**Logistic Regression with scikit-learn**

## Logistic regression in Python with the [scikit-learn module](http://scikit-learn.org/). Dataset

The dataset I chose is of Parole Violation at   
<https://github.com/chicheongweng/MITx-15.071x-The-Analytics-Edge/blob/master/data/parole.csv>

Description of Variables:  
sex: Male 1, Female 0

Race: (1,2,3) Not necessary for our prediction

Age: Integer

State: (1,2,3)Not necessary for our prediction

Max.sentence: integer

Multiple.offenses: binary

Crime: integer

Violator: True(1), False(0)

Since

Problem Statement

Let’s treat this as a classification problem by focusing on three variables sex, multiple.offenses and violater. Since they are in binary form, it will be easier to use logistic regression function on those variables.  
  
Libraries Used:

Scipy

Numpy

Sklearn

Process:

-Read file using numpy

-Categorized X and y with the variables that were chosen to be used

-Instantiated LogisticRegression class available in Sklearn.linear\_model

-Called predict function from LogisticRegression on X

-Calculated the accuracy of the prediction using metrics.classifictionreport() from sklearn

-Ran python file

Result:85% precision rate for 409 samples and 409 test cases.

'precision', 'predicted', average, warn\_for)

precision recall f1-score support

0.0 0.92 1.00 0.96 377

1.0 0.00 0.00 0.00 32

avg / total 0.85 0.92 0.88 409