1230 Python Programming for Intelligent Medicine

Question 1: Heart Disease

Team 8 Members:
洪愷希、劉馨瑄、孫維

Data Preprocess

Original: (918, 12)

1. **Encoding:**

One hot

('Sex','RestingECG','ExerciseAngina','ST_Slope')

 Label encoding

(ChestPainType)

Final: (918, 17)

2. **We don't del data, No NA,**

3. **Normalization**

Model Selection

1. **Feature selection**

Method: Lasso, and filter to 10 variables   `After Feature Selection: (918, 10)`

```
Selected features:

['Age' 'ChestPainType' 'RestingBP' 'Cholesterol' 'MaxHR' 'Oldpeak' 'Sex_M'
 'ExerciseAngina_Y' 'ST_Slope_Flat' 'ST_Slope_Up']
```

2. **Model Selection**

random forest

XGBoost

K-Neighbors

Radom Forest With FS


```
[0.82065217 0.72282609 0.79891304 0.77595628 0.70491803]
[0.76630435 0.77173913 0.78804348 0.75409836 0.67759563]
[0.8423913  0.7826087  0.82065217 0.81420765 0.68852459]
[0.85869565 0.79347826 0.83695652 0.78688525 0.72131148]
[0.86413043 0.81521739 0.83152174 0.80327869 0.71584699]
score of cross validation everytime:   [0.7646531242575434,
max score of cross validation everytime:  0.805999049655500
Accuracy_score: 0.8369565217391305
classification_report
                precision    recall  f1-score   support

            0       0.91      0.78      0.84        50
            1       0.78      0.90      0.84        42

     accuracy                           0.84        92
    macro avg       0.84      0.84      0.84        92
 weighted avg       0.85      0.84      0.84        92
```
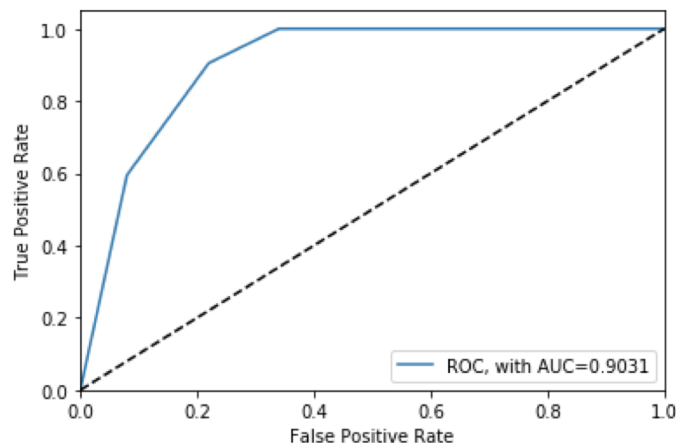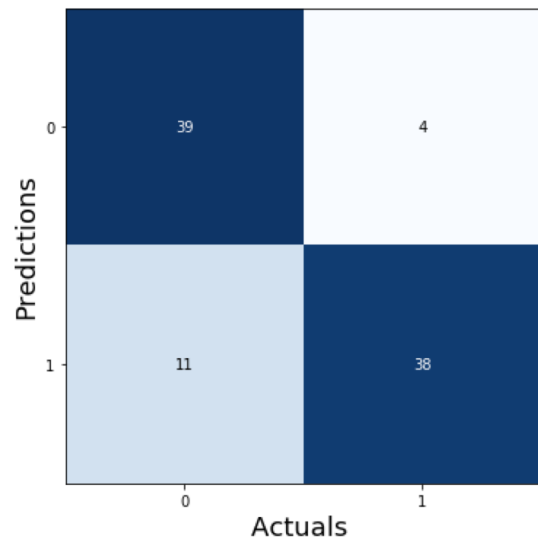
Radom Forest Without FS

```
[0.76086957 0.76086957 0.76630435 0.7704918  0.66120219]
[0.79891304 0.74456522 0.82065217 0.73224044 0.64480874]
[0.875      0.78804348 0.82065217 0.7704918  0.70491803]
[0.82065217 0.75543478 0.79891304 0.80327869 0.67759563]
[0.89130435 0.80978261 0.79891304 0.81420765 0.72677596]
score of cross validation everytime:  [0.7439474934663816,
max score of cross validation everytime:  0.80819672131147
Accuracy_score: 0.7717391304347826
classification_report
              precision    recall  f1-score   support

           0       0.73      0.78      0.75        41
           1       0.81      0.76      0.79        51

    accuracy                           0.77        92
   macro avg       0.77      0.77      0.77        92
weighted avg       0.77      0.77      0.77        92
```


Confusion Matrix