```
import numpy  # linear algebra
import pandas as pd # data processing, CSV file I/O (e.g. pd.read_csv)
import io
# Models:
from sklearn import svm
from sklearn.model_selection import cross_validate
from sklearn.linear_model import LinearRegression
from sklearn.model_selection import KFold, cross_val_score
# Python utilities:
import time
import os
from sklearn.ensemble import GradientBoostingClassifier
from sklearn.metrics import roc_auc_score
from sklearn.metrics import roc_curve
import matplotlib.pyplot as plt
import datetime
%tensorflow_version 2.x
import tensorflow as tf
```

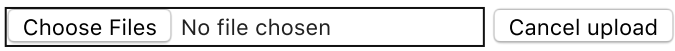
Google Collabrate Import Data

```
from google.colab import drive
drive.mount('/content/gdrive')
```

```
Drive already mounted at /content/gdrive; to attempt to forcibly remount, call
```

Google Collabrate files uploading

```
from google.colab import files
uploaded = files.upload()
for fn in uploaded.keys():
  print('User uploaded file "{name}" with length {length} bytes'.format(
      name=fn, length=len(uploaded[fn])))
```



Train data and test data imported and test data listed

```
train_data=pd.read_csv('gdrive/My Drive/GiveMeSomeCredit/cs-training.csv')
test_data=pd.read_csv('gdrive/My Drive/GiveMeSomeCredit/cs-test.csv')
cv = KFold(n_splits=2,random_state=None, shuffle=False)
```

First 6 data information

```
train_data.head()
```

	Unnamed: 0	SeriousDlqin2yrs	RevolvingUtilizationOfUnsecuredLines	age	59Da
0	1	1	0.766127	45	
1	2	0	0.957151	40	
2	3	0	0.658180	38	
3	4	0	0.233810	30	
4	5	0	0.907239	49	

```
train_data.shape
```

```
(150000, 12)
```

Get overall information more statistical

```
train_data.describe()
```

	Unnamed: 0	SeriousDlqin2yrs	RevolvingUtilizationOfUnsecuredLines	
count	150000.000000	150000.000000	150000.000000	15
mean	75000.500000	0.066840	6.048438	
std	43301.414527	0.249746	249.755371	
min	1.000000	0.000000	0.000000	
25%	37500.750000	0.000000	0.029867	
50%	75000.500000	0.000000	0.154181	
75%	112500.250000	0.000000	0.559046	
max	150000.000000	1.000000	50708.000000	

The ratio of debt to assets

```
train_data.DebtRatio.describe()
```

```
count    150000.000000
mean        353.005076
std        2037.818523
min           0.000000
25%           0.175074
50%           0.366508
75%           0.868254
max      329664.000000
Name: DebtRatio, dtype: float64
```

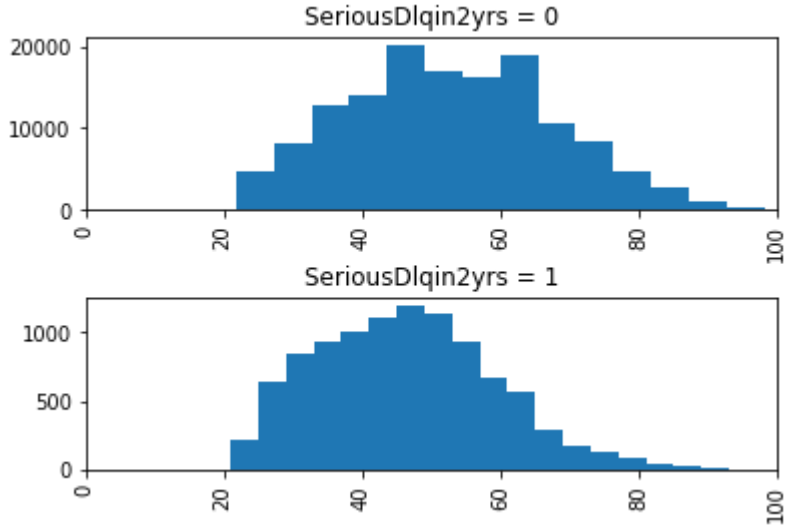
How balanced the data set is

```
train_data.SeriousDlqin2yrs.mean()
```

```
0.06684
```

```
age_hist = train_data['age'].hist(by=train_data['SeriousDlqin2yrs'], bins=20, layou
age_hist[0].set_xlim((0,100))
age_hist[0].set_title('SeriousDlqin2yrs = 0')
age_hist[1].set_xlim((0,100))
age_hist[1].set_title('SeriousDlqin2yrs = 1')
```

```
Text(0.5, 1.0, 'SeriousDlqin2yrs = 1')
```



Investigate further to see if these are outliers or not

```
train_data.DebtRatio.quantile([.975])
```

```
0.975    3489.025
Name: DebtRatio, dtype: float64
```

```
train_data[train_data['DebtRatio'] > 3489.025][['SeriousDlqin2yrs','MonthlyIncome'
```

```
train_data.groupby('NumberOfTimes90DaysLate').NumberOfTimes90DaysLate.count()
```

```
NumberOfTimes90DaysLate
0     141662
1       5243
2       1555
3        667
4        291
5        131
6         80
7         38
8         21
9         19
10         8
11         5
12         2
13         4
14         2
15         2
17         1
96         5
98       264
Name: NumberOfTimes90DaysLate, dtype: int64
```

```
train_data[train_data['NumberOfTimes90DaysLate'] > 95][['SeriousDlqin2yrs',
                                                        'NumberOfTime60-89DaysPastI
  'NumberOfTime30-59DaysPastDueNotWorse',
  'NumberOfTimes90DaysLate']].describe()
```

	SeriousDlqin2yrs	NumberOfTime60-89DaysPastDueNotWorse	NumberOfTime30-59DaysPastDueNotWorse	NumberO:
count	269.000000	269.000000	269.000000	
mean	0.546468	97.962825	97.962825	
std	0.498764	0.270628	0.270628	
min	0.000000	96.000000	96.000000	
25%	0.000000	98.000000	98.000000	
50%	1.000000	98.000000	98.000000	
75%	1.000000	98.000000	98.000000	
max	1.000000	98.000000	98.000000	

choose specific data and get the arg minumum of them

```
def rmDataAndPutArgmin(train_data, first = 20, second = 40):
    New = []
    argminx = train_data.argmin()
    for val in train_data:
        if ((val == first) | (val == second)):
            New.append(argminx)
        else:
```

```
            New.append(val)

    return New


not_missing = train_data.dropna()
target = 'MonthlyIncome'
predictors = [c for c in list(not_missing) if c not in [target, 'Unnamed: 0','Seric
X_data = not_missing[predictors]
y_data = not_missing[target]
regr = LinearRegression().fit(X_data, y_data)
regr.score(X_data, y_data)
```

```
0.02201505632577072
```

```
train_data=train_data.fillna(round(train_data.median()))
train_data.loc[train_data.DebtRatio > 1, 'DebtRatio'] = train_data['DebtRatio'].mec
```

Training Data

```
trainnew = rmDataAndPutArgmin(train_data.SeriousDlqin2yrs)
train_data.SeriousDlqin2yrs=trainnew

x = train_data.drop(['SeriousDlqin2yrs'], axis=1)
y = trainnew

gbscores=[]
for n in [150, 160, 170, 180, 190, 200, 210]:
    gbscores.append(numpy.mean(cross_val_score(GradientBoostingClassifier(learning_
print(gbscores)
start_time = datetime.datetime.now()
print (numpy.mean(cross_val_score(GradientBoostingClassifier(learning_rate=0.1, n_e
print ('Time elapsed:', datetime.datetime.now() - start_time)
```

```
Iter       Train Loss   Remaining Time
   1           0.4575           23.65s
   2           0.4392           23.24s
   3           0.4256           22.91s
   4           0.4154           23.01s
   5           0.4074           22.83s
   6           0.4009           22.79s
   7           0.3954           22.54s
   8           0.3904           22.33s
   9           0.3863           22.11s
  10           0.3828           22.05s
  20           0.3647           20.35s
  30           0.3581           18.69s
  40           0.3546           17.18s
  50           0.3521           15.57s
  60           0.3502           14.01s
  70           0.3486           12.43s
  80           0.3474           10.88s
  90           0.3465            9.32s
 100           0.3455            7.77s
Iter       Train Loss   Remaining Time
   1           0.4571           22.71s
   2           0.4388           22.87s
   3           0.4263           22.71s
   4           0.4162           22.50s
   5           0.4081           22.56s
   6           0.4023           22.34s
   7           0.3970           22.13s
   8           0.3926           21.93s
   9           0.3886           21.78s
  10           0.3860           21.66s
  20           0.3689           20.30s
  30           0.3621           18.60s
  40           0.3583           17.03s
  50           0.3556           15.46s
  60           0.3537           13.89s
  70           0.3523           12.32s
  80           0.3510           10.78s
  90           0.3500            9.24s
 100           0.3492            7.71s
Iter       Train Loss   Remaining Time
   1           0.4575           24.49s
   2           0.4392           24.57s
   3           0.4256           24.23s
   4           0.4154           24.00s
   5           0.4074           23.93s
   6           0.4009           23.77s
   7           0.3954           23.57s
   8           0.3904           23.42s
   9           0.3863           23.26s
  10           0.3828           23.27s
  20           0.3647           21.68s
  30           0.3581           20.13s
  40           0.3546           18.58s
  50           0.3521           17.01s
  60           0.3502           15.46s
  70           0.3486           13.92s
  80           0.3474           12.37s
  90           0.3465           10.82s
 100           0.3455            9.27s
Iter       Train Loss   Remaining Time
```

```
   1           0.4571           24.37s
   2           0.4388           24.85s
   3           0.4263           24.52s
   4           0.4162           24.34s
   5           0.4081           24.18s
   6           0.4023           24.08s
   7           0.3970           23.82s
   8           0.3926           23.74s
   9           0.3886           23.55s
  10           0.3860           23.45s
  20           0.3689           21.84s
  30           0.3621           20.21s
  40           0.3583           18.60s
  50           0.3556           17.09s
  60           0.3537           15.50s
  70           0.3523           13.94s
  80           0.3510           12.38s
  90           0.3500           10.83s
 100           0.3492            9.29s
Iter       Train Loss   Remaining Time
   1           0.4575           25.95s
   2           0.4392           26.02s
   3           0.4256           25.62s
   4           0.4154           25.87s
   5           0.4074           25.65s
   6           0.4009           25.46s
   7           0.3954           25.25s
   8           0.3904           25.06s
   9           0.3863           24.87s
  10           0.3828           24.81s
  20           0.3647           23.23s
  30           0.3581           21.65s
  40           0.3546           20.12s
  50           0.3521           18.58s
  60           0.3502           17.01s
  70           0.3486           15.44s
  80           0.3474           13.89s
  90           0.3465           12.34s
 100           0.3455           10.81s
Iter       Train Loss   Remaining Time
   1           0.4571           26.41s
   2           0.4388           26.62s
   3           0.4263           26.36s
   4           0.4162           26.33s
   5           0.4081           26.13s
   6           0.4023           26.13s
   7           0.3970           25.86s
   8           0.3926           25.68s
   9           0.3886           25.49s
  10           0.3860           25.45s
  20           0.3689           23.68s
  30           0.3621           21.94s
  40           0.3583           20.31s
  50           0.3556           18.74s
  60           0.3537           17.14s
  70           0.3523           15.54s
  80           0.3510           13.96s
  90           0.3500           12.41s
 100           0.3492           10.84s
Iter       Train Loss   Remaining Time
   1           0.4575           27.78s
   2           0.4392           27.61s
```

```
   3           0.4256           27.44s
   4           0.4154           27.19s
   5           0.4074           27.01s
   6           0.4009           27.05s
   7           0.3954           26.86s
   8           0.3904           26.66s
   9           0.3863           26.48s
  10           0.3828           26.52s
  20           0.3647           24.84s
  30           0.3581           23.24s
  40           0.3546           21.67s
  50           0.3521           20.13s
  60           0.3502           18.54s
  70           0.3486           16.98s
  80           0.3474           15.45s
  90           0.3465           13.91s
 100           0.3455           12.36s
Iter       Train Loss   Remaining Time
   1           0.4571           27.23s
   2           0.4388           27.26s
   3           0.4263           27.07s
   4           0.4162           26.91s
   5           0.4081           26.80s
   6           0.4023           26.82s
   7           0.3970           26.59s
   8           0.3926           26.41s
   9           0.3886           26.26s
  10           0.3860           26.22s
  20           0.3689           24.71s
  30           0.3621           23.19s
  40           0.3583           21.63s
  50           0.3556           20.08s
  60           0.3537           18.53s
  70           0.3523           16.96s
  80           0.3510           15.41s
  90           0.3500           13.86s
 100           0.3492           12.32s
Iter       Train Loss   Remaining Time
   1           0.4575           28.79s
   2           0.4392           28.96s
   3           0.4256           28.86s
   4           0.4154           28.60s
   5           0.4074           28.40s
   6           0.4009           28.30s
   7           0.3954           28.10s
   8           0.3904           27.96s
   9           0.3863           27.97s
  10           0.3828           28.04s
  20           0.3647           26.22s
  30           0.3581           24.63s
  40           0.3546           23.08s
  50           0.3521           21.57s
  60           0.3502           20.00s
  70           0.3486           18.52s
  80           0.3474           17.08s
  90           0.3465           15.51s
 100           0.3455           13.95s
Iter       Train Loss   Remaining Time
   1           0.4571           28.50s
   2           0.4388           28.97s
   3           0.4263           28.72s
   4           0.4162           28.44s
```

```
   4           0.4162           28.44s
   5           0.4081           28.57s
   6           0.4023           28.33s
   7           0.3970           28.06s
   8           0.3926           27.88s
   9           0.3886           27.80s
  10           0.3860           27.75s
  20           0.3689           26.23s
  30           0.3621           24.60s
  40           0.3583           23.01s
  50           0.3556           21.47s
  60           0.3537           19.92s
  70           0.3523           18.37s
  80           0.3510           16.84s
  90           0.3500           15.31s
 100           0.3492           13.77s
Iter       Train Loss   Remaining Time
   1           0.4575           30.43s
   2           0.4392           30.48s
   3           0.4256           30.44s
   4           0.4154           30.25s
   5           0.4074           30.02s
   6           0.4009           29.91s
   7           0.3954           29.76s
   8           0.3904           29.71s
   9           0.3863           29.64s
  10           0.3828           29.63s
  20           0.3647           28.04s
  30           0.3581           26.41s
  40           0.3546           24.82s
  50           0.3521           23.24s
  60           0.3502           21.65s
  70           0.3486           20.11s
  80           0.3474           18.56s
  90           0.3465           17.00s
 100           0.3455           15.44s
 200           0.3391            0.00s
Iter       Train Loss   Remaining Time
   1           0.4571           30.39s
   2           0.4388           30.48s
   3           0.4263           30.19s
   4           0.4162           29.97s
   5           0.4081           29.92s
   6           0.4023           29.89s
   7           0.3970           29.62s
   8           0.3926           29.49s
   9           0.3886           29.40s
  10           0.3860           29.34s
  20           0.3689           27.76s
  30           0.3621           26.09s
  40           0.3583           24.55s
  50           0.3556           23.01s
  60           0.3537           21.46s
  70           0.3523           19.95s
  80           0.3510           18.42s
  90           0.3500           16.88s
 100           0.3492           15.41s
 200           0.3431            0.00s
Iter       Train Loss   Remaining Time
   1           0.4575           32.20s
   2           0.4392           32.57s
   3           0.4256           32.06s
```

```
         4           0.4154           31.95s
         5           0.4074           31.72s
         6           0.4009           31.87s
         7           0.3954           31.61s
         8           0.3904           31.40s
         9           0.3863           31.25s
        10           0.3828           31.16s
        20           0.3647           29.38s
        30           0.3581           27.77s
        40           0.3546           26.21s
        50           0.3521           24.64s
        60           0.3502           23.07s
        70           0.3486           21.58s
        80           0.3474           20.03s
        90           0.3465           18.49s
       100           0.3455           16.95s
       200           0.3391            1.54s
      Iter       Train Loss   Remaining Time
         1           0.4571           32.85s
         2           0.4388           32.43s
         3           0.4263           32.09s
         4           0.4162           31.80s
         5           0.4081           31.77s
         6           0.4023           31.54s
         7           0.3970           31.27s
         8           0.3926           31.18s
         9           0.3886           31.00s
        10           0.3860           30.95s
        20           0.3689           29.34s
        30           0.3621           27.73s
        40           0.3583           26.12s
        50           0.3556           24.62s
        60           0.3537           23.02s
        70           0.3523           21.48s
        80           0.3510           19.94s
        90           0.3500           18.41s
       100           0.3492           16.87s
       200           0.3431            1.53s
[0.7701272042736258, 0.7702553260139671, 0.7691237412549932, 0.769091447101051
      Iter       Train Loss   Remaining Time
         1           0.4575           48.09s
         2           0.4392           46.83s
         3           0.4256           46.41s
         4           0.4154           45.77s
         5           0.4074           45.94s
         6           0.4009           45.61s
         7           0.3954           45.52s
         8           0.3904           45.34s
         9           0.3863           45.28s
        10           0.3828           45.17s
        20           0.3647           43.37s
        30           0.3581           41.71s
        40           0.3546           40.28s
        50           0.3521           38.62s
        60           0.3502           37.02s
        70           0.3486           35.42s
        80           0.3474           33.87s
        90           0.3465           32.29s
       100           0.3455           30.75s
       200           0.3391           15.35s
       300           0.3331            0.00s
      Iter       Train Loss   Remaining Time
```

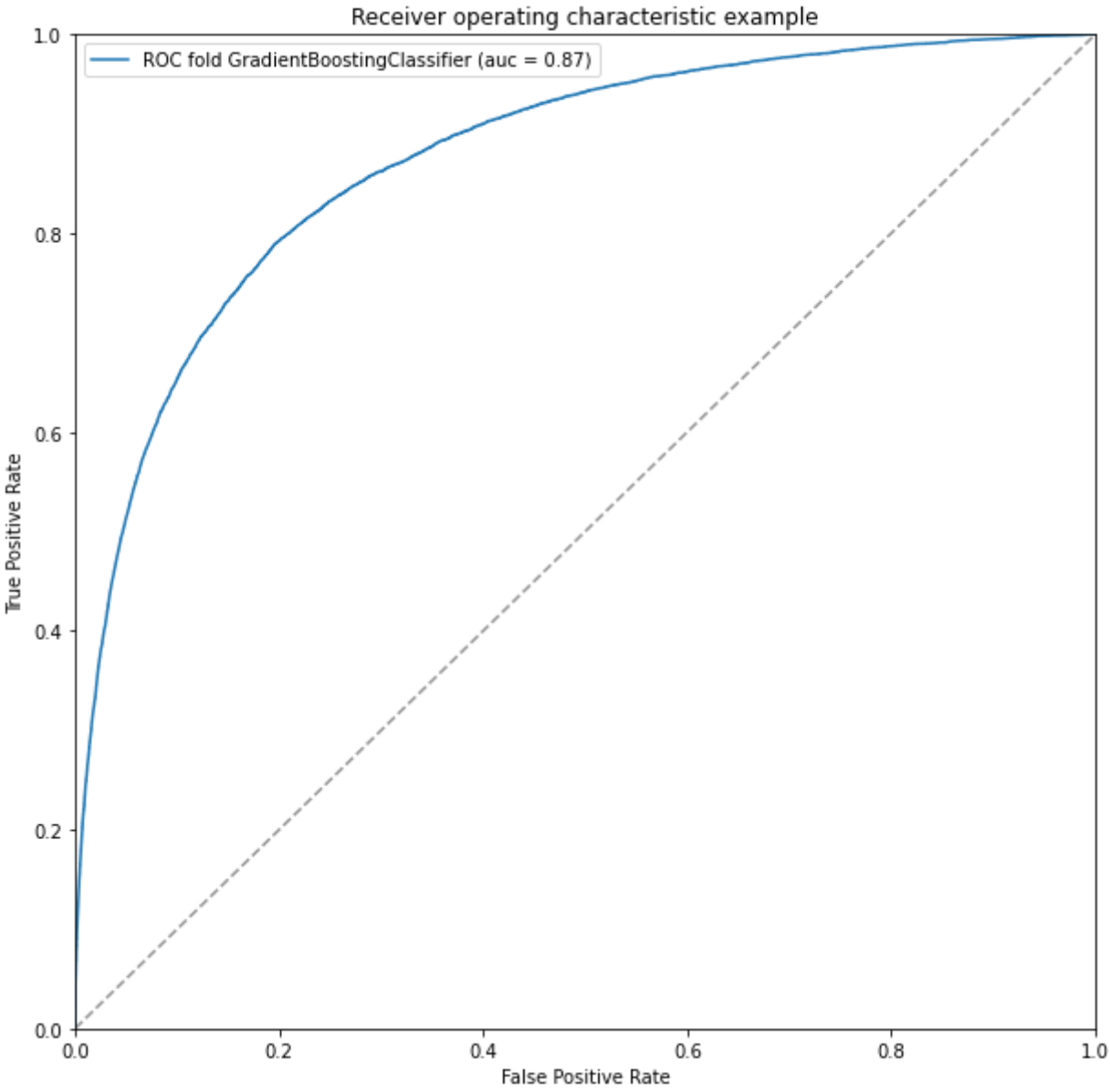
```
          1           0.4571           46.16s
          2           0.4388           46.69s
          3           0.4263           46.81s
          4           0.4162           47.51s
          5           0.4081           47.04s
          6           0.4023           46.70s
          7           0.3970           46.23s
          8           0.3926           45.97s
          9           0.3886           45.84s
         10           0.3860           45.83s
         20           0.3689           44.00s
         30           0.3621           42.19s
         40           0.3583           40.46s
         50           0.3556           38.79s
         60           0.3537           37.11s
         70           0.3523           35.56s
         80           0.3510           33.96s
         90           0.3500           32.38s
        100           0.3492           30.80s
        200           0.3431           15.33s
        300           0.3380            0.00s
0.7657463539285931
Time elapsed: 0:01:32.755747
```

Roc curve of gradient boosting classifier

```
plt.figure(figsize=(10, 10))
models=[]
models.append(GradientBoostingClassifier(learning_rate=0.1, n_estimators=300, verbo
for model in models:
    model.fit(x,y)
    testscore=model.predict_proba(x)[:, 1]
    fpr, tpr, thresholds = roc_curve(y, testscore)
    roc_auc = roc_auc_score(y, testscore)
    md = str(model)
    md = md[:md.find('(')]
    plt.plot(fpr, tpr, label='ROC fold %s (auc = %0.2f)' % (md, roc_auc))

plt.plot([0, 1], [0, 1], '--', color=(0.6, 0.6, 0.6))
plt.xlim([0, 1])
plt.ylim([0, 1])
plt.xlabel('False Positive Rate')
plt.ylabel('True Positive Rate')
plt.title('Receiver operating characteristic example')
plt.legend(loc="best")
plt.show()
```

Iter	Train Loss	Remaining Time
1	0.4578	1.66m
2	0.4391	1.64m
3	0.4260	1.64m
4	0.4164	1.62m
5	0.4082	1.62m
6	0.4024	1.61m
7	0.3966	1.60m
8	0.3922	1.59m
9	0.3881	1.59m
10	0.3854	1.58m
20	0.3680	1.52m
30	0.3618	1.46m
40	0.3585	1.40m
50	0.3562	1.35m
60	0.3548	1.29m
70	0.3536	1.24m
80	0.3528	1.18m
90	0.3520	1.13m
100	0.3514	1.07m
200	0.3469	32.17s
300	0.3435	0.00s



Could not connect to the reCAPTCHA service. Please check your internet connection and reload to get a reCAPTCHA challenge.