	14%	1430	-	train/3zDwi1mVHFahA62IUp0j.byte
s	14%	1433	-	train/3zf9MwdQGUiXa1NYpIoE.byte
s	14%	1435	-	train/3ZvD4c8oNIOTib1JEfMa.byte
s	14%	1437	-	train/3ZxKMsRLlDYwnd4uPi2z.byte
s	14%	1440	-	train/409FLy5cim8qgxjYobET.byte
s	14%	1443	-	train/40as7FtBNoGZTu05XpWq.byte
s	14%	1444	-	train/40aUf0CuXKwHTBvAMq35.byte
s	14%	1449	-	train/400BbdiUSJj8orKvwa72.byte
s	14%	1451	-	train/418XPavxNpslKmeoTUwJ.byte
s	14%	1458	-	train/420SDzWfNamqIiHBYjPu.byte
s	14%	1462	-	train/42j5t00CXfxeumqPUK6p.byte
s	14%	1465	-	train/42Nj0zM5EGmvLJkCpuOx.byte
s	14%	1467	-	train/42QdnyRoXTisYJ8M9hEz.byte
s	14%	1470	-	train/43Pz67mIVtohuYapcLKj.byte
s	14%	1471	-	train/45aTQcuSVrGxZ62vHLbw.byte

s	15%	1474	-	train/45jzUD8TtAe3wCspRHv0.byte
s	15%	1477	-	train/46FsZxPzIVu87hy95CkB.byte
s	15%	1481	-	train/47GqvWtCkjShUVpE1oI0.byte
s	15%	1482	-	train/47mvwgDZ6dTEQcetyPbn.byte
s	15%	1484	-	train/48Dhca7LzbPTWN96xjFY.byte
s	15%	1487	-	train/495fFS8xhen07WqZX6uw.byte
s	15%	1491	-	train/49epoGtYixcgnkfwJ2dv.byte
s	15%	1494	-	train/4ac5qd0rj96X8SglAGZL.byte
s	15%	1500	-	train/4AtRvui0E2FNqp1PX7ZD.byte
s	15%	1501	-	train/4augeM3l7UKbHc6ITnOL.byte
s	15%	1505	-	train/4bdcgZYPyMtlw0hoq7Dx.byte
s	15%	1507	-	train/4BdQMg5PDlWq69ZXwReb.byte
s	15%	1510	-	train/4boIVf8XqkEQ3ZRTjdiz.byte
s	15%	1513	-	train/4BrHgYquho9skZMR06ea.byte
s	15%	1515	-	train/4by5ul19kAEt07CMGPFW.byte
s	15%	1518	-	train/4cPt7qBs9CEIgNkhRXKM.byte
s	15%	1520	-	train/4dqEznNyiFwZVYrC5G6m.byte
s	15%	1523	-	train/4dyK0YWPvuOnFmk5b9rS.byte
s	15%	1524	-	train/4dYWHLCBPVg5xTAzRK0e.byte
s	15%	1528	-	train/4eC9LmdHc6pRoP8BtTGv.byte
s	15%	1534	-	train/4eRJPIsap8c0dZ3hDKfT.byte
s	15%	1536	-	train/4EtCMF7yKHU9ucr0g6io.byte
s	15%	1540	-	train/4Fdkzt7s20ZVYeLIp1vh.byte

s	15%	1542	-	train/4fL5iSGhJDwoWuNvcjBq.byte
s	15%	1545	-	train/4FTihAW9250ds1GwzKYc.byte
s	15%	1551	-	train/4ghaPIpBnJ8zCfruTxID.byte
s	15%	1553	-	train/4GLY2XFSEMMb6CVsAJ9f.byte
s	15%	1556	-	train/4Gp7IWxvR0y0MuS5PNzT.byte
s	15%	1558	-	train/4GrZx0BKvSPauIDVkhjt.byte
s	15%	1562	-	train/4HDpbwyfumP9o3Cx8R0e.byte
s	15%	1564	-	train/4hLqKPS0yopVZMeU3ckY.byte
s	15%	1567	-	train/4hWm7ySRkX0H9KALGp2Q.byte
s	15%	1568	-	train/4i5aXxMQCpbhEkjZYDdy.byte
s	15%	1571	-	train/4iFfbWB7CqmSszKd5gRu.byte
s	16%	1573	-	train/4ilcBJdk3uARvNphIFy8.byte
s	16%	1575	-	train/4iNJaXWGlkyVroOfHMqg.byte
s	16%	1578	-	train/4iXduF6eZKQ1NEkzp2Wn.byte
s	16%	1582	-	train/4Jbygn1sIkMpzAUPBCRa.byte
s	16%	1584	-	train/4JEYXLGnodxWIcv6qZ9w.byte
s	16%	1588	-	train/4JlgupTC2XdIvxE1oHjq.byte
s	16%	1593	-	train/4jVL1kxAIGvb3MBzDYHc.byte
s	16%	1594	-	train/4JyQWUq9cs81pFRPCaBw.byte
s	16%	1596	-	train/4KAdMvSOYoB90Hp1Wnac.byte
s	16%	1600	-	train/4KgGC1w1aNsMP8TWRFio.byte
s	16%	1602	-	train/4kKcuD5YGjy8QtUh6Vm1.byte
s	16%	1605	-	train/4Kq0NxdX39AckSV5F7tm.byte

s 16% 1610 - train/4KW0XqY1owuJMjUShbfr.byte
s 16% 1614 - train/4l7o0adxGiDvStmFVNzb.byte
s 16% 1620 - train/4lS3L8RDheMHYfywNAvr.byte
s 16% 1622 - train/4LT5m3NUgkMhaXQJHfr8.byte
s 16% 1623 - train/4m1Gsx5JvMFgATY7L0z0.byte
s 16% 1627 - train/4md7bUD5MfKjpbq8Px0TZ.byte
s 16% 1631 - train/4MvTXueNz2rS08iER1B0.byte
s 16% 1636 - train/4NcsdhFPKLYWxq7R5AJU.byte
s 16% 1640 - train/4ngIoxi0KRZqMTjepWh2.byte
s 16% 1641 - train/4nmQkq5dyZfwPYaRbD7i.byte
s 16% 1644 - train/4Nwns693pXjobKT1HtFr.byte
s 16% 1646 - train/4o0CMefnH03xUI8XYVbW.byte
s 16% 1647 - train/4ODnoUH5ets62Na11yVx.byte
s 16% 1652 - train/4omjpRdNPCMSfxi3TyEA.byte
s 16% 1653 - train/4onTdkSFqx9VB0EPj50M.byte
s 16% 1657 - train/4P1oqBhybJiSt0a2eFTu.byte
s 16% 1658 - train/4pA56nslGDof7acmqjhw.byte
s 16% 1662 - train/4PHfjxZBM02k7I1DCLcN.byte
s 16% 1664 - train/4PLRF3BAbEtZTXnaGqfD.byte
s 16% 1669 - train/4PUzcORNYF6qx1MvHbtj.byte
s 16% 1673 - train/4qgxb0SvFILhU1RJW5Gf.byte
s 16% 1676 - train/4QlfSpex83Nsh2oBcWXv.byte
s 16% 1680 - train/4QVwiS1jTnsztAvr7GKD.byte

S	16%	1683	-	train/4Rbo7BA6vqyfMcIJhdzk.byte
S	16%	1686	-	train/4rNAjfbGERH3iKk8qoS5.byte
S	16%	1692	-	train/4S5AJVNXiIr-fZG7na0KQ.byte
S	17%	1694	-	train/4SBiOYxhVHqXWgKwtdLG.byte
S	17%	1699	-	train/4skKNvbjsHmzaWHIRQXi.byte
S	17%	1701	-	train/4sLZAv1YX7yBctu3Pkdq.byte
S	17%	1703	-	train/4sPNK2e8VEkXb5t1odf9.byte
S	17%	1707	-	train/4SyfKNtskEM1juUnXZv2.byte
S	17%	1709	-	train/4T32Ml0ctDZmHvw7JkYK.byte
S	17%	1711	-	train/4tdK0k02GsAmbrDepQZi.byte
S	17%	1715	-	train/4ToSq9HwhD3mBVuKJgdF.byte
S	17%	1717	-	train/4txzUmCskIHWrpBgLZM0.byte
S	17%	1722	-	train/4ujAGnaTyJf0ct7sEKHz.byte
S	17%	1724	-	train/4UoAE5RzKCHVNrm2YQnI.byte
S	17%	1727	-	train/4uTz3lUCb8gwtRy02Vk6.byte
S	17%	1729	-	train/4uvhHxowfXi72M8mOLKk.byte
S	17%	1731	-	train/4vA97aYWZbpPSlTuircQ.byte
S	17%	1733	-	train/4vDizRTKFYmXsLA1McnH.byte
S	17%	1737	-	train/4vr2yUw5PiEgN6BaWRho.byte
S	17%	1742	-	train/4wIxhDlpCbJVkZLfHKtP.byte
S	17%	1746	-	train/4wmslOCgr1HfavRdSZ6q.byte
S	17%	1748	-	train/4wqFVbfT1C6Z0o08GXPv.byte
S	17%	1751	-	train/4wvKaeRqlQSBEP9nLOMI.byte

S	17%	1756	-	train/4xli1n8H7ez9aR5vbPSN.byte
S	17%	1759	-	train/4xre6uX7BbkCW0LtSQdn.byte
S	17%	1764	-	train/4YBunKLTsWU38x2vQgwy.byte
S	17%	1767	-	train/4YyxU7pCh31oA0s8kHgm.byte
S	17%	1768	-	train/4z08revYiu6d7aZMTPQL.byte
S	17%	1771	-	train/4Z5WM0r1NqCw2y6tfnlb.byte
S	17%	1774	-	train/4ZMDs6aAF8hWwTJuIPeX.byte
S	17%	1779	-	train/508Uk9z1gvCtucARFWG3.byte
S	17%	1783	-	train/50iq3CaS7H1PDWetAMLR.byte
S	17%	1786	-	train/50NjMBpX8t1kG1zq3wUb.byte
S	17%	1787	-	train/50tF9JeP6UYc0rkmyMvs.byte
S	17%	1790	-	train/51qVyQZvcFaz89onA0BS.byte
S	17%	1793	-	train/52EkXpe6I1yif8SxWDY1.byte
S	17%	1796	-	train/52L3DyEep9tANInoH1lM.byte
S	17%	1798	-	train/52wpQad87Estg1jBLmeW.byte
S	18%	1800	-	train/53cwgavTtGukRl19jM0.byte
S	18%	1802	-	train/53HdP6AB7nv4SuhI9Yci.byte
S	18%	1805	-	train/53r708mKVjxzCtn4i1oW.byte
S	18%	1808	-	train/54F6nDevKmwQazX2hu8j.byte
S	18%	1809	-	train/54jSGL7mYhPV1ArX2BpR.byte
S	18%	1812	-	train/54sFSWh1V8P07qugyLwo.byte
S	18%	1817	-	train/5764RnJYTirWq1utgdkN.byte
S	18%	1820	-	train/57G0byzuhHLAajWRCJaZq.byte

s	18%	1822	-	train/57QmnCGYe0pdIy3zTgJ9.byte
s	18%	1825	-	train/57w4VeFy3QZx8PNk1rv2.byte
s	18%	1827	-	train/581bD2ouMmKpySGQszri.byte
s	18%	1830	-	train/58h7Q0perUzCyg3KHZPN.byte
s	18%	1834	-	train/596okG3UjnycItSXaQzR.byte
s	18%	1839	-	train/59voqhS867GYXDAMzmVj.byte
s	18%	1841	-	train/5a2Ivxte8k3uBCrUEwNq.byte
s	18%	1844	-	train/5a8pEcRyqVKFmYnGsTrb.byte
s	18%	1845	-	train/5aKDs6zNCymfEixZvg2b.byte
s	18%	1849	-	train/5asM8FCGzfhlJI3jBXo0.byte
s	18%	1851	-	train/5aVfOX3lym7ZECUehLD8.byte
s	18%	1855	-	train/5bahrF4B0op6EdVS3JQi.byte
s	18%	1857	-	train/5baZkrT2gEWtsLBSOQPN.byte
s	18%	1859	-	train/5birSkFE6dmgN71BPCJc.byte
s	18%	1861	-	train/5bNgRzHpdGInuxjQBvFt.byte
s	18%	1863	-	train/5BwP4dfo9Wm1csj6XGR3.byte
s	18%	1866	-	train/5ceItS0Y1RjbXrhC9Zdg.byte
s	18%	1869	-	train/5chuo1CpgextP9bqGSBW.byte
s	18%	1872	-	train/5cVTmPKBAp9L1zbh2MtI.byte
s	18%	1876	-	train/5dNsph6uJ1tkZMLzTP0w.byte
s	18%	1880	-	train/5drRtkhwUeEg1WyBCXnF.byte
s	18%	1885	-	train/5dZSWcADY0fCq8hsuJ1L.byte
s	18%	1886	-	train/5edMuN6aRW8wUFPLQnqt.byte

s	18%	1889	-	train/5e1YuDEtgZd3TLHB1jVo.byte
s	18%	1891	-	train/5ePdLKZoCV9nv0xIWc14.byte
s	18%	1894	-	train/5EwhdlmstoBPqi36ZUAN.byte
s	19%	1897	-	train/5F4mbeBZsfwYCQpGgVLa.byte
s	19%	1899	-	train/5fcdwygv9TKr13GW0zMA.byte
s	19%	1901	-	train/5fMDqaRKQBd3AhVPFIkZ.byte
s	19%	1905	-	train/5fsyaH409vx3SUXN2j0q.byte
s	19%	1907	-	train/5FwqQXsTrni4x1KD0dZ0.byte
s	19%	1909	-	train/5FyHhsqWCcxTkAS7wV1D.byte
s	19%	1911	-	train/5ghsiExd6LZpW3GzHyJ1.byte
s	19%	1913	-	train/5goj49D1YHBRu7TEU8AM.byte
s	19%	1918	-	train/5GyTHNsDkWRjvP9KM3F0.byte
s	19%	1920	-	train/5ha8CD4kXfvpQ2oGwNcB.byte
s	19%	1923	-	train/5HbKUA1oyJDeLwhM26Fq.byte
s	19%	1927	-	train/5hnmoKvt7TzDWZ4as12g.byte
s	19%	1931	-	train/5Ictjhg7KpFDfiwq4CbR.byte
s	19%	1936	-	train/5itrSmMjzBw0vcb7JgZe.byte
s	19%	1937	-	train/5j8QqdJ1fUCcNEA49VLY.byte
s	19%	1943	-	train/5Joecs2b0BauQX67HEXW.byte
s	19%	1944	-	train/5jqWz0imIyErK3t1RLv4.byte
s	19%	1948	-	train/5kDwglm7T9UjftPhpExr.byte
s	19%	1951	-	train/5KHFA0cVx0gU94RnZMNi.byte
s	19%	1952	-	train/5Kj6AhNFgwEU9aeCXT0y.byte

S	19%	1956	-	train/5LFfKQjM4c1TNgrEHpX.byte
S	19%	1958	-	train/5LFyp4jAsTSzJD0Y0lCw.byte
S	19%	1964	-	train/5lTYCzQPXqEKr1JkNMD8.byte
S	19%	1969	-	train/5mBNaMtbUno9cTkgieRu.byte
S	19%	1972	-	train/5ME0AcvDafIXHtFJxwri.byte
S	19%	1974	-	train/5MjiBxhL0nSvV9CUedQI.byte
S	19%	1979	-	train/5MSWhldbjbPpEQeqVoPaI.byte
S	19%	1982	-	train/5mVLtUwgRfWi3yeCFJsZ.byte
S	19%	1984	-	train/5MWdAsOpiyk9G0HvDSxu.byte
S	19%	1986	-	train/5nAOq0ri23gH76PDmYch.byte
S	19%	1990	-	train/5NHctFAk0LTw7bIYoq2W.byte
S	19%	1993	-	train/5NMOKeAjmuGHlDnb40QL.byte
S	19%	1995	-	train/5nPlACqyhHz3wOXiQdGV.byte
S	19%	1998	-	train/5nVHoz9ce2AmrSFbRIJg.byte
S	20%	2004	-	train/50oFBKeZaAtRYHmfIcV7.byte
S	20%	2007	-	train/5ov2DWngteLM0GVEySr1.byte
S	20%	2010	-	train/5p2yY0MvAQGXSIuocmbn.byte
S	20%	2012	-	train/5Pi7jUz9V4AaCpbG3RtH.byte
S	20%	2016	-	train/5pQS3Z7bW1c84XRhmJYx.byte
S	20%	2021	-	train/5pzfu3smYkTxHvanAwoj.byte
S	20%	2022	-	train/5PZYiLlxFp6Kgsev0zaI.byte
S	20%	2027	-	train/5QEDgGFSnVXWJLxUouBP.byte
S	20%	2032	-	train/5QkW049HFsDAiwCgKM2c.byte

s	20%	2034	-	train/5QnvTRCteUqcpaL3J7SW.byte
s	20%	2040	-	train/5QY8ZI9CpAKr2o1ctdL3.byte
s	20%	2045	-	train/5RB4kWetM0mahHf6pAuX.byte
s	20%	2051	-	train/5RLpu6X1wbBZz4cmAkh7.byte
s	20%	2053	-	train/5rtfSP91u3LJsbzMVWnw.byte
s	20%	2055	-	train/5RwWjtmMK1LiXqer8fHG.byte
s	20%	2059	-	train/5scq0VdirPRF8DxUKbh1.byte
s	20%	2063	-	train/5sJV4UyRmLc7Mx9a6ldr.byte
s	20%	2066	-	train/5StqsK6dcek7TrWPa4EO.byte
s	20%	2068	-	train/5Su9y2pdWjXihsICkxUF.byte
s	20%	2072	-	train/5t84gFkWcXUQErYsRLcy.byte
s	20%	2077	-	train/5tQ40CNKrDpPhWB7ngou.byte
s	20%	2079	-	train/5UCs70VqrjQ8wIYp1zme.byte
s	20%	2084	-	train/5Um2hXGgeCVPwdtInObq.byte
s	20%	2087	-	train/5uqM1zonRdTQbetDF6wJ.byte
s	20%	2092	-	train/5v0foj1zRqNDHerwuiEg.byte
s	20%	2094	-	train/5vAJ9y0Ceu6UEkbFSsPm.byte
s	20%	2097	-	train/5VE1zp4YCIhwq7UvuAsM.byte
s	20%	2099	-	train/5VgnChfGiXpD0tTRI3W1.byte
s	20%	2103	-	train/5vOq0aC8J1bRBVDXcmFL.byte
s	20%	2108	-	train/5wFNuSn6Gal8VXTAEimB.byte
s	20%	2110	-	train/5wkJpsMIWzDaQ4njKuy7.byte
s	20%	2112	-	train/5Xc0iAYjsG7ugxrZNIoS.byte

s	20%	2117	-	train/5xjZRvEN9KYVuar0GbW0.byte
s	21%	2119	-	train/5Xqlm9iyQrt7CeafjkhS.byte
s	21%	2122	-	train/5xzgeU2r10AauHDEjRZT.byte
s	21%	2123	-	train/5xzWZR10XkKjBTtPYVai.byte
s	21%	2125	-	train/5y4KkCXDrLHbgTwSfdqm.byte
s	21%	2131	-	train/5YFaPQwRb9Z6tMnkyvcd.byte
s	21%	2133	-	train/5yl20B1IoetGVjsQFwP8.byte
s	21%	2138	-	train/5YxvIRpAnTkW7KBDHyJP.byte
s	21%	2143	-	train/5zfkYBOKTD2NjpdI43Eo.byte
s	21%	2145	-	train/5ZmHwgpCs9J8ylMbTSKe.byte
s	21%	2147	-	train/5Zrij7tnYzFRaxI9edvf.byte
s	21%	2148	-	train/5zYVF03NXadb7Gv64PqC.byte
s	21%	2149	-	train/5ZzGq1x9jt4cWUemd62s.byte
s	21%	2152	-	train/60C4Kupa70AXWcjTQIqy.byte
s	21%	2154	-	train/60DUJgelq3snQhyT5bxml.byte
s	21%	2157	-	train/60phCeViuwkrxb1vmglQ.byte
s	21%	2161	-	train/617nfOLPiCpmMhrVNI0l.byte
s	21%	2165	-	train/61SnhUm4jdWiFCZGsPJ7.byte
s	21%	2167	-	train/623eFhxtASpOuCcQviwr.byte
s	21%	2168	-	train/624nasYWOLCkrcqB3HGg.byte
s	21%	2171	-	train/62v1di8prIYZuGkPRFab.byte
s	21%	2174	-	train/63JdbfzDMVIRZE0NUy07.byte
s	21%	2177	-	train/63vS2seA0bRQhG4fugWE.byte

S	21%	2179	-	train/64pFyk0YIsza0wGx2iMA.byte
S	21%	2187	-	train/65cjJpPCUQiLDRyXfWd4.byte
S	21%	2193	-	train/65PjxA3a1NeH8iCBKpfS.byte
S	21%	2195	-	train/65ZYCso0aNJ3fht1Dx4F.byte
S	21%	2200	-	train/68icCkFVRJK5UtxL0qIu.byte
S	21%	2204	-	train/69aKs7pQZjI2tWN80bHy.byte
S	21%	2206	-	train/69jlRk8g7VH4hXypdwU0.byte
S	21%	2208	-	train/69MHioIdGcUTvxEOXJyw.byte
S	21%	2211	-	train/6a3bdiu9jyrXPQHI81pJ.byte
S	21%	2215	-	train/6ADJCKZ35PmTsEnzRINX.byte
S	21%	2218	-	train/6aJfSuh4VDxr0tjGqviC.byte
S	21%	2219	-	train/6AORky8bx20SVDHtdow7.byte
S	21%	2226	-	train/6BIMrSNvawLAUTejQfEJ.byte
S	21%	2230	-	train/6BuxhyM8GH4bfpVrjOzE.byte
S	21%	2232	-	train/6ByMewHIzS0LpsWdXZU1.byte
S	21%	2234	-	train/6c5KqjIDH74gkz0N2V8B.byte
S	22%	2239	-	train/6CegKJDHWYrUi3oEQZxs.byte
S	22%	2241	-	train/6cJ40V1TGifn92SCPNEA.byte
S	22%	2242	-	train/6cIgd5HFw0tKkaLbq9AU.byte
S	22%	2245	-	train/6Cvyc7zMbZq5DfJprmXP.byte
S	22%	2252	-	train/6dgTKemUkyWupoBha0ZV.byte
S	22%	2255	-	train/6DmCqAs0tFKxGe5aPz1U.byte
S	22%	2259	-	train/6E78GVaCMig3mTz1oFb0.byte

s	22%	2261	-	train/6EHP7m4Xj5xUeTzpqZF0.byte
s	22%	2265	-	train/6Ep5PZ3J09edDNo4ahx8.byte
s	22%	2267	-	train/6F3Dj4qM15tKvNTEoQlP.byte
s	22%	2270	-	train/6fmGMN2PtKoAR5XkwZaj.byte
s	22%	2272	-	train/6FPAEXoHTYVfc4ezxKkQ.byte
s	22%	2274	-	train/6fpDY54F1CsirwcZnTeR.byte
s	22%	2278	-	train/6G0EdXzZTPkbhoucn8RK.byte
s	22%	2281	-	train/6GgdCXPL2Bali0f8ZFkA.byte
s	22%	2285	-	train/6GijzkZEy1L1BArf2Q00.byte
s	22%	2286	-	train/6gJf1hQLyeE3K29BSIra.byte
s	22%	2292	-	train/6GXdqZuQkcfbNyJA9gpx.byte
s	22%	2294	-	train/6HcXhAzQiBMUK5en92FY.byte
s	22%	2299	-	train/6hRaK4DkcXoZfI32BzH0.byte
s	22%	2301	-	train/6HTiZ3c0vd2PfC89GRsN.bytes
s	22%	2303	-	train/6I2mCQyaF10JNEkvqHcZ.byte
s	22%	2306	-	train/6IrZu0ynlqf4otPBA7HT.byte
s	22%	2307	-	train/6iSL1WkDHVIUqJzGmnhN.byte
s	22%	2314	-	train/6JIAcqDxegQoLzdNM91w.byte
s	22%	2316	-	train/6Jiw0HZgfQ8PvGCtsxDS.byte
s	22%	2317	-	train/6JOiKa5LrAtu42TpY7kI.byte
s	22%	2320	-	train/6Jylg0kdK1x3sMbihLvw.byte
s	22%	2323	-	train/6KauYyxGt1n8wViqbe5X.byte
s	22%	2330	-	train/6LGpjCQ10TJ1vxRH3m97.byte

S	22%	2335	-	train/6L0gJTo7YduN84r1vPk9.byte
S	22%	2341	-	train/6McvAWZj4285L7XPrwnE.byte
S	22%	2346	-	train/6mGvPyTaWit3wcogY9j0.byte
S	22%	2348	-	train/6mKG5UgpFBOYCdR3Wz8J.byte
S	22%	2350	-	train/6m1M0bsRQHnPwUJkEvN4.byte
S	22%	2353	-	train/6mUHQtcBjzWA0fGIEnP7.byte
S	22%	2356	-	train/6n5xSkDHdUL9BpNwjzfa.byte
S	23%	2358	-	train/6nC13JdMAkKjX0goN8U4.byte
S	23%	2360	-	train/6nfXBNI5d4j0kgmEDF31.byte
S	23%	2363	-	train/6nLCYyu3cTMDSXWxbj4o.byte
S	23%	2366	-	train/6NTZfuU347Ecxcj90rigy.byte
S	23%	2369	-	train/6nw2iAevTqXH1fsatcp7.byte
S	23%	2373	-	train/6OdWLjyZQzgchY2XST1.byte
S	23%	2375	-	train/6oiK1LC8JYenZGc7kgs2.byte
S	23%	2377	-	train/6okBRVQNHGdWXCzgK7MJ.byte
S	23%	2380	-	train/6p0xPcyrDB4R8YNHEsnU.byte
S	23%	2384	-	train/6PG57gUmE2rLebCBTMno.byte
S	23%	2389	-	train/6q2Y5B0x8Ro1uXjA10se.byte
S	23%	2391	-	train/6q9PF1AdMc0IBjhi58CY.byte
S	23%	2393	-	train/6qhjHRGpSb27PvtQ3csx.byte
S	23%	2395	-	train/6qLobnwUp07W1RyKBjvZ.byte
S	23%	2396	-	train/6Qp2b3wMFnzhlKjxLJrk.byte
S	23%	2398	-	train/6QtaMBu0P4kEjTgcxYAn.byte

s	23%	2401	-	train/6qwhD5Z3bsdFkyAzgxQ0.byte
s	23%	2405	-	train/6qzsP5lZoCW3KHIX9Uf7.byte
s	23%	2406	-	train/6r02dN7YEOjsfTQkz1Gh.byte
s	23%	2408	-	train/6R5rAcs0297BDkT1e38U.byte
s	23%	2410	-	train/6rBCZSJ29AP1VQtkwz7u.byte
s	23%	2416	-	train/6RQtX0X42z0e1TDaZnvi.byte
s	23%	2419	-	train/6s3HbFTuzMP9oSwAp2JX.byte
s	23%	2425	-	train/6Soz1jDVE0mYv0wbty7Q.byte
s	23%	2427	-	train/6SpFHDUdJq18rbcZ0YIv.byte
s	23%	2428	-	train/6SQwZ8l0jKCkh2ErRqJP.byte
s	23%	2433	-	train/6T907yrYp4XJsGPK82Kh.byte
s	23%	2435	-	train/6teIPpCrYKw0zFusDSix.byte
s	23%	2438	-	train/6tgGScmbuHN10Tx0QI7r.byte
s	23%	2440	-	train/6tIbeDRa0jfnZYcFqVdL.byte
s	23%	2443	-	train/6tmqe5B8JDsQnyhGo7vP.byte
s	23%	2445	-	train/6TUw1J5DR47dAgmjHLOs.byte
s	23%	2453	-	train/6U7aMPme14gBDkRCnt3E.byte
s	23%	2457	-	train/6ux5VnvYNqdiRP8ZWI92.byte
s	23%	2458	-	train/6v3n9N00dMkCVBA5FI2.byte
s	23%	2461	-	train/6V5BQcGzN4XIv1DCTpPh.byte
s	23%	2464	-	train/6vCzcIJSuGOTmDP2QVLF.byte
s	24%	2467	-	train/6VNXsCD0lueo1nw3K05M.byte
s	24%	2469	-	train/6VTkgzifPd5Z0BXF3cuY.byte

s	24%	2472	-	train/6w02fn5UqNDMd1tISWKc.byte
s	24%	2476	-	train/6Wd9umcsexZNDP7IfJnq.byte
s	24%	2479	-	train/6wHyQ3DVLMBtsxnhGvmo.byte
s	24%	2481	-	train/6wj5FcD7oghz1b0eVIXC.byte
s	24%	2484	-	train/6WoydrPp0u9ZgiwaKnY1.byte
s	24%	2486	-	train/6wQzVGgRDpw5ZuAM00Us.byte
s	24%	2488	-	train/6Wu9TqbGHMPOCwpZNIA5.byte
s	24%	2491	-	train/6XBvnPoFpYk8S3lbCjwW.byte
s	24%	2494	-	train/6XM2qzI9n1mZf0tu8Qlh.byte
s	24%	2496	-	train/6Y7KC0nLeOUjH9mpEtFh.byte
s	24%	2498	-	train/6YJyE0SvstChMfNp1TbG.byte
s	24%	2502	-	train/6yYp0scglSrPHdnh2wJ1.byte
s	24%	2506	-	train/6Zf39ohpMtgQIBFcjKXU.byte
s	24%	2511	-	train/6ZP1VR2mkrjTtxJvS0h3.byte
s	24%	2514	-	train/6ZUjzliNqdfmES5WIgCs.byte
s	24%	2517	-	train/6ZY9A10hSbQBGF1RcLVH.byte
s	24%	2522	-	train/70xE1cmWgS1fChNpniIG.byte
s	24%	2528	-	train/72cG6vkEjAuFBzym5P4e.byte
s	24%	2530	-	train/72KpNMPtbRARxQLszX49.byte
s	24%	2533	-	train/73cMCHeUwStZjy2VoQRm.byte
s	24%	2535	-	train/73VYXPJgsRlKIMiOLwG4.byte
s	24%	2536	-	train/73WUpI48k9P6lrKFCHiV.byte
s	24%	2540	-	train/74SYbIefnAEJpGHZq6md.byte

s	24%	2543	-	train/75DeYlyFgVjhbZ8vSTRI.byte
s	24%	2547	-	train/75qSAohLWYb60Hs24NUF.byte
s	24%	2548	-	train/75xUGovrLMWa4tI0XJNV.byte
s	24%	2550	-	train/7614qKWBOGwx5N8dcgfA.byte
s	24%	2558	-	train/78mbJ90KHRUSCPTk5h1w.byte
s	24%	2559	-	train/79RN6Du2CAiqaOpStbJB.byte
s	24%	2564	-	train/7aEbXfVnzhG6dRMLT58q.byte
s	24%	2566	-	train/7AhtpZDSxMjVEIs8l09W.byte
s	24%	2570	-	train/7AOuExQwmYLvhqZ2T1cN.byte
s	24%	2573	-	train/7aV6QqpmJ100MfNsAw9k.byte
s	24%	2575	-	train/7BFRAjIPKrv3QVbkDx9C.byte
s	24%	2578	-	train/7BNLHFxEAjRS1bwZxzizq.byte
s	25%	2582	-	train/7C3msQWn96F1YTj1wGSc.byte
s	25%	2587	-	train/7cdWTzVjHIh10Mtn85fb.byte
s	25%	2589	-	train/7cnyTvamsXhUi5IdS1Z4.byte
s	25%	2592	-	train/7cUfx46SVDjs9X0qZ1mi.byte
s	25%	2595	-	train/7cyYPthi5KmvXEs2NkRj.byte
s	25%	2597	-	train/7D92cJaF5QIdpA10EhSt.byte
s	25%	2604	-	train/7dLqHjDaurgEQWxXYkyP.byte
s	25%	2609	-	train/7dW2EIJqY86gKjPbtCkw.byte
s	25%	2611	-	train/7e0vxYy6b24g903UKLqC.byte
s	25%	2616	-	train/7EeBbm5jxfwut6nPD0iF.byte
s	25%	2617	-	train/7eicxTZ3pDPf6501RLXg.byte

s	25%	2621	-	train/7EWa5VQz2NB4MgdPCtTp.byte
s	25%	2623	-	train/7Ff4uMQibABUks59xnpV.byte
s	25%	2625	-	train/7FmNSjrBDkVI8g21bHpC.byte
s	25%	2627	-	train/7fTL11UKHqa9rXYuzxgZ.byte
s	25%	2630	-	train/7fXZ2Kd5Ak1IjHo04g9Y.byte
s	25%	2632	-	train/7g1AQZVLIwn8STsYdW5e.byte
s	25%	2635	-	train/7gbpotqxaYGmMzEs2Uyc.byte
s	25%	2639	-	train/7goisLFSJXEZp1kM3rej.byte
s	25%	2643	-	train/7HAqkJSVFhjle8tn412s.byte
s	25%	2645	-	train/7hm0veuHiBCFWcEMPaNd.byte
s	25%	2651	-	train/7HqBvpZyaDWLXsNSAhUi.byte
s	25%	2653	-	train/7hsb8evACNndQgJPSR61.byte
s	25%	2654	-	train/7hv03rkiLUtM0QpfWgn4.byte
s	25%	2656	-	train/7HYhXpwdBAxL6cGn30v0.byte
s	25%	2658	-	train/7iBwyZQ5kxGEjPN9fThM.byte
s	25%	2663	-	train/7IOrsrJqfjV0ixuTbGF9.byte
s	25%	2668	-	train/7IuEmVGSFBQqKrjhcWvn.byte
s	25%	2672	-	train/7JjQBd2CAyvtgY1SL0e3.byte
s	25%	2673	-	train/7Jjs0ZzyXR9ScewIoMTp.byte
s	25%	2676	-	train/7K1zacbZA3HIjStR6k0s.byte
s	25%	2681	-	train/7krBiEAcFbZxTJsah04S.byte
s	25%	2684	-	train/7KVtWPn30iJpvBkIej4o.byte
s	26%	2688	-	train/7LbrmXchDZJled5HnAfM.byte

s	26%	2693	-	train/7LWYyj9uaCoVqJiEck62.byte
s	26%	2694	-	train/71YJNX4TjapibMyvd9sz.byte
s	26%	2696	-	train/7m5GMT1j6VFifne9D8Sq.byte
s	26%	2699	-	train/7mHG3hRAz15I8CSO9Kej.byte
s	26%	2703	-	train/7n2kiMZ9awS0xbNG5XIz.byte
s	26%	2708	-	train/7NhfmldVgt912wn8MYDs.byte
s	26%	2709	-	train/7nL2QiAIcsRxfMHwyTeB.byte
s	26%	2713	-	train/7nvHG46BM95P3SEW1ezy.byte
s	26%	2715	-	train/7NyOEfMq1n1saVHhbpQw.byte
s	26%	2718	-	train/7obPSDvanqmeT4Wiuw9Z.byte
s	26%	2720	-	train/7oKB60iGX8LStVwnqlQI.byte
s	26%	2723	-	train/70oFgDGrS1hqkueAmTC.byte
s	26%	2724	-	train/7otn6PiHDgN4GmayhT2R.byte
s	26%	2726	-	train/7OZ8auLijd26EwXcIGHo.byte
s	26%	2730	-	train/7pdzUjNQPFyw41B0a8gR.byte
s	26%	2737	-	train/7PtsSV1FrYgEfGpaqOn8.byte
s	26%	2741	-	train/7qBpIPJl0DvXnuWAMr6i.byte
s	26%	2745	-	train/7q1MUcpafAL3QFiPWEzJ.byte
s	26%	2746	-	train/7QNAUrmZEql3XR1I42jz.byte
s	26%	2753	-	train/7qvZykasB4CizHm8EIhT.byte
s	26%	2755	-	train/7qXRGMD4gi8VkiWSjyeF.byte
s	26%	2757	-	train/7r3gQfJxjyIdX1YloWuR.byte
s	26%	2759	-	train/7r5DCxQ2nJKN0GOYPmcH.byte

s	26%	2761	-	train/7rcIYafVLBJQ1Hwy2gD0.byte
s	26%	2763	-	train/7RfoFlxTSDhLmqre0zpg.byte
s	26%	2767	-	train/7rLEfBjbiyXDU0Aewa4P.byte
s	26%	2770	-	train/7rvpSAIkeV8TKQDi0lw5.byte
s	26%	2775	-	train/7SAUNTQf8RMmL904jxID.byte
s	26%	2777	-	train/7schXv12FHSujB3aVK9e.byte
s	26%	2779	-	train/7sNR0UFCQniz8vbhVpIo.byte
s	26%	2782	-	train/7sXGVmyjMe1bc3LdOnYx.byte
s	26%	2788	-	train/7tDpNPXUISleJCKzQjr9.byte
s	26%	2797	-	train/7UFMw39SLs8Qn4RuIY6W.byte
s	26%	2800	-	train/7UJTAcmd85jRz0VitSof.byte
s	26%	2802	-	train/7upD1xywIe0mibLNPSO4.byte
s	26%	2807	-	train/7Uzm5a3KiuTxZPwt8sRb.byte
s	27%	2809	-	train/7v8gE2jK9YmzBFyD6xst.byte
s	27%	2811	-	train/7vCGe0MVocZwq5ltXbh8.byte
s	27%	2813	-	train/7VE6hScuodxAvTp0Nrnk.byte
s	27%	2816	-	train/7vIR8Bg5aZuwnkXmC301.byte
s	27%	2818	-	train/7vjsW1X3rQL1dNenwTxV.byte
s	27%	2822	-	train/7Vt2ALvgX0yw3naSxcpz.byte
s	27%	2825	-	train/7vXx2gNhaZedKUtIS4mP.byte
s	27%	2829	-	train/7w95c0pt83SLYuHBeRlG.byte
s	27%	2830	-	train/7w9HgKIImjx0cYnea243.byte
s	27%	2833	-	train/7WkHb3TOoJ4XRYnB1tNS.byte

s	27%	2836	-	train/7wpdECAtSos0ImULRkQW.byte
s	27%	2840	-	train/7xeFLDBbNVWZiM1u2S5t.byte
s	27%	2841	-	train/7Xl1KHeRiAYN4x62hdtD.byte
s	27%	2844	-	train/7xwMNplsj3EhBJ0Zkd9Q.byte
s	27%	2846	-	train/7y28ZxJs9SKVYom13vbA.byte
s	27%	2852	-	train/7YJU9tqj1c14vwrKG053.byte
s	27%	2854	-	train/7ykVQP3IzRgKpN95o1ma.byte
s	27%	2856	-	train/7YrZK0PHmtnEOMfbCdNG.byte
s	27%	2858	-	train/7ywK5kNsz4EXe2xS3acM.byte
s	27%	2860	-	train/7z031VS5IJ8w0CvQyFdZ.byte
s	27%	2861	-	train/7Z01Lsb5Pr6wNjzRDTdH.byte
s	27%	2864	-	train/7ZCB4m2QyOvGaNfUls5q.byte
s	27%	2866	-	train/7zIMcQ35pnFfgE9qr6vs.byte
s	27%	2868	-	train/7zr2DsIWVvgJmx3ik4L8.byte
s	27%	2873	-	train/80G5Qt0zNdR9CVxTJ3Uv.byte
s	27%	2874	-	train/80gD69r3peF1TAGtXzu7.byte
s	27%	2878	-	train/80Zf05UGkLI4Scj3eQwh.byte
s	27%	2881	-	train/81IgT57cy1MUNGbLkiqm.byte
s	27%	2884	-	train/81LeDy7sJGfSQHFAh0cp.byte
s	27%	2887	-	train/81u1Ad7ac20I3FRTXeYr.byte
s	27%	2888	-	train/81ZbxYRAu0SEJOj5NsgC.byte
s	27%	2893	-	train/82kLgCGIRNXduWs1SrKh.byte
s	27%	2897	-	train/82rMDR053qpfnIL4Hi1Y.byte

s	27%	2899	-	train/83dL70vsfp9yCQTMehHD.byte
s	27%	2901	-	train/83up2hzQfPn5BcFaRwX0.byte
s	27%	2904	-	train/84jKcbktfshl69TiErUG.byte
s	27%	2909	-	train/84ViQ93gyPruDSoOn7zw.byte
s	27%	2911	-	train/85K7ypWXba0UjmuxDEz9.byte
s	28%	2916	-	train/85qyLrVc1vhsKCfnNXmx.byte
s	28%	2918	-	train/85SPpVLIiAYX3dqKvyar.byte
s	28%	2920	-	train/85X9KSMHj4TODcx6FAvW.byte
s	28%	2922	-	train/86kMhQHlXa1UW03ifeg2.byte
s	28%	2925	-	train/86QrjZewznD2W3VhpRbm.byte
s	28%	2929	-	train/87FTRrZLn1WozgdYvbxQ.byte
s	28%	2931	-	train/870EybjJHlup9AavITYCg.byte
s	28%	2934	-	train/891Ymzl3VvoJLAMXh4p0.byte
s	28%	2935	-	train/894ejoK3q5aZurSAzUJs.byte
s	28%	2939	-	train/89VdQqEkGB2Fzv4lXnbU.byte
s	28%	2945	-	train/8aBc6ytRQ4oY1Pw1KDhq.byte
s	28%	2947	-	train/8AHh4WNQ5qYokKIXcmM7.byte
s	28%	2949	-	train/8aMHw0Y05ILJ97eduP1C.byte
s	28%	2953	-	train/8bi3cZxuXAhLaoDEKenJ.byte
s	28%	2955	-	train/8bNCQU7PMD6lotyJixVw.byte
s	28%	2957	-	train/8btcXEfH4v9ITwU70Ssg.byte
s	28%	2958	-	train/8BYckA0Zu47Rjm3bgUzp.byte
s	28%	2963	-	train/8cM57th1BVwar2szTENJ.byte

s	28%	2966	-	train/8d3ZxEPril4pz1oTH9Jm.byte
s	28%	2969	-	train/8djFkIgPK56XtYGDsb7R.byte
s	28%	2973	-	train/8dW2hOcgRfSTnapKVHw1.byte
s	28%	2975	-	train/8DxQhtuOKPRZAJgy6Y3E.byte
s	28%	2977	-	train/8E2PuBnfWTCXFGJ7RZcV.byte
s	28%	2980	-	train/8eiOKmj1ER6wzul0bpT5.byte
s	28%	2982	-	train/8ELU07KQbV3FndjHtxON.byte
s	28%	2986	-	train/8F32X71gy4wZrtvmNA6d.byte
s	28%	2989	-	train/8f4bnEXcVDiPwJMU3Yge.byte
s	28%	2991	-	train/8fcWqDA6KsarzG3E9YJ7.byte
s	28%	2993	-	train/8FgPe1bcMsCH1qLVtWKk.byte
s	28%	2996	-	train/8F0KNb1VGi0hYnc2u41s.byte
s	28%	2998	-	train/8fSXU1iqE7jWbB4yne2g.byte
s	28%	3004	-	train/8GAc1VdPixk7WN9gm5F4.byte
s	28%	3009	-	train/8GilsX4Uk7xQdmz3wgyP.byte
s	28%	3010	-	train/8GoaOMPuBA3slq2J4TEd.byte
s	29%	3013	-	train/8gvOauLJb39wNfdAMDt6.byte
es	29%	3018	-	train/8hjiWlatZGceSXoTx14K.by
s	29%	3026	-	train/8I5jZWSKh1V2k9HqpOuB.byte
s	29%	3027	-	train/8i6m0aVAwsdSY2FEfU59.byte
s	29%	3031	-	train/8inLjyQfkReMHmNUE4qg.byte
s	29%	3034	-	train/8iZw1cMaPBr4RJzx1FXH.byte
s	29%	3039	-	train/8JButIKigxSbR9cZ75qv.byte

s	29%	3045	-	train/8jN4P37vE9R0TiHoZaDG.byte
s	29%	3048	-	train/8jrxyaQiGJb5qcs0mMAv.byte
s	29%	3050	-	train/8JUVDQ2tjPlZSRcYpzrw.byte
s	29%	3054	-	train/8Ke0IvQaxRdsU7mu6CrL.byte
s	29%	3056	-	train/8kORlw25hI69eu7jEfVy.byte
s	29%	3059	-	train/8kS1XpfKnNrUbJG60sMD.byte
s	29%	3063	-	train/8KZ4WQGTEfJYnONerwbV.byte
s	29%	3065	-	train/8l5jU3sPg4MomRFwL9zT.byte
s	29%	3070	-	train/8LmZG0EjyfCecP73osAh.byte
s	29%	3073	-	train/8LrpcC15nHjUDGu7Xio0.byte
s	29%	3076	-	train/8LtljiSurqc9QVgkBPwM.byte
s	29%	3078	-	train/8LVijvDmg32TP01b4qAX.byte
s	29%	3082	-	train/8LyrjEcsXh1MGd4wg3eS.byte
s	29%	3084	-	train/8mAqS5PYhyIwls9t1r0Z.byte
s	29%	3088	-	train/8msTxABp5MqdS2PcyCQ7.byte
s	29%	3090	-	train/8mWh570SFDzATaLEeVH6.byte
s	29%	3091	-	train/8N0Drw6gKsimauCZQpWV.byte
s	29%	3092	-	train/8N0hVC6RFuGAdxJ4keKi.byte
s	29%	3096	-	train/8ne26YagdHzwFQtIpTVL.byte
s	29%	3098	-	train/8NJoTALeEyqVa7GpwKFO.byte
s	29%	3104	-	train/8NYXztOroeRpIaS5BdM4.byte
s	29%	3106	-	train/8odIrLKhPvcDWNCVasU1.byte
s	29%	3110	-	train/80QSaZjs2vGYXDg9nbJB.byte

s	29%	3112	-	train/8ovZYfGnP2Htekc7gizT.byte
s	29%	3116	-	train/8pfbAyu34zotUCZwsv2G.byte
s	29%	3120	-	train/8PZWALmTpNjqr3Kt1XFz.byte
s	29%	3122	-	train/8qaEOKmPjZctM6lYu1AQ.byte
s	29%	3123	-	train/8QaobvZKqje1lCnyrXwp.byte
s	30%	3125	-	train/8QoDEdi6Yw2kUmzuNAcp.byte
s	30%	3128	-	train/8QRIkDJS4bgy2nEvMhfa.byte
s	30%	3130	-	train/8Qzo4j0J7MdkR5ZyErVH.byte
s	30%	3134	-	train/8rD4bFYzXTVB5IPH1Mdv.byte
s	30%	3136	-	train/8rKwm13nGVhWAiUjDH19.byte
s	30%	3141	-	train/8s9r3Mxpemd5ySQwP1D7.byte
s	30%	3146	-	train/8sPhMUEiAa3l5NDIcXZn.byte
s	30%	3149	-	train/8swNnb7BmL6dTVPcZK13.byte
s	30%	3152	-	train/8TAOkFtnREagx32HGdKq.byte
s	30%	3156	-	train/8THfDA2iXQdjNx4Lo6ws.byte
s	30%	3157	-	train/8tsSDovBygRFMTwj1C0k.byte
s	30%	3159	-	train/8U7ztLHDNnSmg3f1Qr2A.byte
s	30%	3161	-	train/8UhJWnK6pLDHAP4gt02r.byte
s	30%	3165	-	train/8UNxCZi37AdHnMXy5Pa9.byte
s	30%	3167	-	train/8uvPtcIkb9nRhqUGAzg5.byte
s	30%	3169	-	train/8Vaqh6ofRjLcy2CwpEHZ.byte
s	30%	3172	-	train/8VGXzLxvhH0qIltWc35U.byte
s	30%	3177	-	train/8VlJEcmXepIy9RS3GMTq.byte

s	30%	3178	-	train/8vM01AULDJyi3PK0SWg9.byte
s	30%	3180	-	train/8VrlyPxXiEU2Bwg5ce3Q.byte
s	30%	3181	-	train/8VU3wPKq21BY5bvnHQMS.byte
s	30%	3185	-	train/8w7AFgTmC3buQSGJ5yBR.byte
s	30%	3189	-	train/8WHGOMuhtDdw69FACxEJ.byte
s	30%	3193	-	train/8wMGYTWedtUEPcHJlqa6.byte
s	30%	3195	-	train/8WnPikqEpG1sOmLyBiYQ.byte
s	30%	3198	-	train/8wsldXc6EzeFC4TNn7Lk.byte
s	30%	3201	-	train/8x1oTUZgX2JaQtqBYNld.byte
s	30%	3203	-	train/8xeA1WMrofFtmNvhKaYJ.byte
s	30%	3207	-	train/8XwCLB6D3cMFWT9dtKSA.byte
s	30%	3210	-	train/8y5qQ2Z1pnLCOFWRJMA4.byte
s	30%	3213	-	train/8ygvhiWxIabopwJnXHS9.byte
s	30%	3215	-	train/8YobsnPDUwJJeEHmKB2ZQ.byte
s	30%	3217	-	train/8Z0i2knypSb6crfgaXmB.byte
s	30%	3221	-	train/8Z5zugC0x3cTB6ikpbSq.byte
s	31%	3225	-	train/8ZFKwPESYC9jUzdngyqr.byte
s	31%	3227	-	train/8ZImDg1PqC9RSok2WxQ1.byte
s	31%	3233	-	train/8zvicWu10kG2PJRo9pA3.byte
s	31%	3235	-	train/8zZyd3OPxpcw1NCRJBjo.byte
s	31%	3237	-	train/90bDVByGTFo1h5XW1JiU.byte
s	31%	3242	-	train/92b0mtQduNgIJA5YnFDU.byte
s	31%	3246	-	train/92xGBknJqLXYFcgUmjV1.byte

s	31%	3248	-	train/93DFQGKTLxy8Ej6kw1Pu.byte
s	31%	3250	-	train/93lsPQvIXLHNfCyY2nVr.byte
s	31%	3252	-	train/93QCNLn5FD2pueg7GdXB.byte
s	31%	3258	-	train/94bEZMkLjDp1BXoucFVI.byte
s	31%	3260	-	train/94gomdyfxJAh6lepszW2.byte
s	31%	3264	-	train/95043pwieMPvd7nkf6HN.byte
s	31%	3267	-	train/96q38IGMtXTJOalbAoK2.byte
s	31%	3272	-	train/97SF640J3ZDy2Q1aRUjP.byte
s	31%	3274	-	train/9810xmWynXIJEiSTGhMf.byte
s	31%	3276	-	train/98gRXLdAIUwOG0eor7F1.byte
s	31%	3279	-	train/9AF16gbmVOB8zWPISnX0.byte
s	31%	3281	-	train/9akqIEt8KZy1piAW3gbB.byte
s	31%	3283	-	train/9AN1xYGCFidTnrJc5RUQ.byte
s	31%	3286	-	train/9As0CbJPKyq8MFgmNc17.byte
s	31%	3289	-	train/9aVMjhASKbL6dHF7rfkX.byte
s	31%	3294	-	train/9bNuq0hnWLms5kYjHVoS.byte
s	31%	3298	-	train/9bx5RUkH6CZzoS70XmIY.byte
s	31%	3301	-	train/9CHdKeRJpZ3jnIgxkMB2.byte
s	31%	3307	-	train/9cuvUrSgBCIq6L3p0He7.byte
s	31%	3311	-	train/9d3mFhrqe5gtuLE0ACOf.byte
s	31%	3316	-	train/9DiHIregYkyCL31zNvtJ.byte
s	31%	3320	-	train/9DrhcFgQZV16jndTMEuK.byte
s	31%	3321	-	train/9DvQRmFzYk1IuTEtKU0c.byte

S	31%	3323	-	train/9dYh3w02WUFAH5f8VyKQ.byte
S	31%	3326	-	train/9Ebq0Q5sCRtSVzg3DyMn.byte
S	31%	3330	-	train/9eKQwWVFlonXmSTaG5jH.byte
S	31%	3332	-	train/9EnBSdaVmKxjteDRPJ2U.byte
S	31%	3336	-	train/9f0vLRPdeQOU6V8Dgs3k.byte
S	31%	3337	-	train/9f5wo2qZsPe3YWAvyJjt.byte
S	32%	3342	-	train/9Fj7xr8o43KwHZ6nzIqC.byte
S	32%	3346	-	train/9ftj5a4V8hCr3pTAUiGl.byte
S	32%	3352	-	train/9gdaQ5EJP1AWw3cU18Cb.byte
S	32%	3354	-	train/9GJejM0ZvT8N12gEknId.byte
S	32%	3357	-	train/9gpG7YV3EBUsIXL6wDNP.byte
S	32%	3361	-	train/9hd6DsViftagyG02NCKe.byte
S	32%	3365	-	train/9hibmvMuBATOs1Ukco0t.byte
S	32%	3368	-	train/9hu54PH0G1qoxEmT7AMb.byte
S	32%	3371	-	train/9I6UftSTGjwB8pxQY0MX.byte
S	32%	3373	-	train/9IF01DNV0XYi4mr3yzUs.byte
S	32%	3379	-	train/9Iv1TmfgCB6px7HXE0ho.byte
S	32%	3385	-	train/9JDipK3MheU4cHqs8I0x.byte
S	32%	3387	-	train/9j0k8Tn5JMt4qcXmplfx.byte
S	32%	3392	-	train/9KBhpLPe6o13q0QHMjk4.byte
S	32%	3397	-	train/9kMOTJ1I0c5aU6xzYhXg.byte
S	32%	3404	-	train/9lIdfH73S6PZrp54gVwA.byte
S	32%	3407	-	train/9LQfr8w371dNWSsRI4Xv.byte

s	32%	3408	-	train/9LRnCtUy8FI0IagGPe6B.byte
s	32%	3412	-	train/9lZ3LpNTJbEWGhzMrsYU.byte
s	32%	3418	-	train/9mfSC136RIG0udNwqyKM.byte
s	32%	3420	-	train/9mHqfWkgcJb7utKZ5rCV.byte
s	32%	3424	-	train/9mKvw8t3d6jRVfnzsDIJ.byte
s	32%	3427	-	train/9mNHweEF0uTPbRcyLMG6.byte
s	32%	3430	-	train/9mvaSjcLt435elds1z8B.byte
s	32%	3433	-	train/9N06grZ2fCsGxpvciaoQ.byte
s	32%	3436	-	train/9NeB300anTrd71cizjRS.byte
s	32%	3441	-	train/9nNFOagEhDXlmgLwIRJt.byte
s	32%	3443	-	train/9nR0ZHP7X2YJhQu3Nzsa.byte
s	32%	3445	-	train/9nvIAP8UiY2GxMe5ptSJ.byte
s	32%	3448	-	train/9oCBgaw6WUrrjtbSd5liq.byte
s	32%	3449	-	train/9OE0MJmbVG6FtHC7Uesg.byte
s	32%	3451	-	train/90PGQgfxI4ZcybTsJo5E.byte
s	32%	3454	-	train/9PcpS0d47xHb5aeCVUsQ.byte
s	32%	3457	-	train/9PjuhOrlRqxHeQcEBYMX.byte
s	32%	3459	-	train/9Puj015SAeBy6EZNpanJ.bytes
7D5Rs6XAkGfvuIctZ.bytes	32%	3465	-	train/9qK
	32%	3467	-	train/9qmgZWJTn4jFhKy3LYdM.byte
s	32%	3469	-	train/9qPu1xikDdTmLtB4J0FE.byte
s	33%	3473	-	train/9RaVK3pIfnECFNAkvST6.byte
s	33%	3475	-	train/9RM3upgzNwZhFEPjqecG.byte
s				

s	33%	3477	-	train/9Rp4Mie0LyFmqbdU18oG.byte
s	33%	3480	-	train/9Rwu8Es7InNiLjDkQsXd.byte
s	33%	3485	-	train/9s8NSuY5cHirbLmaWtVe.byte
s	33%	3487	-	train/9shmYlDbkcpt8go6RWvz.byte
s	33%	3489	-	train/9SOPAy1Kjf2HnxEhRC6I.byte
s	33%	3490	-	train/9sPtSV8rx5WNUukQ4q1n.byte
s	33%	3492	-	train/9STjspcOKbrE80PFVAoq.byte
s	33%	3496	-	train/9TjEU3e5urYx2Zwh0My8.byte
s	33%	3498	-	train/9toXzFrSARZ0TJGKMbV6.byte
s	33%	3501	-	train/9U2dK1jG06Y8kNnPZWM5.byte
s	33%	3505	-	train/9ugQ8rA0fRaS0dVHbsYM.byte
s	33%	3507	-	train/9uiRm8ZFyDwsIQegUH7o.byte
s	33%	3510	-	train/9UvoPgMf0tnp6KA7NiCu.byte
s	33%	3514	-	train/9vP6c42rE30zVN1T8AJ0.byte
s	33%	3518	-	train/9vwWB1yIHR0GcA0pU37.byte
s	33%	3524	-	train/9WMOnQwLKEcCFyX6fs4j.byte
s	33%	3525	-	train/9wOVdUL6r0SmjE1kAMxW.byte
s	33%	3533	-	train/9XgT0w1fq2RcJ70xFDhs.byte
s	33%	3535	-	train/9xo84PyBYen5uj1XTiZ0.byte
s	33%	3539	-	train/9XZWSgoYzHci4Q2INbdr.byte
s	33%	3541	-	train/9Y4nW1Ko6FcNVtOG7qTQ.byte
s	33%	3546	-	train/9yEnJoLOWURsBbt2FivP.byte
s	33%	3550	-	train/9YkUqjMJE6d1yxtgs8FW.byte

s	33%	3553	-	train/9yRz8iAj15I2ueLm4PtW.byte
s	33%	3558	-	train/9zAFLHoWJDERpSP3y5Mu.byte
s	33%	3559	-	train/9zAJOr1DSPhRkcHxVB7p.byte
s	33%	3560	-	train/9zbitD6NE10I4FjdgJvV.byte
s	33%	3564	-	train/9zgKyBYkuCJTjlSc7DNA.byte
s	33%	3566	-	train/9ZLhF14N2dUJfDQe0jRS.byte
s	33%	3569	-	train/A0CFJYzbxeu35R2Ktnjy.byte
s	33%	3572	-	train/A0L4e6WxD8IEcjuaHvik.byte
s	33%	3574	-	train/a14ZBu2HThxFVvDmtQ9W.byte
s	33%	3579	-	train/a1mdihQNvLRP4CDr6XyK.byte
s	34%	3583	-	train/a1u9v0bUgGqk5nBwQ2tm.byte
s	34%	3586	-	train/a29m3nUocvzC8hBOWkTL.byte
s	34%	3589	-	train/a2ISAr10X09MkzcEDedq.byte
s	34%	3592	-	train/a2sZItYqWX4B6cPk0Mwo.byte
s	34%	3595	-	train/a3BszoA7hXZc0EuHjmw1.byte
s	34%	3603	-	train/a3Y6cPAM1Q1Vpf8s5S0C.byte
s	34%	3604	-	train/a3ychUzAELtHgXI7kSoC.byte
s	34%	3608	-	train/A4i9wPlygmMQK2ftbXaI.byte
s	34%	3612	-	train/A4nsFofmGel0p9cKxvUr.byte
s	34%	3618	-	train/a5btW1Hh1zjuXiBykZG6.byte
s	34%	3621	-	train/a5dsJH2cv6QprLwWYkBX.byte
s	34%	3625	-	train/a5nMiy8gBFtpuGTxIv0W.byte
s	34%	3628	-	train/A5Rg18PaNvZ4cq0OuDoX.byte

s	34%	3629	-	train/a63qScpsxBFJtAXjd1nV.byte
s	34%	3636	-	train/a6yPugSArOdWK1zQ0c7I.byte
s	34%	3642	-	train/a7iA4S0nBrNKEMC3ZOjv.byte
s	34%	3644	-	train/a7Niv6pD5WPycnQK1TZ8.byte
s	34%	3647	-	train/a7ugQM5sWX6R4fibeJFj.byte
s	34%	3652	-	train/a8gqeKmEjbJIiVXUZ5DH.byte
s	34%	3654	-	train/a8hZBDrjGiMdtzSk1FNV.byte
s	34%	3656	-	train/A8REQ3NzB0dtHFCWDhoc.byte
s	34%	3658	-	train/A90BHbUZYfQvGI1m58Mn.byte
s	34%	3660	-	train/a9B0thJpRMlyj7r5ZSKP.byte
s	34%	3663	-	train/a9r2zHUoK5gXj8E0lqsb.byte
s	34%	3664	-	train/A9SBTC0onUdluM8aqzXf.byte
s	34%	3667	-	train/Aa9zCWtbpOo750I1MRdY.byte
s	34%	3670	-	train/AaKdPBWQD37j9prRxFoZ.byte
s	34%	3672	-	train/aAndF7evKuSX94kxBWDo.byte
s	34%	3675	-	train/AarFhLo7SgGIEU0fHzQn.byte
s	34%	3679	-	train/ab4t071wcCyvkjrpZD3u.byte
s	34%	3682	-	train/ABDGLcYxIh1ef0nq98vg.byte
s	34%	3685	-	train/ABedwUfPmn9jv5YNbykK.byte
s	34%	3687	-	train/ABGeDjKUir1fFwcoMQ5V.byte
s	34%	3692	-	train/ABNUEpLD8G1K6iubzwWe.byte
s	34%	3694	-	train/ABxDq104MENTXmwLK8iU.byte
s	34%	3697	-	train/aC3qABWUp29F0XvZNkIY.byte

s	35%	3701	- train/ACGth0381I7D2XUFVyPp.byte
s	35%	3704	- train/aCmnVPKkA4vrdayHS6uB.byte
s	35%	3706	- train/aCN05F7wbqyDKZd4ti8R.byte
s	35%	3709	- train/ActJLU7ZGYEiHs1raKF8.byte
s	35%	3710	- train/ACTK9tVJunILfsYrwyFh.byte
s	35%	3712	- train/acxojmFTMUAL2HuNfeQd.byte
s	35%	3714	- train/Ad4GW9FZuaM8iqzoV6N5.byte
s	35%	3719	- train/adG3c4XbBMFJvxQ2TKoY.byte
s	35%	3724	- train/ADHSRzr6YevfqojpWCBZ.byte
s	35%	3727	- train/aDMhrzLGy1bKt0s8NvVg.byte
s	35%	3732	- train/ADrM1eNgCO2iIbsKQEpF.byte
s	35%	3734	- train/adS4KHMlGWqDIre07T8n.byte
s	35%	3739	- train/adxTjKLeEt0HFnuJq1BS.byte
s	35%	3742	- train/ae1M5piZAUrqnKc0TYdL.byte
s	35%	3744	- train/aE81yCJwg67YFqOmZPRn.byte
s	35%	3745	- train/ae95Uvr8yCIf0NMKmqZV.byte
s	35%	3746	- train/AEBgp7zcRQZtd1mJbrFw.byte
s	35%	3749	- train/aeEpJKzLA90hHB0QTbPN.byte
s	35%	3751	- train/aEL6vo2BZD8w3QNH0ctS.byte
s	35%	3753	- train/aeNhGKA8io4HrFtDcq31.byte
s	35%	3755	- train/aeoETp7ktY2qAxGcBvfm.byte
s	35%	3757	- train/Aep5gJWMNi4P20LaIV7S.byte
s	35%	3760	- train/aeS2NqfM8tWdDKUsQgR5.byte

s	35%	3763	-	train/aeYfI1yn7sPiR49EW5xp.byte
s	35%	3767	-	train/aF4Nn1KgwE02CiT5uHb6.byte
s	35%	3770	-	train/aFdmMqJ09zSI61pXLoTQ.byte
s	35%	3774	-	train/AFiruP531sbNCjnkoIyM.byte
s	35%	3778	-	train/aFSCmbvwe7HErAuPTQp2.byte
s	35%	3779	-	train/aFshyjLKu86f0pt1Qnzk.byte
s	35%	3784	-	train/AFY7KdVJNL9ng4pjk1mx.byte
s	35%	3786	-	train/Ag04k8rzoVqwtJLWBCET.byte
s	35%	3788	-	train/aG3K4v9cinAygDWmT0Mp.byte
s	35%	3789	-	train/ag50qP4ZzQo1e0VerXdp.byte
s	35%	3791	-	train/AGawszJT86ZI4H23ReK1.byte
s	35%	3793	-	train/aGCAeEkBq0owI8JTZNrn.byte
s	35%	3795	-	train/AgfxJMKRqbt0SnsjN2yQ.byte
s	35%	3799	-	train/AGmSytfUCDr7JBdjT0in.byte
s	36%	3803	-	train/aGS0rWZK7vX8TNFUz9ci.byte
s	36%	3808	-	train/AH3uZ7WMq6RoOyNn89kw.byte
s	36%	3811	-	train/aHDLR3gnyj710kNiAXfC.byte
s	36%	3817	-	train/AhiDaKuve0wcdIkz820L.byte
s	36%	3819	-	train/AHpF5eJXfm0Tcda1uMbt.byte
s	36%	3820	-	train/ahPOWpeHdEkJV4BbDN1T.byte
s	36%	3824	-	train/ahwFAIPcS3yxp1j02KCJ.byte
s	36%	3828	-	train/aI5RJnpEKd70r3VYB4Zc.byte
s	36%	3831	-	train/aIB1DcJ893RjpdgMELfy.byte

s	36%	3833	- train/AiBpmbGDE1zPYOT4usqd.byte
s	36%	3836	- train/Aigal5hnywEbjpgK9vNUC.byte
s	36%	3838	- train/aIGXWDVUyzTcrYAOeBgP.byte
s	36%	3840	- train/aIKU0rgABR7pNQzhZyTu.byte
s	36%	3841	- train/AIKUJVx4slnHOTzyuN5r.byte
s	36%	3843	- train/aIn3UKNJyCvFkD4dG1jR.byte
s	36%	3846	- train/aiRmPYXZnQNA509cpthB.byte
s	36%	3852	- train/aiuJTcU5Sw1g1LzpFjdV.byte
s	36%	3855	- train/AJ2hWQ4bKcS5PF93wEOm.byte
s	36%	3856	- train/AJ3Bb6XsL9250tWe1fUK.byte
s	36%	3857	- train/AJ4B8U0ijaqGzpvnl06M.byte
s	36%	3862	- train/AJBxQN3qer6FYj1RkbdU.byte
s	36%	3870	- train/AjkSRTPFOyGM0imvQ196.byte
s	36%	3873	- train/ajnhEXuNF4kJfOM7S136.byte
s	36%	3877	- train/aJtX1TdUzZGoEwB3i5Pm.byte
s	36%	3880	- train/AK0WapS19V5GYu1HsxLB.byte
s	36%	3883	- train/AK8ZjCh1nXIzygf2s3Tv.byte
s	36%	3888	- train/akGLtUpCrwA6VZT7Y4BN.byte
s	36%	3889	- train/aKH9GvftsjaQzeUwi0bL.byte
s	36%	3892	- train/AkmwZ96MNWGXnObY21SQ.byte
s	36%	3896	- train/aKQHksThE3fdVCizYpGI.byte
s	36%	3900	- train/AkSm1VnbXzwCOrB1IPQZ.byte
s	36%	3903	- train/AKuYXb6dyLtEeiDo1pxs.byte
s			

s	36%	3908	-	train/a12koZ501X8hjATSLqgf.byte
s	36%	3910	-	train/AL87dvhUHD1u1f9MT3rp.byte
s	36%	3916	-	train/a1F0ZAc7xULvhP5Qsr4Y.byte
s	36%	3919	-	train/aLG7QrUD6W0mdNY5tC3y.byte
s	37%	3923	-	train/Alk7nXPtb84eCUWd01Gv.byte
s	37%	3926	-	train/AlNimKRVGWnbq7X64jfs.byte
s	37%	3931	-	train/ALuyE2Q9mjGd0kq8HWoz.byte
s	37%	3932	-	train/ALXK3Sxdgf89a07w4e0k.byte
s	37%	3936	-	train/Am7nEVXHePkCtqsdfzG5.byte
s	37%	3938	-	train/AM9IXU1wJ5PHfNqoVyxa.byte
s	37%	3944	-	train/AMmdfQwU8WeIgXBNhsjp.byte
s	37%	3945	-	train/am04HkjsxM8KcdFBp0h5C.byte
s	37%	3950	-	train/Amuopgi0qHhRsI52G40B.byte
s	37%	3954	-	train/aN8PetkbW4RfxABMh2TE.byte
s	37%	3956	-	train/aNBpfsEkzr29eQC5HKUd.byte
s	37%	3958	-	train/AncaPV65W0ugYNvEohMU.byte
s	37%	3962	-	train/Anhj52SkgdeY1FfuI3GL.byte
s	37%	3966	-	train/anpXsbFtcGLq4oE9KdiS.byte
s	37%	3970	-	train/ANsbQFrM0cvZoxHjS5y9.byte
s	37%	3975	-	train/Ao6pPNzw2BkKsX0hTR7E.by
es	37%	3981	-	train/aobrBYI3k0jSdMNLcUfu.byte
s	37%	3983	-	train/A0cH9YQ80NMUKLarP1h1.byte
s	37%	3987	-	train/aoeMj0FhtA5zDJ174Vv3.byte
s				

s	37%	3991	-	train/A0KiJWGCyavIe0cmMZBh.byte
s	37%	3993	-	train/A0MqLsvZmIi3pK95D4CR.byte
s	37%	3998	-	train/A0VMlEpXa3N2Tj581QGc.byte
s	37%	4001	-	train/Ap1Lj5JHUiDVIoSWFq73.byte
s	37%	4004	-	train/APHtWI0Cm8T2yK9ZeBFM.byte
s	37%	4007	-	train/APTVoki0DMSIyn3r5UuW.byte
s	37%	4012	-	train/apzPxCsQNhE1HbFGBLk3.byte
s	37%	4017	-	train/AQK4rWIdJoFR0LnHaZ7Y.byte
s	37%	4019	-	train/AqKpGPedBXWCia6DNMrm.byte
s	37%	4021	-	train/aQ09Ltn3WzvS7dpZ2XIb.byte
s	37%	4023	-	train/AQTFcnfp8Bqil704UJuk.byte
s	37%	4027	-	train/aqws4GBWlifZPdN5kbXC.byte
s	37%	4028	-	train/aQXjIr1BqHLyG2YEvVkm.byte
s	37%	4032	-	train/aR0gzKL2uEQJNDokWs56.byte
s	37%	4034	-	train/aR10KIrs1cNG8hVYjudg.byte
s	37%	4037	-	train/aRbNvuZ4LQ0JVY0TP2Ax.byte
s	37%	4040	-	train/Arc1RISTQmqu617NfztD.byte
s	37%	4043	-	train/ArgIoRxLXCld0nwN5EGV.byte
s	38%	4046	-	train/aRi6UqPbCVQs1gLWmF4e.byte
s	38%	4049	-	train/ARLikU0rTQmDbZeKaJ0l.byte
s	38%	4052	-	train/Ar0cdDYZkz5QTKh2y7Ms.byte
s	38%	4055	-	train/aRSepuWwyvZLgfXx8m9E.byte
s	38%	4058	-	train/ARUIJ6XPorhKji7SsHg5.byte

s	38%	4061	-	train/ARXaLhsFK8G1bSN2pjwz.byte
s	38%	4063	-	train/AS0PvEt6Vu15aKUif93c.byte
s	38%	4065	-	train/as21YthG7eCEyPbgwv6A.byte
s	38%	4066	-	train/asAFjCoNEKmUnXY0czBf.byte
s	38%	4071	-	train/aShC7zEjIfLm20oM4swq.byte
s	38%	4075	-	train/asKPnzUXjShMc0Tl6Wge.byte
s	38%	4080	-	train/ASwCVlU7bIDm6HjuNOK1.byte
s	38%	4082	-	train/asxd9IrOhnPK1pMVRtX0.byte
s	38%	4085	-	train/ATDpmb5qodJ1VRrc9KNG.byte
s	38%	4087	-	train/AtDZWHR1rcVobv0yp8Yu.byte
s	38%	4090	-	train/ATfrBeFJV9U2NDdCmEcZ.byte
s	38%	4095	-	train/ATKzi05WXfeNpJPcvmaZ.byte
s	38%	4098	-	train/AtpkleqR2JvcjSI3sbdF.byte
s	38%	4100	-	train/ATr6NemOyI2KF0w48BR5.byte
s	38%	4104	-	train/atU9Jyc0SXb5V2xMmnu7.byte
s	38%	4107	-	train/AtWjoFLHmYae1hIfzin9.byte
s	38%	4110	-	train/AU1Y7aLp3ryXsnWN6oHQ.byte
s	38%	4112	-	train/au79s0hkJRX6VYIec8PC.byte
s	38%	4114	-	train/auByF0EAKHrTbtweSXWU.byte
s	38%	4117	-	train/aud9U7Xni4qS0cTrmj0R.byte
s	38%	4122	-	train/AUGwdgy38r6v0XhePCnL.byte
s	38%	4127	-	train/AULFdkoH9T15GIBu1jwp.byte
s	38%	4132	-	train/auQidPm1t1NneSZ7FRjs.byte
s				

s	38%	4137	-	train/Auv3RhZ1U6tXcdDQjg2i.byte
s	38%	4141	-	train/av7N1jxu5MQ3q1b9roUk.byte
s	38%	4143	-	train/AV91H7oUtuadsjy4ZJQe.byte
s	38%	4146	-	train/AVDWo2feMJdbXISwBHm6.byte
s	38%	4149	-	train/AVG9IvBRfaEbZn1DhSNp.byte
s	39%	4152	-	train/avKCprtmoRuXSeDNi5HU.byte
s	39%	4156	-	train/AvSqZN4QKiLxhE2XC0bo.byte
s	39%	4159	-	train/AW0UlrFRqepb4aIKumQv.byte
s	39%	4162	-	train/aW5oFKLR81CiG2ly4gbw.byte
s	39%	4164	-	train/AwGTE6knFv4X571U3Ket.byte
s	39%	4165	-	train/awJH283VpYm7yBDxbOAI.byte
s	39%	4170	-	train/aWNd6erAP9o1iDybv08Y.byte
s	39%	4172	-	train/AwqyMtJmpSKG0jxUBkdn.byte
s	39%	4174	-	train/AWvH1TG5Nym0UCKLRuPo.byte
s	39%	4179	-	train/Ax7iuDgF38YaQMmX9wVG.byte
s	39%	4183	-	train/AxgUXnZu02Wfiwd9Kcam.byte
s	39%	4185	-	train/axJPpKABZEFguIr8l9fS.byte
s	39%	4187	-	train/Ax11rGXDEZ3ftd9agJ47.byte
s	39%	4190	-	train/AxMQBldJRotHf69j32PG.byte
s	39%	4192	-	train/aXnIKb9Cpyd8iPrZ4qm3.byte
s	39%	4195	-	train/Axp8Qjs3IFyWGDH5fwCi.byte
s	39%	4196	-	train/aXqpvlgB2tjeYbSHJiGV.byte
s	39%	4199	-	train/AxTeZ2v16BFVzL5jtSPE.byte
s				

s	39%	4201	-	train/AXvM7m1BfJ8eHVDcspNz.byte
s	39%	4206	-	train/AycKDSBZqFkTC4UVnPRi.byte
s	39%	4209	-	train/AygMUQ9GnsVrSlHt01IJ.byte
s	39%	4214	-	train/AyPm1KkLv0WTuBHtid8I.byte
s	39%	4216	-	train/aYRoe1Nvt4J8Bz6LbSZ5.byte
s	39%	4222	-	train/AYzHFQMeWf250v0tqJSn.byte
s	39%	4226	-	train/aZ90qrU6kB3DzFwTVb7o.byte
s	39%	4230	-	train/Az1MdBSOuma253te0bjX.byte
s	39%	4234	-	train/aZVQvsH0lKUqdEn7oFcf.byte
s	39%	4239	-	train/B0dnNijSbH3fmuCQIL18.byte
s	39%	4242	-	train/B0iZlJ2qwUe9Rvb0SszGC.byte
s	39%	4247	-	train/B0yv5tFCSLkrbfxZTsic.byte
s	39%	4250	-	train/b1Eo267YiRhGCXwjDJTB.byte
s	39%	4252	-	train/B1QctFY2rkoHdzXGnJif.byte
s	39%	4254	-	train/b1s2HyT8eifkmPABoWYp.byte
s	39%	4258	-	train/b1X9uZCg4tSUazoAnyJM.byte
s	40%	4259	-	train/B207sNpJEnMqxdzIbg3e.byte
s	40%	4264	-	train/B2DfZHWl6SrTynJVKCoR.byte
s	40%	4265	-	train/B2G6QYpojbCtZ1DRTcJf.byte
s	40%	4268	-	train/b2taq4gHVORCYDoixMB3.byte
s	40%	4269	-	train/B2Vfj0zqxgmZb7E8QpGK.byte
s	40%	4270	-	train/b3cYmIp50hHkz9ZrwA8V.byte
s	40%	4273	-	train/b3gxz6Za90hDwpY5oFiC.byte

s	40%	4275	-	train/B3q1QtIhPSR2HUyXx5uG.byte
s	40%	4278	-	train/b3XoLNIJ0sU1rzcStWxjF.byte
s	40%	4281	-	train/b4fMnYzOWHxNeGjqRSLp.byte
s	40%	4282	-	train/B4gfk0xmjqw9ldFATahW.byte
s	40%	4286	-	train/B4kQoTYWxbXcEzrvKDZP.byte
s	40%	4288	-	train/b4o8xqvChYE06mZdDtzK.byte
s	40%	4292	-	train/B5bnojrmDVUPvOw7zipS.byte
s	40%	4296	-	train/B5M1QJjfpUbcaG3kYX4m.byte
s	40%	4299	-	train/B5YU1VtdxFgkTM7mN9Sh.byte
s	40%	4303	-	train/B6dzEY3kKIn8me9McfGg.byte
s	40%	4305	-	train/B6fI4wzvJ2tWjSld3HPx.byte
s	40%	4308	-	train/B6nhL5vI2TxV8ODNbRJE.byte
s	40%	4313	-	train/B74NFkymijZMaKAdVqY0.byte
s	40%	4316	-	train/b7eAT01WEJZGYSX3MPNw.byte
s	40%	4319	-	train/b7lZgrI94Bp3XM0f6vuJ.byte
s	40%	4321	-	train/B7QcjYOCZ8RmDH0xtWfU.byte
s	40%	4325	-	train/b8jkVcnpatLKQ062E5m9.byte
s	40%	4328	-	train/b8kXFQJv6aLdDP3fn0tI.byte
s	40%	4332	-	train/b8Zi5NAtq7mysdLGI1Mx.byte
s	40%	4334	-	train/b9AoKd5WisSgXzM1kGCY.byte
s	40%	4337	-	train/b9PJ5cCeBG6N0ymu3pfX.byte
s	40%	4341	-	train/Ba2j9GDwq0H0tNrzbRuI.byte
s	40%	4345	-	train/baFiIOemUNYgJAqsujL0.byte
s				

s	40%	4348	- train/BALQRhwGyW4kzdUqZ3Ir.byte
s	40%	4351	- train/bAPactd8TMeS13U64yIC.byte
s	40%	4353	- train/BAq9kHTFUjw1Ch0RvzXa.byte
s	40%	4355	- train/BAUuZ26kwLNq9a37WmpT.byte
s	40%	4358	- train/BaxZjfsd2AIW13TEPFzi.byte
s	41%	4361	- train/bAZhqgSnCfNdXPDptGyi.byte
s	41%	4363	- train/bazOp8SLuGgisveCP0QI.byte
s	41%	4365	- train/bB4krWM7TCiOD06x2AN1.byte
s	41%	4368	- train/Bb6vTaRMcG5Z39kPCQHg.byte
s	41%	4373	- train/bBw96PZreoAt8FEjJ4Rs.byte
s	41%	4379	- train/BC9hzqbTa00IsePYiND6.byte
s	41%	4381	- train/bcdvUs26L3ofyD05P1Wm.byte
s	41%	4384	- train/BcNRwpZCKQdV8eLhUD1r.byte
s	41%	4385	- train/bcNWH4s60jFiIqk5pt8K.byte
s	41%	4387	- train/bcqIoNTEXGPDUwzCWRx0.byte
s	41%	4390	- train/bcxmnqdC5ApvuW0BgLhi.byte
s	41%	4393	- train/Bd12KAkryHoe4M3UGRti.byte
s	41%	4397	- train/bDfQn1eI6rqp5Nw9GC30.byte
s	41%	4401	- train/Bd01phEQY6LyC xv5MrA4.byte
s	41%	4405	- train/bDUG1rBm50yogVwfuPJR.byte
s	41%	4406	- train/BDuWAEz5GFKaP4Q1k6Jd.byte
s	41%	4409	- train/BdZYrUSpnXvyITwgOi4H.byte
s	41%	4412	- train/be32jsQYGSuXWkz0iPrN.byte
s			

s	41%	4416	-	train/BEdmtYAN01J3Y6Rkxu7c.byte
s	41%	4417	-	train/beDosHq791jFfRE1Vdmi.byte
s	41%	4420	-	train/bEF9IJRgPA5dDU8NevYS.byte
s	41%	4422	-	train/bENI2W7cLoyjuGvZpQ31.byte
s	41%	4426	-	train/BEqMgotbrGVUiXsuFJWw.byte
s	41%	4427	-	train/BeSFz30shNn8uwGKUCTx.byte
s	41%	4429	-	train/BewAOSD3I9tH0s2VLrca.byte
s	41%	4432	-	train/beYsSmB2uZoA19Ek5Mrf.byte
s	41%	4433	-	train/Bf0LG2a8MqCtp3evSrPX.byte
s	41%	4435	-	train/bF5KEHN7mXfvktq0Qxzd.byte
s	41%	4441	-	train/BFIzDn0r5aKum4ibGtHP.byte
s	41%	4446	-	train/bFpJD05AxXrCzntaWvTq.byte
s	41%	4449	-	train/bfRSxLKkthU2yJTgYePo.byte
s	41%	4455	-	train/Bfwzv7WUD5ucy9t0qrMI.byte
s	41%	4456	-	train/bfY2xBJKmhHdVjPu9FLM.byte
s	41%	4459	-	train/bG1zA9Iwqf6gCVSoJ0ZN.byte
s	42%	4464	-	train/bGFAIf31HjuecEwyWCY8.byte
s	42%	4468	-	train/BGLoYf0m4sgbHaOKMEJy.byte
s	42%	4472	-	train/Bgqr8AmHC7WFak9cX0fL.byte
s	42%	4473	-	train/BGqXnfMsI60QDb2Ue35k.byte
s	42%	4477	-	train/BgxjMZfYu1CsoGX7A5bV.byte
s	42%	4481	-	train/Bhp8E1sD7L2xcugonWyf.byte
s	42%	4484	-	train/bHT5pg8MuXLF2UZAwxBc.byte

s	42%	4487	-	train/bHX3c9EA1NTBKSk2aD47.byte
s	42%	4488	-	train/Bi0qfC3cKRzFTNxgjbRj.byte
s	42%	4495	-	train/bIC9f4xQY0m3SMpvgtdWd.byte
s	42%	4500	-	train/BiHmc5bPMrnvClfgEw3h.byte
s	42%	4501	-	train/BiKc6IFPEovX59sgzLp4.byte
s	42%	4504	-	train/biR27No0u03z1PjqGfJw.byte
s	42%	4509	-	train/BizuI15W3K6fAoqSZnyX.byte
s	42%	4513	-	train/BjaLF1KHchGQ1UrY06fn.byte
s	42%	4516	-	train/bjFKyp5JTwGv6QV9gYAE.byte
s	42%	4518	-	train/BjL7pyHh31cqui4UA8Ge.byte
s	42%	4521	-	train/bjnMB75Xxi2VQOR6LSgC.byte
s	42%	4524	-	train/bJvcld8YQ90xDAnwtSfp.byte
s	42%	4528	-	train/bjYnW35luS8H4QGpr92E.byte
s	42%	4531	-	train/Bkc9awJI1G4DZTodR2zx.byte
s	42%	4532	-	train/BKDwzZfquvWotCTXOM15.byte
s	42%	4536	-	train/BkiT8QwUGHn2CqrDuZaX.byte
s	42%	4538	-	train/BKpbxgMPWUNZosdn08Ak.byte
s	42%	4540	-	train/bKRFqsMctGHnH13fw6od.byte
s	42%	4543	-	train/BkUC8L5ziTgyA4Pberfq.byte
s	42%	4546	-	train/BkV8KZjcfSy3DFa5MYre.byte
s	42%	4549	-	train/bkY5l3u0gwSZqzvnaU7t.byte
s	42%	4552	-	train/BL4IWnsa2AbXyoPHQ0kJ.byte
s	42%	4557	-	train/bLGq2tnA8CuxsF4Py9R0.byte
s				

s	42%	4558	-	train/BLGy8Ek5f4JlH0vIXxz2.byte
s	42%	4562	-	train/bLNmOA1hUHcBzatEePVj.byte
s	42%	4566	-	train/BlshLxAuJR0VGyQWmT1w.byte
s	42%	4568	-	train/BLTSJkg1MdceD3sv90oW.byte
s	42%	4570	-	train/b1UxaBNCoSJ3veXVIKr4.byte
s	43%	4572	-	train/blw8c7vA4pDdrTk3Zn0H.byte
s	43%	4574	-	train/Bm1b7W65puvrLdECV kay.byte
s	43%	4577	-	train/BM4bG5xEwohipXWTKt3U.byte
s	43%	4584	-	train/Bmf4CXo9651TrigSb0jk.byte
s	43%	4590	-	train/bMK0Xukv9xUAi2QSGJ6Z.byte
s	43%	4593	-	train/bm0jLhGfn9tIgicHVCEZ.byte
s	43%	4597	-	train/bMwiKy4LQPJor0xkTe83.byte
s	43%	4600	-	train/BMyRb1VjoWUws42eSLgc.byte
s	43%	4605	-	train/Bn3MOp6gHdfLz8PboiJA.byte
s	43%	4608	-	train/bN6ycnsgpUauJroQ493R.byte
s	43%	4614	-	train/bniqS1ycf0AYX8a2Jsw5.byte
s	43%	4618	-	train/bNT4jk68t51Zghy3HPcr.byte
s	43%	4622	-	train/Bnwd1vNx85KI7yzphZHS.byte
s	43%	4623	-	train/BNwV59DaCqRXs3LpTtAx.byte
s	43%	4627	-	train/Bo1NCzWcGTyOHQ39vMga.byte
s	43%	4632	-	train/bo0znJjLCqDKkAT9RB0m.byte
s	43%	4637	-	train/BOYuoPTga1DCSrERI8kj.byte
s	43%	4642	-	train/Bp7t1yMD5zQI2U0gN49E.byte

s	43%	4644	-	train/BpAoClbIMLsxzWyPH0aJ.byte
s	43%	4647	-	train/Bpd4WExsYXfkN3R7gc1T.byte
s	43%	4650	-	train/Bpf5rdhvyHN1AKs16007.byte
s	43%	4654	-	train/bPGRecDZuU0q7TFwr9Jj.byte
s	43%	4656	-	train/bpIA0CvVdmQcEB2oisqj.byte
s	43%	4660	-	train/bp0R4jctxD03aysPq5Qd.byte
s	43%	4663	-	train/bpS0xXv95Z2TmfnuVhiK.byte
s	43%	4667	-	train/BPx2oZdGwasJRukQDzpn.byte
s	43%	4670	-	train/Bq57J0KSC8IMi31dhcL2.byte
s	43%	4673	-	train/bQepsKFSzxt8Pfk6qGaZ.byte
s	43%	4674	-	train/bqhCWjYG2HgFMdxUr01k.byte
s	43%	4678	-	train/bQN1L00n7Tup3VwP9Z68.byte
s	43%	4679	-	train/bQoNdlu0zvWihMaDkZ64.byte
s	43%	4681	-	train/BqRPMt4QY1sHzvF6JK7j.byte
s	43%	4684	-	train/bQVJ7HpnSC01EI4PNzwy.byte
s	43%	4687	-	train/BqYN73Kuyc0ixJPpHrRz.byte
s	43%	4691	-	train/bR2QdKrOsGkLPgTFXi7t.byte
s	44%	4695	-	train/BRcPd4zSt09XVsnayUYw.byte
s	44%	4698	-	train/BrePaE2xAs9fJtqvN1Wp.byte
s	44%	4700	-	train/BRghGEp2P8emtvNHYSTU.byte
s	44%	4702	-	train/BRHGMC6IkeKx4LjfPJ1Y.byte
s	44%	4708	-	train/bR1eFLPfA672tpnWVsdK.byte
s	44%	4711	-	train/bRoM4EpjukSGXwA1Z6zn.byte
s				

s	44%	4715	-	train/brsxDiLaBgj315Emol4K.byte
s	44%	4717	-	train/BrUsW7NEhzVQaljI30Xy.byte
s	44%	4723	-	train/bSEPTYj0DGy2iZfawId3.byte
s	44%	4729	-	train/bSN5cyA8a4TVvLxBGtfD.byte
s	44%	4734	-	train/bSRakDIeOuFog7NHQP3j.byte
s	44%	4737	-	train/bSZgP3NGeYM4c65TjEWn.byte
s	44%	4739	-	train/bT9MIkgOeyVW8cFHS67j.byte
s	44%	4741	-	train/btAy2hf1qLaTpzVPULxe.byte
s	44%	4743	-	train/btfJ456hq8T3vU9DZjsg.byte
s	44%	4745	-	train/BThVXxrfyLW6dQGZS7Um.byte
s	44%	4747	-	train/btk1Yfq0glZevm2dPRsL.byte
s	44%	4748	-	train/bt1Cw00cmiDT493yhXf7.byte
s	44%	4750	-	train/btnjxGM0mzXWZ0ys5N2r.byte
s	44%	4754	-	train/BtVdF1wcQGAKLhZ5Pp13.byte
s	44%	4759	-	train/bu8vjZM902UPcGmXQKSz.byte
s	44%	4761	-	train/bUA016u0gwLvqxBHJjeV.byte
s	44%	4762	-	train/bUBIJjfCnTiQRvzh416F.byte
s	44%	4766	-	train/bUdgzZroIDt51qNWy3jM.byte
s	44%	4771	-	train/buhA0QaykEJtsIZoWjBp.byte
s	44%	4772	-	train/bUhx0WFojmzE5BSZVqTH.byte
s	44%	4776	-	train/BuLVYHbp2AgWN7oQ46cZ.byte
s	44%	4780	-	train/buqKNv0RSThA9sF51MZe.byte
s	44%	4781	-	train/buRJ9zQWwTilSgYckpCK.byte

s	44%	4787	-	train/bv60XkYZsCKHa3NgFUDr.byte
s	44%	4791	-	train/bVDNy1kq1EmIzJfvdcng.byte
s	44%	4793	-	train/BVLP2pKr8yxF9iHd1bEj.byte
s	44%	4799	-	train/BvWnd3X1LYb1QGAyP6Zf.byte
s	44%	4802	-	train/bw3g9UF6dIjnexWSrsY4.byte
s	45%	4804	-	train/bwcMeNJ6du0xi2q0GQ4n.byte
s	45%	4807	-	train/bWHnGtLmyoK0de1vJB3S.byte
s	45%	4811	-	train/bWMTEq0N6d7KRZhvigGQ.byte
s	45%	4815	-	train/bWRYhgXnvw3FLGMCdOur.byte
s	45%	4816	-	train/bwvcfxORTe0Aju6d98Es.byte
s	45%	4819	-	train/BwYh9iT0tILyCW8SvJZf.byte
s	45%	4822	-	train/bx75FsX2VcQ1CkznwIEM.byte
s	45%	4825	-	train/BxcU0qI5b8zn2m1rXSt9.byte
s	45%	4828	-	train/bXGWQEp24u1LweHra3fB.byte
s	45%	4832	-	train/bxj8HFNN0cCiAW7G0gzh.byte
s	45%	4833	-	train/BXKmeITqa7cfsyYEdpFS.byte
s	45%	4836	-	train/BXQNYRIMHrq8LZ7Ay9pc.byte
s	45%	4837	-	train/bxrjz5vEWF9geNm138dH.byte
s	45%	4839	-	train/BXrPSm1KvMJz1guqHFLa.byte
s	45%	4841	-	train/bXWLaKzx1viEY1Gt7cjlw.byte
s	45%	4843	-	train/BXYpT4e7jKgbI95Lqaov.byte
s	45%	4848	-	train/BY8w7dTSS2GgLR5pbMKj.byte
s	45%	4852	-	train/BYhZ8Rd2N17F69DtnMXH.byte

45% 4856 - train/BynGZjrHQXJiKkNt3Y2W.byte
s
45% 4858 - train/BYRrHLwm9QoeXT1zpNVI.byte
s
45% 4861 - train/bYvCdoTMQypIa367GUL1.byte
s
45% 4864 - train/ByY1Q4t83mnODGTj59op.byte
s
45% 4866 - train/ByzLS3siG0pYK6wEut2d.byte
s
45% 4869 - train/bzcsn9Q08SDWutfxMhYm.by
es
45% 4871 - train/bzDB7QwxH0eCMmnRSqUL.byte
s
45% 4876 - train/bZ11kUSLsMrifRo0cnIA.byte
s
45% 4877 - train/Bz19Ln6XW7gsEpKkj2NP.byte
s
45% 4882 - train/BZTLwj0Jkvu3xa4bgf1Y.byte
s
45% 4884 - train/bzVp2rDhy0vx4c37gmuC.byte
s
45% 4886 - train/BZyd9n1Lik5loVQWYgJR.byte
s
45% 4891 - train/C01DeUio3WYBJg45AM2H.byte
s
45% 4894 - train/C0ATrYp62uZIndBXMgmL.byte
s
45% 4897 - train/C0H18uWb1scIEg4XrtJN.byte
s
45% 4900 - train/C0uidNjwV81rPgzt1JSG.byte
s
46% 4902 - train/C0YjaU7beAW2IzcQVtfr.byte
s
46% 4904 - train/c1Bwr4tG0v6IoiP72MgX.byte
s
46% 4906 - train/C1E8ox6VhKuSUIjap2P.byte
s
46% 4908 - train/C1Ik1XQF4DJiB8stvR39.byte
s
46% 4909 - train/C1MR4txhqcAg0vDWSZN1.byte
s
46% 4910 - train/c1nP81vwZ6xB3jLdirMW.byte
s
46% 4911 - train/C1okxVtRdfUWIXY9G3s8.byte
s

s	46%	4912	-	train/C101cN2H8W0YATqZxipn.byte
s	46%	4916	-	train/c1X3j2Lat7Eo4I8GWFD5.byte
s	46%	4919	-	train/C24qXnmbryh08LzGvaTF.byte
s	46%	4921	-	train/c2hn9edSNJKmw0OukrBv.byte
s	46%	4923	-	train/C2qD0rIna1h7LpEgU9sZ.byte
s	46%	4924	-	train/c2sdKG4MLmHBzFPg5IW1.byte
s	46%	4928	-	train/c2WbVqndpY9mhHXgRTti.byte
s	46%	4931	-	train/C3bFzsgkTP1uHDAmxdhcp.byte
s	46%	4933	-	train/C3gK1ERFx4WDm7IskLtb.byte
s	46%	4937	-	train/c3PhRtdH4SuDyLWNQXa6.byte
s	46%	4944	-	train/C3ZfGBVmTh1cWFQDEmw.byte
s	46%	4947	-	train/c4j1D5gS2UPRMh8GWBzm.byte
s	46%	4948	-	train/c4jU5YD9AmPipweFb710.byte
s	46%	4949	-	train/c4RSgNFGuhfKJakmUWCx.byte
s	46%	4952	-	train/C4V0aQKgkyhfcnNI3RsH.byte
s	46%	4954	-	train/C4z1LIqNQDndKphj7etT.byte
s	46%	4958	-	train/c5hXZVENsBA4RvDfjJPw.byte
s	46%	4962	-	train/c5npw2uMISjrah74tDXo.byte
s	46%	4963	-	train/c5oI2wARWuaKLsGXf7i8.byte
s	46%	4964	-	train/C5RGotXVr0xI3Np8g10A.byte
s	46%	4967	-	train/C5UiyDre7dL1XVHxsRIo.byte
s	46%	4969	-	train/C5zRPoai3nALEsq6vJk.byte
s	46%	4972	-	train/c6Da21807vGgAbYnUjPm.byte

s	46%	4975	-	train/C60EvgQ2n5F93Tjhoemt.byte
s	46%	4979	-	train/C6xhRUmd8vVAp4kJJBu5.byte
s	46%	4981	-	train/C6zmBoPQ9y38n20S1hsR.byte
s	46%	4983	-	train/c6ZsLyHf12kpbSd8vTCB.byte
s	46%	4984	-	train/C70ihKqv4L16GDQSrK1J.byte
s	46%	4986	-	train/C74vBKyH2rR1j5hJWePN.byte
s	46%	4990	-	train/C7knyMSfZacwTAdUBjem.byte
s	46%	4993	-	train/C7mXMKDZBFjnNwxYH1et.byte
s	46%	4999	-	train/c8BfFP6iYEIRaUxdGtmX.byte
s	46%	5001	-	train/c8gbt2qHrVm9z30yCIkD.byte
s	47%	5004	-	train/C80aF3weNRMS2vphGUAX.byte
s	47%	5006	-	train/c8Q7CVawB6vyUIoN0PA2.byte
s	47%	5008	-	train/C8XKihk4AVvsQF1zWMql.byte
s	47%	5009	-	train/C8y1gmcFqh1VwLY3nUXz.byte
s	47%	5011	-	train/C9ekdvY7Rx1MIVBGg4Wj.byte
s	47%	5014	-	train/c9gmWqvaiKZANzUreRG4.byte
s	47%	5017	-	train/C9KZx8YrJeJnW5B0sqgQ.byte
s	47%	5022	-	train/c9tj6deKFW4zDf0vE5VG.byte
s	47%	5024	-	train/c9UpXT52VDBvPgAySnqH.byte
s	47%	5026	-	train/C9vsf3VBgmR75y1qGJpH.byte
s	47%	5029	-	train/Ca30jSKDypEY2PLGvnXU.byte
s	47%	5032	-	train/CAeHqYZwpdarU0m0D2tG.byte
s	47%	5036	-	train/cao35U1R4dib9Qj0tnfI.byte

s	47%	5040	- train/cAPwa3zS	Ir71vLV	MPe2h	.byte
s	47%	5044	- train/catrJpdjF4qNw2Ue	PAQL		.byte
s	47%	5046	- train/cAxogRZB2v9lEm	zr100a		.byte
s	47%	5050	- train/cB54S10CNa8Pd3M0Qx	IE		.byte
s	47%	5052	- train/cb72yv	fjMuQkKZERh5UP		.byte
s	47%	5053	- train/cB8s02i	lF5KDxH4XITvb		.byte
s	47%	5056	- train/CbeFcSWlYTV93GN7w0f	U		.byte
s	47%	5059	- train/cBI4bRkZVxHU7f1ug6t	D		.byte
s	47%	5060	- train/Cbj4xGISgF	HtOnPYfKJ2		.byte
s	47%	5061	- train/cbLvyT5HKlrmDP60g	E7W		.byte
s	47%	5064	- train/CBZd1HELuV4rTPSb0s	NM		.byte
s	47%	5066	- train/cc8jta0FUZl	hIGSXmBHf		.byte
s	47%	5072	- train/cCr1JnyTfGsQAR8zU	vmL		.byte
s	47%	5077	- train/cd2W8Fxs	rJDnOHZGEiet		.byte
s	47%	5081	- train/CdBnaqstfPMA9NXro0	Jh		.byte
s	47%	5084	- train/CDFibvHczS19akf0EB	wo		.byte
s	47%	5087	- train/cdk9u0beg4S6Tnf5yh	R2		.byte
s	47%	5089	- train/cdlgXf2EV9sOKvn6D	wt1		.byte
s	47%	5092	- train/CDoRcvHMNxiq40Ij	WTSQ		.byte
s	47%	5095	- train/cDRA2GdrqXYfoNaUM	Jg7		.byte
s	47%	5098	- train/ce2bA7oZP38DCIX9Gh	QV		.byte
s	47%	5101	- train/ce5KGVx3FUMRzmCXQH	Os		.byte
s	47%	5103	- train/cEait2x5FUG64Vz7jnu	q		.byte

s	47%	5107	-	train/cebAsutfF9NkrvLRo60z.byte
s	48%	5111	-	train/ceKaj4rnXfU7QvShuqY3.byte
s	48%	5113	-	train/CetLXMsaqZQR7xy3gjGD.byte
s	48%	5115	-	train/CEv0JwLrsbWrNkIgj4B8.byte
s	48%	5118	-	train/ceyIk4f0BRUiar5bLJ8l.byte
s	48%	5120	-	train/cezkDo9iMP0yXgBJSaWv.byte
s	48%	5122	-	train/CEztMhWprIdasmvRux4j.byte
s	48%	5125	-	train/Cf5Tv2SJ18cXm7NZkr3L.byte
s	48%	5127	-	train/cFa4U6xqeVW7liJzMqk0.byte
s	48%	5131	-	train/CFH2LexhBSApVRurWQD0.byte
s	48%	5133	-	train/CfNFEIb1UPQgiqt8uRy0.byte
s	48%	5135	-	train/cf0jhHAekrZLs7GiCm4W.byte
s	48%	5137	-	train/Cfpes01aH64iYx0BURuJ.byte
s	48%	5140	-	train/CG6sL3Aug4yomNnztEIw.byte
s	48%	5146	-	train/CGg0ZekvTIXAMYxfNLW1.byte
s	48%	5151	-	train/cgkonXYDZtdqiQ4su8bL.byte
s	48%	5153	-	train/cGNfA8C9wh1R5grDKQjX.byte
s	48%	5156	-	train/Cgq3TizwBXpnk7b8a5Qs.byte
s	48%	5158	-	train/CGtMcxADYKyHhnTrebkz.byte
s	48%	5160	-	train/CGtzUByc19fMLP6ioT8n.byte
s	48%	5163	-	train/ch28HSMqUPA1zYWnX7Tw.byte
s	48%	5168	-	train/chDbN10moTRnQ8hZ7ru0.byte
s	48%	5169	-	train/chDLN8VJzG1XuZhrsxyn.byte
s				

s	48%	5171	-	train/chePT1jq7g5Q002ZN9DJ.byte
s	48%	5178	-	train/cHnxoFJRgXl4GSMD6bLU.byte
s	48%	5182	-	train/Chq2jb7PZtFyi9S3eYUA.byte
s	48%	5185	-	train/chVqtrziORSJkCwu2KfQ.byte
s	48%	5188	-	train/cI3dPqwaMLU0opfuNy6g.byte
s	48%	5190	-	train/ci6Su0WdoDLIG5vnJhay.byte
s	48%	5193	-	train/cICg0pMBHFzbWymJo8uL.byte
s	48%	5198	-	train/CIJ1f70KSNDz8ZyoRFEH.byte
s	48%	5201	-	train/cikBKN2eQY8WXfCHvnI7.byte
s	48%	5204	-	train/cIoDLZzPdTiQBpCjW2Ab.byte
s	48%	5206	-	train/Ci0ve16S9fVptuRWaEFm.byte
s	48%	5211	-	train/CIuUf3p6NALZst9ynk7l.byte
s	48%	5214	-	train/Ciz8f6xSLwjMPsmDR0dg.byte
s	49%	5216	-	train/CjBNmFq806khnE10tgA9.byte
s	49%	5220	-	train/Cjid9IAL13h2Rnpvc75w.byte
s	49%	5222	-	train/CJONcsB5ExKb92XA6tT7.byte
s	49%	5225	-	train/cJQZpFegP6ChAVsX7nYG.byte
s	49%	5228	-	train/CJTezukAoORBNh0QxUMa.byte
s	49%	5229	-	train/CJUmezigaEoXRfY6FGT.byte
s	49%	5232	-	train/CJyopa7dqfBG1R6wUgSV.byte
s	49%	5236	-	train/ckA3Pn9iYjL27sqf4SRM.byte
s	49%	5242	-	train/CKIT85YvdjSMoQr09A2L.byte
s	49%	5243	-	train/cKkbQLrgIT01qjEHMDGY.byte
s				

s	49%	5244	- train/CKL9vk4x2IJZoM5sbu1R.byte
s	49%	5248	- train/cKmUqV7zyePDgYIt8i20.byte
s	49%	5250	- train/CkOT69iQsaLAGq8mYEeV.byte
s	49%	5255	- train/CkYbvN6hPw21aWxKeMZG.byte
s	49%	5259	- train/cl0e6QDGkbwLVuSZrHxo.byte
s	49%	5261	- train/CL26qki7n4KRNEH0mAb1.byte
s	49%	5263	- train/CL5Moas9h1iPGv6nYSyJ.byte
s	49%	5264	- train/cL6skgJYz3Uy2T8Pvdw0.byte
s	49%	5267	- train/CL8hrPMFWvzkfb6w1JJIN.byte
s	49%	5270	- train/CLH1PKJvWSrwaMOQ9GIU.byte
s	49%	5271	- train/CLHVv0j2hM4UekxQKw1f.byte
s	49%	5272	- train/cLiu0PNSs3aer8BmoMAJ.byte
s	49%	5273	- train/CLJUMfrb1jF4VliXSutB.byte
s	49%	5276	- train/CLoqY11If0zkmyZ8ONjp.byte
s	49%	5277	- train/CLpFEnA496yoGwrDmz78.byte
s	49%	5280	- train/clQnt9kyP5Tw8SDVfBpE.byte
s	49%	5284	- train/CLw2EfPFqBAke1vu5oDR.byte
s	49%	5289	- train/cM2epyubCXDZQgBKJ0N0.byte
s	49%	5294	- train/cMbvH0ZxLKN9i80fH3nI.byte
s	49%	5300	- train/Cmf516DZaASV43IUQqe7.byte
s	49%	5305	- train/CMKGnmX4D3dQIzgxUAqT.byte
s	49%	5310	- train/CMpESZXjIKbqg7G5Nh1V.byte
s	49%	5315	- train/cMvOmIQpRPnhUTKjHgYA.byte
s			

s	49%	5318	-	train/cMzSVED56G2iCUtF9eb7.byte
s	49%	5320	-	train/cn2u9ljzm6UL7o4q3kAX.byte
s	49%	5323	-	train/CnAw9VxYtrp5Ps2li1ML.byte
s	49%	5325	-	train/cnf03bdvewDVXMWHFT75.byte
s	50%	5325	-	train/cnf03bdvewDVXMWHFT75.byte
s	50%	5329	-	train/cNIGDKm5dUbMZjvnQX8R.byte
s	50%	5333	-	train/CNMG4WYjkvLo351bngdm.byte
s	50%	5336	-	train/Cn09A1jfBmokT5qN7J6S.byte
s	50%	5337	-	train/CnPvMBc9sWK1jrbG4ILT.byte
s	50%	5340	-	train/CNRrJcT3pLw2GAiDVq7Z.byte
s	50%	5346	-	train/CO18w5p9ZeBlqIKPhiyj.byte
s	50%	5348	-	train/co40RmlyQMghD2XAV0Fe.byte
s	50%	5352	-	train/Cocz51rsXakGETheDUBH.byte
s	50%	5354	-	train/COdQXbslFapMY0rnNTLE.byte
s	50%	5361	-	train/cOMbeEqitn16Yh3XRwG5.byte
s	50%	5364	-	train/COntTlwPvaG3p4hMSJqI.byte
s	50%	5366	-	train/COq8Xi1VEaDfj6MFHUP1.byte
s	50%	5370	-	train/cOvAdrLosthuC9paE4UN.byte
s	50%	5372	-	train/COz1qc0Ds7RHkdbQBwvU.byte
s	50%	5375	-	train/cPBMG2HuL1CaNgbSE8iw.byte
s	50%	5377	-	train/CpeIDYlmhsART40c7gqz.byte
s	50%	5380	-	train/CpjlodZOi1La0WtA8QuM.byte
s	50%	5382	-	train/Cpl1foGPHIm9gsve5NLV.byte

s	50%	5387	-	train/CPsXYMZcng0i9521IJVa.byte
s	50%	5390	-	train/cpvSmajetq6923K05MRD.byte
s	50%	5392	-	train/CQ9ATzbP3j20n810GX4d.byte
s	50%	5394	-	train/CQBGfcqo9wWzUVDA2smx.byte
s	50%	5396	-	train/cQd0vTgyk0n1f1rxaCAB.byte
s	50%	5401	-	train/cqH1rY9oAVpyWMKJ8mOF.byte
s	50%	5404	-	train/Cq1FyksTYLVw0eQfJZGD.byte
s	50%	5408	-	train/cqoIY36dhwfDNrn1HJsx.byte
s	50%	5411	-	train/cQpOT5jn9b2FuweAsNx1.byte
s	50%	5412	-	train/CqQ1P6D9YapR71wb2d48.byte
s	50%	5417	-	train/CQS2WYL40q1MpZnNfJU8.byte
s	50%	5421	-	train/CQVMwzgEYXvN4t0oKSis.byte
s	50%	5424	-	train/CQxAJbN0GImuVeyq87js.byte
s	50%	5426	-	train/CqzUdJEnt9LOYk4ex8Mm.byte
s	50%	5430	-	train/crB9AngSLtmYldMvoPuK.byte
s	50%	5432	-	train/CrEM10GaB6feRx3J20iP.byte
s	50%	5434	-	train/Cr-f0YyHJukXG35MldUSV.byte
s	51%	5437	-	train/crKe1xWtHpsz20LdiVMS.byte
s	51%	5439	-	train/crKu0S7kqIQwsCN6Z0az.byte
s	51%	5441	-	train/cRLVYFJTAaoIOWk1dHUy.byte
s	51%	5446	-	train/CrscPL6baYHtUAW2FIR5.byte
s	51%	5448	-	train/cRu8ICh5i4pa1MfKeVTm.byte
s	51%	5451	-	train/cS0k5dW0BjEybfLr1u7N.byte
s				

s	51%	5453	- train/CS8VeJ613FL9spGiuY45.byte
s	51%	5457	- train/CSehsbFLKPBTWo4gnd2R.byte
s	51%	5459	- train/cSEND4vF3fL0qUgy1CZ2.byte
s	51%	5464	- train/CSkITPbvqspWu681aVJn.byte
s	51%	5470	- train/csXTGQtRjwAfxHIho25y.byte
s	51%	5472	- train/Ct97bQeGS5fRmaKANMZ.byte
s	51%	5474	- train/ctClpAGF7i4oa0f8nYb2.byte
s	51%	5477	- train/Ctjp687wq1TJncfI9ZSi.byte
s	51%	5480	- train/cTt2FxBW3NsAd4LhvQpn.byte
s	51%	5482	- train/ctum3yBP7pIQ1W4U96Zk.byte
s	51%	5483	- train/CTv0Q8ujqlZc5dGnfeaw.byte
s	51%	5485	- train/ctv0bTz6W3KFQHsmB80n.byte
s	51%	5489	- train/Cu5NVmSa1JAHvM2dp9Z1.byte
s	51%	5491	- train/CuB6pZ1WVqT9GK3QDHdm.byte
s	51%	5494	- train/cudBgeQFL8z4prNGIsXH.byte
s	51%	5496	- train/CUhuXitp6fL8jM9Row1H.byte
s	51%	5499	- train/cuNE2eL7XtkbJQw50zsW.byte
s	51%	5503	- train/cuQANt4oBpMzmYKRkGf6.byte
s	51%	5508	- train/CuxeASFn8t53UTN17ZOm.byte
s	51%	5511	- train/Cv4pEJPDoxHsByfn2eYT.byte
s	51%	5516	- train/CVBnb3mE7XPhIuor5J14.byte
s	51%	5517	- train/CVcysi9h3kTI1gXj0qHm.byte
s	51%	5521	- train/CvdWjIn9A4GkBbp1HF6T.byte
s			

s	51%	5525	- train/CVIUYfaQjet1154x7hyq.byte
s	51%	5528	- train/cvMhPJWk123gZjFTq4Rd.byte
s	51%	5529	- train/CVmnTbuGZEev5hN30diS.byte
s	51%	5534	- train/cVTwfdkWl2hzAvG3eFlH.byte
s	52%	5536	- train/CVZd3lhxBF0Tqu9K8ymE.byte
s	52%	5538	- train/cW45XHLmtsC3QgpiUvJ0.byte
s	52%	5541	- train/Cw9Ep0inq0sjVhcdHt6L.byte
s	52%	5544	- train/cwAyWxKMPkXELbpBS9if.byte
s	52%	5546	- train/CWDE4xUiRdwmFGAt80pn.byte
s	52%	5551	- train/cwPNTvn3zIrJhXGK1aDS.byte
s	52%	5554	- train/CwtPLqOS2b3mg1rWJhx4.byte
s	52%	5556	- train/cwWFfNAeLChbKD9mRd2u.byte
s	52%	5559	- train/CwYeBdG7J5ozg29XVx1T.byte
s	52%	5561	- train/CX2o0MKvc1Uw3Z7zrqxV.byte
s	52%	5564	- train/cxCf1UQEJT7mNFg08KpM.byte
s	52%	5567	- train/cxIb52FL1vanqM1ykopB.byte
s	52%	5570	- train/cxKg9nzCNJLrdDRHomMj.byte
s	52%	5573	- train/cXp7BCts2Qyb0iNEWhzY.byte
s	52%	5576	- train/cXuqeLQGf6dyh4ZIUgFV.byte
s	52%	5578	- train/CY2nyMb5XBKN0RaLsr9E.byte
s	52%	5582	- train/CYb1XLMZ8v6USIH0q3r4.byte
s	52%	5587	- train/cYIil3ZrukB1pV8Uzo2w.byte
s	52%	5590	- train/CYLbHFkG3R2N1ZAr6cdV.byte

s	52%	5591	-	train/CyIipwE0nFVsDBX8Zqmt.byte
s	52%	5594	-	train/CynZDa3JPgzdGSY6A0mX.byte
s	52%	5597	-	train/cyTsg9fViFZ8uLjUXGSv.byte
s	52%	5599	-	train/cYVKNj6JMAE9gaq0b81w.byte
s	52%	5604	-	train/cZ8wCtJL6unqdGBDsmXx.byte
s	52%	5607	-	train/cZD31Ea9HQfnUx0BzCRb.byte
s	52%	5608	-	train/Czdn32NEL4YfxARaJuM9.byte
s	52%	5611	-	train/Czgw1rG4JPoV5ksNDuEx.byte
s	52%	5616	-	train/CZnRQNX0g8IGEcdMq1lt.byte
s	52%	5617	-	train/czt805XbKR0YTujW4qvE.byte
s	52%	5619	-	train/CZvcR8GBrn5JIPdtqz1Y.byte
s	52%	5621	-	train/czy0Q1I3ebmEYXWjaGkr.byte
s	52%	5623	-	train/D0CpBFrm82d6aqcvnTLz.byte
s	52%	5627	-	train/d0j39TQiauSXbpsINeU1.byte
s	52%	5634	-	train/D14MK7pSBhfer3WxHztw.byte
s	52%	5636	-	train/d1EnrNW57S40FeVhfzPu.byte
s	53%	5639	-	train/D1jnAYav7uQ6GZ9cU5zP.byte
s	53%	5644	-	train/D2Ib3EY0wWHVZokK9u1n.byte
s	53%	5649	-	train/D2U61RktjeuTH3LJXFax.byte
s	53%	5652	-	train/D2ZKSx9I7MarJWeVtjls.byte
s	53%	5655	-	train/d3fFztVhA1OUSc5obTqC.byte
s	53%	5661	-	train/D3ukxpLVATJ8FR5m9Cjd.byte
s	53%	5664	-	train/d4ABW7XyvPzItZuoQ6MT.byte

s	53%	5666	- train/D4HgKrPeUyIIGFMv1tVR.byte
s	53%	5669	- train/d4mxVy1iuz037wDecN5U.byte
s	53%	5673	- train/D4UkKvnxZCswzAahu6iW.byte
s	53%	5675	- train/D51kzLEClKnMejm4VAU7.byte
s	53%	5678	- train/D5gT0IGAk81460VUqlye.byte
s	53%	5682	- train/D5mrsLY0P2V1UIXuJgES.byte
s	53%	5684	- train/D5rOGmu3ePpWqYX7bn6j.byte
s	53%	5685	- train/D5wceyp2lHBF8JP3qz0S.byte
s	53%	5688	- train/d603ro80lAQxKC9wGib1.byte
s	53%	5691	- train/d69TJVnLQBqZCIaY2o3E.byte
s	53%	5693	- train/D6b9fCwnyHXaipBtWrK8.byte
s	53%	5694	- train/d6bu9Ei0UP2lAnk3egIF.byte
s	53%	5702	- train/d7aHABNQz41Sni2XyVUJ.byte
s	53%	5705	- train/d7oLya1XJFWjEpRzmGVI.byte
s	53%	5706	- train/d7QPAmUKbcNh3xFZjivG.byte
s	53%	5709	- train/d7XHlwrYtyQk8CNIa0AU.byte
s	53%	5711	- train/d8A2vi70X5jwgrlU9N4z.byte
s	53%	5714	- train/d8HuwQGmjUDz736eFYrM.byte
s	53%	5716	- train/d8lFkrfZsQom9zgRqpLi.byte
s	53%	5719	- train/d8UcQkjpeM0EnN4YSCLR.byte
s	53%	5722	- train/d98DIzHL6lsYQS2X0URe.byte
s	53%	5727	- train/d9HvRYX1U7fGL0cEieVw.byte
s	53%	5729	- train/d9o6Is37la5eh0cQXtvY.byte

s	53%	5730	- train/d90jb8wkYFnaEDUoilxv.byte
s	53%	5737	- train/da1g3kNFpCPXWMTDrnVb.byte
s	53%	5739	- train/da4QzC5lpXIOi3gy6Fm2.byte
s	53%	5743	- train/dA9ztZeivK5lag017kJj.byte
s	53%	5745	- train/dAabnhJ37zorIS068q4l.byte
s	53%	5748	- train/dAhrESDjXnYI3kfmob7B.byte
s	53%	5751	- train/DaL5EU7cJxGbqos4Tukv.byte
s	54%	5755	- train/DAr5hm7i84gJzEWTUoeG.byte
s	54%	5757	- train/dauhrHMJspEmkvUWL65l.byte
s	54%	5759	- train/DAUzEY0LjqsFep1t4RQd.byte
s	54%	5762	- train/daYJhP9jKDTN5o8Z1b2G.byte
s	54%	5764	- train/daylB4KCKifT73MveWOR.byte
s	54%	5766	- train/DaZMvyFpkg6QBRnHILl5.byte
s	54%	5770	- train/Db6h0B1PFyELXmr95JGI.byte
s	54%	5771	- train/Db6Vg5QfaN97KxzclZjt.byte
s	54%	5773	- train/db7pXDtBiKLxP3J2WuEk.byte
s	54%	5777	- train/DbAU9EVq0yd2C5cw1ihM.byte
s	54%	5779	- train/Dbe0TMIG9gjHfv734XdL.byte
s	54%	5782	- train/DBKHf83CLms5xjqMGraY.byte
s	54%	5784	- train/dbLieaCTZ5rU8mV1uBQx.byte
s	54%	5787	- train/dBo2VNziGhKEm8S01jc6.byte
s	54%	5789	- train/dbpHcR9jZeurykDsAnMT.byte
s	54%	5793	- train/DC1wN6o8fus4iEUtQlJY.byte
s			

s	54%	5796	-	train/DC81WfMxtBvPa4wuTsQc.byte
s	54%	5802	-	train/DcgUNdMzXlah1kbxiYtv.byte
s	54%	5804	-	train/dcHe1LSBskpjf6JYiavT.byte
s	54%	5806	-	train/Dcj8eMZLv0QtymzArUJ5.byte
s	54%	5810	-	train/dCpHJ194WM8ecNxTKGur.byte
s	54%	5812	-	train/DcS9t6ymEZvnNu8rx4C7.byte
s	54%	5814	-	train/dcv0gytK4pFI1xaQSWHE.byte
s	54%	5816	-	train/DCYJfNIVretKzpH6yLxS.byte
s	54%	5818	-	train/dD3WMa15Af62kSHZXGEu.byte
s	54%	5821	-	train/DdHqQwxGnP7ICr1bvJTk.byte
s	54%	5824	-	train/dDqoZ30r2ja7chv8PSQn.byte
s	54%	5828	-	train/De0q3gzMjh2b8d1p9YmW.byte
s	54%	5831	-	train/dEeZW0x1rz8VsJ94BuPg.byte
s	54%	5835	-	train/DeHdPsMiVbvELXWn2w86.byte
s	54%	5836	-	train/dEIhM3zraJs4P1T7FiNw.byte
s	54%	5839	-	train/DEkyP78dbRNJ9Bvm4jVi.byte
s	54%	5842	-	train/deOcZLbwRBp9y10liCj5.byte
s	54%	5847	-	train/DeThaSN5zInYXH0cx2ym.byte
s	54%	5849	-	train/deTXH9Zau7qmM0yfYsRS.byte
s	54%	5851	-	train/dew08WY2tLPQ1VZsB1zS.byte
s	55%	5853	-	train/DEyOUKzN2g8suv96eYoC.byte
s	55%	5854	-	train/DEZyY3jrx1v9zbMASC8L.byte
s	55%	5857	-	train/DFd42XkRm6PnZ9Q8S0y7.byte

s	55%	5860	-	train/dfGtV514uRhiHZKDIE3B.byte
s	55%	5864	-	train/DFJWGx1HYve5iBUp0K46.byte
s	55%	5866	-	train/DfQ8pMRCoZqHSuBIyWnY.byte
s	55%	5868	-	train/DfSa4jlAyFPkXQpTbUoY.byte
s	55%	5872	-	train/dFUABXPZzn3mL060Ejek.byte
s	55%	5875	-	train/dfwmz1qMShiGeVUKYoNc.byte
s	55%	5879	-	train/DG4J93jNTMdyKxPqUzHb.byte
s	55%	5882	-	train/DGM9EySR0r8hLPcelzgi.byte
s	55%	5883	-	train/DGmbwEzu381CyXgAMB2q.byte
s	55%	5887	-	train/DgVEHi9NkMW5TSx7Pd04.byte
s	55%	5889	-	train/DH7q2gbnvTQ6CSFe1a8h.byte
s	55%	5893	-	train/dhFnCxfamsLv kubPZYjN.byte
s	55%	5896	-	train/DHjiEJsSkP0rpdF7RNqB.byte
s	55%	5900	-	train/dHpIeYK9gkS3DWE8ziBU.byte
s	55%	5903	-	train/dhS5TXjYM0yLufrw3xgo.byte
s	55%	5907	-	train/dHUq5yM8w0LND9PaepZY.byte
s	55%	5909	-	train/DhzFERM3B611SmNP2JTZ.byte
s	55%	5912	-	train/Di0ceNElhvZrdMRgbyYX.byte
s	55%	5915	-	train/dI2jS5i1QnP4b1ckxE60.byte
s	55%	5918	-	train/dI89k371NKgURwFqs0By.byte
s	55%	5921	-	train/DIAb8Wom3hCTuYKZpJFs.byte
s	55%	5922	-	train/DibnkXf1Y49wc0po7vyK.byte
s	55%	5925	-	train/dIfivq0V7gPeDQ95cynu.byte

s	55%	5928	-	train/dILobOrVQ9TGuhFU60JD.byte
s	55%	5930	-	train/diM03xh9Zv2Rm4nNVIjc.byte
s	55%	5933	-	train/dIq4Y9QGdh5yJ3lP1U8C.byte
s	55%	5937	-	train/DiT8evB0kIyGsdxqjAhC.byte
s	55%	5939	-	train/dIUsc2eFoGQ05ifjtn3R.byte
s	55%	5942	-	train/dIWe7sY89Bb1M5ifaJ3n.byte
s	55%	5945	-	train/DJ1XTk1KAM6d2YfigaQb.byte
s	55%	5947	-	train/Dj3FqX9IAHco2tG1U0ZS.byte
s	55%	5951	-	train/DJH5ptZyxizQUc0d3q4a.byte
s	55%	5954	-	train/djHpkGZoE0n64ub7WPer.byte
s	56%	5957	-	train/djK3pcVGIDlzCbT2wrYv.byte
s	56%	5960	-	train/DjV4Jdp805avuihmhTBNX.byte
s	56%	5965	-	train/Dk2IW6ub09LKcsHtVNzU.byte
s	56%	5970	-	train/dkbDn1I7TqHbpteJx60c.byte
s	56%	5973	-	train/dKETSOwafR8BIyGir4Qh.byte
s	56%	5977	-	train/dKQLjfbSZGaT5s71VoPY.byte
s	56%	5982	-	train/dKt4HhezElT2nBIP6c5F.byte
s	56%	5983	-	train/dkToRNx05vwZc20sSGuV.byte
s	56%	5986	-	train/DkyjfaKHbhTCYr0px2iQ.byte
s	56%	5988	-	train/dl1Rnu3LjSfcbJ6VT5tZ.byte
s	56%	5992	-	train/dLbutyFNmW7nfhUiE4cr.byte
s	56%	5998	-	train/Dlm7I8KFdwQ6CAMjOXgW.byte
s	56%	6001	-	train/DltQIwzFoACY78VGd1pu.byte

s	56%	6007	- train/dLVM6kPA8ru3EeK1qRZB.byte
s	56%	6012	- train/dm1b0f25FYHjuyveM4K8.byte
s	56%	6015	- train/DmCaN6PGnJpsWuMcKo8t.byte
s	56%	6021	- train/dmN7G3jpYsiuPOvWonh8.byte
s	56%	6022	- train/DMPv11eFd8B40sNICKZa.byte
s	56%	6025	- train/dmrZGyScpfTCRzUW8a5V.byte
s	56%	6029	- train/DMYi0TwJsZ5pt31Iaxz0.byte
s	56%	6031	- train/dN3DR9xeJwyLlHFo5sju.byte
s	56%	6035	- train/dncOwWDxE13Ri7XCMmUz.byte
s	56%	6038	- train/DnJFtCH7gTMiVwY1AxcN.byte
s	56%	6043	- train/DnMyV0BR6Wav1rOPK1xJ.byte
s	56%	6046	- train/DnptPIqdsURyYA5jkKTo.byte
s	56%	6049	- train/DNtJKyn0WV7PeS3q8Fv0.byte
s	56%	6052	- train/dNVnhr0jiTPeY98JGZoQ.byte
s	56%	6054	- train/dNyq0DTYLUsAj0ki2gGZ.byte
s	56%	6057	- train/Do9QfaXw52dYzcFipUeq.byte
s	56%	6061	- train/d0eUI4W0VjhNE7uF81z1.byte
s	56%	6065	- train/dogxTGpvCrYFKZcLOWR5.byte
s	56%	6068	- train/d0iY69mAf1JzMSr3vkcR.byte
s	56%	6075	- train/d0XyrkBugNZY2C8ojpze.byte
s	57%	6076	- train/d0ZeYQCpVaBDEJG2v5zF.byte
s	57%	6078	- train/dp4u8GQi5gJTHY12PsNq.byte
s	57%	6080	- train/dPBsX5TMfCq1KYitkS7E.byte
s			

s	57%	6083	-	train/DPg5ad8fwEq2jAWCF4o0.byte
s	57%	6088	-	train/DpNvYX1qedHLiQnRhF98.byte
s	57%	6091	-	train/DPVTHpkfLrABWaqgoyjw.byte
s	57%	6095	-	train/DQ5fK17GCZbnJs6oNWva.byte
s	57%	6098	-	train/DQ98CkXM6J3tY5m1hAfK.byte
s	57%	6099	-	train/Dq9nwWYMF50xf1PGuaZe.byte
s	57%	6104	-	train/DQJf10CT3gvt8pbMqHyX.byte
s	57%	6107	-	train/dQNpqtaDIbLOPH511kMw.byte
s	57%	6110	-	train/DqS4bCWK9kuvpN1Xgo37.byte
s	57%	6114	-	train/dr2jSJMq0sBozfZPLN18.byte
s	57%	6120	-	train/DRdEN1t1oex93HIvs6L2.byte
s	57%	6121	-	train/DReqVt7Hxa5p1h8v0QP9.byte
s	57%	6124	-	train/Drg6AkshP8zYZ1Jo0wxf.byte
s	57%	6126	-	train/Dri7s2veJT4hj0R3aHkf.byte
s	57%	6131	-	train/DrN06fVRJLoqYK4Bkdy0.byte
s	57%	6133	-	train/drq576zAkIsjDpa0CLQ0.byte
s	57%	6136	-	train/DrSkvafcTm2A4BNzdVUt.byte
s	57%	6142	-	train/DRw7V8B0HGF1mpjkeadN.byte
s	57%	6145	-	train/dRYGQSA8TrmIPtabWNeu.byte
s	57%	6147	-	train/drZagbnxePDH7JqYt2w0.byte
s	57%	6149	-	train/ds2LojgFNKU7aCTM8VYD.byte
s	57%	6150	-	train/DS3nJy1YKdXQp7wbghk8.byte
s	57%	6152	-	train/dS6twLfzX3iyNv8EBnam.byte

s	57%	6154	-	train/dS6yJEktRlWQsBaIFYCP.byte
s	57%	6157	-	train/dsB2UgC4XFQ1cj8DMyhT.byte
s	57%	6159	-	train/dSCoRme1bUOs3t9Bp8g5.byte
s	57%	6160	-	train/DSd3HQnFN7iUAxbYBEju.byte
s	57%	6162	-	train/dSIY6QxZ4tXnoP8lsqmb.byte
s	57%	6165	-	train/dSPzTKHQtaUNgG5w2Luj.byte
s	57%	6170	-	train/DSwecTduKPtZNQoVR8qh.byte
s	57%	6172	-	train/DszCbghZwq70Rp8f62J0.byte
s	57%	6174	-	train/DT2t5JMrWnp3gmuhLBc8.byte
s	57%	6175	-	train/dt3fRl4c06CpjsLHNWgn.byte
s	57%	6176	-	train/dT4zQLE6Z2kCapGr1FYy.byte
s	57%	6180	-	train/dtcBSuoU8MZ0wilHFe7n.byte
s	57%	6184	-	train/dTExc1o357Z9Wghz8VIq.byte
s	58%	6187	-	train/DtHRwmIsBeconYmWh836.byte
s	58%	6189	-	train/dTI7mpvZHQsJNjl0g6Gk.byte
s	58%	6191	-	train/DtnYlsgyQ654Xiw3dkoR.byte
s	58%	6194	-	train/dtrU9fswTWbzBmAK7onk.byte
s	58%	6197	-	train/DTUYnFKf1buGgsAecxL1.byte
s	58%	6200	-	train/dtyLupxMfEQVX10JwvWl.byte
s	58%	6203	-	train/DU2wypXWRjvbh4oAd0ex.byte
s	58%	6207	-	train/duDYQ8vkb29jNryIchzX.byte
s	58%	6212	-	train/duiRH6TLGrhgXSBmfvEt.byte
s	58%	6215	-	train/DUsyJ2AwpZHS5Yuxobei.byte
s				

s	58%	6216	-	train/DUVqFkRTd3MnipSGWgbX.byte
s	58%	6221	-	train/dv4yexuAzXwGUENnBhro.byte
s	58%	6223	-	train/dV8HAvzFer2naZEwu1Pk.byte
s	58%	6224	-	train/dV8TQtBk3AK6nFq4zihw.byte
s	58%	6226	-	train/DVBG6mNt8U0iye7MgIPK.byte
s	58%	6231	-	train/dvgWEDFtuNPLUfc8ST5e.byte
s	58%	6233	-	train/dVmW97NCRysfpkMxSuBi.byte
s	58%	6236	-	train/dvqB8LFJpKMkZU19mAg0.byte
s	58%	6238	-	train/DVsfp6pU89nBdhZHqxI5.byte
s	58%	6243	-	train/dvVcMR0BzWxa8hAyZNHg.byte
s	58%	6246	-	train/dw1szJLjHfcmE49uZKvW.byte
s	58%	6248	-	train/Dw2sXhq6imtArjTZo3YC.byte
s	58%	6250	-	train/Dw4btEPMk3reBuNsZ15S.byte
s	58%	6252	-	train/DW4JTXzdAu08aLeQOFPt.byte
s	58%	6255	-	train/dW7PwfxAsy8HLFn3NXtB.byte
s	58%	6257	-	train/dwBC17xMSWvqFiZcJ1am.byte
s	58%	6260	-	train/DWFmpGk9T0cqB1f3YCjy.byte
s	58%	6262	-	train/DWixC3uFh1E6J4K0rfsZ.byte
s	58%	6264	-	train/DWnB8cNMprEmkXTtHu90.byte
s	58%	6267	-	train/DwOFyEfhlGeMU2q0vbZa.byte
s	58%	6272	-	train/DwtbW2PVfAcyQMTE5IoR.byte
s	58%	6275	-	train/dXB9U7M1pgt81PGhyNuo.byte
s	58%	6280	-	train/dxMk1zvUyh8IPglO25ZN.byte
s				

s	58%	6282	-	train/dxq8SW04gJvZH2NBfPpD.byte
s	58%	6284	-	train/dXrumv84P3aSAcHfgZhT.byte
s	58%	6287	-	train/DxXaHhzNWCijY0bQsPry.byte
s	58%	6288	-	train/dxXyVcRlfr3t6skMUPCK.byte
s	59%	6291	-	train/DY3QpwMcCOHsSWuVe89P.byte
s	59%	6294	-	train/DYh2UtEr7A0JsjmFQZvB.byte
s	59%	6296	-	train/DYifKtF806I7JjkLHZgP.byte
s	59%	6300	-	train/DyMjiCx7R4JG5nIgfNKz.byte
s	59%	6303	-	train/DyOK7Wh5IpPq4bZ62oGH.byte
s	59%	6305	-	train/dyTqfGCNl7Hv21oZLMK5.byte
s	59%	6308	-	train/DYX2E3otnwLK1cIebNuJ.byte
s	59%	6311	-	train/Dz6QkSVB4ax1siyK8fGP.byte
s	59%	6312	-	train/Dz8kXNIW71w91dKMAEY2.byte
s	59%	6315	-	train/DzbAP8aBvgfkwVn29yTX.byte
s	59%	6318	-	train/DZebn8mxWJYjA14adQgF.byte
s	59%	6322	-	train/dZj7fy0liEU4QkcvH0YJ.byte
s	59%	6324	-	train/dZJi4Ayf1xpbrXz9s6Nh.byte
s	59%	6325	-	train/DZpbRaEg4YIA69j7Wmot.byte
s	59%	6330	-	train/dZTvVqYPXm49Mf3uksbl.byte
s	59%	6336	-	train/DZvJ1nBiqal34YSwo8k0.byte
s	59%	6340	-	train/e0fdQp9N2PYB5ShZy4GD.byte
s	59%	6344	-	train/e0okgCIVZru7T1Av5qz6.byte
s	59%	6346	-	train/E0qNPVLshkYD0gyto91H.byte
s				

s	59%	6347	-	train/e0rtoZ7S9LHuzhwyxcsi.byte
s	59%	6350	-	train/E0UuPpkYH zrZnSM6XbhC.byte
s	59%	6352	-	train/e0Yr26MQJovX8cdCV1tU.byte
s	59%	6353	-	train/e0Z7VP4EL1n1BH2haFo6.byte
s	59%	6356	-	train/e14oxuL6Cb8H07d3zSqn.byte
s	59%	6360	-	train/E1hdL4PQ513Gc82vyeAT.byte
s	59%	6362	-	train/E1lymL5AhVWeUwCtv68I.byte
s	59%	6365	-	train/e1Pxcv2KRUSyDWmAasid.byte
s	59%	6367	-	train/e1RMvSd83AgaCQY4PT2u.byte
s	59%	6371	-	train/e285vEry1U9RkNdZhG3B.byte
s	59%	6374	-	train/E2kSi8TAfjsCg0Hcto1r.byte
s	59%	6377	-	train/e2r6IncxE1LQ0KFGgphj.byte
s	59%	6379	-	train/e2taZQLKCS nZTV7d0oh.byte
s	59%	6381	-	train/E3by0trSIqhfnYPQmLk1.byte
s	59%	6383	-	train/e3HSBfTA9MXzhEN2qIpL.byte
s	59%	6385	-	train/E3SfzdYktZBm8ov65e4P.byte
s	59%	6386	-	train/E3sij5M1NGadWfcSDvQw.byte
s	60%	6392	-	train/E4eSC8pqYdrkWfB0tFv5.byte
s	60%	6397	-	train/e4jdKn6hHpiNF1Rtv7xJ.byte
s	60%	6398	-	train/E41BhCXvziDf2rU3GM7t.byte
s	60%	6403	-	train/e4UyrvS8H0X5tFqGo6jd.byte
s	60%	6405	-	train/E4wCHstIQWNDe81Z90kp.byte
s	60%	6407	-	train/E54k3zBMYb1V9FHd7TGD.byte

s	60%	6409	-	train/e5apYjv4ZGERk20zF1y6.byte
s	60%	6412	-	train/e5fn4GutJ0601CgUSZF8.byte
s	60%	6414	-	train/e5iHh49LgopGV8sJ2fq1.byte
s	60%	6417	-	train/E5LXhtYsZMWfA4ebSVUk.byte
s	60%	6420	-	train/E5UkBdwnH0T3yjeIf0Pv.byte
s	60%	6423	-	train/e5wb09ugRASHGF13Zj7y.byte
s	60%	6429	-	train/e6G1p0jLSDA3wxbnPsm0.byte
s	60%	6432	-	train/E6PX3bYws7GnvQ02JHBS.byte
s	60%	6436	-	train/E6vzaC2gPsTfyd03Lmo8.byte
s	60%	6438	-	train/e6woXQkDrjFVxIu0tdfh.byte
s	60%	6439	-	train/e6xiQwfrT10IJGVYhnqD.byte
s	60%	6443	-	train/e7BFP34ZIgJaWMYXhL1S.byte
s	60%	6445	-	train/E7edRP1GY9vVUS5pZ8Lk.byte
s	60%	6447	-	train/E7JvkZyTKF6fhWBUrosV.byte
s	60%	6450	-	train/E7QxgmVGInW4LC1byA5w.byte
s	60%	6452	-	train/e7vxkP0zwIYT9N4ChXuf.byte
s	60%	6453	-	train/E7yDMZh286B51eVJ4wzg.byte
s	60%	6458	-	train/e8UthBGf0zAkSP2iVr3l.byte
s	60%	6461	-	train/E8YKtDkh004AxmGayQ1c.byte
s	60%	6465	-	train/e9a1oKnLSH8JGDjQ2B0T.byte
s	60%	6468	-	train/e9d614aPtXTmKu7Jjf0E.byte
s	60%	6471	-	train/e9Jm5ZnFL3IS6ANtf1q4.byte
s	60%	6474	-	train/e9M8Qjno0aqN3HhLZky4.byte

s	60%	6477	-	train/E9wm8zARbH1j4TW7VMrK.byte
s	60%	6479	-	train/e9yfU12D1pbTARXYdckn.byte
s	60%	6481	-	train/ea4vnMoYFy8D03BEKbwt.byte
s	60%	6484	-	train/eagtG7IQYUKLZ3qs9yTE.byte
s	60%	6487	-	train/EAPboVp78v6st050m3BL.byte
s	60%	6489	-	train/Eat0QHsRMXTLP8jYxu04.byte
s	60%	6494	-	train/Eazj0IX9b3PmLoA7SHg4.byte
s	61%	6498	-	train/EbBqJ7r0tZ45nIlxhKfc.byte
s	61%	6501	-	train/eBD0vM4fUQmy0XEhracH.byte
s	61%	6503	-	train/ebI3LHRqBUwco8aQtW15.byte
s	61%	6509	-	train/eBST0wXfIAqj86r9VY5b.byte
s	61%	6510	-	train/ebT5zBZIQDVuxP8HRFyc.byte
s	61%	6516	-	train/EC1WfFmvIi0sx8lzwDAZ.byte
s	61%	6517	-	train/EC2yymbNV8rSzRu4qZA1.byte
s	61%	6520	-	train/EC82QvRNxyZ0iAKh1o0V.byte
s	61%	6526	-	train/ECiA7GPQj6MNZtSJvRqL.byte
s	61%	6530	-	train/ECM0gid48UH3tfPyJW1m.byte
s	61%	6532	-	train/Ecn1RwFdJBsYNQo8Vz01.byte
s	61%	6534	-	train/ecnym1Q0Xg3BADTJf2WH.byte
s	61%	6537	-	train/EcurA2WMeGk5Tv9V4LH1.byte
s	61%	6541	-	train/ECzKJTQ1pj6PDIMnXx40.byte
s	61%	6545	-	train/Ed4GPSVXx8yneMLZ1hKj.byte
s	61%	6547	-	train/edabs9nDHqGtNkgQ1uzm.byte

s	61%	6552	-	train/eDgCZMUwHsh2az0yLb9o.byte
s	61%	6554	-	train/EdoAPfeszUw7qQGYIOa1.byte
s	61%	6559	-	train/EdSkji9WtJ3wTIHB8hQM.byte
s	61%	6562	-	train/EdwQ1my8754NBkMaJxGA.byte
s	61%	6569	-	train/Ee80BjQvgWns9XKqV40r.byte
s	61%	6571	-	train/eECbifLAG9D1S8tMyIQY.byte
s	61%	6575	-	train/EeY2GNUFngl7CdaDjZf0.byte
s	61%	6576	-	train/eEZIGWQA2g79rmuShpd6.byte
s	61%	6581	-	train/EFBARE5CnfmPuWThjKUZ.byte
s	61%	6584	-	train/eFK1IWckUmuM3oCAjON0.byte
s	61%	6587	-	train/EfScYk5oVPw4Jxni2s3b.byte
s	61%	6592	-	train/efuhvT79QdsU0DMgJXma.byte
s	61%	6596	-	train/eFz1q0CTHycYP8v3mxqQ.byte
s	61%	6598	-	train/eG1bTK5AB7Mryunti3m0.byte
s	61%	6601	-	train/EgBSwQp9j7F02cZnehvm.byte
s	61%	6605	-	train/EgdZfMrUuFpcJYmS3HVA.byte
s	61%	6607	-	train/EgeBpshOLYbwyu0ZdrvX.byte
s	61%	6609	-	train/egFM4CjVl8Bz5pbN1Ux0.byte
s	61%	6615	-	train/eGMTAm0csZLI9QbJpCUY.byte
s	61%	6618	-	train/egnhwcvsVjwONCU8t2Ju.byte
s	62%	6621	-	train/Egr3Y2WxNbavsDdj1MVS.byte
s	62%	6624	-	train/egUM629t8ACVhj3Bp5rR.byte
s	62%	6628	-	train/EgymHkwfrqWdp28M7dJe.byte

s	62%	6631	- train/eh6DbBL3XNnE11cxtKkY.byte
s	62%	6632	- train/EH7DRYeINXPfJUAG0Lqx.byte
s	62%	6638	- train/EhHWtoiAX4kZp3s7wVUc.byte
s	62%	6640	- train/ehIMXRTHGQibpNKkuZmx.byte
s	62%	6645	- train/EhR93mQKyvaSF5euIlik.byte
s	62%	6651	- train/eI053W17vGE249i0YyHK.byte
s	62%	6654	- train/Ei1t7BmK2Yh0db4QNXzL.byte
s	62%	6656	- train/EI4Jgyofj8CHR0ts59Ur.byte
s	62%	6657	- train/eI5Fqz3790SqaiRHpscX.byte
s	62%	6663	- train/eigPHkQWn1yJ8tpoAXYM.byte
s	62%	6670	- train/EiN0dCvKSfymUpfhsbu1.byte
s	62%	6673	- train/eiR2pUnKVjlrMEdPcBDx.byte
s	62%	6678	- train/EJ2ZomyQW0zLdbxTGVvj.byte
s	62%	6680	- train/ej60US115DPCKGpvrFyw.byte
s	62%	6684	- train/ejK11rfCdNmypPw5suUk.byte
s	62%	6687	- train/EJOGkzrY3FVAthgmKicu.byte
s	62%	6690	- train/EjSPcs4ifvgNxIZB9Q0Y.byte
s	62%	6692	- train/EjWBahQA6X7R0n5foySV.byte
s	62%	6694	- train/ejXdxvps1Wo5gckEh1Ha.byte
s	62%	6695	- train/eJyCVaGpDWgqL4uXv18K.byte
s	62%	6697	- train/EJZfWS6XwhotmT71g0jM.byte
s	62%	6701	- train/eKbBCxVTnXJWySaDrLwG.byte
s	62%	6707	- train/EKHpgnWV7uGJId048wXB.byte

s	62%	6709	-	train/EKmgShQsf6a9vY0znN1U.byte
s	62%	6711	-	train/eKQW3oFCzAEyGumBlgh5.byte
s	62%	6714	-	train/ekSMGjq4FE0L3tux1Uic.byte
s	62%	6717	-	train/eKTudHf0rpmbDUExgSn7.byte
s	62%	6719	-	train/eky5W7FjRS1uVH2M0PIb.byte
s	62%	6722	-	train/EL48d6uXjZxrHF5bsB1A.byte
s	62%	6724	-	train/ELanoVlRXbYv26US0T8r.byte
s	62%	6728	-	train/ElkwpXW31RdCmVbI0Var.byte
s	62%	6729	-	train/ELmatIK0qT8kuJWFjcwR.byte
s	63%	6734	-	train/ELuV9FA7IjKXR52NrokU.byte
s	63%	6736	-	train/ELxoB14Hci00ry2nz9qt.byte
s	63%	6739	-	train/eM39XrI0GpTDhjSyQdx8.byte
s	63%	6743	-	train/emB5kWB EAGX2YQLS6jpg.byte
s	63%	6744	-	train/Emc01jZdFVvC17bWX2Ae.byte
s	63%	6751	-	train/EMmQTRgzWJ7FLbuZAaoP.byte
s	63%	6755	-	train/emUFPnYtVzvKxo42iQqE.byte
s	63%	6757	-	train/eMZYXFS0jiHTV5y742UC.byte
s	63%	6760	-	train/enD0hZTwMtLNIW2rB6GY.byte
s	63%	6763	-	train/eNhamfpF47A0WXjSiTqk.byte
s	63%	6766	-	train/Enm8Drf9FPs6yjkHvowi.byte
s	63%	6769	-	train/enqtU6j094EJFaY3MQoS.byte
s	63%	6772	-	train/eNvtBzPm6MnGE8rJQcLF.byte
s	63%	6777	-	train/E0eh6tPNzBog5vQRMF0r.byte

s	63%	6780	- train/EoLuvjUwHXh2saIBzQb0.byte
s	63%	6781	- train/eoMlxiaXtB1gGJ9D2qSV.byte
s	63%	6783	- train/E0poZ2Li1jz5ebtgsaIV.byte
s	63%	6787	- train/EoUzWsPd9JBVSyTM2cvX.byte
s	63%	6792	- train/EoxXzTHVP6QkujRtm1ns.byte
s	63%	6794	- train/E0z1VPtvNhC07QugnxW3.byte
s	63%	6796	- train/Ep1cRwezPyQsAL7GMm1F.byte
s	63%	6798	- train/EP5gD2SpuBbczrkQXKTm.byte
s	63%	6801	- train/epboI43NYGPHLySKiAs1.byte
s	63%	6808	- train/epj8anqLcrRxHGVWT1Bg.byte
s	63%	6811	- train/EPKSp8Nue3tGUL6mRxd.byte
s	63%	6814	- train/eplt1BgqxYo6VuTAwf2n.byte
s	63%	6819	- train/EPuBR4UHTaSW0lemC9QV.byte
s	63%	6822	- train/EPz8I26gZmC0iGbWXwhx.byte
s	63%	6825	- train/ePzxjFayC48I0Xb9kuS3.byte
s	63%	6829	- train/EqdNe3DySZfhFlrkcmTu.byte
s	63%	6830	- train/EqGiNdOX79rVz8RsM0Cv.byte
s	63%	6832	- train/eQiJI9UaoyCMudbDKz0s.byte
s	63%	6835	- train/eqMtk743YWi6Dpm2Hab1.byte
s	63%	6838	- train/eQUEjktxdVNPqmlw3nJW.byte
s	63%	6842	- train/EQZuSxaCGw4mpAKycjLV.byte
s	64%	6844	- train/er3fFEimPIGX0dHL4tx1.byte
s	64%	6846	- train/Er6sn81fNIRqa2Db5pTu.byte
s			

s	64%	6850	- train/ErCJD5UsMnpP3oVQk41z.byte
s	64%	6851	- train/eRf9nXMwWl80G0TxbgKv.byte
s	64%	6854	- train/ERjWvo0hJSgF4pUb9drB.byte
s	64%	6859	- train/eRLxnaXMK0zUtEjvi9H2.byte
s	64%	6862	- train/ERUu09IMtDwszYTj6ckm.byte
s	64%	6869	- train/Es1glGAM8aT7uXW0yNfc.byte
s	64%	6873	- train/esaXkil2E04B7Y1ypNwK.byte
s	64%	6876	- train/EsBRxv5MTp296yaKkwjW.byte
s	64%	6878	- train/esfmSy1PQMkGNb0BDxZR.byte
s	64%	6883	- train/eSJGtd8D90oRFyPbufl6.byte
s	64%	6886	- train/ESncNgeGJIQujo6v0TBa.byte
s	64%	6888	- train/esPBa5nQFc0qIyZ3VMhr.byte
s	64%	6890	- train/EsQSLYTmVKNxr56Z41R7.byte
s	64%	6891	- train/ESqUten2YJiswjZodR40.byte
s	64%	6897	- train/esXbGRDMpEzTQV5gIiaU.byte
s	64%	6901	- train/ET2AayjsGuUHKY43mFx9.byte
s	64%	6903	- train/eT5zyaVgSRMwGc2oQfHW.byte
s	64%	6907	- train/EtGBOKD2bpQI3ZwJRPmc.byte
s	64%	6911	- train/etiXx5SskyzfIu0aqWJ7g.byte
s	64%	6914	- train/ETlu0SjYRQnJ9r3pLDKq.byte
s	64%	6917	- train/eTQhnbRXam27Kwqj1MUy.byte
s	64%	6919	- train/eTt2HsXmqIJRhvaEfN05.byte
s	64%	6921	- train/eTUdK5jomy9xn3kB20WZ.byte
s			

s	64%	6924	-	train/Etxjz6KuaSJNY7TeBA2L.byte
s	64%	6925	-	train/EU2ISJQQR87vw1t5opAu.byte
s	64%	6927	-	train/eu5p80LQma3sw20hkiZ1.byte
s	64%	6932	-	train/EudLvJYRpI7K3r5MBPnq.byte
s	64%	6937	-	train/euLCGs3vYh5nwgZQt2H4.byte
s	64%	6940	-	train/EUW4FKIai0y15NrJv0RM.byte
s	64%	6942	-	train/euYEOdm6IyNdvHa5xcMi.byte
s	64%	6944	-	train/EuyYZNBLqI2maPtzxFsQ.byte
s	64%	6946	-	train/EV5qiMnPCxDZIRgmwk0b.byte
s	64%	6947	-	train/EV5y3k40fPat2g0TAbHl.byte
s	64%	6948	-	train/ev8WLIbF5yHECBUz9rZw.byte
s	65%	6950	-	train/EvIHo01arDd5U3unLVes.byte
s	65%	6952	-	train/EvMY086fGxSLUTKyX1n5.byte
s	65%	6956	-	train/EVxmNT7dKfBMebuyaRkg.byte
s	65%	6962	-	train/ew87VkSudbX1Qzqc6AUF.byte
s	65%	6966	-	train/ewGVqv4Ris50oU7X6acb.byte
s	65%	6967	-	train/EWhzrqv2D3kBYb6SKpU1.byte
s	65%	6971	-	train/EWjvFU5GXeSaHzxBh0f9.byte
s	65%	6972	-	train/EwlC6ehLbWF0VgJd0fs2.byte
s	65%	6974	-	train/Ewn8395kMeDfBKxXJQd4.byte
s	65%	6981	-	train/EXapH4diUyGeJ95vf1YT.byte
s	65%	6984	-	train/exC0jIpaqfdWOPwgrKoV.byte
s	65%	6987	-	train/eXGRpaHzil5q0PQdoSb6.byte

s	65%	6990	- train/EXJvSznUcQjwyD4rGCNq.byte
s	65%	6992	- train/EXm7GqWzb8pMQ2Ucyx0a.byte
s	65%	6994	- train/EXN8IoYcZs1qaCMBPtWA.byte
s	65%	6995	- train/EXOGAWFx2bptk1Nmwlui.byte
s	65%	7000	- train/eXu9bAoQ5UWBfCz3KsJN.byte
s	65%	7003	- train/exuybVnG2JL6rKshkRg4.byte
s	65%	7006	- train/EY4FNVtg08Q75rCvcI19.byte
s	65%	7009	- train/eY9E3uQM1JhbocItT5R2.byte
s	65%	7010	- train/EYaB9R8TS2gtmQNJvuWc.byte
s	65%	7011	- train/eYcJ0kLMFuNyoAXi91dK.byte
s	65%	7016	- train/eyM13bXGZU1ERs8pxWwm.byte
s	65%	7019	- train/Eyp1hkveqJKYwHmdXViM.byte
s	65%	7026	- train/ez3yDwEmqdTGY5SMv9Z0.byte
s	65%	7033	- train/eZjWiTlI85zsSycCfGMV.byte
s	65%	7034	- train/EZlzyM9sapy4uWKgcQ15.byte
s	65%	7037	- train/ezSnoZwWGUQfIDF3k58b.byte
s	65%	7040	- train/EzwfYMsrq7AnU3t5hCDy.byte
s	65%	7044	- train/EzZBAe2N4tSm0UIHcC3n.byte
s	65%	7047	- train/F0N7nEHia0fxj2kJzGXd.byte
s	65%	7051	- train/f13zcuZU92T6Md0ynX4r.byte
s	65%	7054	- train/F1Gznj82acfQMmHNRLAE.byte
s	65%	7058	- train/f1oeHaTRizBkKXEnLGcF.byte
s	65%	7060	- train/f1Zy0M3PrxVHzEks0aG5.byte

s	65%	7062	-	train/f2104Ztv6hQPDnaYz9B7.byte
s	65%	7067	-	train/F2MhkaU7vXxQrqCpHtoE.byte
s	66%	7070	-	train/f3Cmly2RMnFq57Yis891.byte
s	66%	7072	-	train/f3ESKUTJrxZQ5nBjVPgG.byte
s	66%	7076	-	train/F3LZkyqGQV79ARhuE2HY.byte
s	66%	7082	-	train/f4BqH0J1xLkYKI3TEQWm.byte
s	66%	7084	-	train/f4FGZxJb07UsA1LvYpew.byte
s	66%	7086	-	train/f4jUEqOubyIYo3Fa0PVC.byte
s	66%	7087	-	train/f4SBYHe3GaDiykbJtLdX.byte
s	66%	7091	-	train/f5DGWZn73yvJkLPYQihK.byte
s	66%	7094	-	train/F5nJC0Ii49ZHxkAY2bBz.byte
s	66%	7100	-	train/f6FPy1iIWB4swGUQDT8X.byte
s	66%	7102	-	train/f6kY4VAMsg3iJ90IeE2Q.byte
s	66%	7106	-	train/F6PujHJGnIyNLqEzbXvB.byte
s	66%	7111	-	train/F7jRh2B8TXWfnut3VdkA.byte
s	66%	7114	-	train/F7MWGPqic0rbd1KCImEk.byte
s	66%	7116	-	train/f7NWzt8HU0bgdZF14DaS.byte
s	66%	7120	-	train/f7YPy6jIpSWrhD1zwFXQ.byte
s	66%	7123	-	train/F8bN6Gx0wTRZWsHI3XL0.byte
s	66%	7124	-	train/F8csP5rZEx346TXkojt1.bytes
s	66%	7129	-	train/F8Vlm6LJMgyE50a1xbSX.byte
s	66%	7133	-	train/f9E4HBYXeghm8ViJaKbr.byte
s	66%	7136	-	train/F9nT2NgCsbrhX8eqOp01.byte

s	66%	7141	- train/fadiNiZQsxUMBElZgc97.byte
s	66%	7144	- train/fak1b0n71SJFw5WceiKP.byte
s	66%	7146	- train/FaKJdlpSU10sgw7jmQhe.byte
s	66%	7148	- train/FAMI9aEG2hk8BKeNYzmZ.byte
s	66%	7150	- train/faOmpIAWF9ZURkYqg8vr.byte
s	66%	7154	- train/FawW0xhrVe4ysCXLkA8u.byte
s	66%	7157	- train/FB7P8fucig2os1TCEept.byte
s	66%	7159	- train/Fbdqw108uxIcZ6rDMgmY.byte
s	66%	7163	- train/FBHvx5A1IchTEYDPkU0V.byte
s	66%	7165	- train/FBJC0htji5egyV1RPnHS.byte
s	66%	7168	- train/fBr3Rkmu1KAbXEQ4nvzg.byte
s	66%	7170	- train/fBRjL0xWbPyreYn4hvkU.byte
s	66%	7171	- train/FbtjwxYNJlpdC764rITZ.byte
s	66%	7175	- train/fbvWPV2U0xQ3TpqohAFy.byte
s	66%	7181	- train/fc9yYUTNz0qWibKtwR2j.byte
s	67%	7185	- train/fcCVKbupXtDHZkMisaJw.byte
s	67%	7187	- train/fCgvNxUcn71sZi00EjPI.byte
s	67%	7188	- train/fcId811TeFim0jMzXo6R.byte
s	67%	7191	- train/FCV1U4vTjmHoun5QE9s7.byte
s	67%	7195	- train/fCyAx0nr81FhsXuNeEkt.byte
s	67%	7196	- train/fcZDXW9nd11QeiPHjIC0.byte
s	67%	7199	- train/FdBUjIT8P2iN00Dw3Jxt.byte
s	67%	7202	- train/fdDxaCeIbvKowFXhMGQJ.byte
s			

s	67%	7203	-	train/fdGV3HjRu4nBSMNibIkX.byte
s	67%	7207	-	train/fDJXgvz2cAtPopmIO0hi.byte
s	67%	7210	-	train/FdlaszJrb8ZjBEKXo0nc.byte
s	67%	7212	-	train/fDLvOn85VGbX6BNsTtaY.byte
s	67%	7215	-	train/fdMKx46g3GqtF0oLsPNU.byte
s	67%	7219	-	train/fdSHkFm8b2XBevTGCYM4.byte
s	67%	7221	-	train/FdTt7i63rIxeSv1scBn0.byte
s	67%	7223	-	train/fDvcA15FYzlk7Jr3I64X.byte
s	67%	7226	-	train/fdYytF1l3QMXAP5axJnV.byte
s	67%	7228	-	train/fe3ahn9iVGtqZkWc84AC.byte
s	67%	7231	-	train/fe9snrXGxKw1L28tDpRv.byte
s	67%	7234	-	train/feCLEnBkHFvKpWirShuV.byte
s	67%	7239	-	train/fEMdwkrczmt7j0RTL14J.byte
s	67%	7243	-	train/fEsMVSGolZeiqgz1pPKy.byte
s	67%	7245	-	train/FEWwOKmxAqVZPX6r1eLg.byte
s	67%	7252	-	train/ffTm4jSh6IQ0toiY09qw.byte
s	67%	7254	-	train/FfvCzj8YXUbDsLVARQmG.byte
s	67%	7260	-	train/Fg8vNxDcn4oRuG3MZ0d1.byte
s	67%	7263	-	train/fgD5k647jbdSEWBZCQnh.byte
s	67%	7268	-	train/fGL0FWUZNaCElQ8nIiJ3.byte
s	67%	7270	-	train/fgnYk4dvDjt0VoSH8BE6.byte
s	67%	7273	-	train/fGwBYrXVc8oWC7uNnbag.byte
s	67%	7276	-	train/fgYxPlSK2MeTiGzjbsI8.byte

s	67%	7278	-	train/fh1V8wnr0JAZGzsRQB60.byte
s	67%	7279	-	train/fH2ajwFimY14pR8NbLDT.byte
s	67%	7282	-	train/Fh9HCK50ZId4PbcQDTJr.byte
s	67%	7284	-	train/fHELzXgxKRC4TFikt1GB.byte
s	67%	7290	-	train/FhpBuDLoktwm9v3grXd4.byte
s	67%	7293	-	train/fhTWnxRPbIcSqtz10CoD.byte
s	68%	7297	-	train/FhxiaMwrVAFxKq7NYkvU.byte
s	68%	7304	-	train/FiGCD4zX3Nr1o8QPsk6c.byte
s	68%	7306	-	train/fIGuejtwV8C3bQ6FEno0.byte
s	68%	7313	-	train/FitUnvqQ59PElo4uLph0.byte
s	68%	7318	-	train/fiXI4tZdP6M2gKW0L1bj.byte
s	68%	7321	-	train/fjbbqGwVH6ANh5CRluirI.byte
s	68%	7323	-	train/FjdkOVL0pvfCyJKhiEat.byte
s	68%	7325	-	train/fJh0PBLs1RbxtymGZUfr.byte
s	68%	7329	-	train/fJrQ8Y32EzevA5D0itRN.byte
s	68%	7331	-	train/FjS1eW6GDVdh2K0xnz37.byte
s	68%	7334	-	train/fJYw618qML1RsU0BeZGp.byte
s	68%	7336	-	train/FjzLT1pN9kgZuJP5se2o.byte
s	68%	7340	-	train/Fkha6WrBdNbMje70q0T4.byte
s	68%	7346	-	train/FkThbmQlgnjGPdV3p8vB.byte
s	68%	7347	-	train/FKx08J75rj3Cs1od0yTX.byte
s	68%	7350	-	train/FL235qi0sXfm9WlcCgzd.byte
s	68%	7353	-	train/FLBerci0hxEK1fYtUon5.byte

s	68%	7357	-	train/fLHpYZzIj3MT2vAXr54o.byte
s	68%	7362	-	train/FlnCzbtLd1mwWJf6pIEZ.byte
s	68%	7365	-	train/FLumNoC2IxqDP1HSj3zh.byte
s	68%	7367	-	train/fLvDejnz5S0RQbJON8CV.byte
s	68%	7369	-	train/fLZPWHz4YsoCp0Fh9k6G.byte
s	68%	7373	-	train/fMCEj5tSwasRNbJepnuY.byte
s	68%	7374	-	train/fmcNPq3vhgr5Tx8ZaRWI.byte
s	68%	7378	-	train/fMeiEQ4BrOm0dCgY71Fu.byte
s	68%	7383	-	train/FmLkTo9MiyNfSdH5e12Q.byte
s	68%	7389	-	train/FmV4kRyzJewN19aEjArv.byte
s	68%	7390	-	train/fmXTOnYxKoQUHVRDL742.byte
s	68%	7394	-	train/FN51MWj0kPY3SBRcoHvz.byte
s	68%	7399	-	train/Fnda3PuqJT6Ep5vj0Wck.byte
s	68%	7401	-	train/fnHDty72dWwjKXgYqUPA.byte
s	68%	7405	-	train/FNkr36t4qTZoy5veUDRm.byte
s	68%	7409	-	train/fnPJ4EFtWhxVsIS3bMjz.byte
s	68%	7411	-	train/FNRvlymt10cLEBAD0j9u.byte
s	68%	7414	-	train/FnXVjLgJDclYUw9h213B.byte
s	68%	7418	-	train/Fo3l7eAEM4XDmKyIV9tj.byte
s	68%	7419	-	train/fo7WkX580Ta4VpFd1Kwh.byte
s	69%	7424	-	train/FoD16RIxbOn70KasSQmC.byte
s	69%	7425	-	train/FoEDHqjGNb24vXAOPpKc.byte
s	69%	7429	-	train/fojWIsY21tQG8K6BCKgV.byte

s	69%	7432	-	train/f0oBREwIcNmD8u6Kts75.byte
s	69%	7435	-	train/FoS47rVe0NgkwqDayJ3v.byte
s	69%	7440	-	train/foWjCIUpHk4tEruJlBGc.byte
s	69%	7442	-	train/foZJnHz62LA1GKIvVF0h.byte
s	69%	7446	-	train/FPBh0UAvt278Dop9armX.byte
s	69%	7449	-	train/FpfoxH3CNAgl9iBY5naM.byte
s	69%	7452	-	train/FPIdpNxGE3QZlzbvMwC9.byte
s	69%	7455	-	train/FpJWSLn8DkwZKyAG9XxM.byte
s	69%	7459	-	train/FpRwW15YzghCByLxi897.byte
s	69%	7461	-	train/fpub5iWnCYX8oeFvElkK.byte
s	69%	7465	-	train/FPX8CW9No1mqKYbh3nVt.byte
s	69%	7467	-	train/FPZgaCXimE53GWD6UjJY.byte
s	69%	7470	-	train/fQ5YydnBoJM7gmr19AIH.byte
s	69%	7473	-	train/FqbIWSJMyRK1Ni7tVAu0.byte
s	69%	7475	-	train/FqEZWbA3x1Sf178hYjMn.byte
s	69%	7478	-	train/fqKj2wJk7rsb1yHFmoQd.byte
s	69%	7481	-	train/fQmbIDTrZBds8Uu45HR0.byte
s	69%	7484	-	train/fQp8mha3FKG0M19P2Yje.byte
s	69%	7487	-	train/fQRdkX2ahe0t0c5ysNU8.byte
s	69%	7491	-	train/fQtXNr5Jk4HK13U8EosW.byte
s	69%	7494	-	train/FQvbKakrJtVAs1oqd8cS.byte
s	69%	7497	-	train/fqzNxg17suJb1KL082V5.byte
s	69%	7500	-	train/fr197DA51ehQgZULyW2z.byte

s	69%	7502	- train/FR491eoK7WDw3QtHLU6k.byte
s	69%	7504	- train/fr7HG9tZAaI1q62XnKyB.byte
s	69%	7506	- train/fr9UPVj4SbmQ7eaJYgoc.byte
s	69%	7508	- train/frcDlXdshQYGCyKa7LjA.byte
s	69%	7510	- train/FrfWM00lI5XwyoQBnutV.byte
s	69%	7511	- train/Frgfw47Iky03ZBXlsQWc.byte
s	69%	7518	- train/fRLS3aKkiJP4GH0Ds6Pv.byte
s	69%	7523	- train/FrQHSDf7tNREowYmchbg.byte
s	69%	7525	- train/fRTmEnxQVBYjkUZ0C3t4.byte
s	69%	7529	- train/FrWpo1U800u9H3d1cq6x.byte
s	70%	7532	- train/FRzN6ZiS5KU3Iyql2sc7.byte
s	70%	7536	- train/FShG9xVbwLlSDTYBdJMj.byte
s	70%	7538	- train/FsirpP0oDwXeAzC2KndG.byte
s	70%	7542	- train/fs1D2jBe0w5tbhr7QF8G.byte
s	70%	7545	- train/FsQeNCSYAb9jIP7D0BByz.byte
s	70%	7548	- train/fSWvxIs0aV0YjKB185ut.byte
s	70%	7549	- train/FSZsJwdxRCKPAeB2Trf4.byte
s	70%	7550	- train/Ft0dsjHpwqcWoXK6ST7n.byte
s	70%	7554	- train/fT6574JMzaXF5OmJE1Sr.byte
s	70%	7556	- train/ftaSA2odeIb3PhFY9Els.byte
s	70%	7559	- train/ftboX4CmgIYJ3UTpAqS7.byte
s	70%	7562	- train/fTdzEIkhXGAJiV7BqxsW.byte
s	70%	7566	- train/FtmwqCbojMpvJDKB7H8r.byte

s	70%	7568	-	train/FTofSxMVXG9wtPrRuIzQ.byte
s	70%	7569	-	train/FTP1W3bHICJZwrxia7hf.byte
s	70%	7571	-	train/Ftqd7rZi148x0zIJMOPY.byte
s	70%	7575	-	train/fU4l7Lc8jnyxRewu1okV.byte
s	70%	7578	-	train/FuajKf6MUmELZ304sxlw.byte
s	70%	7580	-	train/fug584IGjQdWTSY3pke2.byte
s	70%	7583	-	train/FupJW071GZ9U51VSEcak.byte
s	70%	7585	-	train/FUR0EhC7x01QW8f6DorS.byte
s	70%	7591	-	train/fuwAE7WyhXbHInmlgxZa.byte
s	70%	7594	-	train/fuys9cVz0Q3idvakHBSP.byte
s	70%	7598	-	train/fvbowPJXxAKQsVt7MiRL.byte
s	70%	7601	-	train/FvIYS1W07BTpjMeH6UbK.byte
s	70%	7602	-	train/FVm0980BKzIHYiGUscDx.byte
s	70%	7604	-	train/FvobzqOYfnQ0k1x1tZLc.byte
s	70%	7607	-	train/fVTrpQLuNzd7Z0xX3K0v.byte
s	70%	7610	-	train/fVvJS4yEYIobqw1LH06F.byte
s	70%	7611	-	train/fVWAAt17QsixdmThDBu4C.byte
s	70%	7612	-	train/FW0vdghTM15JtBqVkp3r.byte
s	70%	7618	-	train/fwC3QFVbKxvhS02mnTWe.byte
s	70%	7621	-	train/FWhXBxORyrDzpVAg41Tj.byte
s	70%	7624	-	train/fwMv183WJ5KB01UN4ckC.byte
s	70%	7626	-	train/fwtRXAaBD60Wbu4JEjgc.byte
s	70%	7629	-	train/FwXLJhKd382fCi4umt00.byte

s	70%	7633	-	train/FxaARS1BKUQr47njqN09.byte
s	70%	7635	-	train/FXGvem2ls1xYbt7rkN3u.byte
s	70%	7638	-	train/FX0ALRt8SHUp2naVJksy.byte
s	70%	7641	-	train/fXPaL8v6bgNqRIcn1BZO.byte
s	71%	7645	-	train/FXUJWnbGiCERH9Mcg8L1.byte
s	71%	7647	-	train/fy0LNRhDO6mHbB148onM.byte
s	71%	7650	-	train/Fy7gh1lciNL4MAdaETDG.byte
s	71%	7658	-	train/FyIcXN7Z6SQojdgntUD9.byte
s	71%	7661	-	train/fYpFPVOnCAKS80qRtiH9.byte
s	71%	7664	-	train/FYPXNW2S6e09zMVR3HaK.byte
s	71%	7666	-	train/FYsIBi6yv7zdqw1b5uHE.byte
s	71%	7667	-	train/fYsopzlmCDGNAd72Ivxq.byte
s	71%	7670	-	train/fYUD3GcsHp2JBE4Meg0Q.byte
s	71%	7674	-	train/fz6TBt9AbedVuG4vrnma.byte
s	71%	7680	-	train/fzgZyTMiq8lmuKkhdS27.byte
s	71%	7682	-	train/fZhTampMLxV1DU64tyIJ.byte
s	71%	7686	-	train/fzLv3FCPoyHwONVkBp8.byte
s	71%	7689	-	train/fzpmKQC0AFSw8hWcyHXx.byte
s	71%	7690	-	train/FzQi9DwaKgWsAb1cRXHY.byte
s	71%	7694	-	train/Fzu6tvTilbXUEVdM7oHL.byte
s	71%	7700	-	train/fZzkQN1XKvy7cVJ584hB.byte
s	71%	7702	-	train/g08oBJNFLwTvYI2tZeOp.byte
s	71%	7705	-	train/g0DCo2kvMYdPAh4UN0jT.byte
s				

s	71%	7707	-	train/G010hNpMKqBQb9VrkTim.byte
s	71%	7710	-	train/G0r4zLJ6yBPK1T2s9bNt.byte
s	71%	7712	-	train/G0xmgghzaP9n2HUCN1KFZ.byte
s	71%	7715	-	train/G19q0jYAKMamw0rLgHUv.byte
s	71%	7719	-	train/G1McmpbnaBkjwtIZzys1Y.byte
s	71%	7724	-	train/g265T09zY4b3ZExKnr1R.byte
s	71%	7726	-	train/G29t8jq3S5susQkfeL4zn.byte
s	71%	7730	-	train/g2kbzAYTnm8WeHxNhOMR.byte
s	71%	7732	-	train/G2NCXj6ZW0yTFALv7Hqd.byte
s	71%	7738	-	train/G2w1N4BuxnLUHiXQeSZR.byte
s	71%	7740	-	train/g2xul18TswtIEb7SqfXD.byte
s	71%	7745	-	train/g3Dv6s51iZR497FAVmxI.byte
s	71%	7750	-	train/G3td1uFSQnHxrZ9jPE4X.byte
s	71%	7754	-	train/g46jFCHzY1oqtws0yx05.byte
s	71%	7758	-	train/g4HxIkheCATBOSbrQzjR.byte
s	71%	7762	-	train/g4oqfIJhBG3N7p0tnYUw.byte
s	72%	7765	-	train/g5aTwB2PojoyI9D1bHh1C.byte
s	72%	7770	-	train/g509ufJCeLd4nQVF0RiM.byte
s	72%	7771	-	train/G5s0wMUhuWJi1ckDSzn4.byte
s	72%	7776	-	train/g6Hfkw1VaE8Aj5dGS0ZN.byte
s	72%	7780	-	train/g6YuB4D2RCwjUHMw31GZ.byte
s	72%	7784	-	train/G7DQnb0fcE62usXeAmIN.byte
s	72%	7787	-	train/G7JMvWm1nwiVpjcNhSdL.byte

s	72%	7790	-	train/g7sDFNXIqYmCMkWOpPrA.byte
s	72%	7795	-	train/G80aDFlyJqitT9vBsQMZ.byte
s	72%	7800	-	train/G8bXrIAFQ0HYlnZ7eDym.byte
s	72%	7803	-	train/g8k7WSGDwyBaJAQKXhUL.byte
s	72%	7807	-	train/G8unSowFszcjkpdafJWe.byte
s	72%	7809	-	train/G8w1LJNegZjoD7YFkiS6.byte
s	72%	7811	-	train/g92uBMP7wqGFvszXjQot.byte
s	72%	7814	-	train/G9AfUZY5xLuDFISqKyQs.byte
s	72%	7817	-	train/g9izzfbqECNJw7RojkIY.byte
s	72%	7819	-	train/g9jFi4GeR7SZkTt6uovW.byte
s	72%	7822	-	train/G9MXfcb4iVKa3BeN1SLg.byte
s	72%	7825	-	train/g9X5q6kEmFb1H3Z2ydPe.byte
s	72%	7828	-	train/ga3WqXeTNF519BCMGx2U.byte
s	72%	7831	-	train/gA94Fpfcq5lRMvmkKXYJ.byte
s	72%	7838	-	train/GAFXJNzhQ50rMLdt8p39.byte
s	72%	7840	-	train/GAHqOv5KajTcR648pnbw.byte
s	72%	7843	-	train/gakqGme850Xcf76A9SnF.byte
s	72%	7845	-	train/GAOqVercpKCZJX41lwzh.byte
s	72%	7849	-	train/GAvItLYkN1s5nPyE6drh.byte
s	72%	7851	-	train/gAyPsqEjndt4xoGc2J9Z.byte
s	72%	7855	-	train/gBceKDhWdI6jفزpsMNM5.byte
s	72%	7858	-	train/GbgDPpJ7jy6Y9ncT0wA1.byte
s	72%	7860	-	train/gbkKyeR2uhS05GzsIr0T.byte

s	72%	7863	-	train/gBNdz3vVibKA1aYDCwek.byte
s	72%	7865	-	train/gbNWT5xaD1MJQKFRG2oP.byte
s	72%	7867	-	train/gBPu5wieVs9jhbcmrIMl.byte
s	72%	7869	-	train/GBQtWjTxb4v21NCDAaoI.byte
s	73%	7872	-	train/gBSqh0U39mH4z2IZ1WFR.byte
s	73%	7874	-	train/GBUMvwuR3QHfLSip0ohk.byte
s	73%	7877	-	train/gBv02rS0FyHbzcR5Qp76.byte
s	73%	7879	-	train/gbYsj2FL7BQpaxX9310P.byte
s	73%	7882	-	train/gC6s297ohTDUN03IzpKM.byte
s	73%	7884	-	train/Gc7fIKEtrPNZ2A48gHRy.byte
s	73%	7888	-	train/gCBJQKMq14Atfe3ZSRX9.byte
s	73%	7891	-	train/gchAJP4keCm0fiXr1tDn.byte
s	73%	7894	-	train/gCM2jaAJyE046GQusbiT.byte
s	73%	7896	-	train/GCoiXh8BY0N14j01P9zL.byte
s	73%	7898	-	train/GCp1DME8xZNUogB5aYWh.byte
s	73%	7904	-	train/GCtriz82s6SZXB0HDRc1.byte
s	73%	7905	-	train/GCuDFKgQY2f1WH0B9VrE.byte
s	73%	7908	-	train/gD0iXSVtJzPQhsfGvNar.byte
s	73%	7909	-	train/gD2jWCFQ0m7POR5uElfN.byte
s	73%	7914	-	train/GdA8RwInXaH5J10fSEpN.byte
s	73%	7918	-	train/gDkj1Q9xSVqKFcte6dp3.byte
s	73%	7921	-	train/gdmNQEs1b8Hut9fyw5VS.byte
s	73%	7923	-	train/gdMSLJ5UZmtPpN8W3ja6.byte

s	73%	7926	-	train/gdpyWF9KBx3NHDri0cuv.byte
s	73%	7930	-	train/gDWiFA4jvZnJQ9H15Xxk.byte
s	73%	7933	-	train/gDYlpwAxoP1MLSV0Fi6d.byte
s	73%	7935	-	train/GDZJ2rT0NmUsgPAkzoMu.byte
s	73%	7939	-	train/GEaJRH0xlpr03C8ShcQK.byte
s	73%	7943	-	train/gedRiDzKZq09BMX12Lhy.byte
s	73%	7947	-	train/ge1hQLoiRE0HZT0I2fYv.byte
s	73%	7954	-	train/gEtM2pIuKP0xB3NXvi6n.byte
s	73%	7959	-	train/geY4cHBupKaLMrT38A9b.byte
s	73%	7964	-	train/GF8K2TW3s9QUjRzxbr5v.byte
s	73%	7966	-	train/GfCSdcknV4PzMv6Ui9W3.byte
s	73%	7968	-	train/GFdzM4pukx7wtXIKrS6g.byte
s	73%	7969	-	train/gfj1rRXDkia740U9bMcJ.byte
s	73%	7972	-	train/GFLxvaBSrfoN5uUA3l9s.byte
s	73%	7974	-	train/gFNGYSsVA0x58JEI0laR.byte
s	73%	7978	-	train/GfTjcoAWk01IQDvXrgiU.byte
s	74%	7979	-	train/gfu5Owkyns4CVh0G8S9N.byte
s	74%	7983	-	train/GFXkasEv1gfuNP1R0jQM.byte
s	74%	7985	-	train/gG23n7f1CLM6mZSrjIHb.byte
s	74%	7988	-	train/gGbArPzUsntR6fBWjYu5.byte
s	74%	7991	-	train/gGfFjZ8Xab24eQ6zADN1.byte
s	74%	7995	-	train/GgNDAxu4rSEsTMaoUczR.byte
s	74%	7997	-	train/gGrBPR4c8KkNWHuA61UY.byte

s	74%	8000	-	train/GgZ3AWheHi7BjuKDmF08.byte
s	74%	8003	-	train/Gh06DE8T1zeKQ3vxjbpo.byte
s	74%	8005	-	train/gh38VeyNnLw0ucFzAKdH.byte
s	74%	8006	-	train/Gh4gVXZBCHSkjTDLsMlt.byte
s	74%	8009	-	train/GHfp03Ls1VE2bjXu1TqZ.byte
s	74%	8014	-	train/GhqXt1KSQAj7D4oL30VY.byte
s	74%	8019	-	train/ghtBsXwAKYrkLy76bPR9.byte
s	74%	8021	-	train/GhtqnW4gc1QXTKBRvbjz.byte
s	74%	8023	-	train/gHUIeo45cqWZX308K6km.byte
s	74%	8024	-	train/ghvVNmusq9JQicTP5kHD.byte
s	74%	8029	-	train/gI0oq52JWkxB1M8NFiuk.byte
s	74%	8033	-	train/gieTJZzmLhf082bXVxkD.bytes
s	74%	8034	-	train/gih3sG9TYD41VZOFRmpb.byte
s	74%	8039	-	train/gIONM8E513LYstAjyzWl.byte
s	74%	8041	-	train/gitPaXxIMeKOZRTW2k34.byte
s	74%	8047	-	train/GJc10auMEA70XU6nv8g3.byte
s	74%	8053	-	train/GjJRLwAZuYEroCO1TQVa.byte
s	74%	8054	-	train/Gjl3p5gNZxYU7uQH9hSf.byte
s	74%	8056	-	train/gj0Iy9sRbtFGoWASY16c.byte
s	74%	8058	-	train/gjpTJB8G2bqynsVr15kW.byte
s	74%	8060	-	train/gJQLAK983BsIvVwCHXRD.byte
s	74%	8064	-	train/gJspjOHk7zXSGMB2Y5Tm.byte
s	74%	8069	-	train/gjwWSlnoxq3Et27IeZD5d.byte

s	74%	8071	-	train/GK3DuZAoU1YkPTWnzC7V.byte
s	74%	8074	-	train/GkA3vhfoHbwdeKPB14ZR.byte
s	74%	8077	-	train/gkE2CmMqQtSP3JI6ZoHU.byte
s	74%	8080	-	train/GKhapcnCHF2qxI3XkbyR.byte
s	74%	8083	-	train/gkJ361WfpDU0TiMmIBse.byte
s	74%	8087	-	train/gKlJ0CFirsA3c0ZUmWDe.byte
s	75%	8089	-	train/gkMcsUACjoBPnFY0bQfa.byte
s	75%	8090	-	train/gkMCUhdEl6p0LXbyPItu.byte
s	75%	8094	-	train/GkQctIpwN58dSs407UhY.byte
s	75%	8097	-	train/gKvtaoZ7Lk11V6yxq3Fc.byte
s	75%	8100	-	train/GLbHYz1tI5Z0F6vc1eEx.byte
s	75%	8103	-	train/Gld6X9LuQw58tmRvo1qc.byte
s	75%	8106	-	train/gLEvBR41T9HcjW2zCAni.byte
s	75%	8109	-	train/gLhwt14Dd2XY1BsIS0Em.byte
s	75%	8112	-	train/gLmvzTR3BJhWD9kaw0nA.byte
s	75%	8114	-	train/gLNB2kuE1wasXd0jMQ7J.byte
s	75%	8117	-	train/Gl08VUEpCsqcwXioP7zu.byte
s	75%	8119	-	train/glq6WfYSiamR1KbeMIyp.byte
s	75%	8124	-	train/GLW9Re00VAhCr4TUqkZD.byte
s	75%	8126	-	train/GLXFv4PfV6dD9eaI03h.byte
s	75%	8128	-	train/GLzSZuahXjcxv4fQPosm.byte
s	75%	8130	-	train/GmEAFpP5Z00suwaNIeqx.byte
s	75%	8135	-	train/Gmk6HR4iIWJDEubOU3xe.byte

s	75%	8138	-	train/gMnlfq2FJuTr0dexUN07.byte
s	75%	8141	-	train/GmPti3pxguLDKAZTFowQ.byte
s	75%	8147	-	train/GN8ZQjzWfKBYUtH1ehM0.byte
s	75%	8150	-	train/GneEPpmtLjvhTW0HIsNc.byte
s	75%	8151	-	train/gnfdqj5h8rkxUDzJ4obe.byte
s	75%	8155	-	train/GNHUoTAd3etcxL1M0WbS.byte
s	75%	8161	-	train/gN0cktmwv3FMZ4TH02EJ.byte
s	75%	8164	-	train/gnQaT4Gdq810LlsR06tA.byte
s	75%	8167	-	train/gNtLHRhP5depSUoqJ9mk.byte
s	75%	8169	-	train/gnuITJ1G6aZPskYvm3FX.byte
s	75%	8173	-	train/GNZvEY5fDAqxhpjdysn6.byte
s	75%	8175	-	train/GoAF7XTjK2HmwJzDhdkQ.byte
s	75%	8178	-	train/g0Ej2AmqLZSwkGIYsonJ.byte
s	75%	8181	-	train/Go1UKPreSvZL6pHkM14t.byte
s	75%	8184	-	train/GOVHJNpe2hzRrPc1LFZw.byte
s	75%	8187	-	train/goYPdUatw3BQE16iCI0D.byte
s	75%	8192	-	train/GP9MBktKncadV1gF7N25.byte
s	75%	8194	-	train/GPAypQZ0LH7rD1safnMI.byte
s	75%	8198	-	train/GPEC53g1KvWksBcYhuef.byte
s	76%	8202	-	train/GPifvoxhun3rNgwZ7VQS.byte
s	76%	8206	-	train/gpOUTtWIXsvLdewJmfY9.byte
s	76%	8213	-	train/GpuwW62nZ7VgK1b3X1A8.byte
s	76%	8215	-	train/gpYNi0jEzrAoxWyu12s6.byte

s	76%	8216	-	train/gPz42XwWb17JVAQytYxT.byte
s	76%	8217	-	train/gq0Y8ZH1eU2NuIo5tOA6.byte
s	76%	8221	-	train/gq9cv71PXyzBDsY8SfVe.byte
s	76%	8225	-	train/GQh8kxUDtRVwzEP0Wvfs.byte
s	76%	8230	-	train/gqlPiFkcetLs6w09HQb5.byte
s	76%	8234	-	train/gQ0Itc9hCLYWjk6TKHom.byte
s	76%	8236	-	train/GQpTRMvrez5Z2hqykUwJ.byte
s	76%	8238	-	train/gqr10PzwFNxSRnX7y9t6.byte
s	76%	8242	-	train/gQuxjNVCZ0b7Jn6c24kw.byte
s	76%	8246	-	train/gr4aMsZW3fHNbiP7eI5d.byte
s	76%	8248	-	train/Gr8pcIbXheNmzaBiHgFR.byte
s	76%	8250	-	train/gRBocU81iGy75lemTdsa.byte
s	76%	8253	-	train/Grf06bXRtzV1o78SWmC9.byte
s	76%	8256	-	train/grm1u5UkAvVC8obIKqi7.byte
s	76%	8259	-	train/gR0mPjLGQ34WNVr8ZtiX.byte
s	76%	8263	-	train/GRXZaphsokTrql4u2xMd.byte
s	76%	8270	-	train/GseuqVvdHoYCySXkJpEL.byte
s	76%	8272	-	train/gSHu3cdNKLOWke7hDipC.byte
s	76%	8276	-	train/GsPr60aJXyNVFQA5eTHK.byte
s	76%	8279	-	train/gsqykQJmjZ9FdXMeYoNa.byte
s	76%	8280	-	train/gSrYQn4hJRdj1xEc26wM.byte
s	76%	8283	-	train/gsxM1eBSQobH75AZFjOR.byte
s	76%	8286	-	train/gSz13hKvMnYtFL8Tj6ey.byte

s	76%	8288	-	train/gt21j5QSnq7DbXKoBYm1.byte
s	76%	8290	-	train/gtdaw5AUufWFskKmGPpL.byte
s	76%	8291	-	train/GTdj9ktEmJCRqwfnC0Qg.byte
s	76%	8295	-	train/Gtg83nF6ri2sIjAPMVhC.byte
s	76%	8297	-	train/gtiXlab7VqBfK2PYzdx6.byte
s	76%	8300	-	train/gtoDIM0zURwQ6ZdLx1JN.byte
s	76%	8304	-	train/gtsQZdU0r159I68nfBjL.byte
s	76%	8306	-	train/Gtu7doCvKfENQVp1bw4F.byte
s	76%	8310	-	train/GTyr8o3mQDug7cb0eCMz.byte
s	77%	8312	-	train/GTZP073Y4JwN5oRLveDx.byte
s	77%	8314	-	train/GU32g0fBF91xp7JqNXHY.byte
s	77%	8316	-	train/Gu5cz8Ur0o0QDfmsZkA6.byte
s	77%	8317	-	train/guaL1dwOMlUT5VesBNH7.byte
s	77%	8319	-	train/GUFWALVuS2Ndn5YJfHjw.byte
s	77%	8322	-	train/GU0CE4g7Wk2QPIhxJnYF.byte
s	77%	8324	-	train/gUPvp2kiFac1EGMQZL0A.byte
s	77%	8329	-	train/guSsk0jAbhJXcxPnHGQ0.byte
s	77%	8332	-	train/gutpXNfdSFbzG9wB7KM1.byte
s	77%	8336	-	train/gV6kHaR3DyELvIBhTnmF.byte
s	77%	8340	-	train/gvDKqkdM71EHLUouyzjY.byte
s	77%	8341	-	train/gVDuTfNzseWn96HYR23y.byte
s	77%	8342	-	train/GVeSwJETaiBybNsg87xC.byte
s	77%	8346	-	train/gvhcKkI6twz5QeJ1BjEx.byte

s	77%	8350	-	train/GvQTe0XSqnhKLWz85DP4.byte
s	77%	8354	-	train/gVXOW67NHqmsncU9FGza.byte
s	77%	8358	-	train/Gw6ekSBuImx1052yTaVq.byte
s	77%	8360	-	train/GWbiQstavLpfhYqDE800.byte
s	77%	8361	-	train/GWBugsn5QySAo76NDE0H.byte
s	77%	8363	-	train/gWCuk7BUzdH01whK68eY.byte
s	77%	8365	-	train/gwEG0RDkZB6Xfru7JpVQ.byte
s	77%	8369	-	train/gwjhnyxt8EsiPb3mkCL.byte
s	77%	8373	-	train/GWQYvtI8a2rouyEOPcHb.byte
s	77%	8376	-	train/GwusdVTRkYeZOaAgcpHm.byte
s	77%	8380	-	train/gX5HMyEwdP19Bb0ZwOV8.byte
s	77%	8384	-	train/GxB6dX7o5E0cP9rTnb2g.byte
s	77%	8387	-	train/Gxe2usYqQgwz789UpckN.byte
s	77%	8388	-	train/GxEYhcJlkd30iKQeSuNy.byte
s	77%	8392	-	train/Gxk6Z5edAo4tUgFD90sR.byte
s	77%	8396	-	train/GxnCmdTYLaqXAc7F9zUI.byte
s	77%	8399	-	train/gXZ5G3kB0dTV7UbEpR29.byte
s	77%	8402	-	train/Gy8uob0sY93wZzCINP1B.byte
s	77%	8405	-	train/gYcty9D6xrqwZ5ChJBQs.byte
s	77%	8408	-	train/GYFfQ042ls3rtZLUJe5x.byte
s	77%	8410	-	train/GyN1TW59jIfadSXxUZFb.byte
s	77%	8413	-	train/GYo8tD760Wx0jkyIB1iL.byte
s	78%	8418	-	train/Gysrvz2mYQ9K0tJHP4T3.byte

s	78%	8419	-	train/gYtamL8EyZN9Bp3wTkQF.byte
s	78%	8423	-	train/gYZueKo9wDtQ4FRbB2lf.byte
s	78%	8427	-	train/GZDBtHrSVFvM2cC5dEL9.byte
s	78%	8429	-	train/gZDSNVG40yutMcw3Fs8r.byte
s	78%	8431	-	train/gzfdolrL4kHKb0GaxJnZ.byte
s	78%	8433	-	train/GZJ3hp4v8AzVB6imf1I1.byte
s	78%	8436	-	train/GZMUBgvzt3brcfsh84iR.byte
s	78%	8440	-	train/gzsqAHCC8nFy59j7hodk.byte
s	78%	8445	-	train/h08Yo3NOLxdpHgMG5zB6.byte
s	78%	8450	-	train/H0evJ1QxsltOCmkDaTGr.byte
s	78%	8453	-	train/H0ifOVETDIjMrgGWSa4k.byte
s	78%	8455	-	train/H0qUX7nT8yhsKD11gMxR.byte
s	78%	8457	-	train/h0VaLmr69gd0iyRwStZQ.byte
s	78%	8461	-	train/H1B87Apkso2gm0Fjd4TZ.byte
s	78%	8463	-	train/h1Dy5qSUbYtfl7AM3a2l.byte
s	78%	8465	-	train/H1MY0ctT2Pn4bpUSQLmo.byte
s	78%	8466	-	train/h1nZq0BX5LruizPw3EUm.byte
s	78%	8471	-	train/h2bGqACH5pKZtoOzMVRJ.byte
s	78%	8472	-	train/h2cdZou5y1lgEQaUiJvSD.byte
s	78%	8474	-	train/h2dm1EMoA0I3fPxZ0Bsc.byte
s	78%	8479	-	train/H2toBzike7AxGYNnZMXq.byte
s	78%	8485	-	train/h3sORpAu1ScivEPIrBl0.byte
s	78%	8488	-	train/h3V2WxD9LUz60iOnGIrN.byte

s	78%	8489	-	train/h3wJc9sbFx8AMane7uqY.byte
s	78%	8495	-	train/H4m05vLkBgUd1eFixXoS.byte
s	78%	8498	-	train/H56g8029xo1ciyfQXFZS.byte
s	78%	8501	-	train/h5gMG6Y3duSDeNHvFLzA.byte
s	78%	8506	-	train/h5vYpXNJjs0qRDVyF8Z7.byte
s	78%	8508	-	train/h5ZBNR8Ag9YMIdk0Plt4.byte
s	78%	8512	-	train/h605H92YWwJ7r1pRELge.byte
s	78%	8515	-	train/H6tgJ4jTyak8VsANRFud.byte
s	78%	8519	-	train/H753kd0g6fvyzQLJlFbK.byte
s	78%	8523	-	train/H7i091zy2AOYnJvPwfxc.byte
s	78%	8525	-	train/h7Kbin4elq0PXJo2pvrR.byte
s	78%	8527	-	train/h7MY4l3UxgNjiHR1C5Qf.byte
s	79%	8530	-	train/H7Z6WfoCrxzSjMVd04Dy.byte
s	79%	8533	-	train/H8dnb7uFRMm01jqDYWC3.byte
s	79%	8536	-	train/H8P4U0bGjAITting3h6c.byte
s	79%	8539	-	train/h8VyAJq3RfgoQIciCT9u.byte
s	79%	8540	-	train/h8wnNSFs62cI7axUq0iK.byte
s	79%	8544	-	train/h9amYlEipK0JtCN25uvB.byte
s	79%	8550	-	train/H9VJ5DiCrbTFq7k0sfGN.byte
s	79%	8552	-	train/HA2qnVEF4t31ceLkiTbK.byte
s	79%	8554	-	train/Ha85TktZVbeLdCOXN7vj.byte
s	79%	8557	-	train/HAI3gUuLk0jmQwtZ5G0o.byte
s	79%	8558	-	train/HaI3lhUAXN6Qv0jW9FMg.byte

s	79%	8563	-	train/HaOw84bYfkn2yxpiRFPM.byte
s	79%	8568	-	train/hATF04dxIqGvZSkmK8ru.byte
s	79%	8571	-	train/hAu9FPNJlWKS0DpSnbR7.byte
s	79%	8573	-	train/HaviCoRt0lZ4spd1Qx36.byte
s	79%	8575	-	train/HaWuY5IXN7gOmJG4ZCkd.byte
s	79%	8577	-	train/Haxpwg4SmfWQLEPyZ5jK.byte
s	79%	8581	-	train/HB6l3Esd2mZGrOWtfYiI.byte
s	79%	8583	-	train/HBbWV1wZoJz2sKU18e3r.byte
s	79%	8584	-	train/hbdQuIOsXgM4iLNlGTpn.byte
s	79%	8586	-	train/HBGj1awLFEUZ7MYNgbQX.byte
s	79%	8588	-	train/HBmJ8DWMVSpdY43605Ao.byte
s	79%	8589	-	train/HBoK6P8VSAEZhr5z4Qq2.byte
s	79%	8593	-	train/HbSOMm8u5lgAnPYLDWJy.byte
s	79%	8595	-	train/hBTCW1VyKHjzUkncqpeo.byte
s	79%	8597	-	train/HbWgYw9MFDceJX7msTR6.byte
s	79%	8600	-	train/HC140EyUzrK9PLfADTlq.byte
s	79%	8603	-	train/hC4p1vnRPZlyMtsFVrqe.byte
s	79%	8606	-	train/HCaGgWTI2fY0qNdZzeVF.byte
s	79%	8608	-	train/HCbGVZFI5800xptuS7sm.byte
s	79%	8610	-	train/HCd3B9trqLsEMDFc2jNa.byte
s	79%	8612	-	train/HcfcB9zY8QAsTaJE2uV3.byte
s	79%	8615	-	train/hCH2ztpK1XUakfjqrX5P.byte
s	79%	8617	-	train/HCItVjnoUXrq30JcdNQk.byte

s	79%	8619	- train/hCoF1baLwU4QytIcXxJV.byte
s	80%	8621	- train/hCP3n1Ws5IRjMAoYBufJ.byte
s	80%	8623	- train/HCS2eUx40QTG5nRY3vMf.byte
s	80%	8626	- train/hcwtORMJqx6kPiQzVjbX.byte
s	80%	8632	- train/Hd2RSo5Vklez7KAWpaF1.byte
s	80%	8636	- train/hDiAgUca06yjEXeWn8mT.byte
s	80%	8639	- train/HdkjoqVQLg08cRDvpuGb.byte
s	80%	8642	- train/hdmGkzIWx89uo1T6PCXY.byte
s	80%	8644	- train/hDNY1kuFK3c0tqEyWM6f.byte
s	80%	8646	- train/hDVjFuAliwxn2No6XkT0.byte
s	80%	8648	- train/hDxgImyk6HKLi0YuaZ29.byte
s	80%	8653	- train/HE9280WkP4RJtdsNL7bZ.byte
s	80%	8654	- train/HeBQriWlXgyVJcI31Y7f.byte
s	80%	8659	- train/HELCZNMau0bmr-fTIj1Av.byte
s	80%	8662	- train/hePHMV0Uspcnd23R4Fyz.byte
s	80%	8663	- train/HESbjuQ2Zo1n3XL5WGvN.byte
s	80%	8668	- train/hf20ZuRmJbxks0Xv9itV.byte
s	80%	8670	- train/hf6rMnsUwcFdSR8J4QKy.byte
s	80%	8672	- train/HFa7lQeGbUMP8hwKTr9J.byte
s	80%	8674	- train/HFbakUnGgEIfcYjZLAvi.byte
s	80%	8676	- train/hFEc3IvTKgdysL7etM1r.byte
s	80%	8678	- train/HfLoa5txyimVwPgs4NvF.byte
s	80%	8684	- train/HFQTfjJp98VMhYeBrkmd.byte
s			

s	80%	8691	-	train/hgADmCjLzU17Vw5S4Gfu.byte
s	80%	8695	-	train/hGD0efK53nrvzcdx42tY.byte
s	80%	8697	-	train/Hginatj9xKRWTLD3Jyc.byte
s	80%	8701	-	train/hgmnRpGx5NY601MwFPvs.byte
s	80%	8707	-	train/hGt2BrujJ6TfQEDz1FnM.byte
s	80%	8708	-	train/HhDF2X80mRbwZjBVQPxC.byte
s	80%	8715	-	train/hI7BciQG1fT1HLJWmZX6.byte
s	80%	8718	-	train/hIDK0p1PAiJdL87CzG69.byte
s	80%	8724	-	train/Hi0Ww8CRpay9fEM63KLs.byte
s	80%	8727	-	train/hiQNlaFHXJsLzdkgu5DA.byte
s	80%	8729	-	train/HiRp1L9qaYo81yeE234G.byte
s	80%	8731	-	train/HisQ9bLAevYNUuPnx5aD.byte
s	80%	8734	-	train/HiXUclzIL2rEutjB74Nm.byte
s	81%	8736	-	train/Hj0CXE8hPiFvztq6mnp3.byte
s	81%	8741	-	train/HJ68L4XrwdscnZ37bei.byte
s	81%	8745	-	train/hjdkwfeHXnWoUsJPY2y3.byte
s	81%	8749	-	train/HjgSxn8pJ2ITBsmohaLr.byte
s	81%	8751	-	train/hjI9pimf05cxRBaKw18s.byte
s	81%	8756	-	train/HJQiyIRqr6FPeBcoaEsk.byte
s	81%	8758	-	train/HJt1Mc578nYFBDdKNkVu.byte
s	81%	8765	-	train/hK7qMc3ZHmF5e4bLfC0j.byte
s	81%	8768	-	train/HKD6kbY482MvpV9sa0hi.byte
s	81%	8773	-	train/HKgJMPYD1cUmF74G0pis.byte

s	81%	8775	-	train/HkL72qMuOcp0T3gWahnN.byte
s	81%	8778	-	train/HkrQFoitPN0jaUVqWIux.byte
s	81%	8780	-	train/hkscPWGaIw0ALpHuNKr8.byte
s	81%	8783	-	train/HkzCqPSGNEh8Z2g7oXse.byte
s	81%	8785	-	train/Hl2P1kxawM74dKtuoA9c.byte
s	81%	8788	-	train/hl5wNixcfnCgEjp324sT.byte
s	81%	8790	-	train/Hl76G9NEXne5FzjwYkUJ.byte
s	81%	8793	-	train/Hlep9aJIioFTQcrmX04P.byte
s	81%	8795	-	train/hLfixdwFNU9sm4pkWbvZ.byte
s	81%	8797	-	train/hLFrZkvCVugAjG1TcQ1n.byte
s	81%	8800	-	train/HLIpzufArQKh9XcBPVoT.byte
s	81%	8806	-	train/HlRnpWyuP37okUqjdhC2.byte
s	81%	8808	-	train/hlTCsg3kPWYGzJAcLBK8.byte
s	81%	8813	-	train/HlZExoNdpYg4kMqfLAin.byte
s	81%	8815	-	train/HM8jyTb7ioG9sSDn6htE.byte
s	81%	8818	-	train/hMC14VJ78NDsKSotHYeu.byte
s	81%	8822	-	train/hmgHSWv8RYUidfGkn6V3.byte
s	81%	8823	-	train/hmjczH7rQ4qkL1bDzePy.byte
s	81%	8825	-	train/hm1jZYQoVN82BXJOSEtz.byte
s	81%	8828	-	train/hmnKR1fgGFeidy50HP9C.byte
s	81%	8830	-	train/hmPkWqAdWUyQHjFsacCJ.byte
s	81%	8833	-	train/hmW0kZVbo2URyzB0iale.byte
s	81%	8835	-	train/hmxjukz4UBqMPf9Jec6y.byte
s				

s	81%	8838	- train/hMyn4mwJjsLI5l03PdV8.byte
s	81%	8839	- train/hmYPrWvVt04Deu6J7E5i.byte
s	81%	8841	- train/HmznEa6Xt9ebTjq0J0Cg.byte
s	81%	8842	- train/Hn2koKsPizIVUq1bWL5g.byte
s	82%	8842	- train/Hn2koKsPizIVUq1bWL5g.byte
s	82%	8843	- train/hNA1XZpxS5QV8R2YTkaU.byte
s	82%	8845	- train/HNawLdV5YIytUe0pFXCA.byte
s	82%	8846	- train/HnAYP2v0CXfpmdB0Iqeh.byte
s	82%	8847	- train/hNboYglQSecVZ1LntIGf.byte
s	82%	8848	- train/HnCIJEd3foNTButSv7wK.byte
s	82%	8849	- train/hnCWyLtgETFlkJ84DqXS.byte
s	82%	8850	- train/HndZocCNMY04QB1pr7gk.byte
s	82%	8851	- train/hnEAVHz5I7tQcjGWvRi1.byte
s	82%	8852	- train/HnfBWZ6J3L8SwVNdYIgr.byte
s	82%	8853	- train/hnGHbT23JFgl90RqPKuk.byte
s	82%	8854	- train/HngN1QsVwobWqJB3RXCz.byte
s	82%	8857	- train/HNL7GobQ1xzMhkgga3uy0.byte
s	82%	8858	- train/hnMRIAswZOPolcxpUJe3.byte
s	82%	8859	- train/HNMWsf6ShPknjGm04Re.byte
s	82%	8861	- train/HN0m6L7wSiGhjp1U3stq.byte
s	82%	8862	- train/HNpYIj1rkg9oMPc2b3C8.byte
s	82%	8867	- train/HnZ5MNaCpu0k7d3vPwfE.byte
s	82%	8868	- train/h05E7Bs6r02fpo4UGZNx.byte

s	82%	8869	-	train/h07VJNqLIXrfFmZAyeMd.byte
s	82%	8870	-	train/ho9d5i06MgDQnw0zUWyC.byte
s	82%	8871	-	train/h01L91KTQSUNmYyBqApe.byte
s	82%	8873	-	train/HorXtKLc4EdzP8gYmheV.byte
s	82%	8874	-	train/Hos7LWmZR926V05KcgpD.byte
s	82%	8875	-	train/Hostb5jC4vASg0GupIZl.byte
s	82%	8876	-	train/hoTWF75dP83y1kMLAtSQ.byte
s	82%	8877	-	train/h0tXLq9on7svPe4NTpAd.byte
s	82%	8879	-	train/Hov1QtF67wa9ELyuJmnS.byte
s	82%	8884	-	train/Hp4J1c7MEvW03wuFgTbi.byte
s	82%	8885	-	train/hP6lGov9R0eVci5QyuYS.byte
s	82%	8888	-	train/hpbVtc5U0IS8LuGz3d2C.byte
s	82%	8890	-	train/HPFuImzhBojSq8EKiRn0.byte
s	82%	8891	-	train/HPgEAKLnivQmBkXh6jFc.byte
s	82%	8894	-	train/HPJWqBhI7v6zQ9DeLn02.byte
s	82%	8895	-	train/hpKeXGQLyJzdtm09u1MR.byte
s	82%	8896	-	train/HpL0D0wY6RN7hcGeCylb.byte
s	82%	8897	-	train/hp1xP3Se6N40oMBqY81v.byte
s	82%	8900	-	train/hpntqKQa1RVmfJGu1B48.byte
s	82%	8902	-	train/HPqv0cpKDx2yzeRiB854.byte
s	82%	8907	-	train/hQ3EP5rHNeKgWXGIaZUC.byte
s	82%	8909	-	train/hq3MXxLtUF6aiJpc7bsD.byte
s	82%	8912	-	train/Hq6B27bhV01gvwaLGECx.byte

s	82%	8913	-	train/HqahLV8QnY0UscZz9D61.byte
s	82%	8915	-	train/HQCNdPn3ytMDrB96b8kX.byte
s	82%	8921	-	train/hQnAcOfHYisDkINaod7L.byte
s	82%	8922	-	train/Hq00wnbzfENMpoXFEmxR.byte
s	82%	8928	-	train/HQYDIEBFLi8hM6RP5kWw.byte
s	82%	8931	-	train/hR1KatoGrMUNXi0fHw90.byte
s	82%	8933	-	train/HRabWnUoX7zv8h16F3Q2.byte
s	82%	8935	-	train/HRE9TjDJgZl6LcuV7yKn.byte
s	82%	8939	-	train/hRKfPJTs2kzBviCADwuo.byte
s	82%	8942	-	train/hrMLJ7DFIfmUGPB2q8y1.byte
s	82%	8945	-	train/hRVBH2z6buFoNA7kZvS8.byte
s	82%	8947	-	train/hs3jvI9Jx7iTzKPYEXcl.byte
s	82%	8950	-	train/hS8AvgErJHmtR0y0iPdX.byte
s	82%	8954	-	train/hSATyqd2Ixzgf4D3jM0Q.byte
s	82%	8956	-	train/HsBP1S0KdTuUW8Y3nF6C.byte
s	82%	8959	-	train/HsjaJr36kD2E1dSIp4Fm.byte
s	83%	8962	-	train/hsNbP67uBJjwyXVpcRqY.byte
s	83%	8966	-	train/HSpxv7XiuwNj2ceELTnJ.byte
s	83%	8970	-	train/hStvL36JYZfqk1XWr2dN.byte
s	83%	8974	-	train/HSvpkeVLni193WImYs52.byte
s	83%	8979	-	train/HsZwBUFC37T9q4dc1Pkv.byte
s	83%	8980	-	train/ht0Lp1XYaVveSyInJx97.byte
s	83%	8983	-	train/ht5xf1Fq0HX0p1Us7PoZ.byte
s				

s	83%	8986	-	train/htDPZ1Bi0IAayWJYwOUN.byte
s	83%	8991	-	train/HtmVeG1qSNXLg1F9vDyk.byte
s	83%	8993	-	train/htuGR4dWY0eTLVOKF xv9.byte
s	83%	8996	-	train/hTWBq5QXgOm0eKV9ktRu.byte
s	83%	8999	-	train/HU9Fp58PbXvVDqIm4t7E.byte
s	83%	9001	-	train/hUc3sa5YXmb2EoZ1pGrD.byte
s	83%	9004	-	train/HuD3AtnJlpCVLdkIW sZS.byte
s	83%	9006	-	train/HueRQzwrp20PDBYMITGx.byte
s	83%	9007	-	train/HUJqQ50FyM1hw7ABZ4Ni.byte
s	83%	9011	-	train/hundC7q5ZQLoY1NvWt4S.byte
s	83%	9015	-	train/hur17S8UwEACeDgOBMYx.byte
s	83%	9019	-	train/huUAxkrTVYFZco27Bba5.byte
s	83%	9022	-	train/hUWD9x6XdNuJrMSg8pqa.byte
s	83%	9024	-	train/Huz3U7hA5kVnJ6q2QcgM.byte
s	83%	9026	-	train/hv70Fi3rwKS1csq6DmCE.byte
s	83%	9029	-	train/hvFfuWd0X1eJ9gUYkqZ4.byte
s	83%	9031	-	train/HvLeFD0ARQ8bmGzcgaP6.byte
s	83%	9033	-	train/HVlorUn2MsXN6BdZghv4.byte
s	83%	9038	-	train/hVxz0Pnb19TRZ3t7mMs2.byte
s	83%	9041	-	train/hW0uMVwbTds1PIxXZCvr.byte
s	83%	9043	-	train/hwb7W1KfRcgA2CVXHPYv.byte
s	83%	9046	-	train/hwekzyp1YmGI3joFdLaP.byte
s	83%	9049	-	train/hwHqFn9SxcKJCr8LUaz2.byte

s	83%	9050	- train/HWhSIz2mFVokMc4E7CaQ.byte
s	83%	9055	- train/hwPFV83iJa9Ek5HQSCzd.byte
s	83%	9057	- train/hWrFmiajeZ8bENA2sydM.byte
s	83%	9060	- train/hWxgEPKIclqL97uVA2m4.byte
s	83%	9063	- train/HwZeBiCNoxp1rRFKG7av.byte
s	84%	9064	- train/hwZX9Eiumyqjdt06xrkF.byte
s	84%	9066	- train/HX5UK3f2skd0m9ZSr4DE.byte
s	84%	9068	- train/Hx6nJfolcVd7qeXyC0QB.byte
s	84%	9071	- train/hxDQjadonWvu0qbgRrHz.byte
s	84%	9074	- train/hXgY32Qk1VpyuWBa4xIP.byte
s	84%	9075	- train/HxiW01uaVGkR65QoAwYD.byte
s	84%	9078	- train/hXkaQsWRA4yLtomnPTUx.byte
s	84%	9081	- train/HXoz2NOQtTgJ04RsL5ME.byte
s	84%	9086	- train/hXWmVrauve08ACsMNqZn.byte
s	84%	9090	- train/Hy8L53hUREkZMs1Pn1SY.byte
s	84%	9092	- train/HyAR3BXGISmoYQ7J9CLF.byte
s	84%	9095	- train/HyG52Uc8FNxgk0wnXpKY.byte
s	84%	9097	- train/HyITQWvZXan3zBoDpCA0.byte
s	84%	9100	- train/hYs7QuD49gNVbiWemn1z.byte
s	84%	9102	- train/HyUaQDRpwS9GT5mC3xjZ.byte
s	84%	9104	- train/hyx7urYs1ViGUA3PT0XK.byte
s	84%	9107	- train/HZ1WSp9dEVFumCsKg8Rv.byte
s	84%	9112	- train/Hzek5JnwTtjLCOUqMBo9.byte

s	84%	9115	-	train/hzH5vLburkG8simTiae	.byte
s	84%	9120	-	train/hzOwRqLEsDnCQMYrpF9N	.byte
s	84%	9122	-	train/HzP3pTQ57cAULx6d0k91	.byte
s	84%	9127	-	train/hZty2fVr5ATnjewGvPIl	.byte
s	84%	9129	-	train/HzvWS4IuPE9idXmIgOAw	.byte
s	84%	9132	-	train/I09WMFrplYEUNljGZeBt	.byte
s	84%	9134	-	train/I0aTJkCwgMY8P6h9vWSt	.byte
s	84%	9136	-	train/i0gZtoA0mxsID4cKHau1	.byte
s	84%	9139	-	train/i0JrWpyNfu5vV9tzahSY	.byte
s	84%	9142	-	train/i0OLrWBtKpRQPv4J9fhN	.byte
s	84%	9144	-	train/I0tk2asbMQNo9G8Bw6Sr	.byte
s	84%	9147	-	train/i0XRVkoPMwK95hfgWAmJ	.byte
s	84%	9152	-	train/i1nGx3SaCDYVLFqgu2r5	.byte
s	84%	9154	-	train/i1QF5bIjKc6AYl7u2fxe	.byte
s	84%	9157	-	train/I21HaMnK9rSlpNvoyUhg	.byte
s	84%	9163	-	train/I2Rtqw5TSvPfYjFkXGKr	.byte
s	84%	9167	-	train/I2ZH0sW5hT4aDKwo9gR3	.byte
s	85%	9168	-	train/i2zSbNFKQtWAYH8n7Cfx	.byte
s	85%	9170	-	train/I3E9CHzBKNQp8j0L0lmF	.byte
s	85%	9172	-	train/I3hyQ0BZM0Kxpmj6a74D	.byte
s	85%	9174	-	train/I3m9yx15jsrv71N6EBdH	.byte
s	85%	9176	-	train/i30bMKnYJjsxV1ECAD6F	.byte
s	85%	9178	-	train/I3QveCRHion2FwyzaUb0	.byte

s	85%	9182	-	train/i4asgqmJlfer58VvktWz.byte
s	85%	9187	-	train/i4mPZXrJRahKbxDCTIvu.byte
s	85%	9193	-	train/I5bhsB7CcpUDVn8Yglwy.byte
s	85%	9198	-	train/i5qAV936TsMzYyWbw0HX.byte
s	85%	9201	-	train/I5v28ycbZiHdqzKXDxMJ.byte
s	85%	9205	-	train/i6HCJxaZV5cBzW1NUuof.byte
s	85%	9207	-	train/i6KyXHtn09sfC7BZ2FDN.byte
s	85%	9210	-	train/i6SI0yPxJMh1LORK7Qe5.byte
s	85%	9213	-	train/i6ulTyEvMWfmbOdC02B4.byte
s	85%	9215	-	train/I6zJbjtT7gxKcwVWf29C.byte
s	85%	9217	-	train/i7aIBcR9ADrmsUyhQJSw.byte
s	85%	9221	-	train/I7LJAB2rbs1XKMYVmvdx.byte
s	85%	9224	-	train/I7uOWp8Ra1sZHt4mb59h.byte
s	85%	9226	-	train/i7Zea2uqM6KfAHtICFEo.byte
s	85%	9228	-	train/i86xH3JvoKDSFXOYLcUm.byte
s	85%	9231	-	train/i8AI7QqlSXMaEzPCZRFu.byte
s	85%	9232	-	train/i8bxZAEuLCsRG0WIyOBv.byte
s	85%	9236	-	train/I8GxNTCH0XVoJ6FamwBD.byte
s	85%	9239	-	train/I8nKdJA3Lz6Q0mliuqwR.byte
s	85%	9242	-	train/I8vgEN70XoeaSCf3rd0h.byte
s	85%	9246	-	train/i965bg2xFhQCzfv7TZ3I.byte
s	85%	9249	-	train/i9JrvbXVAD6RBT57kQLq.byte
s	85%	9251	-	train/I9l0RUzi68vFf2510QjP.byte

s	85%	9256	- train/I9zJ0pktX2G8iL6vYMKT.byte
s	85%	9259	- train/ia4wxFc6n0RdAuKXz23Y.byte
s	85%	9261	- train/iA6lTGX7Lsy4FNHuDb5k.byte
s	85%	9264	- train/IABzVoh0WJZtRQXrcdC4.byte
s	85%	9266	- train/iAG4k1vc8m0l7RXSYpNr.byte
s	85%	9271	- train/IAX6uCwTffUH8vOKkzcV.byte
s	86%	9274	- train/IB4bRZv7dGhs102piuDV.byte
s	86%	9279	- train/ibDT9XVyqJNrcztvEHgF.byte
s	86%	9282	- train/IBEqZCidDj8LlUGg6u9.byte
s	86%	9285	- train/iBhngrmHEGAuytQvfDeW.byte
s	86%	9292	- train/IBwigS1GXp6RxhV8qTEM.byte
s	86%	9296	- train/iBz3MaqYb0Pp1xDhgTCN.byte
s	86%	9299	- train/icaSMQIhTNrDURZyoLBG.byte
s	86%	9303	- train/iCIffsNxYkKvZzV8cGma.byte
s	86%	9305	- train/iCjLYXNmqrRhtbeWFFG5.byte
s	86%	9306	- train/iCkJe3xvjfQanEOUywFo.byte
s	86%	9312	- train/iCnEzomH19dKjvfZQ4g8.byte
s	86%	9314	- train/iCPn8QUoEZ14IHp3jSkM.byte
s	86%	9318	- train/iCvAxEGIW5wnpmhfNK2y.byte
s	86%	9321	- train/ICzpyf6g71v5FmNZtVEx.byte
s	86%	9324	- train/IDd8vCHY5Ak0hERZfSpb.byte
s	86%	9325	- train/idDSEvTsVqZhI7MlKnGu.byte
s	86%	9328	- train/iDHxeowbV73huM1G90Lj.byte

s	86%	9331	-	train/idjesBrKybwkzPXMaiUL.byte
s	86%	9335	-	train/idMAY1L4VTpuk9cC8IjW.byte
s	86%	9338	-	train/IDqzFwbkhjylEV35KGmv.byte
s	86%	9340	-	train/Idry1KoQ2uFm0aLeABDG.byte
s	86%	9343	-	train/iDx4TemwdyfNER52l0Mn.byte
s	86%	9345	-	train/IDZ3l1WMPGcxm78SeQvr.byte
s	86%	9354	-	train/iEDwlh1scqnJ4aeRTbQ7.byte
s	86%	9355	-	train/IeEJzAM1aZKSo8vYsCLp.byte
s	86%	9360	-	train/IehXbMUBZRNgHu3pvAm0.byte
s	86%	9362	-	train/ieLs93AGTlXK6jBkQc2x.byte
s	86%	9365	-	train/ieORvyxW5dSpQtVqULX6.byte
s	86%	9369	-	train/IETGJWiuU7As6SPewt0k.byte
s	86%	9370	-	train/ieTyx3pGN70aXrcqwFu4.byte
s	86%	9371	-	train/ietZ2N7g30pfdSvn60xI.byte
s	86%	9374	-	train/iEVTh37Mo4NWlAaqCFkU.byte
s	86%	9376	-	train/IEwuJXUx0QSMLFyH7k2t.byte
s	86%	9383	-	train/If74PKxzUtEXw05kChoB.byte
s	87%	9384	-	train/If7D6pQ05xPYckjg41LB.byte
s	87%	9393	-	train/IfJGXu1STEWvZDoMV4Q0.byte
s	87%	9397	-	train/IFRGbmc7WwDiHNtvC1pU.byte
s	87%	9398	-	train/IFs5zKLk0MqNtubS69XB.byte
s	87%	9401	-	train/ifUj190Nxy3D72bKTWzY.byte
s	87%	9406	-	train/Ig7mTKfQJ13ap9xriVjM.byte

s	87%	9408	- train/Ig9ZEaGmMrT1lHJhRPQx.byte
s	87%	9411	- train/iGaW9vzZUxsODf2bud6n.byte
s	87%	9413	- train/IGdmzs15LgubiAjXWfBR.byte
s	87%	9416	- train/igjezBrTh4SbGtIKZm1A.byte
s	87%	9419	- train/IG1j5HAEq1FUhbC9JWvr.byte
s	87%	9421	- train/iguJ6WF7SPzs9evTr12n.byte
s	87%	9425	- train/Igx4oNafhQKyeMPi0VSA.byte
s	87%	9429	- train/IH03dnjiQsCLYX6yoWkg.byte
s	87%	9434	- train/iH6v0dnEp7RDSGh2VuCZ.byte
s	87%	9438	- train/iHE9eFjSh5PsLXYNG2u7.byte
s	87%	9440	- train/IHiArX1xcBZgv69o4s0a.byte
s	87%	9442	- train/IhL9pebrmG171FXCkyQo.byte
s	87%	9448	- train/IhRZjKUCrAsfvkJQcyuB.byte
s	87%	9451	- train/IHvh8pS76tGXR13kOrL9.byte
s	87%	9458	- train/iIeGdtEAXlyvJfKoQ5xu.byte
s	87%	9460	- train/iIMmk8tNvLweAr2dfn9F.byte
s	87%	9464	- train/ij6XhzqMWJFI8t4SUvQy.byte
s	87%	9467	- train/IjAlEaeHV2M4GkFnRKmS.byte
s	87%	9470	- train/ijCY3HDqVnPdAZEtFRv9.byte
s	87%	9471	- train/ijgoCyseRQAaErFUHZMY.byte
s	87%	9473	- train/IJkRMnXj2Af1NqE0Hac4.byte
s	87%	9477	- train/IJP4m1KraznpiCAh50cD.byte
s	87%	9482	- train/IjSo1vLdV1387MF5YNZK.byte
s			

s	87%	9486	- train/iJwSVZNB6Qf3jdC0oAXI.byte
s	87%	9489	- train/iK503zvsIBnMQAwmgHux.byte
s	87%	9491	- train/iK9ovZhBJCDH0T34dtI5.byte
s	87%	9495	- train/IkHEuqzV2fRptB3TnsOM.byte
s	87%	9496	- train/Ikm7ZVy3DBPx00LMv1tQ.byte
s	87%	9499	- train/ikNw2nIvYAHPaod4Dqsp.byte
s	87%	9501	- train/Ikpcu3oC1JwGKQMdztrb.byte
s	87%	9504	- train/iks3Y5jQmILCpn806wR7.byte
s	87%	9506	- train/IkzqRMZW71YALpX6VtCf.byte
s	87%	9509	- train/IL4NksaVyYtQEKAxNUSJ.byte
s	88%	9513	- train/ILcigar0eZ13Xtyf8MVn.byte
s	88%	9515	- train/iLEPJ7GwKoDNnLH0mkeQ.byte
s	88%	9521	- train/ilogAd4QsU38IFBcuwDp.byte
s	88%	9523	- train/ILSeoJ68B1bWqwd7kvM2.byte
s	88%	9529	- train/IMA4VZEdYtJ5Pg0bRwLH.byte
s	88%	9532	- train/ImBUE37WLcf9PJ4kgZn1.byte
s	88%	9534	- train/IMDdcNTh4YbE8X0kyfWZ.byte
s	88%	9536	- train/iMfxGbk126VRgczUIp03.byte
s	88%	9539	- train/IMhunfg5toLaeisP003J.byte
s	88%	9541	- train/iMIveqXKnkP3NAc4uC1d.byte
s	88%	9543	- train/imns4fDL9QbS7y60hwac.byte
s	88%	9545	- train/imqELVF2Hf3ZG1tgBAov.byte
s	88%	9548	- train/IMx1nqKaWEz6UdVsuewA.byte

s	88%	9549	-	train/ImxloiQ92Xvz8ah6peYM.byte
s	88%	9553	-	train/inc4uW6VxpDb3TtN5jQH.byte
s	88%	9555	-	train/IncSU1t5HBKZW9j476pX.byte
s	88%	9558	-	train/Ine2sQjBCd4oqpcbSix0.byte
s	88%	9560	-	train/infCxsk45eNY6QbKt23u.byte
s	88%	9564	-	train/InrbAZj9cgpitWdaUHAVN.byte
s	88%	9568	-	train/iNwtL07ncSHTDa2RC8Mr.byte
s	88%	9571	-	train/IO12QehGcoYTmpvb8PUE.byte
s	88%	9576	-	train/ioEn2sGfdQ0z87DKeF4B.byte
s	88%	9579	-	train/iOIdv3emQfZN5jS9XFAu.byte
s	88%	9581	-	train/iOJebqWZLA7lXSktgc2N.byte
s	88%	9586	-	train/IopqlNQASCdjT8eOWRHw.byte
s	88%	9590	-	train/iouyXgUJX7pjI1LBTsDF.byte
s	88%	9595	-	train/iP7SnkW0I6OzuVJq9erE.byte
s	88%	9600	-	train/Iph0GU7e9aVLo1s8XkC2.byte
s	88%	9601	-	train/iPIBLrzewX8YydsA0tnu.byte
s	88%	9604	-	train/IpiTQaNwzv7hdmWcb9P.byte
s	88%	9608	-	train/iPKzLTFx0a5GWwQo1YIq.byte
s	88%	9610	-	train/ipnL25WtfDEumwS9hNkU.byte
s	88%	9614	-	train/IPS2LtDTghvn1ledpo9f.byte
s	88%	9616	-	train/IPtfXUCL4eJc7xs0jv1i.byte
s	88%	9619	-	train/ipU1FvuxcN4l2M9zZmEs.byte
s	88%	9621	-	train/iPV0Q8oez4GynBhEfg3S.byte

s	89%	9623	-	train/Iq1eU0Vbgd6Bm5XZYfxk.byte
s	89%	9625	-	train/Iq568cLUZv7RVxy2dr0f.byte
s	89%	9628	-	train/iQDcR430KXIJzT5WV8AH.byte
s	89%	9630	-	train/iqIar6PN0tQDBcZV41u8.byte
s	89%	9633	-	train/IQk0AncFSwEV0Zh2efJy.byte
s	89%	9636	-	train/iqS2Jm1Iv3GCEUpof5ua.byte
s	89%	9639	-	train/IQuKqwWjoiG301Aeb1VD.byte
s	89%	9642	-	train/iQy8or0f7tFLkGpZ0CBj.byte
s	89%	9643	-	train/IQZ1HSzefy0jUGF8VA1x.byte
s	89%	9645	-	train/IQzSymvVCNgMr2qWZUh8.byte
s	89%	9649	-	train/IR2aS8pG4mbN59ZjCixk.byte
s	89%	9651	-	train/IR9gt00weU7DWVqhJryu.byte
s	89%	9654	-	train/IrD2NqBzMT4euJQbXyv0.byte
s	89%	9655	-	train/iRDBIyTSnfmr93j864dc.byte
s	89%	9658	-	train/ireBgWXwZGyvju9aR7P6.byte
s	89%	9661	-	train/Irm1DPzxU72SNhFbWjB8.byte
s	89%	9665	-	train/IrptWDXsmNchY1V9wGuy.byte
s	89%	9669	-	train/irUj2KVPLFBa4EY1bxbth.byte
s	89%	9671	-	train/iRwMLz8VmZTpseJbKc0g.byte
s	89%	9674	-	train/Is7XAaJBcrSQNj2dmK1G.byte
s	89%	9676	-	train/iSD0zmncxkoV39KgPQLy.byte
s	89%	9678	-	train/isFh4lrPwI5vYeqQN7Jj.byte
s	89%	9680	-	train/iShnob7ANLg4E13UVFcT.byte

s	89%	9681	- train/IsiGAwHTrJx49BcFjNLq.byte
s	89%	9684	- train/isNPnYgfp4AFoH9RBL2Z.byte
s	89%	9686	- train/IsSfm9LGyi5c8uP3p7hJ.byte
s	89%	9689	- train/ISusbpVaCTQH4913cXA7.byte
s	89%	9692	- train/IT8nb7Mv05QZptGjNasH.byte
s	89%	9695	- train/ITEUR4DewcnMkQhsWqNA.byte
s	89%	9699	- train/IThBzvVbMReu248oK3UJ.byte
s	89%	9701	- train/itKaznkWYBjsPd2obgJq.byte
s	89%	9703	- train/ITPGq9AXn4EZKmYeFpuB.byte
s	89%	9710	- train/iTYV7QcEhq15JGZKWzk3.byte
s	89%	9711	- train/ITZQmUtjYaHDxMonFrKC.byte
s	89%	9712	- train/itZsB0pnP7yx4ocgGShV.byte
s	89%	9716	- train/IuAnZRqFaM15eYbPwXvz.byte
s	89%	9722	- train/Iukne97MAiE0oY6CSh5y.byte
s	89%	9724	- train/IumK1fce0qnGjAzPDrSR.byte
s	89%	9727	- train/iUPd9pr2ebcqtCXR1g3n.byte
s	90%	9730	- train/Iur3cKshVeHqib6C9tnv.byte
s	90%	9736	- train/iuy39zbPM8EZ4NIpOUTf.byte
s	90%	9741	- train/iV7KRuTdvaMeXbByh0pq.byte
s	90%	9743	- train/iVAMDtFgjQHusEGmWJCT.byte
s	90%	9747	- train/IvgyiZN1ULu7BQ09xYop.byte
s	90%	9751	- train/IVMFepgX5rASctxiL7y8.byte
s	90%	9755	- train/IVQkUgrW24Bmbtfvwqo7.byte

s	90%	9762	-	train/Ivwjph2yx11gQAjGiDU0.byte
s	90%	9767	-	train/IwBULXyFlnWZOrsm1Yh2.byte
s	90%	9769	-	train/iWd5sag08IpDQFHKbV1t.byte
s	90%	9771	-	train/IwgDdxrbUkiXpfAMy6J1.byte
s	90%	9772	-	train/IwgVeC9qRbUkfE6Q3HTh.byte
s	90%	9775	-	train/iwJ0hQfj1vm6qbGcE0ZY.byte
s	90%	9777	-	train/iWkQeLjbAGmI4onsrZNU.byte
s	90%	9779	-	train/IwNUoFX01y2aBL9ZmObh.byte
s	90%	9782	-	train/IWryLwtbuK130sG04Tzl.byte
s	90%	9785	-	train/iwTV1BuUW5Rjd0C4oqzx.byte
s	90%	9788	-	train/IWxQPzmgwUTV0tauq0hZ.byte
s	90%	9791	-	train/Ix1zV6v9hTQNPnk0aR7m.byte
s	90%	9795	-	train/IXcCRNnWtATu1fsQ2qaG.byte
s	90%	9799	-	train/IxEuUnjarmv241T8pofQ.byte
s	90%	9801	-	train/iXgHCQ6bTzvFPout92dn.byte
s	90%	9804	-	train/IxN43MpQlgV0ZrPdGjk9.byte
s	90%	9807	-	train/IxQyphN8rDiOn3dYsjS6.byte
s	90%	9809	-	train/IxVpXTyfqWoj8ri7sMSc.byte
s	90%	9811	-	train/iXvTztNsUj9dJw5rqAWo.byte
s	90%	9815	-	train/iY8OV2AcxubldjFhTBSp.byte
s	90%	9820	-	train/IYFTBN6LCn7SEQgbPOaU.byte
s	90%	9822	-	train/iygdLJ12UmnPhGqtsSCo.byte
s	90%	9827	-	train/iypLG5f97Y2ZTKcBuUxW.byte

s	90%	9831	-	train/iyS0Cw3xsHdq8efXvL2t.byte
s	90%	9832	-	train/IyTm2g1Xst0oAW5uLBQR.byte
s	90%	9835	-	train/IyVlH08cxfzMLvjFbr0X.byte
s	90%	9836	-	train/IYXhFeyzUjrs6RNZcWAd.byte
s	90%	9840	-	train/Iza3Zouq4wNDSCvgKcLr.byte
s	90%	9841	-	train/izAB3JHdR0Iwy8UmKbY1.byte
s	90%	9846	-	train/IzNgeE6msMS5X0ODRrWA.byte
s	91%	9851	-	train/IZWuOMidGxg1HYNP9fJr.byte
s	91%	9855	-	train/j06GEwJCPfkXusRh1c8D.byte
s	91%	9857	-	train/J0DAPdNqMUiQaThtKyGZ.byte
s	91%	9859	-	train/J0fY305lUmuNqVPBDRKn.byte
s	91%	9863	-	train/j1bM0nfqksmQ3X5epgau.byte
s	91%	9866	-	train/J1LCS4Foa2jbA0YGWQhz.byte
s	91%	9870	-	train/J1pPTHfbe7c2I6zFNLj0.byte
s	91%	9871	-	train/J1pSxIYHluoVsGXdU0rc.byte
s	91%	9874	-	train/J1X7hSrCk2YDdszA0xBj.byte
s	91%	9880	-	train/j2IBarGQysHh3Cfz6M8i.byte
s	91%	9884	-	train/J2PUnDWVabBmrFxEeYG.byte
s	91%	9888	-	train/j2wLH4fBoVF7rJK0ks5t.byte
s	91%	9890	-	train/J31r0hE2xFveWf6CNwGa.byte
s	91%	9892	-	train/j375bQGwkDU9X18WA2IS.byte
s	91%	9896	-	train/j3F0SiYcJwQPbeRAIaVZ.byte
s	91%	9902	-	train/J3yrHmKZTv5tpB1ExFkh.byte

s	91%	9903	-	train/J3Z47eiRfVBUtdbrODSK.byte
s	91%	9906	-	train/j4dgfzhlaw1exCvbJAQS.byte
s	91%	9908	-	train/j4G3L8yMzNPkwhmE9vVg.byte
s	91%	9910	-	train/J4LdiPEnt2yDVqhCofc3.byte
s	91%	9914	-	train/j514c8CPoS9fIyKmn0aH.byte
s	91%	9919	-	train/j50yPphcxSMDXJo9LI4C.byte
s	91%	9921	-	train/J5RbToUgXr2wAm3uGZ0e.byte
s	91%	9925	-	train/J6eCDilQ2sMaTugckmB7.byte
s	91%	9928	-	train/j6msrVaoBEJ8cWxgZeAH.byte
s	91%	9931	-	train/J6TicZvrREo1B7CnU1Se.byte
s	91%	9933	-	train/j6zenOwUWpPruio3Xk4b.byte
s	91%	9935	-	train/j7EMTLk2BV1upcU9tgqa.byte
s	91%	9939	-	train/j7UqgpTCiRr1s6kWHvKB.byte
s	91%	9941	-	train/J80lTkouaSw0y9RWHeZC.byte
s	91%	9945	-	train/J8EvIQbVSY0XxKC4wgd5.byte
s	91%	9947	-	train/j8FwWsmQaJ6ryRVPY37G.byte
s	91%	9949	-	train/j8nmPzbFQ3ovqrXu9JtV.byte
s	91%	9952	-	train/J8t7spr0xIUbZyg1cmBu.byte
s	91%	9954	-	train/J9BpcFqysLwr0YHeNt2o.byte
s	91%	9958	-	train/j9fqwK78kFxmBWclQh2N.byte
s	92%	9963	-	train/J9vHglnKXhiYD4716Gwy.byte
s	92%	9967	-	train/ja7NE82XwHYceyg6o0vf.byte
s	92%	9971	-	train/JahiBscoyeldwpEV0mxW.byte

s	92%	9973	-	train/JajfhmMDxzBX8OKb9o4k.byte
s	92%	9975	-	train/Ja08p2qFXjiNARcbIvGH.byte
s	92%	9976	-	train/jAPy1K002qwrC4VDgUGt.byte
s	92%	9978	-	train/JaQydTjesbikFpx6m9Zv.byte
s	92%	9981	-	train/JatHiWwG3Amc6VTXuSqC.byte
s	92%	9983	-	train/jav8x0cXCqs0ofl9MPyw.byte
s	92%	9988	-	train/jBAq9uw7KzWcE85yJMhd.byte
s	92%	9991	-	train/JBEzM0ZsuWaDt8bLypPv.byte
s	92%	9994	-	train/JBhPuigxYtENMs4yToa1.byte
s	92%	9995	-	train/JbHYMfgWXkjsx7iVA0Za5.byte
s	92%	9998	-	train/jbkwtqSZ6HavpN7Rr8e1.byte
s	92%	10000	-	train/jBqmxJCU0dZT03kSyhMI.byte
s	92%	10003	-	train/Jbv0zUjc0hsALd5waN8F.byte
s	92%	10005	-	train/JBwr0MNAVT4z13kvifYb.byte
s	92%	10006	-	train/JBXrCHbwSvZ9c5y06701.byte
s	92%	10009	-	train/jC6QJ5RarVUNMAfpsnbw.byte
s	92%	10011	-	train/JCA6mY29clo8LOGxaMnT.byte
s	92%	10014	-	train/JcFgaI0058Am1EVtPxY4.byte
s	92%	10015	-	train/jCfTR40xbaJmMz2knoLF.byte
s	92%	10023	-	train/jCrTVsM6f9S0mRP8p14i.byte
s	92%	10026	-	train/jCXhwV04ZdNqA9fHauDM.byte
s	92%	10031	-	train/jcZCrbDqHQusThgn0eV5.byte
s	92%	10032	-	train/JCzoBIAiqnR09N3Pumvb.byte

s	92%	10034	-	train/JCzVMiHYuK13LZgo6Nct.byte
s	92%	10038	-	train/jD6G4s3pinYJQTMxmgfZ.byte
s	92%	10041	-	train/jdcAqB7IF9bflz0VQMLY.byte
s	92%	10043	-	train/jDfyvw6lL31RQYMBHGzq.byte
s	92%	10046	-	train/jDlW6ZoIc00Nh5L4sqa2.byte
s	92%	10049	-	train/jdqclF2YAsm4LJwZUxPN.byte
s	92%	10052	-	train/jdUNHTV5DrnJSZqW2MPp.byte
s	92%	10055	-	train/je0Ds6Gax5oiTCmPAcBR.byte
s	92%	10057	-	train/je4HWtx9M5ySpPB00wgf.byte
s	92%	10061	-	train/JecsGrMgn5xWXVkyj2Io.byte
s	92%	10065	-	train/jepVLK14FMHN90rgJCzs.byte
s	93%	10070	-	train/jEVAIFWvTw1hu3yrizUa.byte
s	93%	10071	-	train/jExBgXmKlha4URac86kQ.byte
s	93%	10074	-	train/JEzvM430SnZQcV901IBR.byte
s	93%	10076	-	train/JF2417UDdQBcfryqLz3g.byte
s	93%	10077	-	train/jF8Qxdo3MyVEc9l602B1.byte
s	93%	10080	-	train/jFcM7da1kw3GCRBZnHrQ.byte
s	93%	10082	-	train/JfgmjCtoTKyM6pcIrBY7.byte
s	93%	10083	-	train/JfHLpzI7Za8nXSy0lGj0.byte
s	93%	10086	-	train/JF1Nr7Vo458dMxwQz6nh.byte
s	93%	10091	-	train/JfUstdXDhbQZ0xuBcG60.byte
s	93%	10093	-	train/JfvwSRiN008Z4X7jK1qd.byte
s	93%	10095	-	train/JFW8DGfRUPbNo3YHz2xq.byte

s	93%	10098	-	train/JG1LeckMOuCUwIzbn5Wg.byte
s	93%	10103	-	train/JGaqTC0sRIciH5n3kDUX.byte
s	93%	10106	-	train/Jgc17BNL9qFovTjD65hY.byte
s	93%	10107	-	train/jgc7JuLdUTfKEIiC32v5.byte
s	93%	10109	-	train/jgeLHoN0sITmWJ8bpnZi.byte
s	93%	10115	-	train/jg03mkayeMX8CuLbxFpV.byte
s	93%	10117	-	train/jg0s7KiB0aTEzvSUJVPP.byte
s	93%	10118	-	train/Jgp1wWtT2sOZKb5FSAy7.byte
s	93%	10121	-	train/JGqMAsU8BNzPpT1cIF2n.byte
s	93%	10124	-	train/jGt4cydZi2DSfu3qbxHV.byte
s	93%	10126	-	train/jGtigsJRXU0zaMbcmpHl.byte
s	93%	10128	-	train/jgvqdXTZ6pcAeb043NxF.byte
s	93%	10131	-	train/jGXx165NzmIQff802LRC.byte
s	93%	10133	-	train/jh371q0WXKwZP8gynubU.byte
s	93%	10135	-	train/jh5EgYDdqxsan190RW8U.byte
s	93%	10136	-	train/JH5XUDg8MOCWt7xQESRv.byte
s	93%	10139	-	train/JhBYHQsTMGWqaf240u0p.byte
s	93%	10143	-	train/jHfu9AM11GkNWch0Yv53.byte
s	93%	10148	-	train/JHQWofBYhqSF8V12ktU0.byte
s	93%	10151	-	train/JhufrNciMQKET0UGZ8Ls.byte
s	93%	10153	-	train/JhV2BbPKtQixTCMFzId0.byte
s	93%	10156	-	train/jHY0wUV8zrCoKBgXkG2l.byte
s	93%	10158	-	train/JI3ESrZoxtwm8kfdQ2UC.byte

s	93%	10162	-	train/jIDdt1WG4JV3zMa0XSb2.byte
s	94%	10165	-	train/JINMqhxUa3Efc8QF9vCd.byte
s	94%	10168	-	train/JiS9zt40D5u1xFqckbU0.byte
s	94%	10172	-	train/JitysIgm3CB0nDK1W4ko.byte
s	94%	10175	-	train/JIX9YcvfnEK054aHWphi.byte
s	94%	10176	-	train/JIy6ZgaGtHBAmv2krdKx.byte
s	94%	10179	-	train/jIzgiMTcuZPt4eJNr2Yw.byte
s	94%	10181	-	train/Jj7eK6AhkmDRB53WS04P.byte
s	94%	10184	-	train/JjOt1WnEXeH8Mhd2461K.byte
s	94%	10188	-	train/jk94DfpsQo7X1HqSrhZi.byte
s	94%	10192	-	train/JkcvVWjUdD00TuSmzA4q.byte
s	94%	10193	-	train/jKdpYN8VETvaHs2RqmQS.byte
s	94%	10198	-	train/JKNVsPwYM4pojQL127q3.byte
s	94%	10202	-	train/JKprfw8QS7yketdHLFEU.byte
s	94%	10203	-	train/jKs2xGwT7nCOQ5hYraFU.byte
s	94%	10206	-	train/jKT8Gepv10VC72PYzWXF.byte
s	94%	10210	-	train/JkzWI2MAvRi64g5QGUEf.byte
s	94%	10211	-	train/jL2VfCmgMHhoaG94PS6k.byte
s	94%	10213	-	train/jL9VyspPbfEoQ38lCFZG.bytes
s	94%	10216	-	train/jLcX3o0ivMTafFeSq19b.byte
s	94%	10221	-	train/Jlhb9akq5NpXnvIg4DG8.byte
s	94%	10223	-	train/JLiNd1ksMDOB4P9cWY2b.byte
s	94%	10225	-	train/jlPuQh3av94snwycJEZT.byte

s	94%	10229	-	train/JLrOfGem93RVtBySioNx.byte
s	94%	10230	-	train/JLs2jrmke9d0h0ZgITwf.byte
s	94%	10237	-	train/jLuXBKEVC9ngUadxi2Nw.byte
s	94%	10240	-	train/JLZtxyXbmRfe8qi0QHig.byte
s	94%	10246	-	train/JMdWHZWNibQcFoSPG0s0.byte
s	94%	10251	-	train/JmKxR1BwiHDqnyskEA4p.byte
s	94%	10255	-	train/Jm07iq6nVLPrjH4z2UC9.byte
s	94%	10258	-	train/JmRCrDBG EzfgISdvNsF7.byte
s	94%	10260	-	train/jMtLuy8lJKkiomc39Ibv.byte
s	94%	10266	-	train/JN6fQoCvEeImVXjk5usc.byte
s	94%	10267	-	train/JnCS80h4fUmciFgN5IZ3.byte
s	94%	10270	-	train/JNEaK80h4meUkGLnzsox.byte
s	94%	10272	-	train/JNez4Z10tfKa52BDckVb.byte
s	95%	10273	-	train/JNfrkZYIEmpVDBzn1kvL.byte
s	95%	10277	-	train/jnKkElPsxcIaUFhQb8Jd.byte
s	95%	10279	-	train/jnP3ldrIwpWfEg80DV9K.byte
s	95%	10281	-	train/JNr1KDpRFXhP6wcQjMna.byte
s	95%	10285	-	train/jNUWxT9ZE3AzQHJcf1Rg.byte
s	95%	10289	-	train/JnX9eYfjBik8KzyGE7sM.byte
s	95%	10291	-	train/JnysZFLMzgC4rQDtW7pB.byte
s	95%	10293	-	train/Jo234aEArMdiRlQSD8g0.byte
s	95%	10294	-	train/JO2iskzU1BVGj6Y4PEZc.byte
s	95%	10296	-	train/j03ktp4CTZ8Voy2cHJUB.byte
s				

s	95%	10297	-	train/J03YyqSKEfcd0RN4hLFu.byte
s	95%	10301	-	train/j0aZK2toAgbxChzNvPmp.byte
s	95%	10303	-	train/JoDLpkCx5ziwtAU73NVF.byte
s	95%	10306	-	train/JokgKM4C3RYe0Atmv7j9.byte
s	95%	10309	-	train/JonIk6ScNpZM7WgFzm1D.byte
s	95%	10312	-	train/j0Ttmr8uJvc1SIBf1xHo.byte
s	95%	10314	-	train/j0VlvmtXnWKg6JFrZLTG.byte
s	95%	10316	-	train/JOWv6HNY1xwuqPrICEdR.byte
s	95%	10318	-	train/joxy4nGdLbYR187qJMAN.byte
s	95%	10323	-	train/JP7BMu0e5I0yYmSCbTNd.byte
s	95%	10325	-	train/jpdhY15v8mkeftGiF4VA.byte
s	95%	10328	-	train/jpkIrHe201TamDoXqEQi.byte
s	95%	10330	-	train/JPNtw3z5pD06EAd94oxh.byte
s	95%	10332	-	train/JpsS8BAHtDOMNUIyYkax.byte
s	95%	10333	-	train/jpT74BziaEYAtIfhnyW0.byte
s	95%	10339	-	train/JPz2sCNRcjVUDrwLh8S3.byte
s	95%	10341	-	train/JQ7wRLlnb6WqEkd28mgF.byte
s	95%	10344	-	train/JQbyODV7raAtZHldLfS0.byte
s	95%	10347	-	train/JQHf0c52agNmervIpCkT.byte
s	95%	10348	-	train/jqIy4VhtwkTvF1ZLYHAa.byte
s	95%	10352	-	train/jqNt0mQMAOPZDHWxc6sY.byte
s	95%	10355	-	train/JqRujNbmhKvsxnfrTPwI.byte
s	95%	10363	-	train/jrAH5I14hpeni2TbcPOR.byte

s	95%	10369	-	train/jrKTfD3VAXsFJ9vn0SMe.byte
s	95%	10372	-	train/JRp gnDye94P8crMsVNHv.byte
s	95%	10374	-	train/jRqJ7xPQF2ia6tTsbKUY.byte
s	95%	10379	-	train/Jrx0687zPbmCVTacWq9d.byte
s	96%	10384	-	train/JS8ZjyGstceXaxVgH5bE.byte
s	96%	10388	-	train/JSEN00RduDWZzG3lHcey.byte
s	96%	10395	-	train/JsR2Akf5xer89PYDCjF6.byte
s	96%	10399	-	train/jsTnFQZN0zuGqAgc0faS.byte
s	96%	10403	-	train/JSYxG6FyPnqQzEp73B9e.byte
s	96%	10408	-	train/Jt9wHff0M8PNQKS04mZy.byte
s	96%	10409	-	train/jta5QoJ6BOMDnvhIu2Uq.byte
s	96%	10413	-	train/jTbRSLsJ3a1IByZnYx4D.byte
s	96%	10415	-	train/jtEkniTysxuqG6DhAX57.byte
s	96%	10417	-	train/jTgCQwsHRJMLbVrZYxe3.byte
s	96%	10421	-	train/Jtp04ZqNwST65MdH7xhD.byte
s	96%	10423	-	train/JTQc1mK6P8bI7zF4f90r.byte
s	96%	10425	-	train/jTw10q9xuU6eoa82WkdZ.byte
s	96%	10429	-	train/ju23GQl4DYFVeNwc85ka.byte
s	96%	10434	-	train/jUDRevNaIGil1Tp7S0ks.byte
s	96%	10438	-	train/juEShDrt6lY0wJCp9icX.byte
s	96%	10439	-	train/JufSg9mpGi1dVZ0C7zWA.byte
s	96%	10444	-	train/jULaFyfMxw9nPWX6uK12.byte
s	96%	10445	-	train/jumeizJhVvb4d1Sn3M9Z.byte
s				

s	96%	10448	-	train/JU03pfywZnC4e9xHLBMA.byte
s	96%	10452	-	train/JuUnQZ5ef8BW4aG1Vbkr.byte
s	96%	10455	-	train/jUy3cSnq2h9ICJkZ1HEX.byte
s	96%	10456	-	train/juYn1MbR7SmyqDIPzZX0.byte
s	96%	10461	-	train/jv9q0Ae8pCmL7iGRDv6o.byte
s	96%	10464	-	train/jvdbriZknaRP9emq2Lp0.byte
s	96%	10467	-	train/jVEk2G0fyQUiNWbxAzch.byte
s	96%	10470	-	train/jviZUT0mHFqDPVzrKXLu.byte
s	96%	10472	-	train/JVPnwpfeExcNDt2ZTAQm.byte
s	96%	10474	-	train/JvsW2yaUokpXVcbh10Mu.byte
s	96%	10478	-	train/JVYGACXbeMWgP9hHsocD.byte
s	96%	10479	-	train/JVykPqaQI54ZfizrLeeF.byte
s	96%	10481	-	train/jw1wp9ARJTIFdD5816V7.byte
s	96%	10485	-	train/jw6dqJau5vzf7Vy1sb9D.byte
s	96%	10487	-	train/JWczYA5Isj7MdnQhq2Lv.byte
s	96%	10492	-	train/jwgxPFMNGtXKTYkio0rI.byte
s	96%	10497	-	train/jWNEHhAV6QLM9bTPeFyS.byte
s	96%	10500	-	train/jwQGe74vFSE3bVoxdCIZ.byte
s	96%	10501	-	train/jwQPIaD0Urhgv3FH2Cnb.byte
s	96%	10503	-	train/JwsCip5tZ1Rhk9ga1KB3.byte
s	97%	10507	-	train/JWYBQVADRKy1xCaXIfu8.byte
s	97%	10509	-	train/JwzA3tXdgc7eNYxjCy5L.byte
s	97%	10512	-	train/Jx4hzsGNkP71Wd5qRXYS.byte
s				

s	97%	10515	-	train/jxDGoTHUWQeLfakytD05.byte
s	97%	10521	-	train/JxhBKmpVSNiTybZEoLYD.byte
s	97%	10523	-	train/JXkn1EudrFKLeSxT35It.byte
s	97%	10527	-	train/Jxo85BDNHMavFUSnp1ZO.byte
s	97%	10529	-	train/jx0WbXpVtGuyQI5DSUCB.byte
s	97%	10534	-	train/Jy0NTuqF1px6LnhIXWK4.byte
s	97%	10537	-	train/jy3BzhR9UfMtuVeICmkQ.byte
s	97%	10540	-	train/JyBvdherH16jbUktYL79.byte
s	97%	10542	-	train/JycBQ9bRKZChEMI4DWPe.byte
s	97%	10548	-	train/jynSzuQrYfI7eGHWKmPZ.byte
s	97%	10551	-	train/Jyr4P9SzRXWmEgFwhfub.byte
s	97%	10554	-	train/Jyt0pUNjThR5knZgifGQ.byte
s	97%	10559	-	train/jz61IKlwh8tLe27qWudk.byte
s	97%	10561	-	train/Jzbqx3eGnATaDo4IVWjL.byte
s	97%	10563	-	train/jZCp8IYzrv5TXsS0wEV3.byte
s	97%	10565	-	train/JzF18ftmDC3V0LnIAhWP.byte
s	97%	10569	-	train/jzgSsMvr4QtI5i2LaN1J.byte
s	97%	10573	-	train/JZ1EihvwTBmPUzQreyHc.byte
s	97%	10576	-	train/jztINZ6PGKThubQ78fWk.byte
s	97%	10582	-	train/JZX2B0zdS7MAVmjLUkx6.byte
s	97%	10585	-	train/JZyx7pL2ueEr1Nw1j8aD.byte
s	97%	10587	-	train/K05YJpcfyTCU4EdShQI1.byte
s	97%	10591	-	train/k0LehV4n0dwpWZ7DmSbM.byte

s	97%	10593	-	train/K0Z2FrjYgcqoQaUuOGWP.byte
s	97%	10595	-	train/K107ZH09Xzc5aquiuxUmo.byte
s	97%	10597	-	train/K17ifkYqXgN0Hav8uzCL.byte
s	97%	10600	-	train/K1c0hi76tkCvbNXPfLmD.byte
s	97%	10604	-	train/K1lFzHEd9BrbaILcmMsJ.byte
s	97%	10608	-	train/K1TBRacdPrvq0XCjDw2J.byte
s	97%	10612	-	train/k2Bues6ACEpX4WjGylo0.byte
s	97%	10614	-	train/K2jS3nvYxiVstye7IcC0.byte
s	98%	10616	-	train/k2mxrqNg1JzRiVsIbytQ.byte
s	98%	10618	-	train/K20ouiFyb1xSIVLRNfqE.byte
s	98%	10620	-	train/K2QNGrnBZP7I5MiF9kUb.byte
s	98%	10625	-	train/k2XTme5cZwf3Y9VShBn1.byte
s	98%	10627	-	train/k37eIpQ2ARSHZlYhnM0d.byte
s	98%	10628	-	train/K3BiLDx5eFqoG2tETujV.byte
s	98%	10632	-	train/K3noAT9Sb0V4R1DMf2Zk.byte
s	98%	10634	-	train/k3Vna8Fm1WodsDx1TH4N.byte
s	98%	10638	-	train/k4fumEIZvt1KUbJ32c0H.byte
s	98%	10642	-	train/K4M2P0Em1I7FxfSdXQD3.byte
s	98%	10646	-	train/k4U7bfep0EJzgK8S231X.byte
s	98%	10650	-	train/k4zuJ9BFX08KM3xIiDVh.byte
s	98%	10654	-	train/K5tfseBz7vxwAhCPI6Rq.byte
s	98%	10658	-	train/k6lGB2z9JmIVMAnsKZvP.byte
s	98%	10661	-	train/K6XgyIA40cPJemU1CV7G.byte

s	98%	10665	-	train/K786iNwZCzFfscS0h5u9.byte
s	98%	10672	-	train/K7va1ZC6DzFHqBVgWY2R.byte
s	98%	10676	-	train/k89IEjZqBQTxAyLGWCwi.byte
s	98%	10680	-	train/k8FHSmDTwC3eIrYsU46v.byte
s	98%	10684	-	train/K8zhErbvBDQ9FdYyT30R.byte
s	98%	10692	-	train/k9xRmJIPX3h68arYZ5Hj.byte
s	98%	10695	-	train/kaCP4Mghom8e1ujB6HY9.byte
s	98%	10700	-	train/Kakd056B2GhiC3szw7TI.byte
s	98%	10705	-	train/kAtomObScfxVW1DKXU1a.byte
s	98%	10709	-	train/kaz4GuJLsES3IDtnXc10.byte
s	98%	10714	-	train/kbilTy4z7H8PmpW20QXx.byte
s	98%	10717	-	train/kBoNGqDtARYyC6uz5V42.byte
s	98%	10721	-	train/KBV7bhMZ0tNApmdnyYqF.byte
s	98%	10724	-	train/kbZrNaLliTWFB3SM5K9e.byte
s	98%	10726	-	train/Kc1XJamkjDW41B7VtvLZ.byte
s	98%	10730	-	train/kcCrogv0Dhe6A52B8URa.byte
s	98%	10735	-	train/KCJpmMH7aG4c3T0eEdxQ.byte
s	98%	10739	-	train/kcwP62orCTg9MFEVzIus.byte
s	98%	10743	-	train/Kd1CHyERprf9NTQM1sWL.byte
s	98%	10750	-	train/kdmNfXGws1xUa3yATSfw.byte
s	98%	10754	-	train/KDW8jkt3SfaENRxPv1pa.byte
s	98%	10758	-	train/ken1LRYK7B4ryGmT3MfC.byte
s	98%	10763	-	train/kfigr70cURVxjFCabY4n.byte

s	98%	10767	-	train/Kfn9jdUWrz5cm03hiu7S.byte
s	98%	10768	-	train/KFNC4Dl2LmXM0u8iAdYT.byte
s	99%	10772	-	train/kFSnHGV65dtl0u0AExJm.byte
s	99%	10776	-	train/kg24YRJTB8DNdKMxpwOH.byte
s	99%	10780	-	train/kgfPdJYxa5GtleWVrvEM.byte
s	99%	10786	-	train/KGorN9J6XAC4b0Ekmyup.byte
s	99%	10791	-	train/KH8I6ft57iXUTmOnvZVP.byte
s	99%	10795	-	train/KHmRYShqZacoQE0Jx862.byte
s	99%	10801	-	train/kI2PQoudeFEDY0BTliUf.byte
s	99%	10804	-	train/KIdHiQmorRG90vEcnp7.byte
s	99%	10806	-	train/KiI6X4afklWuORnqBtz5.byte
s	99%	10810	-	train/ki0ADfHlLYqyt9dUz4eN.byte
s	99%	10814	-	train/kiv3fagQwzY9u4rOdET1.byte
s	99%	10817	-	train/KIxWlZ5Py8GaRj0SOrAD.byte
s	99%	10818	-	train/kJ1a9c6NoCv4DqA7Hh8R.byte
s	99%	10824	-	train/kJNLp6CyAKfgVMFwUInG.byte
s	99%	10827	-	train/kKBjnDycpQmzPol5OU6Z.byte
s	99%	10831	-	train/kKzsPq9NQLSZA0ewUgBo.byte
s	99%	10837	-	train/kLPUQYKrA1jGsJdtS0bf.byte
s	99%	10841	-	train/KMAcej7qgX5u9RbTUZ0x.byte
s	99%	10847	-	train/KmFdoSG4XLMVIHewYRC8.byte
s	99%	10851	-	train/kMLGcpf0ht689UwexvZm.byte
s	99%	10852	-	train/KMyXSocsnm0ZiI0N1Pg2.byte
s				

```
Files: 10931
Size:      51652298914
Compressed: 18810691091
```

Mounted at /content/drive

```
In [ ]: import IPython
        from google.colab import output

        display(IPython.display.Javascript('''
            function ClickConnect(){
                btn = document.querySelector("colab-connect-button")
                if (btn != null){
                    console.log("Click colab-connect-button");
                    btn.click()
                }

                btn = document.getElementById('ok')
                if (btn != null){
                    console.log("Click reconnect");
                    btn.click()
                }
            }

            setInterval(ClickConnect,60000)
        '''))

        print("Done.")
```

Done.

```
In [ ]: import scipy
        from scipy.sparse import load_npz
        from scipy.sparse import vstack
        root_dir = "/content/drive/My Drive/final_features/"
        bi_vector1 = load_npz(root_dir+"bytebigram1.npz")
        bi_vector2 = load_npz(root_dir+"bytebigram2.npz")
        final_bi_vector = vstack((bi_vector1[:5434],bi_vector2[:5434]))
        print(final_bi_vector.shape)
```



```
In [ ]: import pandas as pd
import pickle
import numpy as np
import seaborn as sns
import matplotlib.pyplot as plt
from sklearn.preprocessing import normalize
from sklearn.metrics import confusion_matrix
from sklearn.model_selection import train_test_split
from sklearn.neighbors import KNeighborsClassifier
from sklearn.linear_model import LogisticRegression
from sklearn.ensemble import RandomForestClassifier
from xgboost import XGBClassifier
from sklearn.model_selection import RandomizedSearchCV
from sklearn.calibration import CalibratedClassifierCV
from sklearn.metrics import log_loss
from sklearn import preprocessing
```

Extracting Byte Bigram feature

```
In [ ]: df = pd.read_csv("/content/drive/My Drive/Microsoft_Malware_Detection_Files/trainLabels.csv")
y_data = df["Class"]
y_data.shape
```

```
Out[ ]: (10868,)
```

```
In [ ]: y_data
```

```
Out[ ]: 0      1
1      1
2      1
3      1
4      1
..
10863  9
10864  9
10865  9
10866  9
10867  9
Name: Class, Length: 10868, dtype: int64
```

```
In [ ]: byte_vocab = "00,01,02,03,04,05,06,07,08,09,0a,0b,0c,0d,0e,0f,10,11,12,13,14,15,16,17,18,19,1a,1b,1c,1d,1e,1f,20,21,22,23,
24,25,26,27,28,29,2a,2b,2c,2d,2e,2f,30,31,32,33,34,35,36,37,38,39,3a,3b,3c,3d,3e,3f,40,41,42,43,44,45,46,47,48,49,4a,4b,4
c,4d,4e,4f,50,51,52,53,54,55,56,57,58,59,5a,5b,5c,5d,5e,5f,60,61,62,63,64,65,66,67,68,69,6a,6b,6c,6d,6e,6f,70,71,72,73,74,
75,76,77,78,79,7a,7b,7c,7d,7e,7f,80,81,82,83,84,85,86,87,88,89,8a,8b,8c,8d,8e,8f,90,91,92,93,94,95,96,97,98,99,9a,9b,9c,9
d,9e,9f,a0,a1,a2,a3,a4,a5,a6,a7,a8,a9,aa,ab,ac,ad,ae,af,b0,b1,b2,b3,b4,b5,b6,b7,b8,b9,ba,bb,bc,bd,be,bf,c0,c1,c2,c3,c4,c5,
c6,c7,c8,c9,ca,cb,cc,cd,ce,cf,d0,d1,d2,d3,d4,d5,d6,d7,d8,d9,da,db,dc,dd,de,df,e0,e1,e2,e3,e4,e5,e6,e7,e8,e9,ea,eb,ec,ed,e
e,ef,f0,f1,f2,f3,f4,f5,f6,f7,f8,f9,fa,fb,fc,fd,fe,ff,??"
byte_bigram_vocab = []
for i, v in enumerate(byte_vocab.split(',')):
    for j in range(0, len(byte_vocab.split(','))):
        byte_bigram_vocab.append(v + ' ' + byte_vocab.split(',')[j])
len(byte_bigram_vocab)
```

Out[]: 66049

```
In [ ]: byte_bigram_vocab[:5]
```

Out[]: ['00 00', '00 01', '00 02', '00 03', '00 04']

```
In [ ]: from tqdm import tqdm
from sklearn.feature_extraction.text import CountVectorizer
import scipy
```

```
In [ ]: byte_file_id= df['Id']
```

```
In [ ]: def create_bygram(strt,end,proc_num):
    vector = CountVectorizer(lowercase=False,ngram_range=(2,2), vocabulary=byte_bigram_vocab)
    bytebigram_vect = scipy.sparse.csr_matrix((len(byte_file_id)//2,len(byte_bigram_vocab)))
    print(strt,end)
    for i, file in tqdm(enumerate(os.listdir("/content/bytes"))[strt : end]):
        f = open('/content/bytes/' + file)
        bytebigram_vect[i,:]+= scipy.sparse.csr_matrix(vector.fit_transform([f.read().replace('\n', ' ').lower()]))
        f.close()
    scipy.sparse.save_npz('/content/gdrive/My Drive/Microsoft_Malware_Detection_Files/bytebigram'+str(proc_num)+'.npz', byte
bigram_vect)
```

```
In [ ]: from multiprocessing import Process
P1 = Process(target =create_bygram,args =(0,(len(byte_file_id)//2),1))
P2 = Process(target = create_bygram,args = ((len(byte_file_id)//2),(len(byte_file_id)),2))
P1.start()
P2.start()
P1.join()
P2.join()
```

0 5434

5434 10868

0it [00:00, ?it/s]/usr/local/lib/python3.6/dist-packages/scipy/sparse/_index.py:118: SparseEfficiencyWarning: Changing the sparsity structure of a csr_matrix is expensive. lil_matrix is more efficient.

self._set_arrayXarray_sparse(i, j, x)

1it [00:02, 2.03s/it]/usr/local/lib/python3.6/dist-packages/scipy/sparse/_index.py:118: SparseEfficiencyWarning: Changing the sparsity structure of a csr_matrix is expensive. lil_matrix is more efficient.

self._set_arrayXarray_sparse(i, j, x)

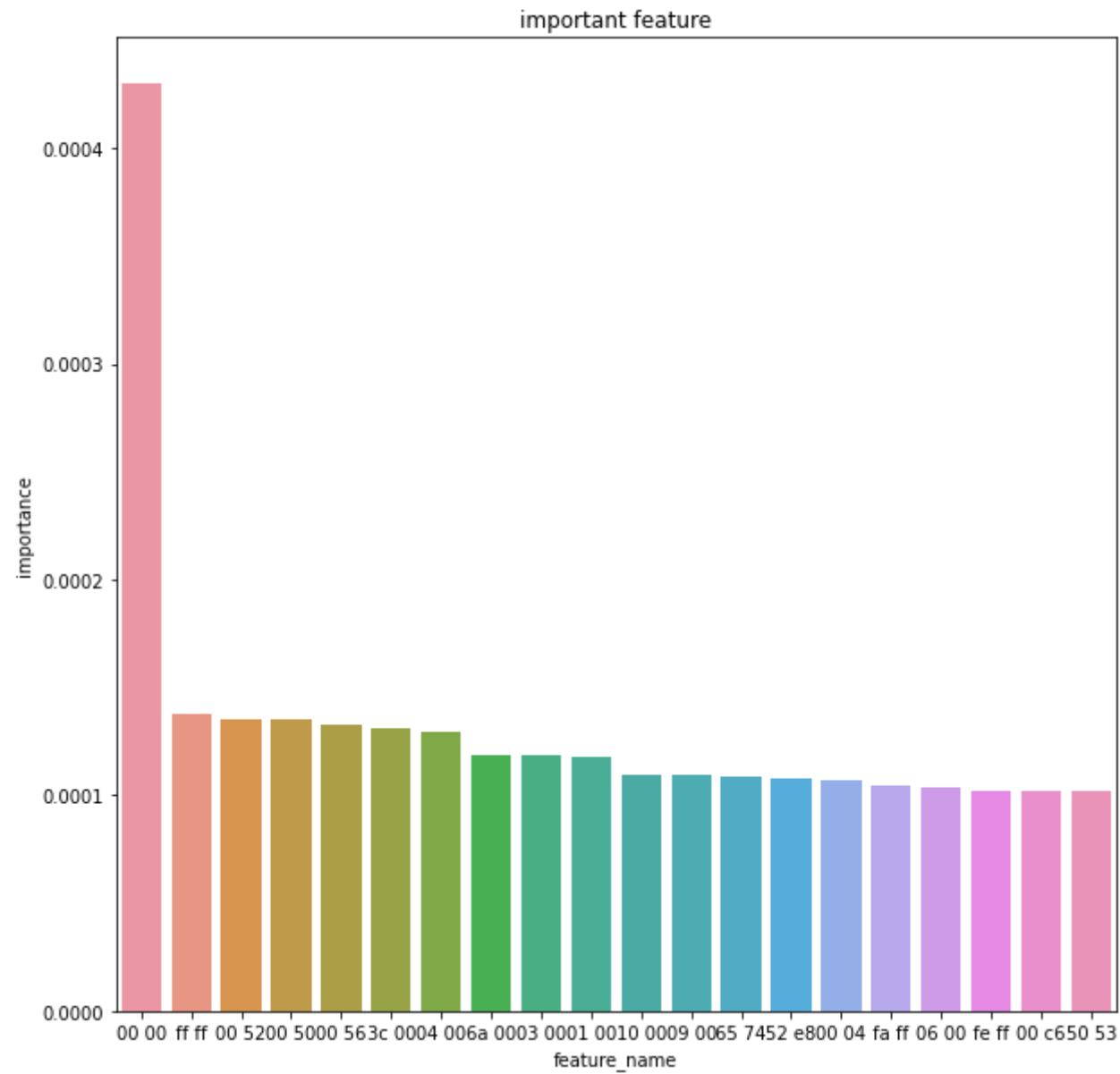
5434it [10:18:40, 6.83s/it]

5434it [10:24:17, 6.89s/it]

```
In [ ]: from sklearn.ensemble import RandomForestClassifier
```

```
def imp_feature(x_data,y_data,vocab):
    clf = RandomForestClassifier(n_estimators=100,n_jobs=-1)
    clf.fit(x_data,y_data)
    imp_feature_index = np.argsort(clf.feature_importances_)[::-1]
    feature = np.take(clf.feature_importances_,imp_feature_index[:20])
    feature_name = np.take(vocab,imp_feature_index[:20])
    plt.figure(figsize=(10,10))
    sns.barplot(x=feature_name,y=feature)
    plt.title("important feature")
    plt.xlabel("feature_name")
    plt.ylabel("importance")
    return imp_feature_index[:300]
```

```
In [ ]: normalized_bigram = normalize(final_bi_vector,axis=0)
top_features = imp_feature(normalized_bigram,y_data,byte_bigram_vocab)
```



```
In [ ]: np.save("top_features",top_features)
```

```
In [ ]: top_features = np.load("top_features.npy")
```

```
In [ ]: top_byte_features = np.zeros((10868,0))
        for i in top_features:
            copy = final_bi_vector[:,i].todense()
            top_byte_features = np.hstack([top_byte_features,copy])
```

```
In [ ]: byte_feature_csv = pd.DataFrame(top_byte_features,columns=np.take(byte_bigram_vocab,top_features))
        byte_feature_csv.to_csv("byte_feature.csv")
```

```
In [ ]: byte_feature_df = pd.read_csv("/content/drive/My Drive/final_features/byte_feature.csv")
        byte_feature_df = byte_feature_df.drop('Unnamed: 0',axis=1)
```

```
In [ ]: byte_feature_df[5000:6000]
```

Out[]:

	00 00	ff ff	00 5a	00 50	00 e8	fe ff	ff 00	01 00	ff 52	00 5e	00 03	07 00	65 00	00 52	00 c0	51 6a	ff 8d	6a 00	56 57	65 73	00 c8
5000	13019.0	2268.0	145.0	953.0	1344.0	192.0	2975.0	608.0	94.0	109.0	141.0	168.0	215.0	120.0	524.0	24.0	322.0	545.0	16.0	14.0	123.0
5001	29649.0	5046.0	71.0	195.0	377.0	738.0	1074.0	1136.0	44.0	68.0	389.0	133.0	82.0	73.0	176.0	7.0	123.0	298.0	105.0	31.0	82.0
5002	5735.0	204.0	26.0	30.0	86.0	27.0	278.0	120.0	34.0	12.0	80.0	50.0	26.0	35.0	21.0	24.0	40.0	53.0	10.0	20.0	13.0
5003	7856.0	19.0	5.0	21.0	9.0	3.0	3.0	58.0	0.0	4.0	0.0	3.0	13.0	7.0	6.0	11.0	3.0	39.0	11.0	10.0	2.0
5004	6089.0	111.0	68.0	48.0	54.0	31.0	292.0	116.0	34.0	22.0	81.0	49.0	38.0	55.0	32.0	32.0	50.0	82.0	11.0	26.0	31.0
...
5995	7781.0	2420.0	6.0	90.0	190.0	234.0	212.0	383.0	15.0	8.0	33.0	36.0	37.0	24.0	48.0	6.0	92.0	126.0	90.0	33.0	10.0
5996	15319.0	73.0	10.0	102.0	87.0	14.0	24.0	230.0	29.0	12.0	107.0	13.0	20.0	91.0	8.0	0.0	3.0	4.0	2.0	22.0	4.0
5997	25960.0	10445.0	12.0	278.0	367.0	622.0	1860.0	1068.0	171.0	502.0	254.0	262.0	48.0	32.0	291.0	3.0	527.0	329.0	145.0	37.0	21.0
5998	27645.0	10999.0	14.0	94.0	261.0	365.0	4609.0	1016.0	235.0	19.0	83.0	149.0	40.0	44.0	194.0	12.0	221.0	126.0	76.0	34.0	30.0
5999	11012.0	2430.0	208.0	946.0	1393.0	7458.0	2877.0	552.0	105.0	141.0	155.0	155.0	147.0	181.0	545.0	17.0	325.0	547.0	28.0	8.0	151.0

1000 rows × 300 columns

Extracting asm image feature

```

In [ ]: import subprocess
from atpbar import atpbar
from multiprocessing import process
from tqdm import tqdm
from csv import writer
import os
import array
import numpy as np
def sevenzip(df,strt,end,dest_dir,num):
    csv_file=dest_dir+str(num)+".csv"
    rows =[]
    with open(csv_file,'w') as f:
        fw = writer(f)
        column_names =['filename']+[( 'asm_{:s}'.format(str(x)))for x in range(1000)]
        fw.writerow(column_names)
        for i in tqdm(range(strt,end)):
            file_name = str(df[i])+'.asm'
            system = subprocess.Popen(['7z','e','train.7z','-o'+dest_dir,file_name,'-r'])
            system.communicate()
            file_id = file_name.split('.')[0]
            image_data = read_image(os.path.join(dest_dir,file_name))
            rows.append([file_id]+image_data)
            fw.writerows(rows)
            os.remove(os.path.join(dest_dir,file_name))
            rows=[]
            system2 = subprocess.Popen(['cp'])
def read_image(file):
    f = open(file,'rb')
    ln =os.path.getsize(file)
    width = 256
    rem =ln%width
    a = array.array('B')
    a.fromfile(f,ln-rem)
    f.close()
    g=np.reshape(a,(int(len(a)/width),width))
    g = np.uint8(g)
    g =np.resize(g,(1000,))
    return list(g)

```

```
In [ ]: from multiprocessing import Process
import time
p1 = Process(target = sevenzip,args=(name,6001,7001,"/content/asm_image_features",1))
p2 = Process(target = sevenzip,args=(name,7001,8001,"/content/asm_image_features",2))
strt = time.time()
print(strt)
p1.start()
p2.start()
p1.join()
p2.join()
end = time.time()
print(end)

print("time_taken ",strt-end)
```

1602945961.6316867

100%|██████████| 1000/1000 [3:51:54<00:00, 13.91s/it]

100%|██████████| 1000/1000 [3:58:30<00:00, 14.31s/it]

1602960272.057986

time_taken -14310.426299333572

```
In [ ]: asm_image_fea = pd.read_csv("/content/drive/My Drive/final_features/final_asm_image_features.csv")
```

```
In [ ]: asm_image_fea
```

Out[]:

	Unnamed: 0	filename	asm_0	asm_1	asm_2	asm_3	asm_4	asm_5	asm_6	asm_7	asm_8	asm_9	asm_10	asm_11	asm_12	asm_13
0	0	01kcPWA9K2BOxQeS5Rju	72	69	65	68	69	82	58	49	48	48	48	48	48	48
1	1	04EjldbPV5e1XroFOpiN	72	69	65	68	69	82	58	52	68	70	51	48	48	48
2	2	05EeG39MTRrI6VY21DPd	72	69	65	68	69	82	58	49	48	48	48	48	48	48
3	3	05rJTUWYAKNegBk2wE8X	72	69	65	68	69	82	58	55	68	70	48	48	48	48
4	4	0AnoOZDNbPXIr2MRBSCJ	72	69	65	68	69	82	58	49	48	48	48	48	48	48
...
10863	862	KFrZ0Lop1WDGwUtkusCi	72	69	65	68	69	82	58	48	48	52	48	48	48	48
10864	863	kg24YRJTB8DNdKMXpwOH	72	69	65	68	69	82	58	48	48	52	48	48	48	48
10865	864	kG29BLiFYPgWtpb350sO	72	69	65	68	69	82	58	48	48	52	48	48	48	48
10866	865	kGTL4OJxYMWEQ1bKBiP	72	69	65	68	69	82	58	48	48	52	48	48	48	48
10867	866	KGorN9J6XAC4bOEkmyup	72	69	65	68	69	82	58	48	48	52	48	48	48	48

10868 rows × 1002 columns

```
In [ ]: y = np.array(df.iloc[:,1])
x = asm_image_fea.drop(['Unnamed: 0','filename'],axis=1)
```



```
In [ ]: x
```

Out[]:

	asm_0	asm_1	asm_2	asm_3	asm_4	asm_5	asm_6	asm_7	asm_8	asm_9	asm_10	asm_11	asm_12	asm_13	asm_14	asm_15	asm_16	asm_17
0	72	69	65	68	69	82	58	49	48	48	48	48	48	48	48	9	9	
1	72	69	65	68	69	82	58	52	68	70	51	48	48	48	48	9	9	
2	72	69	65	68	69	82	58	49	48	48	48	48	48	48	48	9	9	
3	72	69	65	68	69	82	58	55	68	70	48	48	48	48	48	9	9	
4	72	69	65	68	69	82	58	49	48	48	48	48	48	48	48	9	9	
...
10863	72	69	65	68	69	82	58	48	48	52	48	48	48	48	48	9	9	
10864	72	69	65	68	69	82	58	48	48	52	48	48	48	48	48	9	9	
10865	72	69	65	68	69	82	58	48	48	52	48	48	48	48	48	9	9	
10866	72	69	65	68	69	82	58	48	48	52	48	48	48	48	48	9	9	
10867	72	69	65	68	69	82	58	48	48	52	48	48	48	48	48	9	9	

10868 rows × 1000 columns

```
In [ ]: from sklearn.feature_selection import SelectPercentile,chi2
imp_fea = SelectPercentile(chi2,50)
new_asm_fea = imp_fea.fit_transform(x,y)
```

```
In [ ]: new_asm_fea.shape
```

Out[]: (10868, 500)

```
In [ ]: selected_fea = imp_fea.get_support(indices = True)
selected_fea.shape
```

Out[]: (500,)

```
In [ ]: selected_asm_fea = x.iloc[:,selected_fea]
```

In []:

selected_asm_fea

Out[]:

	asm_1	asm_3	asm_4	asm_14	asm_20	asm_21	asm_23	asm_24	asm_25	asm_26	asm_28	asm_29	asm_31	asm_32	asm_33	asm_34	asm_
0	69	68	69	48	9	9	13	10	72	69	68	69	58	49	48	48	
1	69	68	69	48	9	9	13	10	72	69	68	69	58	52	68	70	
2	69	68	69	48	9	9	13	10	72	69	68	69	58	49	48	48	
3	69	68	69	48	9	9	13	10	72	69	68	69	58	55	68	70	
4	69	68	69	48	9	9	13	10	72	69	68	69	58	49	48	48	
...	
10863	69	68	69	48	9	9	13	10	72	69	68	69	58	48	48	52	
10864	69	68	69	48	9	9	13	10	72	69	68	69	58	48	48	52	
10865	69	68	69	48	9	9	13	10	72	69	68	69	58	48	48	52	
10866	69	68	69	48	9	9	13	10	72	69	68	69	58	48	48	52	
10867	69	68	69	48	9	9	13	10	72	69	68	69	58	48	48	52	

10868 rows × 500 columns

◀

▶

```

In [ ]: def plot_confusion_matrix(test_y, predict_y):
    C = confusion_matrix(test_y, predict_y)
    print("Number of misclassified points ", (len(test_y)-np.trace(C))/len(test_y)*100)
    # C = 9,9 matrix, each cell (i,j) represents number of points of class i are predicted class j

    A = (((C.T)/(C.sum(axis=1))).T)
    #divid each element of the confusion matrix with the sum of elements in that column

    # C = [[1, 2],
    #      [3, 4]]
    # C.T = [[1, 3],
    #        [2, 4]]
    # C.sum(axis = 1)  axis=0 corresonds to columns and axis=1 corresponds to rows in two dimensional array
    # C.sum(axix =1) = [[3, 7]]
    # ((C.T)/(C.sum(axis=1))) = [[1/3, 3/7]
    #                             [2/3, 4/7]]

    # ((C.T)/(C.sum(axis=1))).T = [[1/3, 2/3]
    #                               [3/7, 4/7]]
    # sum of row elements = 1

    B = (C/C.sum(axis=0))
    #divid each element of the confusion matrix with the sum of elements in that row
    # C = [[1, 2],
    #      [3, 4]]
    # C.sum(axis = 0)  axis=0 corresonds to columns and axis=1 corresponds to rows in two dimensional array
    # C.sum(axix =0) = [[4, 6]]
    # (C/C.sum(axis=0)) = [[1/4, 2/6],
    #                       [3/4, 4/6]]

    labels = [1,2,3,4,5,6,7,8,9]
    cmap=sns.light_palette("green")
    # representing A in heatmap format
    print("-"*50, "Confusion matrix", "-"*50)
    plt.figure(figsize=(10,5))
    sns.heatmap(C, annot=True, cmap=cmap, fmt=".3f", xticklabels=labels, yticklabels=labels)
    plt.xlabel('Predicted Class')
    plt.ylabel('Original Class')
    plt.show()

    print("-"*50, "Precision matrix", "-"*50)
    plt.figure(figsize=(10,5))
    sns.heatmap(B, annot=True, cmap=cmap, fmt=".3f", xticklabels=labels, yticklabels=labels)
    plt.xlabel('Predicted Class')
    plt.ylabel('Original Class')

```

```

plt.show()
print("Sum of columns in precision matrix",B.sum(axis=0))

# representing B in heatmap format
print("-"*50, "Recall matrix"      , "-"*50)
plt.figure(figsize=(10,5))
sns.heatmap(A, annot=True, cmap=cmap, fmt=".3f", xticklabels=labels, yticklabels=labels)
plt.xlabel('Predicted Class')
plt.ylabel('Original Class')
plt.show()
print("Sum of rows in precision matrix",A.sum(axis=1))

```

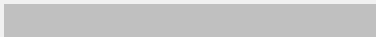
```
In [ ]: selected_asm_fea.to_csv("imp_asm_fea.csv",index=False)
```

```
In [ ]: asm_fea = pd.read_csv("/content/drive/My Drive/final_features/imp_asm_fea.csv")
asm_fea = asm_fea.drop("Unnamed: 0",axis=1)
asm_fea.head()
```

```
Out[ ]:
```

	asm_1	asm_3	asm_4	asm_14	asm_20	asm_21	asm_23	asm_24	asm_25	asm_26	asm_28	asm_29	asm_31	asm_32	asm_33	asm_34	asm_40
0	69	68	69	48	9	9	13	10	72	69	68	69	58	49	48	48	9
1	69	68	69	48	9	9	13	10	72	69	68	69	58	52	68	70	9
2	69	68	69	48	9	9	13	10	72	69	68	69	58	49	48	48	9
3	69	68	69	48	9	9	13	10	72	69	68	69	58	55	68	70	9
4	69	68	69	48	9	9	13	10	72	69	68	69	58	49	48	48	9

5 rows × 500 columns



```
In [ ]: asm_and_byte_fea = pd.concat([byte_feature_df,asm_fea],axis=1,join='inner')
asm_and_byte_fea
```

Out[]:

	00 00	ff ff	00 5a	00 50	00 e8	fe ff	ff 00	01 00	ff 52	00 5e	00 03	07 00	65 00	00 52	00 c0	51 6a	ff 8d	6a 00	56 57	65 73
0	33081.0	4877.0	81.0	210.0	471.0	757.0	1008.0	1118.0	44.0	78.0	404.0	146.0	111.0	74.0	199.0	6.0	136.0	289.0	131.0	32.0
1	24301.0	2396.0	8.0	210.0	88.0	96.0	561.0	565.0	9.0	11.0	127.0	70.0	280.0	38.0	82.0	2.0	83.0	102.0	55.0	43.0
2	13253.0	2291.0	148.0	932.0	1376.0	207.0	2966.0	575.0	65.0	116.0	108.0	185.0	187.0	117.0	488.0	23.0	325.0	552.0	19.0	22.0
3	2155.0	24.0	2.0	14.0	6.0	2.0	1.0	76.0	0.0	3.0	102.0	2.0	154.0	9.0	4.0	5.0	2.0	16.0	3.0	15.0
4	3814.0	43.0	4.0	4.0	14.0	2.0	2.0	37.0	1.0	2.0	2.0	2.0	11.0	6.0	2.0	13.0	0.0	11.0	1.0	4.0
...
10863	9.0	16.0	12.0	20.0	5.0	13.0	12.0	11.0	10.0	13.0	11.0	12.0	9.0	11.0	14.0	18.0	15.0	9.0	21.0	10.0
10864	64616.0	675564.0	62.0	146.0	93.0	94.0	1190.0	1299.0	22.0	28.0	239.0	120.0	245.0	73.0	98.0	7.0	35.0	152.0	210.0	457.0
10865	5652.0	75.0	33.0	23.0	38.0	15.0	284.0	84.0	19.0	17.0	78.0	41.0	26.0	21.0	11.0	12.0	30.0	36.0	14.0	16.0
10866	41031.0	1253.0	16.0	53.0	22.0	148.0	289.0	764.0	12.0	6.0	50.0	65.0	47.0	27.0	65.0	4.0	36.0	49.0	19.0	2.0
10867	8193.0	343.0	1824.0	1879.0	1840.0	57.0	0.0	2183.0	1.0	1.0	1.0	3.0	65.0	1823.0	2279.0	0.0	5.0	1390.0	0.0	69.0

10868 rows × 800 columns

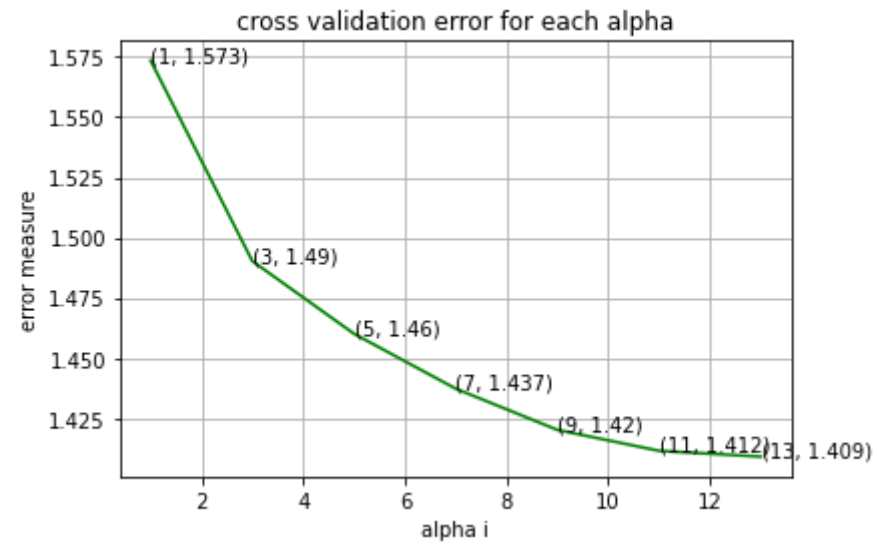
machine learning model on combined bytes and asm image feature

```
In [ ]: x_train,x_test,y_train,y_test = train_test_split(asm_and_byte_fea,y_data)
x_train,x_cv,y_train,y_cv= train_test_split(x_train,y_train)
```

K nearest neighbours

```
In [ ]: #KNN
alpha =[x for x in range(1,15,2)]
cv_log_error_array=[]
for i in alpha:
    knn = KNeighborsClassifier(n_neighbors=i)
    knn.fit(x_train,y_train)
    sig_clf = CalibratedClassifierCV(knn,method="sigmoid")
    sig_clf.fit(x_train,y_train)
    predict_y = sig_clf.predict_proba(x_cv)
    cv_log_error_array.append(log_loss(y_cv,predict_y,labels=knn.classes_,eps=1e-15))
for i in range(len(cv_log_error_array)):
    print("log_loss for k : ",alpha[i],"is",cv_log_error_array[i])
best_alpha = np.argmin(cv_log_error_array)
fig,ax = plt.subplots()
ax.plot(alpha,cv_log_error_array,c='g')
for i,txt in enumerate(np.round(cv_log_error_array,3)):
    ax.annotate((alpha[i],np.round(txt,3)),(alpha[i],cv_log_error_array[i]))
plt.grid()
plt.title("cross validation error for each alpha")
plt.xlabel("alpha i")
plt.ylabel("error measure")
plt.show()
```

log_loss for k : 1 is 1.5734498672275266
log_loss for k : 3 is 1.4902515203687345
log_loss for k : 5 is 1.4600857595550643
log_loss for k : 7 is 1.4373075593909381
log_loss for k : 9 is 1.4202006306571073
log_loss for k : 11 is 1.4115185506740884
log_loss for k : 13 is 1.4091009108628423



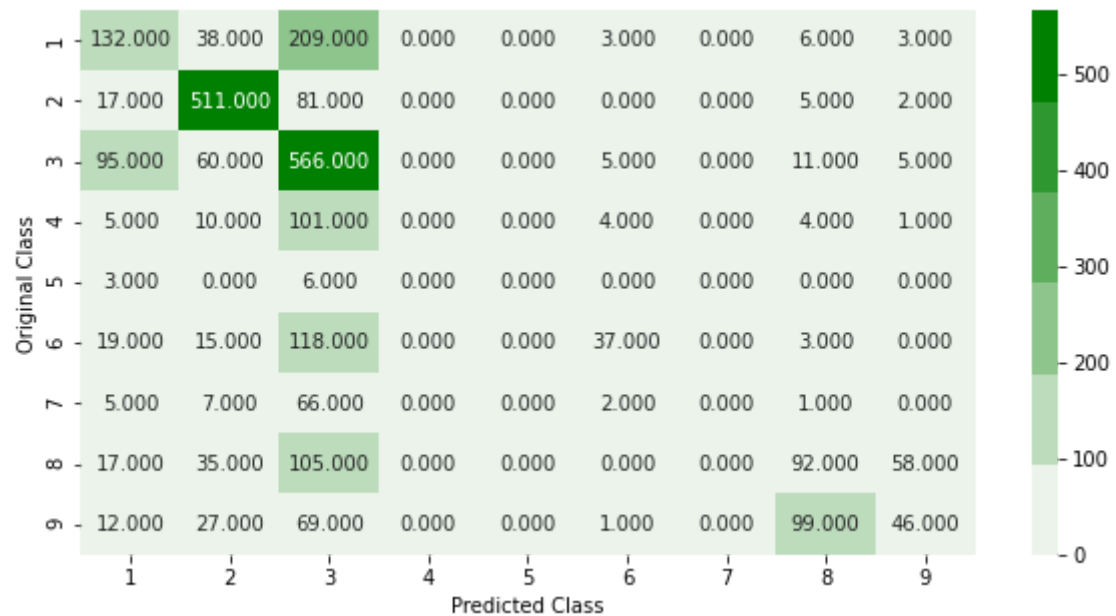
```
In [ ]: best_knn = KNeighborsClassifier(n_neighbors= alpha[best_alpha])
best_knn.fit(x_train,y_train)
sig_clf = CalibratedClassifierCV(best_knn,method="sigmoid")
sig_clf.fit(x_train,y_train)

predict_y = sig_clf.predict_proba(x_train)
print("for values of best alpha = ",alpha[best_alpha],"the train log loss is",log_loss(y_train,predict_y))
predict_y = sig_clf.predict_proba(x_cv)
print("for values of best alpha = ",alpha[best_alpha],"the cross validation log loss is",log_loss(y_cv,predict_y))
predict_y = sig_clf.predict_proba(x_test)
print("for values of best alpha = ",alpha[best_alpha],"the test log loss is",log_loss(y_test,predict_y))
plot_confusion_matrix(y_test,sig_clf.predict(x_test))
```


for values of best alpha = 13 the train log loss is 1.244030303583797
for values of best alpha = 13 the cross validation log loss is 1.4091009108628423
for values of best alpha = 13 the test log loss is 1.3945843768019437
Number of misclassified points 49.061464850938535

----- Confusion matrix -----

/usr/local/lib/python3.6/dist-packages/ipykernel_launcher.py:23: RuntimeWarning: invalid value encountered in true_divide

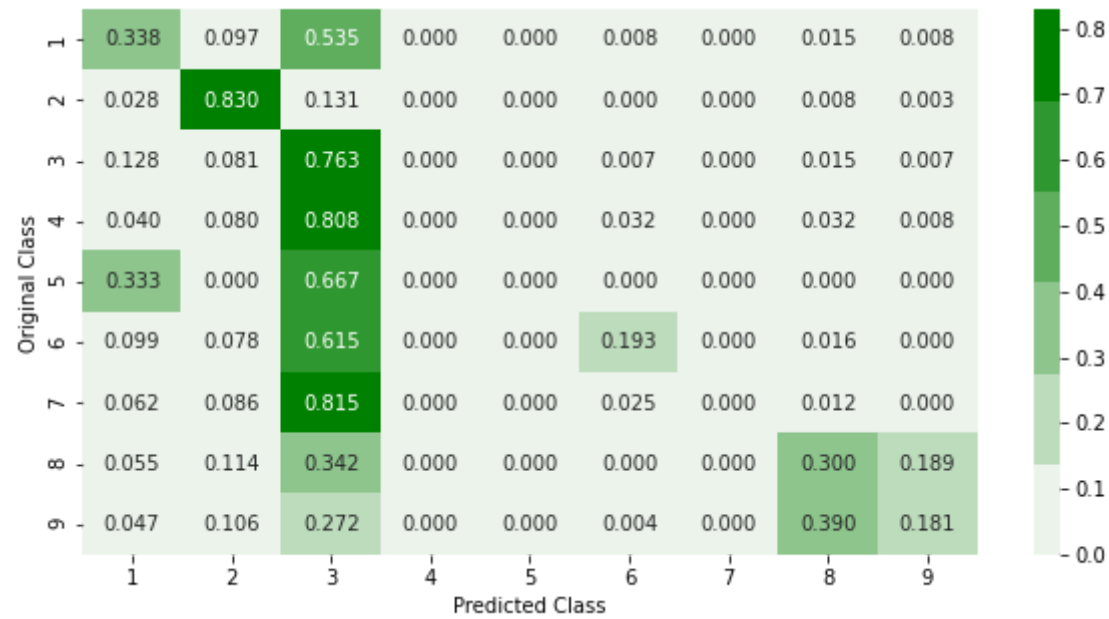


----- Precision matrix -----



Sum of columns in precision matrix [1. 1. 1. nan nan 1. nan 1. 1.]

----- Recall matrix -----



Sum of rows in precision matrix [1. 1. 1. 1. 1. 1. 1. 1. 1.]

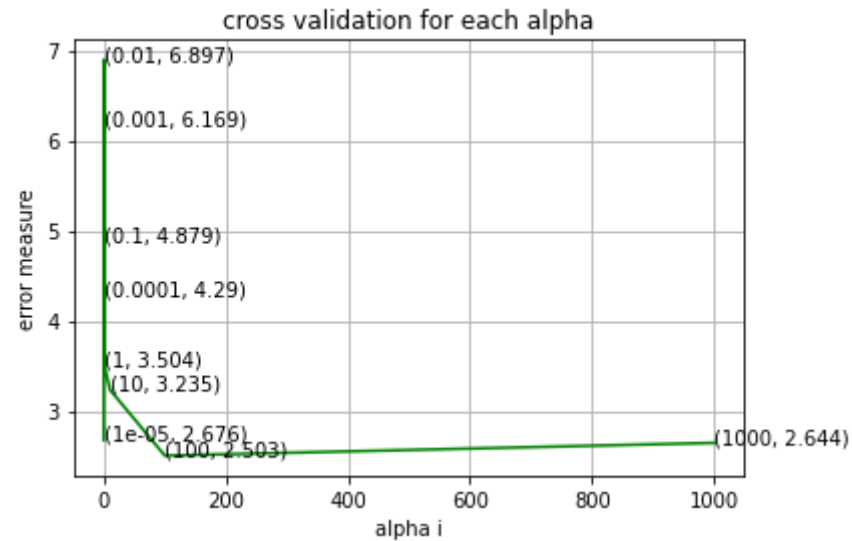

```

In [ ]: import warnings
warnings.filterwarnings("ignore")
alpha = [10**x for x in range(-5,4)]
cv_log_error_array = []
x_train_logis = preprocessing.scale(x_train)
for i in alpha:
    logis = LogisticRegression(penalty='l2',C=i,class_weight='balanced',max_iter = 200)
    logis.fit(x_train_logis,y_train)
    sig_clf = CalibratedClassifierCV(logis,method="sigmoid")
    sig_clf.fit(x_train_logis,y_train)
    predict_y = sig_clf.predict_proba(x_cv)
    cv_log_error_array.append(log_loss(y_cv,predict_y,labels=logis.classes_))
for i in range(len(cv_log_error_array)):
    print("logg loss for c = ",alpha[i],"is",cv_log_error_array[i])
best_alpha = np.argmin(cv_log_error_array)

fig,ax = plt.subplots()
ax.plot(alpha,cv_log_error_array,c="g")
for i,txt in enumerate(np.round(cv_log_error_array,3)):
    ax.annotate((alpha[i],np.round(txt,3)),(alpha[i],cv_log_error_array[i]))
plt.grid()
plt.title("cross validation for each alpha")
plt.xlabel("alpha i")
plt.ylabel("error measure")
plt.show()

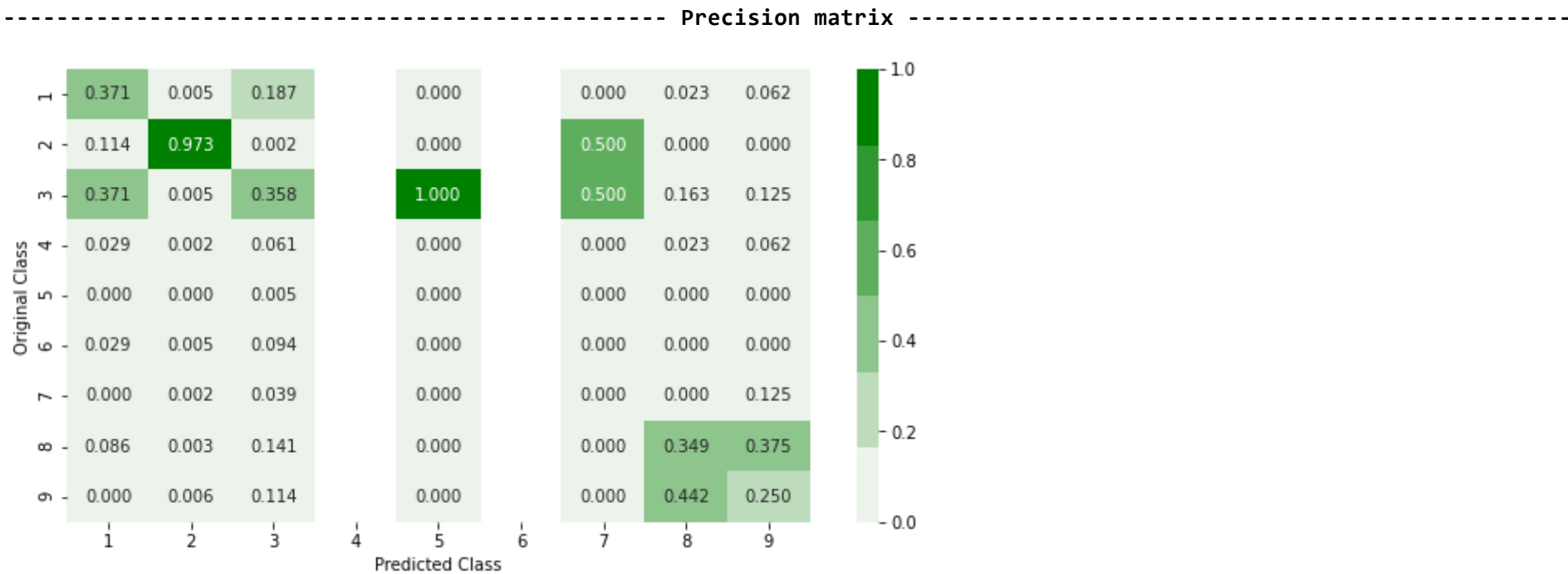
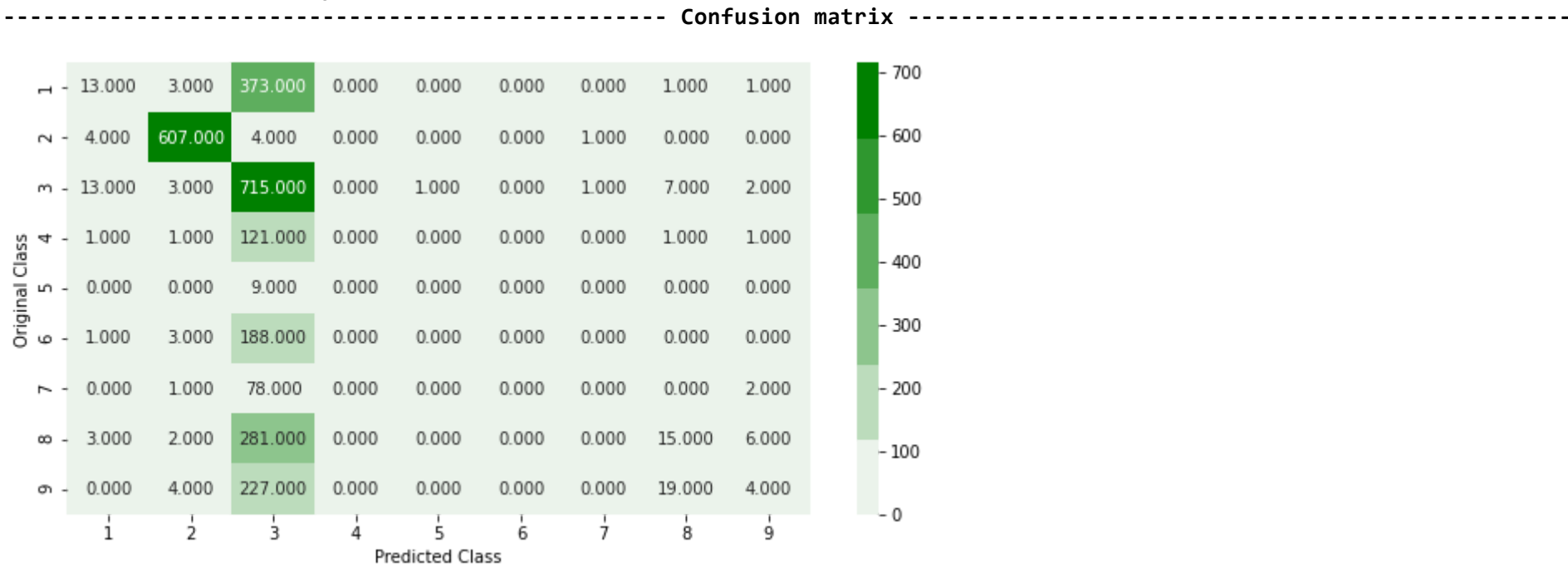
```

logg loss for c = 1e-05 is 2.6764015532691134
logg loss for c = 0.0001 is 4.289540363109524
logg loss for c = 0.001 is 6.169453638438878
logg loss for c = 0.01 is 6.8967886701578065
logg loss for c = 0.1 is 4.879319097785262
logg loss for c = 1 is 3.5039999830670707
logg loss for c = 10 is 3.234563220381588
logg loss for c = 100 is 2.502539053708165
logg loss for c = 1000 is 2.6441868743220005

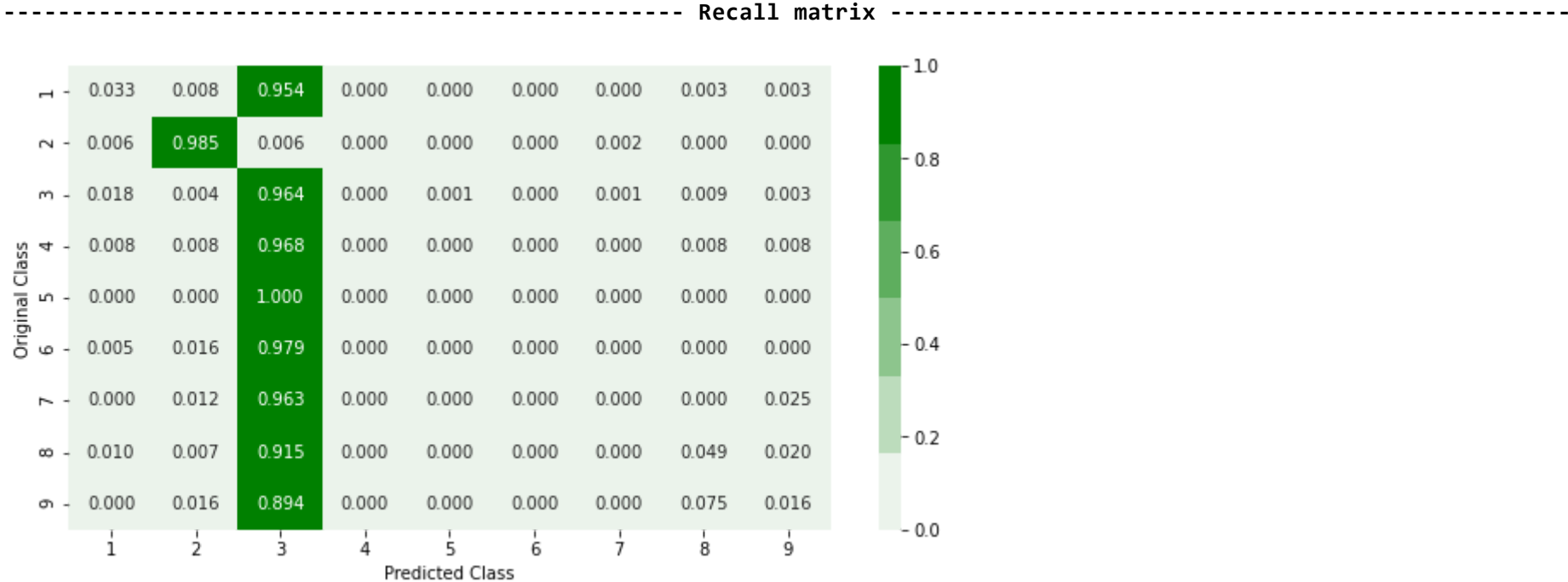


```
In [ ]: best_logist = LogisticRegression(penalty="l2",C=alpha[best_alpha],class_weight="balanced")
best_logist.fit(x_train,y_train)
sig_clf = CalibratedClassifierCV(best_logist,method="sigmoid")
sig_clf.fit(x_train,y_train)
predict_y = sig_clf.predict_proba(x_train)
print("log loss for train data",log_loss(y_train,predict_y,labels =best_logist.classes_))
predict_y = sig_clf.predict_proba(x_cv)
print("log loss for train data",log_loss(y_cv,predict_y,labels = best_logist.classes_))
predict_y = sig_clf.predict_proba(x_test)
print("log loss for test data ",log_loss(y_test,predict_y,labels = best_logist.classes_))
plot_confusion_matrix(y_test,sig_clf.predict(x_test))
```

log loss for train data 1.3302972904841353
log loss for train data 1.3637987266905214
log loss for test data 1.3534189307075528
Number of misclassified points 50.165623849834375



Sum of columns in precision matrix [1. 1. 1. nan 1. nan 1. 1. 1.]

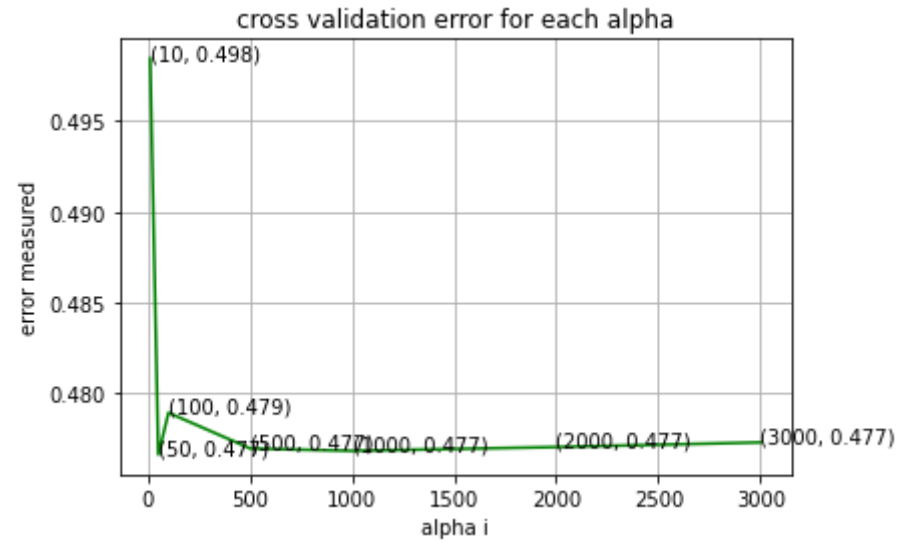


Sum of rows in precision matrix [1. 1. 1. 1. 1. 1. 1. 1. 1.]

Random Forest Classifier

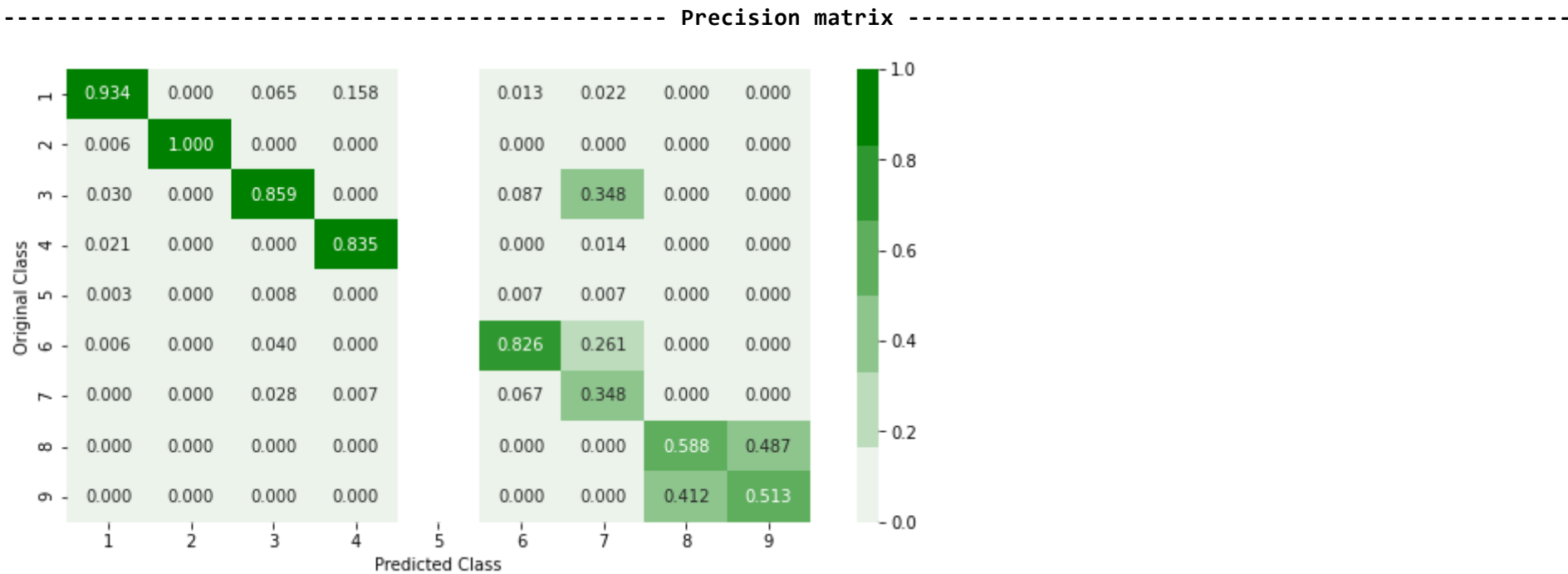
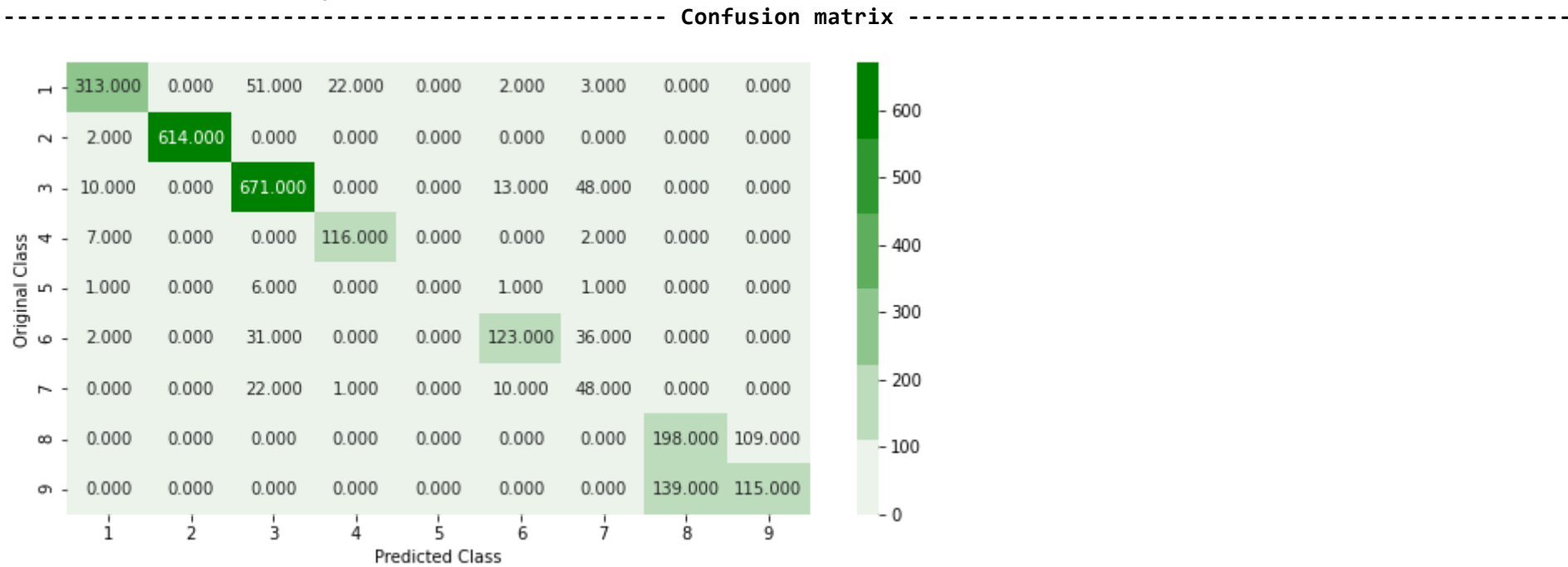

```
In [ ]: alpha=[10,50,100,500,1000,2000,3000]
cv_log_error_array =[]
train_log_error_array=[]
for i in alpha :
    rf = RandomForestClassifier(n_estimators=i,random_state=42,n_jobs=-1)
    rf.fit(x_train,y_train)
    sig_clf = CalibratedClassifierCV(rf,method="sigmoid")
    sig_clf.fit(x_train,y_train)
    predict_y = sig_clf.predict_proba(x_cv)
    cv_log_error_array.append(log_loss(y_cv,predict_y,labels = rf.classes_,eps=1e-15))
for i in range(len(cv_log_error_array)):
    print("log loss for c= ",alpha[i],cv_log_error_array[i])
best_alpha = np.argmin(cv_log_error_array)
fig,ax = plt.subplots()
ax.plot(alpha,cv_log_error_array,c='g')
for i,txt in enumerate(np.round(cv_log_error_array,3)):
    ax.annotate((alpha[i],np.round(txt,3)),(alpha[i],cv_log_error_array[i]))
plt.grid()
plt.title("cross validation error for each alpha")
plt.xlabel("alpha i")
plt.ylabel("error measured")
plt.show()
```

log loss for c= 10 0.49843278158492793
log loss for c= 50 0.476646444975307
log loss for c= 100 0.4789536996052057
log loss for c= 500 0.4769865275294863
log loss for c= 1000 0.4768550707385527
log loss for c= 2000 0.477064897358575
log loss for c= 3000 0.4773188584055862

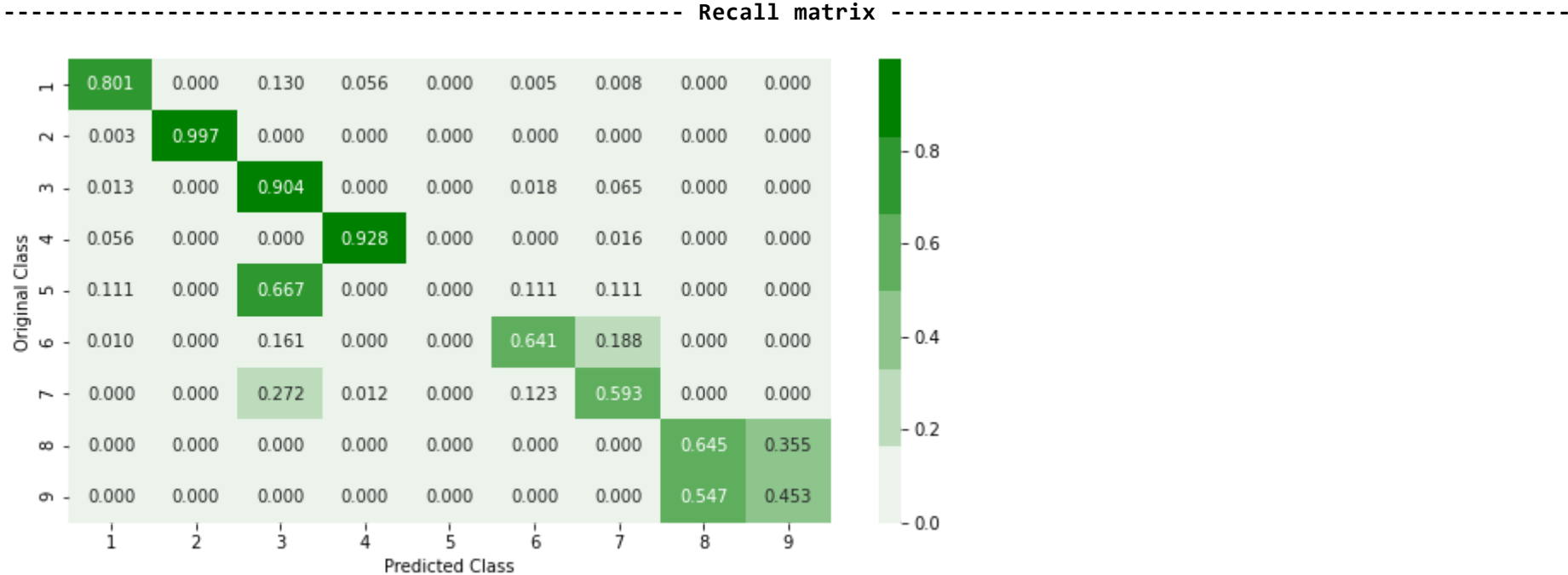


```
In [ ]: best_rf = RandomForestClassifier(n_estimators=alpha[best_alpha],random_state=42,n_jobs=-1)
best_rf.fit(x_train,y_train)
sig_clf=CalibratedClassifierCV(best_rf,method="sigmoid")
sig_clf.fit(x_train,y_train)
predict_y = sig_clf.predict_proba(x_train)
print("for values of best alpha = ",alpha[best_alpha],"the train log loss is",log_loss(y_train,predict_y,labels=best_rf.classes_))
predict_y = sig_clf.predict_proba(x_cv)
print("for values of best alpha = ",alpha[best_alpha],"the cross validation log loss is",log_loss(y_cv,predict_y,labels = best_rf.classes_))
predict_y = sig_clf.predict_proba(x_test)
print("for values of best alpha = ",alpha[best_alpha],"the test log loss is",log_loss(y_test,predict_y,labels = best_rf.classes_))
plot_confusion_matrix(y_test,sig_clf.predict(x_test))
```

for values of best alpha = 50 the train log loss is 0.1421860438461342
for values of best alpha = 50 the cross validation log loss is 0.476646444975307
for values of best alpha = 50 the test log loss is 0.45037533669333607
Number of misclassified points 19.10195068089805



Sum of columns in precision matrix [1. 1. 1. 1. nan 1. 1. 1. 1.]



Sum of rows in precision matrix [1. 1. 1. 1. 1. 1. 1. 1. 1.]

XGboost classifier

```
In [ ]: alpha=[10,50,100,500,1000,2000]
cv_log_error_array=[]
for i in alpha:
    xg = XGBClassifier(n_estimators=i,nthreds=-1)
    xg.fit(x_train,y_train)
    sig_clf = CalibratedClassifierCV(xg,method="sigmoid")
    sig_clf.fit(x_train,y_train)
    print("completed for alpha value :",i)
    predict_y = sig_clf.predict_proba(x_cv)
    cv_log_error_array.append(log_loss(y_cv,predict_y,labels=xg.classes_,eps=1e-15))
for i in range(len(cv_log_error_array)):
    print("log loss for c ",alpha[i],"is",cv_log_error_array[i])
best_alpha = np.argmin(cv_log_error_array)
```

```
completed for alpha value : 10
completed for alpha value : 50
completed for alpha value : 100
completed for alpha value : 500
completed for alpha value : 1000
completed for alpha value : 2000
log loss for c 10 is 0.4384181554906904
log loss for c 50 is 0.42283782967620637
log loss for c 100 is 0.4461897007239115
log loss for c 500 is 0.544324566111328
log loss for c 1000 is 0.5831719324409649
log loss for c 2000 is 0.6038912186429386
```

```
In [ ]: best_xg = XGBClassifier(n_estimators = alpha[best_alpha],nthreads=-1)
best_xg.fit(x_train,y_train)
sig_clf = CalibratedClassifierCV(best_xg,method="sigmoid")
sig_clf.fit(x_train,y_train)
predict_y = sig_clf.predict_proba(x_train)
print("for values of best alpha = ",alpha[best_alpha],"the train log loss is",log_loss(y_train,predict_y,labels=best_xg.classes_))
predict_y = sig_clf.predict_proba(x_cv)
print("for values of best alpha = ",alpha[best_alpha],"the cross validation log loss is",log_loss(y_cv,predict_y,labels = best_xg.classes_))
predict_y = sig_clf.predict_proba(x_test)
print("for values of best alpha = ",alpha[best_alpha],"the test log loss is",log_loss(y_test,predict_y,labels = best_xg.classes_))
plot_confusion_matrix(y_test,sig_clf.predict(x_test))
```

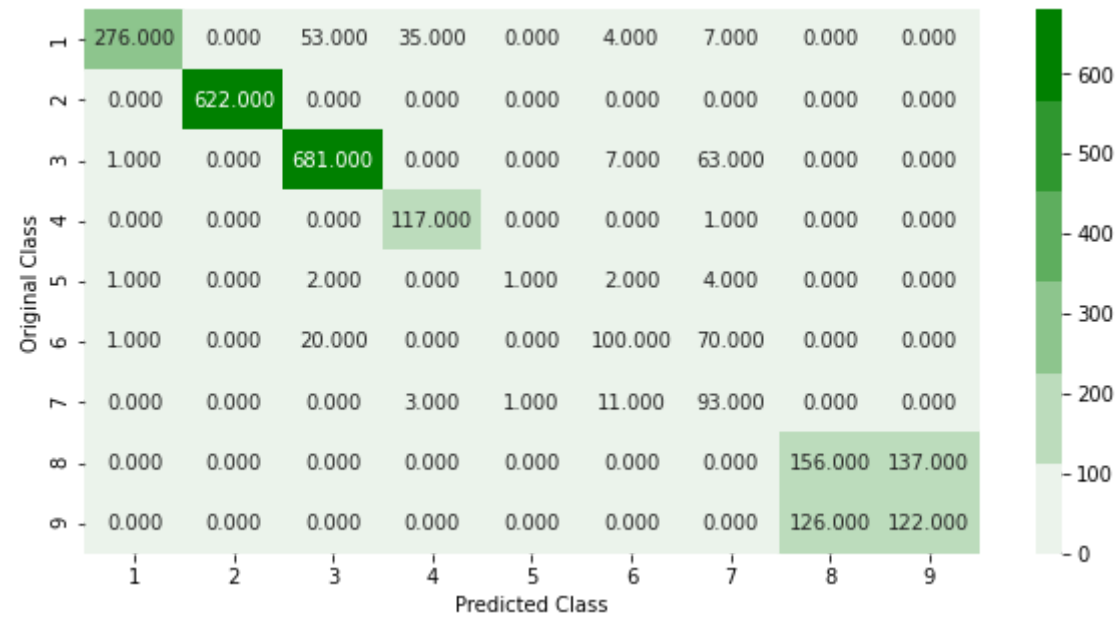
for values of best alpha = 50 the train log loss is 0.34265836146349005

for values of best alpha = 50 the cross validation log loss is 0.42283782967620637

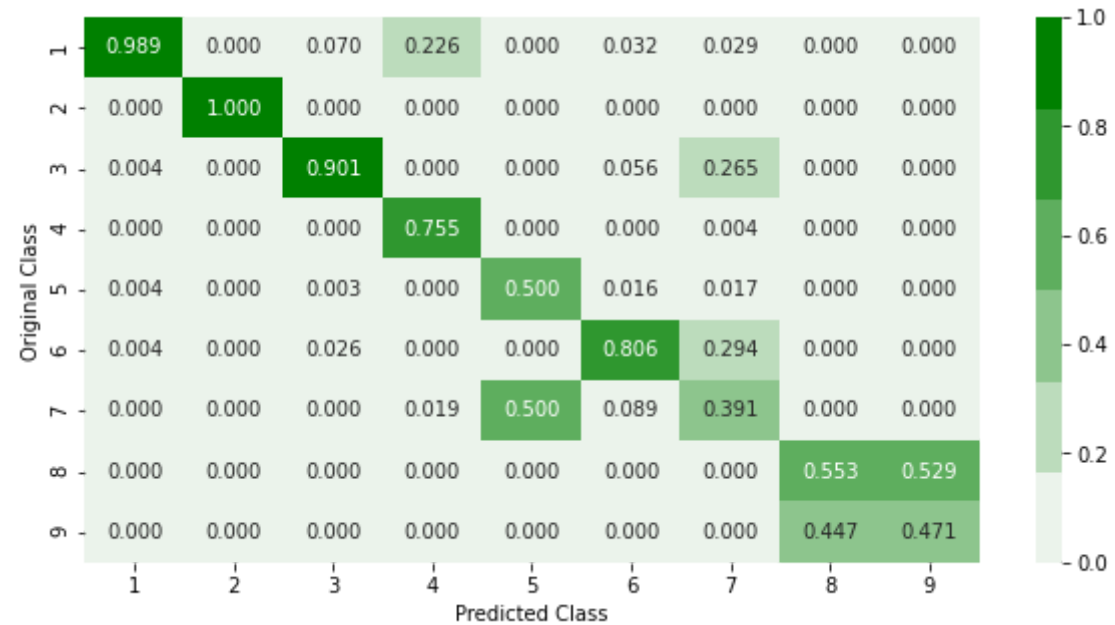
for values of best alpha = 50 the test log loss is 0.4239023897272152

Number of misclassified points 20.20610967979389

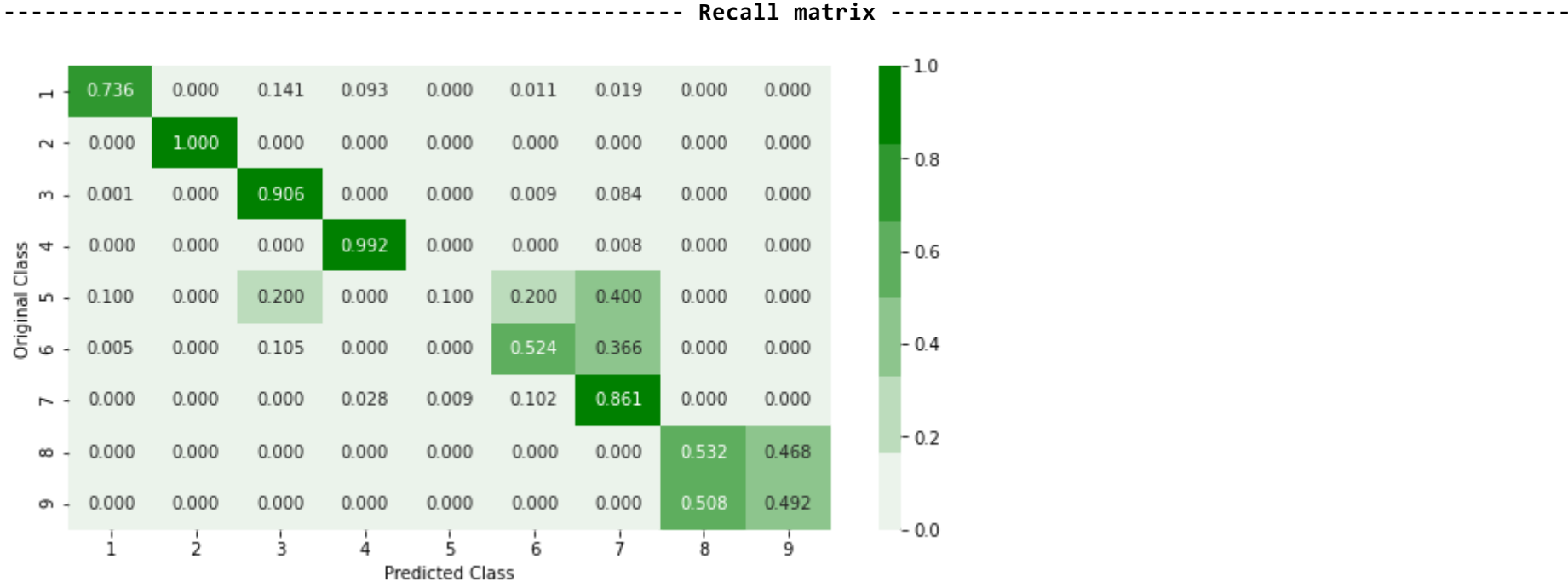
Confusion matrix



Precision matrix



Sum of columns in precision matrix [1. 1. 1. 1. 1. 1. 1. 1. 1.]



Sum of rows in precision matrix [1. 1. 1. 1. 1. 1. 1. 1. 1.]

XGBoost with hyper parameter tuning using random search

```
In [ ]: xgboost= XGBClassifier()
params ={
    'learning_rate':[0.01,0.03,0.05,0.1,0.15,0.2],
    'n_estimators' :[100,200,500,1000,2000],
    'max_depth':[3,5,10],
    'colsample_bytree':[0.1,0.3,0.5,1],
    'subsample':[0.1,0.3,0.5,1]
}
random_cfl1=RandomizedSearchCV(xgboost,param_distributions = params,verbose=10,n_jobs=-1)
random_cfl1.fit(x_train,y_train)
```

Fitting 5 folds for each of 10 candidates, totalling 50 fits

```
[Parallel(n_jobs=-1)]: Using backend LokyBackend with 4 concurrent workers.
[Parallel(n_jobs=-1)]: Done   5 tasks      | elapsed:   8.9min
[Parallel(n_jobs=-1)]: Done  10 tasks      | elapsed:  22.3min
[Parallel(n_jobs=-1)]: Done  17 tasks      | elapsed:  28.3min
[Parallel(n_jobs=-1)]: Done  24 tasks      | elapsed:  29.8min
[Parallel(n_jobs=-1)]: Done  33 tasks      | elapsed:  41.9min
[Parallel(n_jobs=-1)]: Done  42 tasks      | elapsed:  75.4min
[Parallel(n_jobs=-1)]: Done  50 out of  50 | elapsed:  96.0min finished
```

```
Out[ ]: RandomizedSearchCV(cv=None, error_score=nan,
                          estimator=XGBClassifier(base_score=0.5, booster='gbtree',
                                                  colsample_bylevel=1,
                                                  colsample_bynode=1,
                                                  colsample_bytree=1, gamma=0,
                                                  learning_rate=0.1, max_delta_step=0,
                                                  max_depth=3, min_child_weight=1,
                                                  missing=None, n_estimators=100,
                                                  n_jobs=1, nthread=None,
                                                  objective='binary:logistic',
                                                  random_state=0, reg_alpha=0,
                                                  reg_lambda=1...,
                                                  seed=None, silent=None, subsample=1,
                                                  verbosity=1),
                          iid='deprecated', n_iter=10, n_jobs=-1,
                          param_distributions={'colsample_bytree': [0.1, 0.3, 0.5, 1],
                                              'learning_rate': [0.01, 0.03, 0.05, 0.1,
                                                              0.15, 0.2],
                                              'max_depth': [3, 5, 10],
                                              'n_estimators': [100, 200, 500, 1000,
                                                              2000],
                                              'subsample': [0.1, 0.3, 0.5, 1]},
                          pre_dispatch='2*n_jobs', random_state=None, refit=True,
                          return_train_score=False, scoring=None, verbose=10)
```

```
In [ ]: print(random_cfl1.best_params_)
```

```
{'subsample': 0.3, 'n_estimators': 1000, 'max_depth': 3, 'learning_rate': 0.01, 'colsample_bytree': 0.3}
```

```
In [ ]: best_xgb = XGBClassifier(n_estimators=1000,learning_rate = 0.01,colsample_bytree=0.3,max_depth=3,subsample=0.3)
best_xgb.fit(x_train,y_train)
clf = CalibratedClassifierCV(best_xgb,method="sigmoid")
clf.fit(x_train,y_train)
predict_y = clf.predict_proba(x_train)
print("the train log loss is",log_loss(y_train,predict_y,labels=best_xgb.classes_))
predict_y = clf.predict_proba(x_cv)
print("the cross validation log loss is",log_loss(y_cv,predict_y,labels = best_xgb.classes_))
predict_y = clf.predict_proba(x_test)
print("the test log loss is",log_loss(y_test,predict_y,labels = best_xgb.classes_))
```

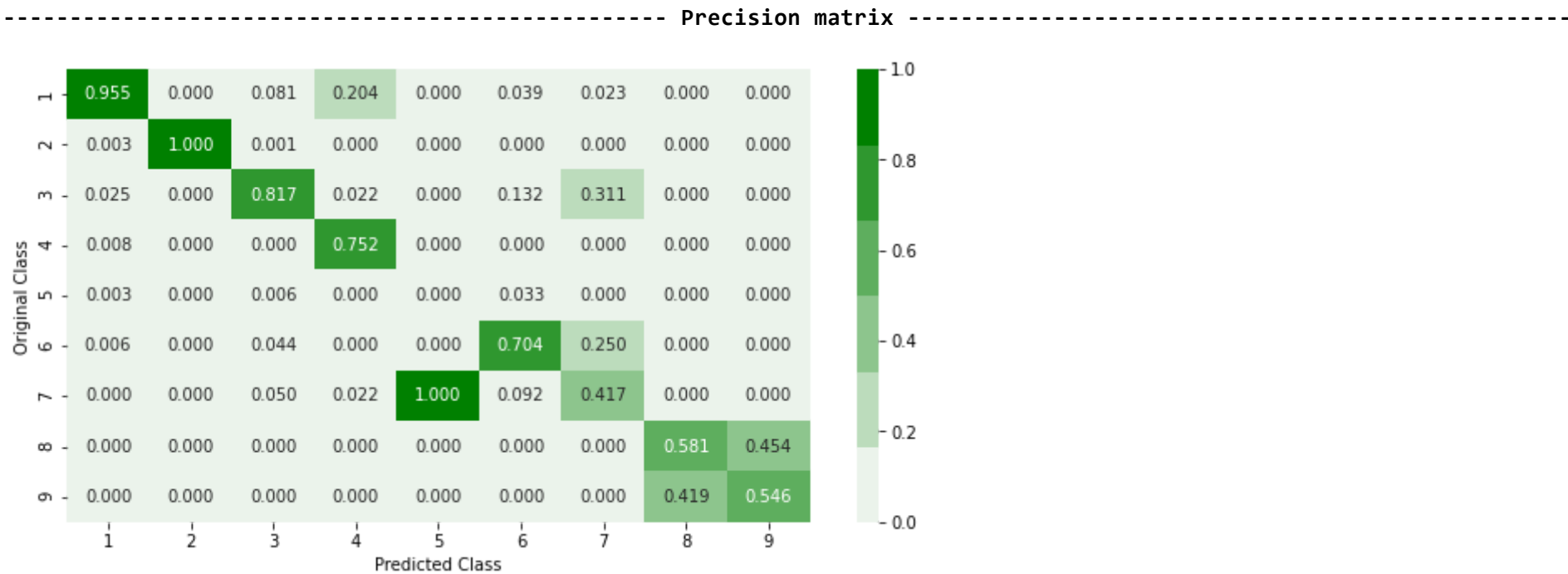
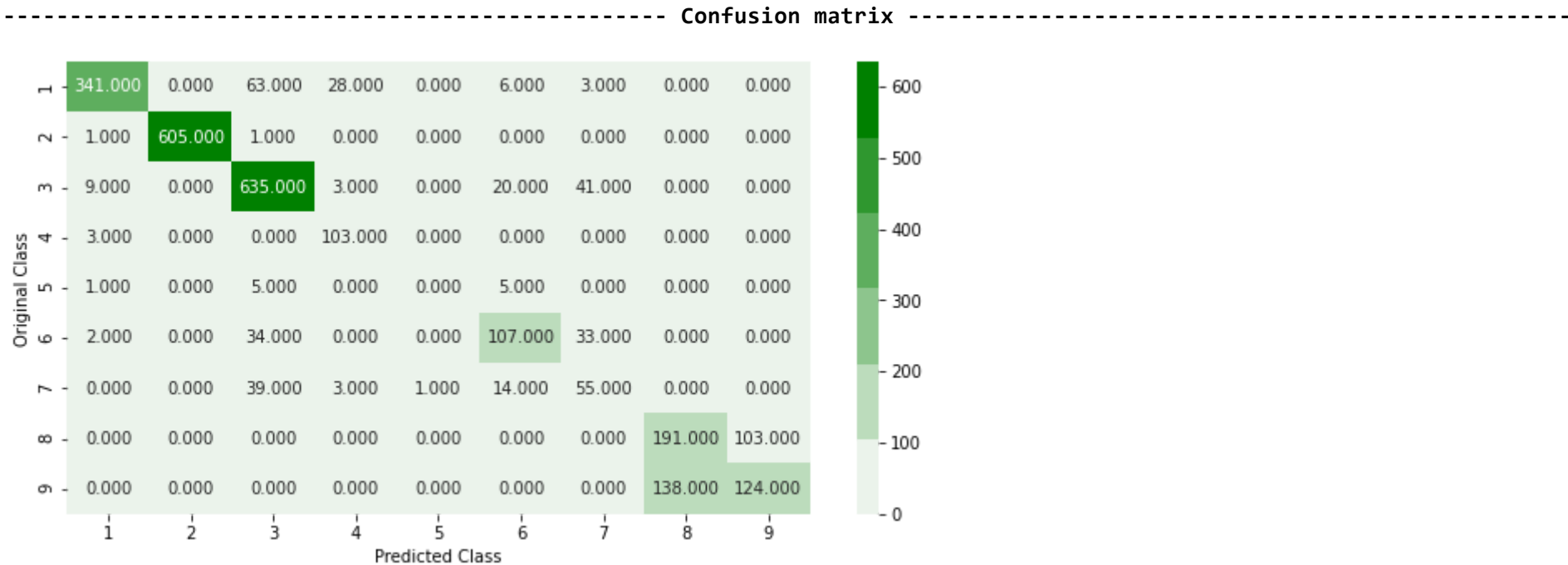
```
the train log loss is 0.2557203719765019
```

```
the cross validation log loss is 0.44861480225105205
```

```
the test log loss is 0.4862796487185741
```

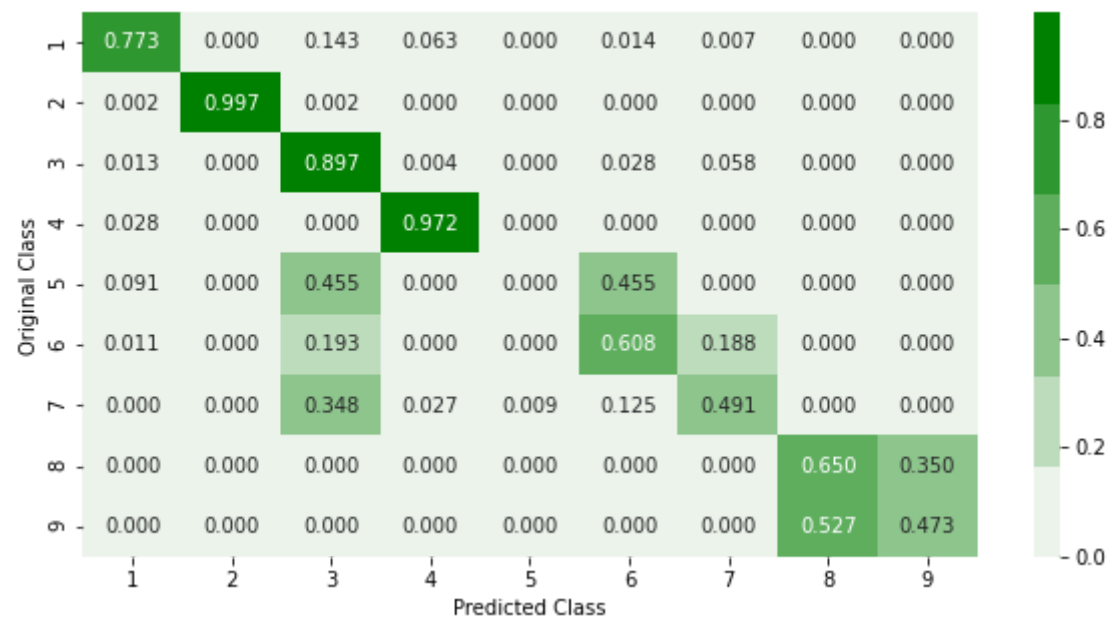
```
In [ ]: plot_confusion_matrix(y_test,clf.predict(x_test))
```

Number of misclassified points 20.463746779536255



Sum of columns in precision matrix [1. 1. 1. 1. 1. 1. 1. 1. 1.]

----- Recall matrix -----



Sum of rows in precision matrix [1. 1. 1. 1. 1. 1. 1. 1. 1.]

Byte file size as feature

```
In [ ]: size_file = pd.read_csv("/content/drive/My Drive/Microsoft_Malware_Detection_Files/result_with_size.csv")
size_file.head()
```

```
Out[ ]:
```

	Unnamed: 0	ID	0	1	2	3	4	5	6	7	8	9	0a	0b	0c	0d	0e	0f	10	11
0	0	01azqd4lnC7m9JpocGv5	601905	3905	2816	3832	3345	3242	3650	3201	2965	3205	3211	3546	4038	4096	3218	3032	3269	2740
1	1	01lsoiSMh5gxyDYTI4CB	39755	8337	7249	7186	8663	6844	8420	7589	9291	358	340	6649	8660	447	218	6869	8869	228
2	2	01jsnpXSAlgW6aPeDxrU	93506	9542	2568	2438	8925	9330	9007	2342	9107	2457	2655	2669	9113	2584	2788	2487	2782	2611
3	3	01kcPWA9K2BOxQeS5Rju	21091	1213	726	817	1257	625	550	523	1078	473	516	445	808	432	403	705	1067	407
4	4	01SuzwMJEIXsK7A8dQbl	19764	710	302	433	559	410	262	249	422	223	237	226	406	643	213	272	447	242

5 rows × 261 columns

```
In [ ]: size_file.rename(columns={"size": "byte_size"}, inplace=True)
size_file = size_file["byte_size"]
size_file
```

```
Out[ ]: 0          4.234863
1          5.538818
2          3.887939
3          0.574219
4          0.370850
...
10863      1.878174
10864      0.215332
10865      0.215332
10866      0.550293
10867      0.502441
Name: byte_size, Length: 10868, dtype: float64
```

asm file size as feature

```
In [ ]: asm_size = pd.read_csv("/content/drive/My Drive/Microsoft_Malware_Detection_Files/asm_with_size.csv")
asm_size.head()
```

```
Out[ ]:
```

	Unnamed: 0	ID	size_asm	Class
0	0	01azqd4InC7m9JpocGv5	56.229886	9
1	1	01IsoiSMh5gxyDYTI4CB	13.999378	2
2	2	01jsnpXSAlgW6aPeDxrU	8.507785	9
3	3	01kcPWA9K2BOxQeS5Rju	0.078190	1
4	4	01SuzwMJEIXsK7A8dQbl	0.996723	8

```
In [ ]: asm_size = asm_size['size_asm']
```

```
In [ ]: advanced_fea1 = pd.concat([byte_feature_df,asm_fea,size_file,asm_size],axis=1,ignore_index=False)
advanced_fea1
```

Out[]:

	00 00	ff ff	00 5a	00 50	00 e8	fe ff	ff 00	01 00	ff 52	00 5e	00 03	07 00	65 00	00 52	00 c0	51 6a	ff 8d	6a 00	56 57	65 73
0	33081.0	4877.0	81.0	210.0	471.0	757.0	1008.0	1118.0	44.0	78.0	404.0	146.0	111.0	74.0	199.0	6.0	136.0	289.0	131.0	32.0
1	24301.0	2396.0	8.0	210.0	88.0	96.0	561.0	565.0	9.0	11.0	127.0	70.0	280.0	38.0	82.0	2.0	83.0	102.0	55.0	43.0
2	13253.0	2291.0	148.0	932.0	1376.0	207.0	2966.0	575.0	65.0	116.0	108.0	185.0	187.0	117.0	488.0	23.0	325.0	552.0	19.0	22.0
3	2155.0	24.0	2.0	14.0	6.0	2.0	1.0	76.0	0.0	3.0	102.0	2.0	154.0	9.0	4.0	5.0	2.0	16.0	3.0	15.0
4	3814.0	43.0	4.0	4.0	14.0	2.0	2.0	37.0	1.0	2.0	2.0	2.0	11.0	6.0	2.0	13.0	0.0	11.0	1.0	4.0
...
10863	9.0	16.0	12.0	20.0	5.0	13.0	12.0	11.0	10.0	13.0	11.0	12.0	9.0	11.0	14.0	18.0	15.0	9.0	21.0	10.0
10864	64616.0	675564.0	62.0	146.0	93.0	94.0	1190.0	1299.0	22.0	28.0	239.0	120.0	245.0	73.0	98.0	7.0	35.0	152.0	210.0	457.0
10865	5652.0	75.0	33.0	23.0	38.0	15.0	284.0	84.0	19.0	17.0	78.0	41.0	26.0	21.0	11.0	12.0	30.0	36.0	14.0	16.0
10866	41031.0	1253.0	16.0	53.0	22.0	148.0	289.0	764.0	12.0	6.0	50.0	65.0	47.0	27.0	65.0	4.0	36.0	49.0	19.0	2.0
10867	8193.0	343.0	1824.0	1879.0	1840.0	57.0	0.0	2183.0	1.0	1.0	1.0	3.0	65.0	1823.0	2279.0	0.0	5.0	1390.0	0.0	69.0

10868 rows × 802 columns

byte_feature+asm_image_feature+asm_file_size+byte_file_size

```
In [ ]: x_train,x_test,y_train,y_test = train_test_split(advanced_fea1,y_data)
x_train,x_cv,y_train,y_cv= train_test_split(x_train,y_train)
```

XGboost classifier


```
In [ ]: alpha=[10,50,100,500,1000,2000]
cv_log_error_array=[]
for i in alpha:
    xg = XGBClassifier(n_estimators=i,nthreads=-1)
    xg.fit(x_train,y_train)
    sig_clf = CalibratedClassifierCV(xg,method="sigmoid")
    sig_clf.fit(x_train,y_train)
    print("completed for alpha value :",i)
    predict_y = sig_clf.predict_proba(x_cv)
    cv_log_error_array.append(log_loss(y_cv,predict_y,labels=xg.classes_,eps=1e-15))
for i in range(len(cv_log_error_array)):
    print("log loss for c ",alpha[i],"is",cv_log_error_array[i])
best_alpha = np.argmin(cv_log_error_array)
```

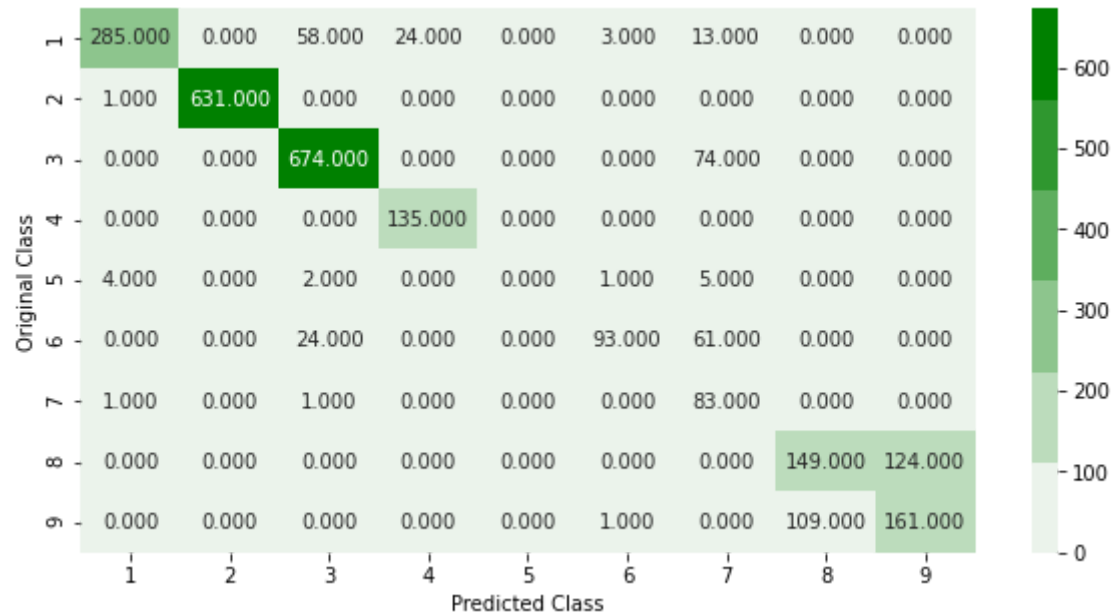
```
completed for alpha value : 10
completed for alpha value : 50
completed for alpha value : 100
completed for alpha value : 500
completed for alpha value : 1000
completed for alpha value : 2000
log loss for c 10 is 0.45911594556733726
log loss for c 50 is 0.45911594556733726
log loss for c 100 is 0.45911594556733726
log loss for c 500 is 0.45911594556733726
log loss for c 1000 is 0.45911594556733726
log loss for c 2000 is 0.45911594556733726
```

```
In [ ]: best_xg = XGBClassifier(n_estimators = alpha[best_alpha],nthreads=-1)
best_xg.fit(x_train,y_train)
sig_clf = CalibratedClassifierCV(best_xg,method="sigmoid")
sig_clf.fit(x_train,y_train)
predict_y = sig_clf.predict_proba(x_train)
print("for values of best alpha = ",alpha[best_alpha],"the train log loss is",log_loss(y_train,predict_y,labels=best_xg.classes_))
predict_y = sig_clf.predict_proba(x_cv)
print("for values of best alpha = ",alpha[best_alpha],"the cross validation log loss is",log_loss(y_cv,predict_y,labels = best_xg.classes_))
predict_y = sig_clf.predict_proba(x_test)
print("for values of best alpha = ",alpha[best_alpha],"the test log loss is",log_loss(y_test,predict_y,labels = best_xg.classes_))
plot_confusion_matrix(y_test,sig_clf.predict(x_test))
```

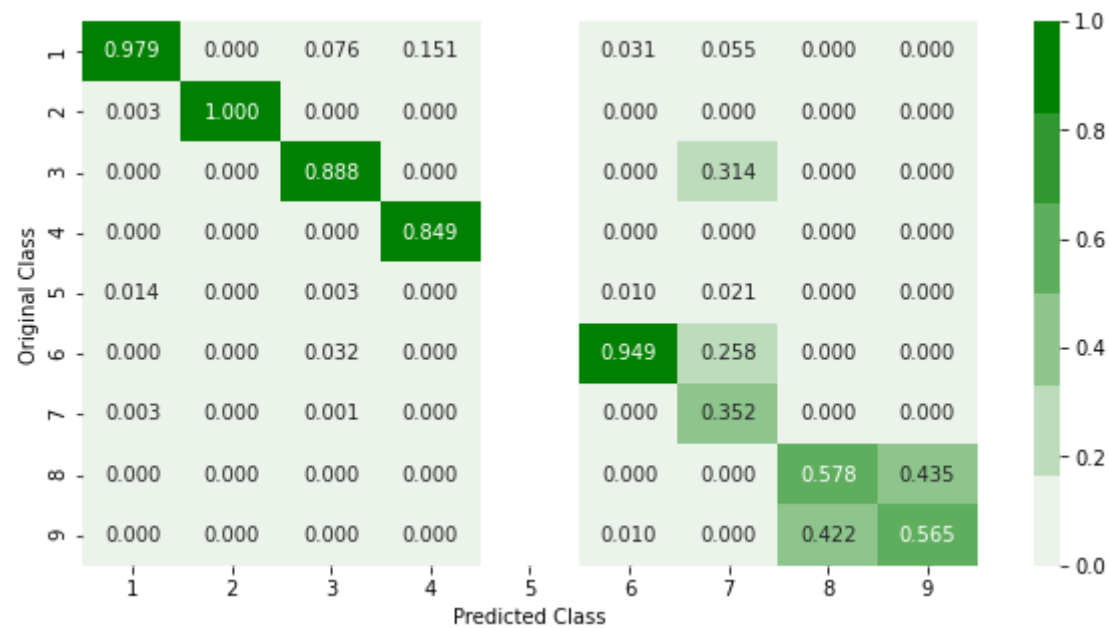
for values of best alpha = 10 the train log loss is 0.39718313782370285
for values of best alpha = 10 the cross validation log loss is 0.4311096817100026
for values of best alpha = 10 the test log loss is 0.4260286331562497
Number of misclassified points 18.62348178137652

----- Confusion matrix -----

/usr/local/lib/python3.6/dist-packages/ipykernel_launcher.py:23: RuntimeWarning: invalid value encountered in true_divide

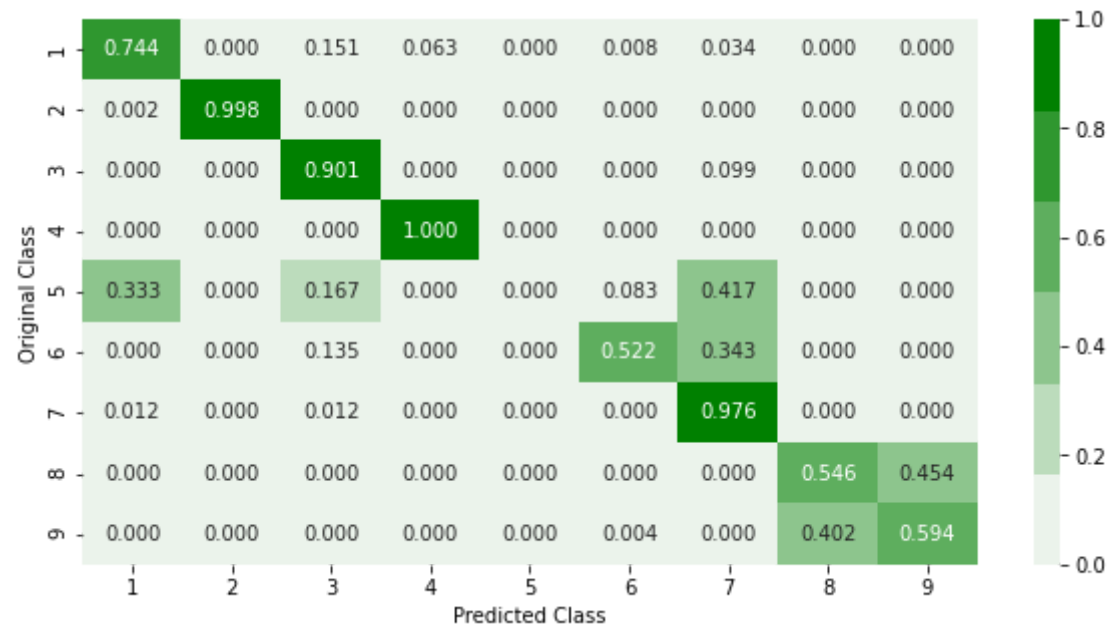


----- Precision matrix -----



Sum of columns in precision matrix [1. 1. 1. 1. nan 1. 1. 1. 1.]

----- Recall matrix



Sum of rows in precision matrix [1. 1. 1. 1. 1. 1. 1. 1. 1.]

XGBoost with hyper parameter tuning using random search

```
In [ ]: xgboost= XGBClassifier()
params ={
    'learning_rate':[0.01,0.03,0.05,0.1,0.15,0.2],
    'n_estimators' :[100,200,500,1000,2000],
    'max_depth':[3,5,10],
    'colsample_bytree':[0.1,0.3,0.5,1],
    'subsample':[0.1,0.3,0.5,1]
}
random_cfl1=RandomizedSearchCV(xgboost,param_distributions = params,verbose=10,n_jobs=-1)
random_cfl1.fit(x_train,y_train)
```

Fitting 5 folds for each of 10 candidates, totalling 50 fits

[Parallel(n_jobs=-1)]: Using backend LokyBackend with 4 concurrent workers.

[Parallel(n_jobs=-1)]: Done 5 tasks | elapsed: 20.2min

[Parallel(n_jobs=-1)]: Done 10 tasks | elapsed: 21.3min

[Parallel(n_jobs=-1)]: Done 17 tasks | elapsed: 33.0min

```
In [ ]: print(random_cfl1.best_params_)
```

```
In [ ]: best_xgb = XGBClassifier(n_estimators=1000,learning_rate = 0.01,colsample_bytree=1,max_depth=5,subsample=1)
best_xgb.fit(x_train,y_train)
clf = CalibratedClassifierCV(best_xgb,method="sigmoid")
clf.fit(x_train,y_train)
predict_y = clf.predict_proba(x_train)
print("the train log loss is",log_loss(y_train,predict_y,labels=best_xgb.classes_))
predict_y = clf.predict_proba(x_cv)
print("the cross validation log loss is",log_loss(y_cv,predict_y,labels = best_xgb.classes_))
predict_y = clf.predict_proba(x_test)
print("the test log loss is",log_loss(y_test,predict_y,labels = best_xgb.classes_))
```

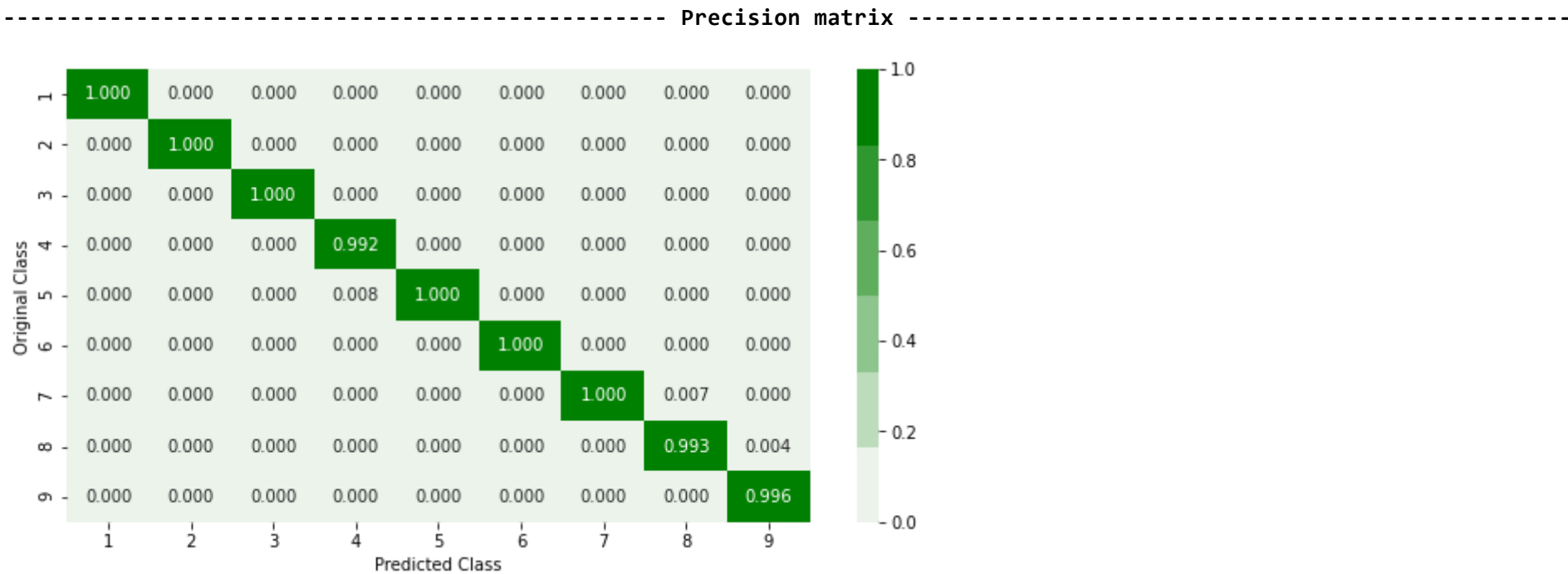
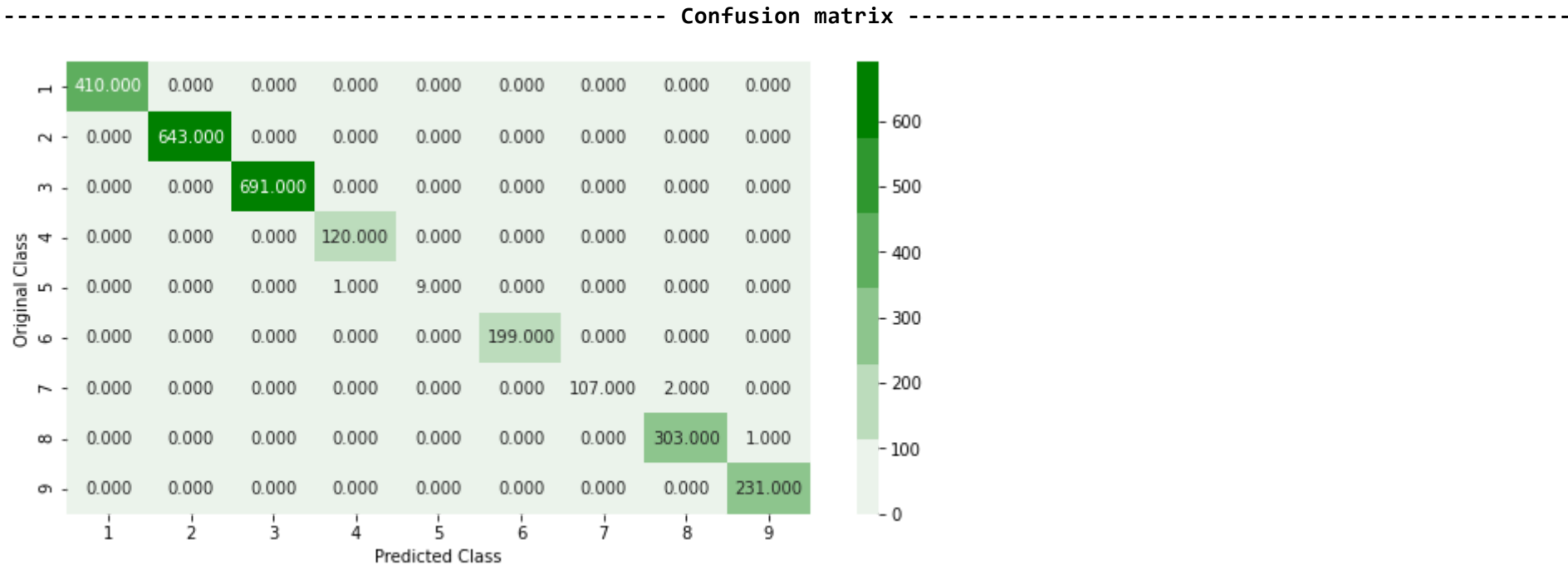
the train log loss is 0.008368743129319869

the cross validation log loss is 0.0216607553755203

the test log loss is 0.01104043768911403

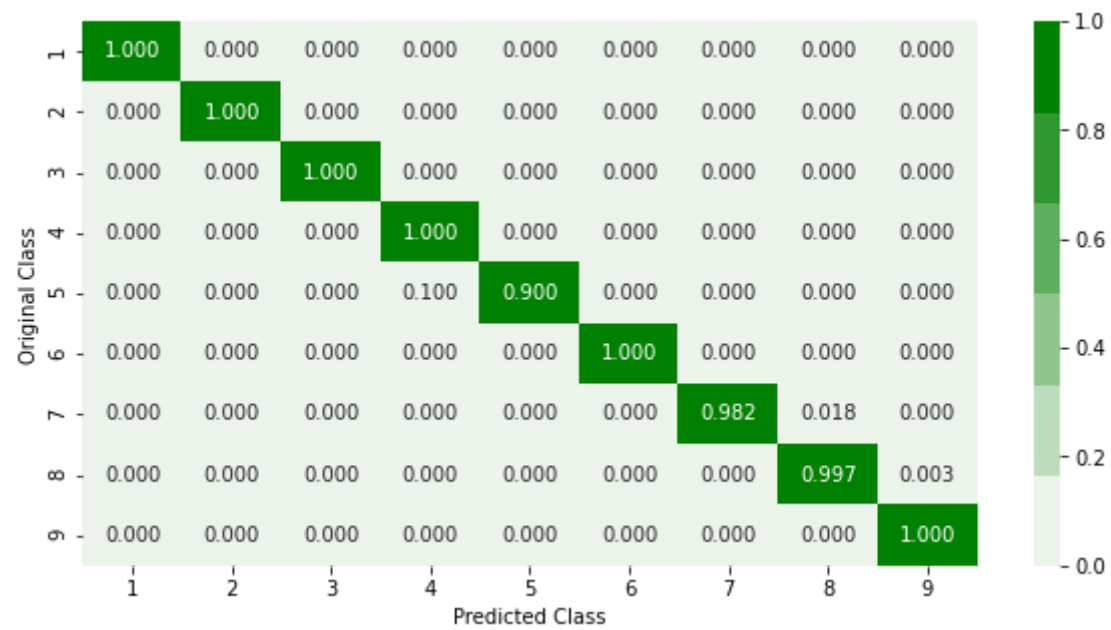
```
In [ ]: plot_confusion_matrix(y_test,clf.predict(x_test))
```

Number of misclassified points 0.1472211998527788



Sum of columns in precision matrix [1. 1. 1. 1. 1. 1. 1. 1. 1.]





Sum of rows in precision matrix [1. 1. 1. 1. 1. 1. 1. 1. 1.]

Extracting Bigram for opcode


```
In [ ]: import subprocess
from multiprocessing import process
from tqdm import tqdm
from csv import writer
import codecs
import os
import array
import numpy as np
def asm_opcode_feature(df,strt,end,dest_dir,num):
    pid = os.getpid()
    print("pid=",pid)

    opcode_file = open("/content/drive/My Drive/opcode__"+str(num)+".txt","w+")
    for i in tqdm(range(strt,end)):
        file_name = str(df[i]+'.asm')
        system = subprocess.Popen(['7z', 'e', 'train.7z', '-o'+dest_dir,file_name, '-r'])
        system.communicate()
        opcode=''
        with codecs.open(dest_dir+"/"+file_name,encoding='cp1252',errors='replace') as f:

            for lines in f:
                line= lines.rstrip().split()
                for each_line in line:
                    if each_line in opcodes:
                        opcode+=each_line+" "
                os.remove(os.path.join(dest_dir,file_name))
            opcode_file.write(opcode+"\n")
    opcode_file.close()
```

```
In [ ]: from multiprocessing import Process
import time
p1 = Process(target = asm_opcode_feature,args=(name,0,2717,"/content/asm_image_features",1))
p2 = Process(target = asm_opcode_feature,args = (name,2717,5434,"/content/asm_image_features",2))
p3 = Process(target = asm_opcode_feature,args=(name,5434,8151,"/content/asm_image_features",3))
p4 = Process(target = asm_opcode_feature,args = (name,8151,10868,"/content/asm_image_features",4))

p1.start()
p2.start()
p3.start()
p4.start()

p1.join()
p2.join()
p3.join()
p4.join()
```

25%|██████| 676/2717 [3:32:20<10:00:30, 17.65s/it]

```
In [ ]: opcodes = ['jmp', 'mov', 'retf', 'push', 'pop', 'xor', 'retn', 'nop', 'sub', 'inc', 'dec', 'add','imul', 'xchg', 'or', 'shr', 'cmp', 'call', 'shl', 'ror', 'rol', 'jnb','jz','rtn','lea','movzx']
```

```
In [ ]: opcode_bigram = []
for i, v in enumerate(opcodes):
    for j in range(0, len(opcodes)):
        opcode_bigram.append(v + ' ' + opcodes[j])
len(opcode_bigram)
```

Out[]: 676

```
In [ ]: file1 = open("/content/drive/My Drive/final_features/opcode__1.txt").read().splitlines()
file2 = open("/content/drive/My Drive/final_features/opcode__2.txt").read().splitlines()
file3 = open("/content/drive/My Drive/final_features/opcode__3.txt").read().splitlines()
file4 = open("/content/drive/My Drive/final_features/opcode__4.txt").read().splitlines()
print(len(file1),len(file2),len(file3),len(file4),sep="\n")
```

2717
2717
2717
2717

```
In [ ]: from sklearn.feature_extraction.text import CountVectorizer
from scipy.sparse import save_npz
from tqdm import tqdm
import scipy
def process1():
    vect = CountVectorizer(ngram_range=(2, 2), vocabulary = opcode_bigram)
    opcodebivect = scipy.sparse.csr_matrix((2717, len(opcode_bigram)))
    raw_opcode = open('/content/drive/My Drive/final_features/opcode__1.txt').read().split('\n')
    for indx in tqdm(range(2717)):
        opcodebivect[indx, :] += scipy.sparse.csr_matrix(vect.transform([raw_opcode[indx]]))
    scipy.sparse.save_npz('/content/drive/My Drive/final_features/opcodebigram1.npz', opcodebivect)

def process2():
    vect = CountVectorizer(ngram_range=(2, 2), vocabulary = opcode_bigram)
    opcodebivect = scipy.sparse.csr_matrix((2717, len(opcode_bigram)))
    raw_opcode = open('/content/drive/My Drive/final_features/opcode__1.txt').read().split('\n')
    for indx in tqdm(range(2717)):
        opcodebivect[indx, :] += scipy.sparse.csr_matrix(vect.transform([raw_opcode[indx]]))
    scipy.sparse.save_npz('/content/drive/My Drive/final_features/opcodebigram2.npz', opcodebivect)

def process3():
    vect = CountVectorizer(ngram_range=(2, 2), vocabulary = opcode_bigram)
    opcodebivect = scipy.sparse.csr_matrix((2717, len(opcode_bigram)))
    raw_opcode = open('/content/drive/My Drive/final_features/opcode__1.txt').read().split('\n')
    for indx in tqdm(range(2717)):
        opcodebivect[indx, :] += scipy.sparse.csr_matrix(vect.transform([raw_opcode[indx]]))
    scipy.sparse.save_npz('/content/drive/My Drive/final_features/opcodebigram3.npz', opcodebivect)

def process4():
    vect = CountVectorizer(ngram_range=(2, 2), vocabulary = opcode_bigram)
    opcodebivect = scipy.sparse.csr_matrix((2717, len(opcode_bigram)))
    raw_opcode = open('/content/drive/My Drive/final_features/opcode__1.txt').read().split('\n')
    for indx in tqdm(range(2717)):
        opcodebivect[indx, :] += scipy.sparse.csr_matrix(vect.transform([raw_opcode[indx]]))
    scipy.sparse.save_npz('/content/drive/My Drive/final_features/opcodebigram4.npz', opcodebivect)
```

```
In [ ]: from multiprocessing import Process
```

```
p1 =Process(target = process1)
p2 =Process(target = process2)
p3 =Process(target = process3)
p4 =Process(target = process4)
```

```
p1.start()
p2.start()
p3.start()
p4.start()
```

```
p1.join()
p2.join()
p3.join()
p4.join()
```

```
0%|          | 0/2717 [00:00<?, ?it/s]/usr/local/lib/python3.6/dist-packages/scipy/sparse/_index.py:118: SparseEfficiency
Warning: Changing the sparsity structure of a csr_matrix is expensive. lil_matrix is more efficient.
```

```
self._set_arrayXarray_sparse(i, j, x)
```

```
0%|          | 0/2717 [00:00<?, ?it/s]/usr/local/lib/python3.6/dist-packages/scipy/sparse/_index.py:118: SparseEfficiency
Warning: Changing the sparsity structure of a csr_matrix is expensive. lil_matrix is more efficient.
```

```
self._set_arrayXarray_sparse(i, j, x)
```

```
12%|██        | 334/2717 [00:02<00:24, 99.12it/s]/usr/local/lib/python3.6/dist-packages/scipy/sparse/_index.py:118: Spars
eEfficiencyWarning: Changing the sparsity structure of a csr_matrix is expensive. lil_matrix is more efficient.
```

```
self._set_arrayXarray_sparse(i, j, x)
```

```
0%|          | 0/2717 [00:00<?, ?it/s]/usr/local/lib/python3.6/dist-packages/scipy/sparse/_index.py:118: SparseEfficiency
Warning: Changing the sparsity structure of a csr_matrix is expensive. lil_matrix is more efficient.
```

```
self._set_arrayXarray_sparse(i, j, x)
```

```
100%|██████████| 2717/2717 [00:47<00:00, 57.72it/s]
```

```
100%|██████████| 2717/2717 [01:18<00:00, 34.66it/s]
```

```
100%|██████████| 2717/2717 [01:33<00:00, 29.19it/s]
```

```
100%|██████████| 2717/2717 [02:23<00:00, 18.98it/s]
```

```
In [ ]: from scipy.sparse import load_npz
file1 = load_npz("/content/drive/My Drive/final_features/opcodebigram1.npz")
file2 = load_npz("/content/drive/My Drive/final_features/opcodebigram2.npz")
file3 = load_npz("/content/drive/My Drive/final_features/opcodebigram3.npz")
file4 = load_npz("/content/drive/My Drive/final_features/opcodebigram4.npz")

print(file1.shape,file2.shape,file3.shape,file4.shape,sep="\n")
```

```
(2717, 676)
(2717, 676)
(2717, 676)
(2717, 676)
```

```
In [ ]: from scipy.sparse import vstack
opcode_bigram_full = vstack((file1,file2,file3,file4))
opcode_bigram_full.shape
```

```
Out[ ]: (10868, 676)
```

```
In [ ]: len(opcode_bigram)
```

```
Out[ ]: 676
```

```
In [ ]: type(opcode_bigram_full)
```

```
Out[ ]: scipy.sparse.csr.csr_matrix
```

```
In [ ]: import scipy
opcode_bigram_df = pd.DataFrame(scipy.sparse.csr_matrix.todense(normalize(opcode_bigram_full,axis=0)), columns = opcode_bi
gram)
opcode_bigram_df.to_csv("opcode_bigram_df.csv")
```

```
In [ ]: opcode_file= pd.read_csv("/content/drive/My Drive/final_features/opcode_bigram_df.csv")
opcode_file.head()
```

Out[]:

	Unnamed: 0	jmp jmp	jmp mov	jmp retf	jmp push	jmp pop	jmp xor	jmp retn	jmp nop	jmp sub	jmp inc	jmp dec	jmp add	jmp imul	jmp xchg	jmp or
0	0	0.000000	0.000101	0.0	0.000070	0.000000	0.000279	0.000703	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.0	0.000000
1	1	0.000506	0.013873	0.0	0.015550	0.009357	0.012011	0.007730	0.000000	0.040649	0.017186	0.017751	0.003449	0.000000	0.0	0.005949
2	2	0.000000	0.001237	0.0	0.001401	0.000000	0.001955	0.016865	0.000000	0.000000	0.000000	0.000000	0.000493	0.000000	0.0	0.000000
3	3	0.004772	0.041357	0.0	0.058068	0.008234	0.045810	0.000703	0.008338	0.044751	0.045227	0.040575	0.016752	0.008063	0.0	0.070204
4	4	0.000000	0.002008	0.0	0.001261	0.000374	0.001257	0.000000	0.000000	0.000746	0.000000	0.000000	0.005912	0.000000	0.0	0.000595

5 rows × 677 columns

```
In [ ]: opcode_file = opcode_file.drop("Unnamed: 0",axis=1)
```

byte_feature+asm_image_feature+asm_file_size+byte_file_size+opcode_bigram

```
In [ ]: advanced_fea2 = pd.concat([byte_feature_df,asm_fea,size_file,asm_size,opcode_file],axis=1,ignore_index=False)
advanced_fea2
```

Out[]:

	00 00	ff ff	00 5a	00 50	00 e8	fe ff	ff 00	01 00	ff 52	00 5e	00 03	07 00	65 00	00 52	00 c0	51 6a	ff 8d	6a 00	56 57	65 73
0	33081.0	4877.0	81.0	210.0	471.0	757.0	1008.0	1118.0	44.0	78.0	404.0	146.0	111.0	74.0	199.0	6.0	136.0	289.0	131.0	32.0
1	24301.0	2396.0	8.0	210.0	88.0	96.0	561.0	565.0	9.0	11.0	127.0	70.0	280.0	38.0	82.0	2.0	83.0	102.0	55.0	43.0
2	13253.0	2291.0	148.0	932.0	1376.0	207.0	2966.0	575.0	65.0	116.0	108.0	185.0	187.0	117.0	488.0	23.0	325.0	552.0	19.0	22.0
3	2155.0	24.0	2.0	14.0	6.0	2.0	1.0	76.0	0.0	3.0	102.0	2.0	154.0	9.0	4.0	5.0	2.0	16.0	3.0	15.0
4	3814.0	43.0	4.0	4.0	14.0	2.0	2.0	37.0	1.0	2.0	2.0	2.0	11.0	6.0	2.0	13.0	0.0	11.0	1.0	4.0
...
10863	9.0	16.0	12.0	20.0	5.0	13.0	12.0	11.0	10.0	13.0	11.0	12.0	9.0	11.0	14.0	18.0	15.0	9.0	21.0	10.0
10864	64616.0	675564.0	62.0	146.0	93.0	94.0	1190.0	1299.0	22.0	28.0	239.0	120.0	245.0	73.0	98.0	7.0	35.0	152.0	210.0	457.0
10865	5652.0	75.0	33.0	23.0	38.0	15.0	284.0	84.0	19.0	17.0	78.0	41.0	26.0	21.0	11.0	12.0	30.0	36.0	14.0	16.0
10866	41031.0	1253.0	16.0	53.0	22.0	148.0	289.0	764.0	12.0	6.0	50.0	65.0	47.0	27.0	65.0	4.0	36.0	49.0	19.0	2.0
10867	8193.0	343.0	1824.0	1879.0	1840.0	57.0	0.0	2183.0	1.0	1.0	1.0	3.0	65.0	1823.0	2279.0	0.0	5.0	1390.0	0.0	69.0

10868 rows × 1478 columns

```
In [ ]: x_train,x_test,y_train,y_test = train_test_split(advanced_fea2,y_data)
x_train,x_cv,y_train,y_cv= train_test_split(x_train,y_train)
```

XGboost classifier

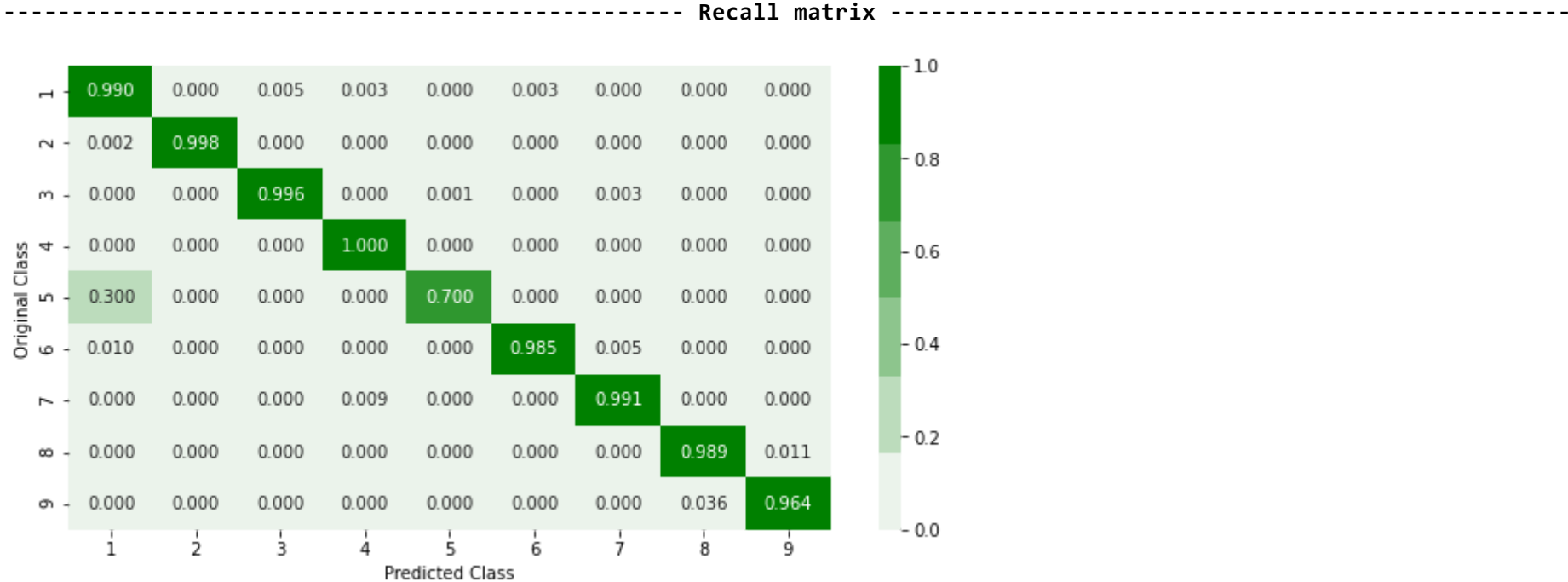
```
In [ ]: alpha=[10,50,100,500,1000,2000]
cv_log_error_array=[]
for i in alpha:
    xg = XGBClassifier(n_estimators=i,nthreads=-1)
    xg.fit(x_train,y_train)
    sig_clf = CalibratedClassifierCV(xg,method="sigmoid")
    sig_clf.fit(x_train,y_train)
    print("completed for alpha value :",i)
    predict_y = sig_clf.predict_proba(x_cv)
    cv_log_error_array.append(log_loss(y_cv,predict_y,labels=xg.classes_,eps=1e-15))
for i in range(len(cv_log_error_array)):
    print("log loss for c ",alpha[i],"is",cv_log_error_array[i])
best_alpha = np.argmin(cv_log_error_array)
```

```
completed for alpha value : 10
completed for alpha value : 50
completed for alpha value : 100
completed for alpha value : 500
completed for alpha value : 1000
completed for alpha value : 2000
log loss for c 10 is 0.033030721035860355
log loss for c 50 is 0.033030721035860355
log loss for c 100 is 0.033030721035860355
log loss for c 500 is 0.033030721035860355
log loss for c 1000 is 0.033030721035860355
log loss for c 2000 is 0.033030721035860355
```



```
In [ ]: best_xg = XGBClassifier(n_estimators = alpha[best_alpha],nthreads=-1)
best_xg.fit(x_train,y_train)
sig_clf = CalibratedClassifierCV(best_xg,method="sigmoid")
sig_clf.fit(x_train,y_train)
predict_y = sig_clf.predict_proba(x_train)
print("for values of best alpha = ",alpha[best_alpha],"the train log loss is",log_loss(y_train,predict_y,labels=best_xg.classes_))
predict_y = sig_clf.predict_proba(x_cv)
print("for values of best alpha = ",alpha[best_alpha],"the cross validation log loss is",log_loss(y_cv,predict_y,labels = best_xg.classes_))
predict_y = sig_clf.predict_proba(x_test)
print("for values of best alpha = ",alpha[best_alpha],"the test log loss is",log_loss(y_test,predict_y,labels = best_xg.classes_))
plot_confusion_matrix(y_test,sig_clf.predict(x_test))
```


Sum of columns in precision matrix [1. 1. 1. 1. 1. 1. 1. 1. 1.]



Sum of rows in precision matrix [1. 1. 1. 1. 1. 1. 1. 1. 1.]

XGBoost with hyper parameter tuning using random search

```
In [ ]: xgboost= XGBClassifier()
params ={
    'learning_rate':[0.01,0.03,0.05,0.1,0.15,0.2],
    'n_estimators' :[100,200,500,1000,2000],
    'max_depth':[3,5,10],
    'colsample_bytree':[0.1,0.3,0.5,1],
    'subsample':[0.1,0.3,0.5,1]
}
random_cfl1=RandomizedSearchCV(xgboost,param_distributions = params,verbose=10,n_jobs=-1)
random_cfl1.fit(x_train,y_train)
```

Fitting 5 folds for each of 10 candidates, totalling 50 fits

```
[Parallel(n_jobs=-1)]: Using backend LokyBackend with 4 concurrent workers.
[Parallel(n_jobs=-1)]: Done   5 tasks      | elapsed: 13.5min
[Parallel(n_jobs=-1)]: Done  10 tasks      | elapsed: 18.9min
[Parallel(n_jobs=-1)]: Done  17 tasks      | elapsed: 26.1min
[Parallel(n_jobs=-1)]: Done  24 tasks      | elapsed: 34.9min
[Parallel(n_jobs=-1)]: Done  33 tasks      | elapsed: 60.5min
[Parallel(n_jobs=-1)]: Done  42 tasks      | elapsed: 82.6min
[Parallel(n_jobs=-1)]: Done  50 out of  50 | elapsed: 92.6min finished
```

```
Out[ ]: RandomizedSearchCV(cv=None, error_score=nan,
                          estimator=XGBClassifier(base_score=0.5, booster='gbtree',
                                                  colsample_bylevel=1,
                                                  colsample_bynode=1,
                                                  colsample_bytree=1, gamma=0,
                                                  learning_rate=0.1, max_delta_step=0,
                                                  max_depth=3, min_child_weight=1,
                                                  missing=None, n_estimators=100,
                                                  n_jobs=1, nthread=None,
                                                  objective='binary:logistic',
                                                  random_state=0, reg_alpha=0,
                                                  reg_lambda=1...,
                                                  seed=None, silent=None, subsample=1,
                                                  verbosity=1),
                          iid='deprecated', n_iter=10, n_jobs=-1,
                          param_distributions={'colsample_bytree': [0.1, 0.3, 0.5, 1],
                                              'learning_rate': [0.01, 0.03, 0.05, 0.1,
                                                              0.15, 0.2],
                                              'max_depth': [3, 5, 10],
                                              'n_estimators': [100, 200, 500, 1000,
                                                              2000],
                                              'subsample': [0.1, 0.3, 0.5, 1]},
                          pre_dispatch='2*n_jobs', random_state=None, refit=True,
                          return_train_score=False, scoring=None, verbose=10)
```

```
In [ ]: print(random_cfl1.best_params_)
```

```
{'subsample': 1, 'n_estimators': 100, 'max_depth': 5, 'learning_rate': 0.1, 'colsample_bytree': 1}
```

```
In [ ]:
```

```
In [ ]: best_xgb = XGBClassifier(n_estimators=100,learning_rate = 0.1,colsample_bytree=1,max_depth=5,subsample=1)
best_xgb.fit(x_train,y_train)
clf = CalibratedClassifierCV(best_xgb,method="sigmoid")
clf.fit(x_train,y_train)
predict_y = clf.predict_proba(x_train)
print("the train log loss is",log_loss(y_train,predict_y,labels=best_xgb.classes_))
predict_y = clf.predict_proba(x_cv)
print("the cross validation log loss is",log_loss(y_cv,predict_y,labels = best_xgb.classes_))
predict_y = clf.predict_proba(x_test)
print("the test log loss is",log_loss(y_test,predict_y,labels = best_xgb.classes_))
```

the train log loss is 0.013363326490262784

the cross validation log loss is 0.032749888583227735

the test log loss is 0.027624610991505854

```
In [ ]: best_xgb = XGBClassifier(n_estimators=100,learning_rate = 0.01,colsample_bytree=1,max_depth=5,subsample=1)
best_xgb.fit(x_train,y_train)
clf = CalibratedClassifierCV(best_xgb,method="sigmoid")
clf.fit(x_train,y_train)
predict_y = clf.predict_proba(x_train)
print("the train log loss is",log_loss(y_train,predict_y,labels=best_xgb.classes_))
predict_y = clf.predict_proba(x_cv)
print("the cross validation log loss is",log_loss(y_cv,predict_y,labels = best_xgb.classes_))
predict_y = clf.predict_proba(x_test)
print("the test log loss is",log_loss(y_test,predict_y,labels = best_xgb.classes_))
```

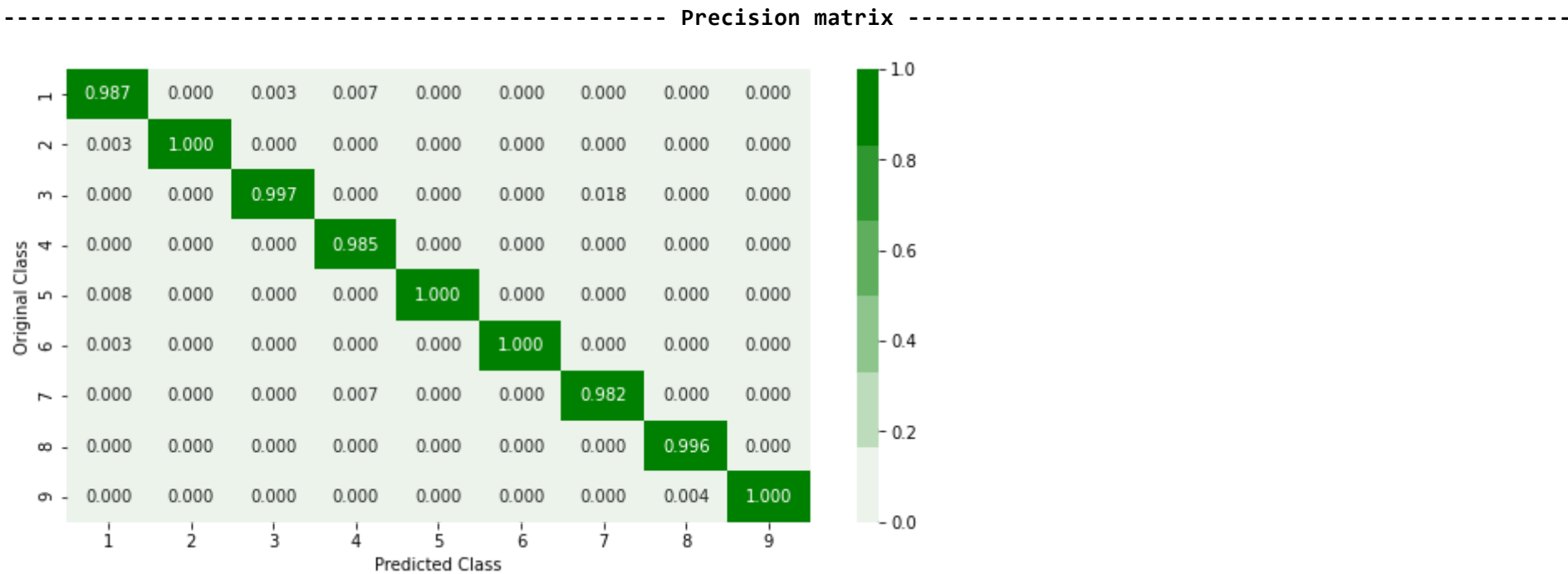
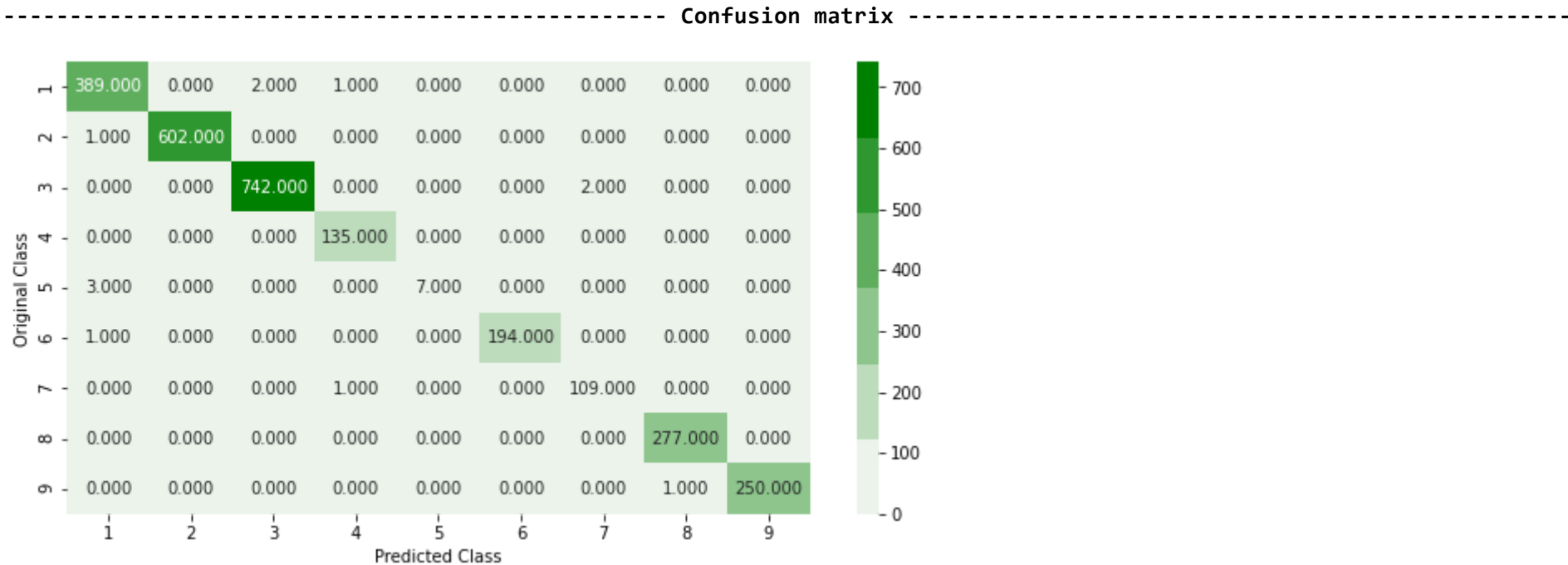
the train log loss is 0.008368743129319869

the cross validation log loss is 0.0216607553755203

the test log loss is 0.01104043768911403

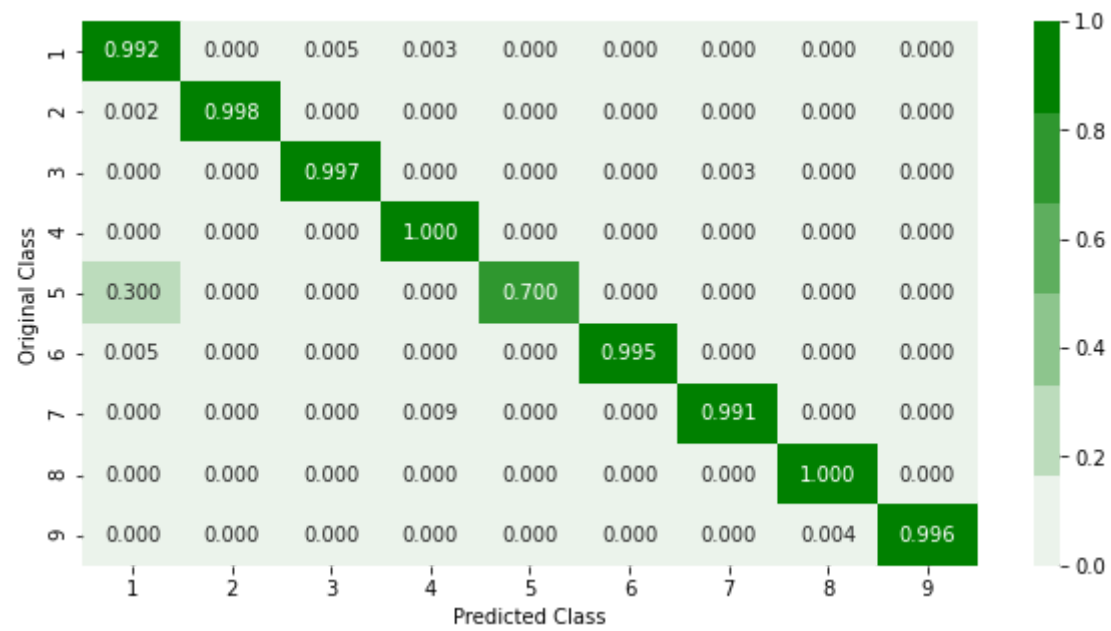
```
In [ ]: plot_confusion_matrix(y_test,clf.predict(x_test))
```

Number of misclassified points 0.44166359955833645



Sum of columns in precision matrix [1. 1. 1. 1. 1. 1. 1. 1. 1.]

----- Recall matrix -----



Sum of rows in precision matrix [1. 1. 1. 1. 1. 1. 1. 1. 1.]

Conclusion


```
In [ ]: from prettytable import PrettyTable
ptable = PrettyTable()
ptable.field_names = ["Model", 'Features', 'log loss']
ptable.add_row(["knn", "Byte_bigram + asm_image", "1.39"])
ptable.add_row(["LR", "Byte_bigram+asm_image", "1.35"])
ptable.add_row(["RF", "Byte_bigeam+asm_image", "0.45"])
ptable.add_row(["XGB", "Byte_bigram+asm_image", "0.42"])
ptable.add_row(["XGB_with_random_Search", "Byte_bigram+asm_image", "0.48"])
ptable.add_row([" ----- ", " ----- "])
ptable.add_row(['XGB', 'byte_bigram+asm_image+file_size', "0.42"])
ptable.add_row(['XGB_with_ransom_search', 'byte_bigram+asm_image+file_size', '0.01'])
ptable.add_row([" ----- ", " ----- "])
ptable.add_row(['XGB', 'byte_bigram+asm_image+file_size+opcode_bigram', '0.05'])
ptable.add_row(['XGB', 'byte_bigram+asm_image+file_size+opcode_bigram', '0.01'])
print(ptable)
```

Model	Features	log loss
knn	Byte_bigram + asm_image	1.39
LR	Byte_bigram+asm_image	1.35
RF	Byte_bigeam+asm_image	0.45
XGB	Byte_bigram+asm_image	0.42
XGB_with_random_Search	Byte_bigram+asm_image	0.48

XGB	byte_bigram+asm_image+file_size	0.42
XGB_with_ransom_search	byte_bigram+asm_image+file_size	0.01

XGB	byte_bigram+asm_image+file_size+opcode_bigram	0.05
XGB	byte_bigram+asm_image+file_size+opcode_bigram	0.01