112-1 Machine Learning

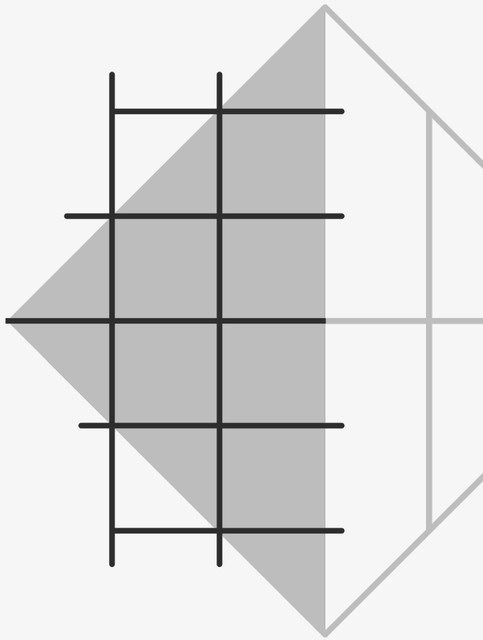
HW2_indiv_1(k)

Advisor: Chih-Hua Hsu

Student: Pao-Chang Chiu

Present Date: 2023/10/11

Contents



Introduction

- All the variables correlated to Y

In [31]:
```
import numpy as np
Y_corr = corr_matrix[ "Y" ][ (-np.abs(corr_matrix[ "Y" ])).argsort() ]
Y_corr
```

Out[31]:
```
Y      1.000000
X11   -0.747957
X14    0.709622
X15   -0.709226
X10   -0.667452
X13   -0.640153
X12   -0.607864
X7    -0.520050
X8    -0.448213
X6    -0.376066
X9    -0.179513
X16    0.054002
Name: Y, dtype: float64
```

Introduction

- The index of all the variables correlated to Y

```
In [33]: Y_corr.index

Out[33]: Index(['Y', 'X11', 'X14', 'X15', 'X10', 'X13', 'X12', 'X7', 'X8', 'X6', 'X9',
                'X16'],
               dtype='object')
```

- The length of all the variables correlated to Y

```
In [34]: len(Y_corr.index)

Out[34]: 12
```

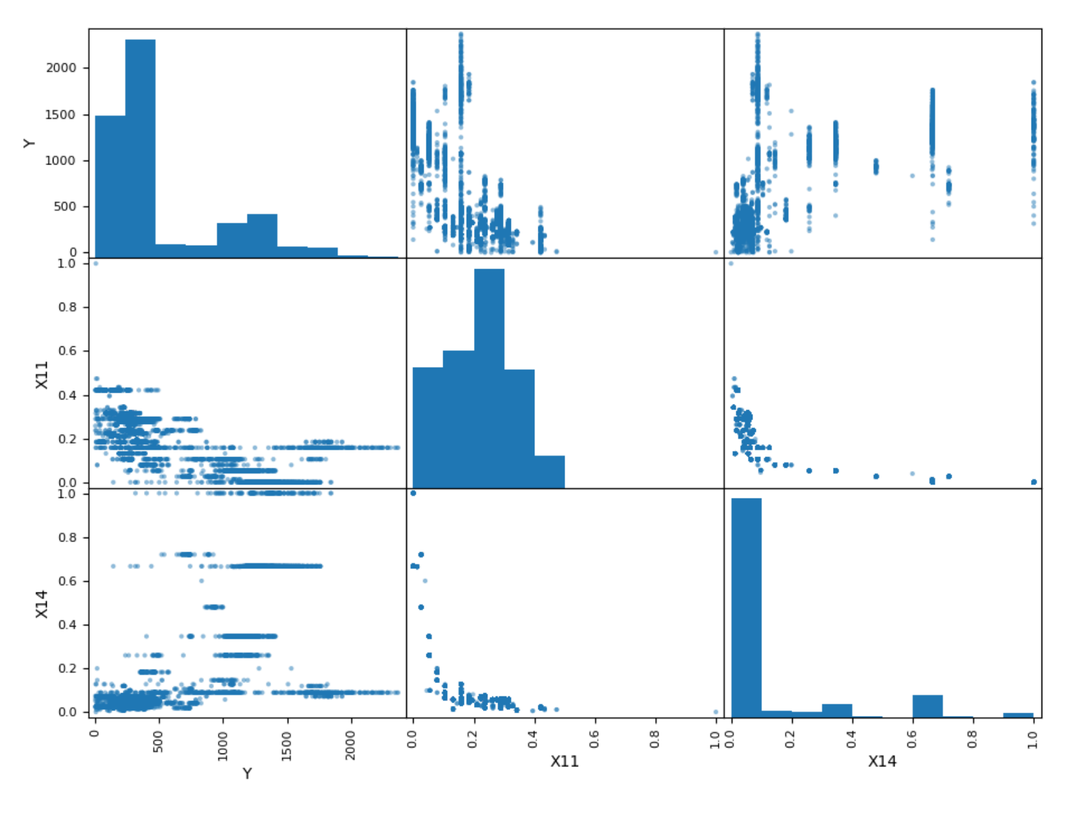
Problems

- What you want to solve
  - *The total number of features is odd*: plot one correlation scatter matrix with *Y*
  - *The total number of features is even*: plot two correlation scatter matrices with *Y*
- How many times do you need to run in *for loop*

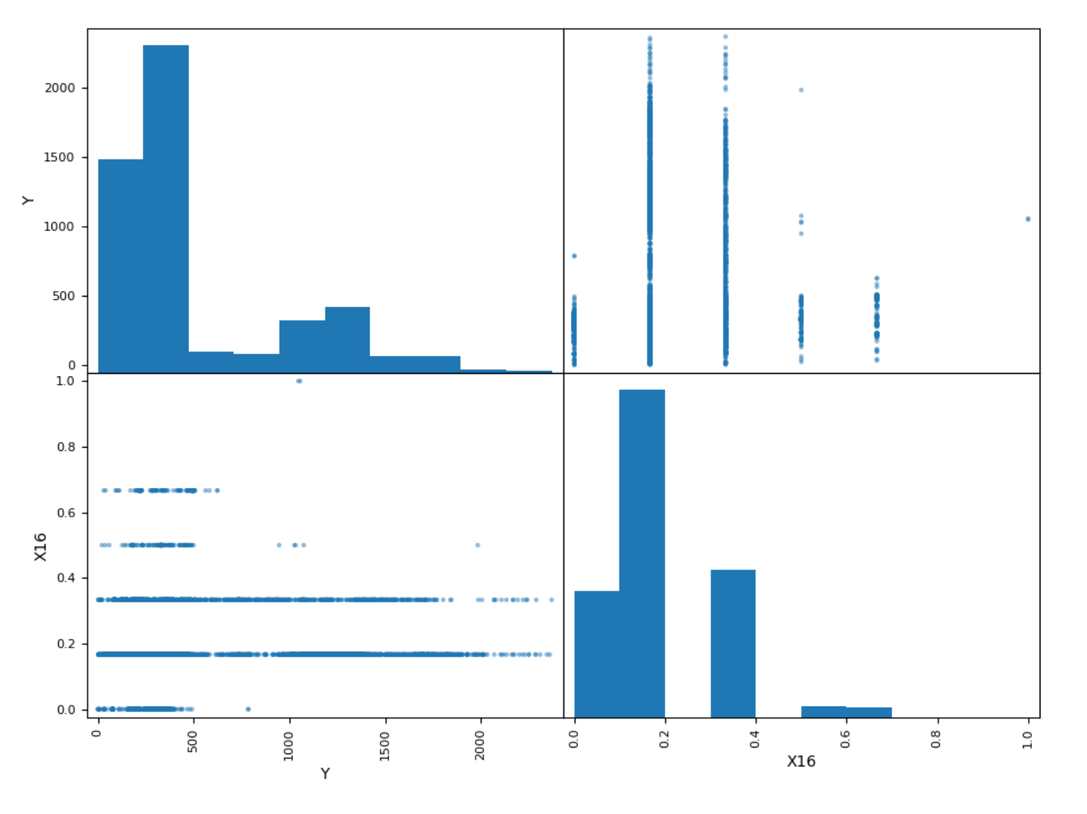
In [35]:
```
import os
import matplotlib.pyplot as plt
from pandas.plotting  import scatter_matrix
if not os.path.exists('HW2_output'): # if the directory HW2_output does not exist
    os.makedirs('HW2_output')        # will make (create) a new directory HW2_output

for i in range( len(Y_corr.index) //    ): # We will choose 2 features in each iteration.
    if 2*i+2 >= len(Y_corr.index): # In case that the total number of features is an odd number.
        scatter_matrix(train_data[ [Y_corr.index[0], Y_corr.index[  ]]], figsize=(11, 8))
    else: # choose 2 features
        scatter_matrix(train_data[ [Y_corr.index[0], Y_corr.index[  ], Y_corr.index[  ]]], figsize=(11, 8))

    plt.savefig("HW2_output/scatter_matrix_plot" + str(i+1))
    plt.show()
    plt.close('all') # close all pictures
```

The expective result of the figures

- *The total number of features is odd*: plot 2*2 correlation scatter matrix
- *The total number of features is even*: plot 3*3 correlation scatter matrix



The total number of features is even



The total number of features is odd

Setting conditional statement in *for loop*

- Counter: i
- Variables: X_1, X_2

i	X_1	X_2
0	1	2
1	3	4
2	5	6
…	…	…
m	n - 1	n

The total number of features is even

i	X_1	X_2
0	1	2
1	3	4
2	5	6
…	…	…
m	n	

The total number of features is odd

Hints

- Think about those hints in the relative place
- Comparison with the previous slide

In [35]:

```
import os
import matplotlib.pyplot as plt
from pandas.plotting  import scatter_matrix
if not os.path.exists('HW2_output'): # if the directory HW2_output does not exist
    os.makedirs('HW2_output')        # will make (create) a new directory HW2_output

for i in range( len(Y_corr.index) //   ): # We will choose 2 features in each iteration.
    if 2*i+2 >= len(Y_corr.index): # In case that the total number of features is an odd number.
        scatter_matrix(train_data[ [Y_corr.index[0], Y_corr.index[  ]]], figsize=(11, 8))
    else: # choose 2 features
        scatter_matrix(train_data[ [Y_corr.index[0], Y_corr.index[  ], Y_corr.index[  ]]], figsize=(11, 8))

    plt.savefig("HW2_output/scatter_matrix_plot" + str(i+1))
    plt.show()
    plt.close('all') # close all pictures
```

X_1 X_2